Issues in Science and Technology Librarianship Summer 2011
DOI: 10.5062/F42Z13FT

Refereed article

Comparison of the Contributions of CAPLUS and MEDLINE to the Performance of SciFinder in Retrieving the Drug Literature

Svetla Baykoucheva
Head, White Memorial Chemistry Library
University of Maryland
College Park, Maryland
sbaykouc@umd.edu

Abstract

SciFinder (SF) is a platform that provides access to two large databases, the Chemical Abstracts database (CAPLUS) and MEDLINE. This article analyzes and compares the individual and combined contributions of these two databases to the performance of SF in retrieving the drug literature. Test searches in which the names of two individual drugs (lisinopril and lovastatin) and a group of drugs (SSRI antidepressants) were used as keywords retrieved document sets that were analyzed for total and annual literature output, document types, journal coverage, and language of publication. While the total literature output from CAPLUS was larger than the output from MEDLINE (which was attributed to the presence of patents), MEDLINE performed significantly better than CAPLUS in retrieving the non-patent literature. The overlap of documents between CAPLUS and MEDLINE was found to be only 20-24%, depending on the name of the drug used to perform the searches. This article analyzes the strengths and the weaknesses of CAPLUS and MEDLINE and shows how these two databases, when searched together in SF, complement each other in covering the drug literature. In addition to the extended coverage of the literature, SF provides sophisticated (but easy-to-use) refining and analytical tools not available on some other platforms.

Introduction

Retrieving literature on interdisciplinary topics often requires using several databases. MEDLINE has been the most widely used database for retrieving the biomedical literature (Bianchi 2002; Weiner 2009). Freely available through PubMed, it can also be searched through some fee-based services. A previous study demonstrated that PubMed (provided by the National Institutes of Health) when used alone does not always satisfy users' needs, especially if a comprehensive literature retrieval is essential (Suarez-Almazor et al. 2000). Other studies, which examined the strengths and weaknesses of several databases in covering the biomedical literature, found that PubMed contained fewer documents than Scopus and Web of Science (WoS) (Falagas et al. 2008). Similar results were reported in a recent article, which showed that Scopus and WoS retrieved significantly more literature on several drugs than did MEDLINE (Baykoucheva 2010).

Services that integrate two or more databases on a single platform allow searching these databases at the same time and from one entry point. A recent article reported that researchers at the University of California Santa Cruz preferred using interdisciplinary databases such as Web of Science rather than subject-specific ones like PubMed (Hightower and Caldwell 2010). Vendors of such services are competing in providing new sophisticated refining and analytical tools that significantly improve search efficiency (Oprea and Tropsha 2006; Bandyopadhyay 2010).

DiscoveryGate (DG) (from Accelrys), for example, aggregates on one platform many databases that can be searched either together or individually (Baykoucheva 2007). A search performed in DG can be expanded to external databases such as PubChem, a free property database provided by the National Institutes of Health. The same search can be expanded even further, as PubChem links the records of the chemical compounds to articles indexed in PubMed (Baykoucheva 2008). A new platform from Elsevier, SciVerse, integrates two large databases -- ScienceDirect and Scopus. A third one, Reaxys, will be added to SciVerse in the near future. The first two are literature databases, while the third one is a platform that aggregates a patent database with two large property databases, the former CrossFire databases Beilstein and Gmelin.

SciFinder (SF), from the Chemical Abstracts Service (CAS), is another platform that integrates two large databases, CAPLUS and MEDLINE (Ridley 2009; Bolek 2000). While MEDLINE mostly covers journal literature, CAPLUS also covers patents. Another database included in SF is the CAS Registry File, the largest property database available today. SF has been used extensively by researchers in the area of drug discovery (Haldeman et al. 2005), but in academia it has been used mostly by chemists. Many librarians and users have indicated that this valuable resource has not gained popularity among students and researchers in the life sciences and the biomedical field mainly because this audience is not aware that SF searches MEDLINE.

The purpose of this study was to evaluate and compare the contributions of MEDLINE and CAPLUS to the performance of SF in retrieving drug literature and to show how users involved in life sciences/biomedical research could benefit from using it. The availability of sophisticated refining and analytical tools and the option of using natural language queries in SF could be very attractive to such users. In this study the names of two individual drugs (lisinopril and lovastatin) and the name of a group of drugs, "SSRI antidepressants" (which stands for selective serotonin reuptake inhibitors), were used as keywords to perform test searches in SF. The results presented here demonstrate how CAPLUS and MEDLINE contributed to the overall performance of SF in retrieving the drug literature.

Methods

Search Strategy

SciFinder Web (SF) (Chemical Abstracts Service) was used to search CAPLUS (Chemical Abstracts Service) and MEDLINE (National Library of Medicine, National Institutes of Health), either individually or at the same time. Test searches were performed using the names of two individual drugs (lisinopril and lovastatin) and the name of a group of drugs (SSRI antidepressants) as keywords. References containing the "concept" of the terms used to perform the searches were selected for further analyses. All searches in this study were performed on October 11 and October 15, 2010. The following document sets were obtained and analyzed for lisinopril, lovastatin, and SSRI antidepressants:

- CAPLUS: Document set obtained from searching SF (searches were performed in CAPLUS and MEDLINE at the same time, which is the default setting in SF) and limiting the retrieved documents only to those found in CAPLUS, using the "Refine by database" feature.
- MEDLINE: Document set obtained from searching SF (searches were performed in CAPLUS and MEDLINE at the same time, which is the default setting in SF) and limiting the retrieved documents only to those found in MEDLINE, using the "Refine by database" feature.
- CAPLUS & MEDLINE: Document set obtained from searching SF (searches were performed in CAPLUS and MEDLINE at the same time, which is the default setting in SF). Any overlapping documents were removed using the "Remove duplicates" feature before analyzing this set further. The set thus obtained contained all documents found in CAPLUS and MEDLINE, minus any duplicated records resulting from coverage overlap between the two databases; only one copy of each document was retained in this set. The strategy and the analyses performed in this study are illustrated in figures 1-3.

Database evaluation

The document sets described above were analyzed and compared in the following aspects:

- Total number of documents retrieved
- Output of non-patent and patent literature
- Coverage overlap between CAPLUS and MEDLINE
- Annual output of journal records
- Number of journals covered

- Document types
- Top 20 journals with highest number of documents retrieved by the databases
- Distribution of documents by language of publication

Selection of drugs

The names of the drugs listed below were used as keywords to perform test searches in SF:

- Lisinopril (Prinivil) is a drug from the angiotensin-converting enzyme (ACE) inhibitor class used alone or in combination with other medications to treat high blood pressure, congestive heart failure, heart attacks and in preventing renal and retinal complications of diabetes
- Lovastatin is a cholesterol-lowering drug belonging to the group of statins
- SSRI antidepressants (also called SSRIs), selective serotonin reuptake inhibitors or serotonin-specific reuptake inhibitors, are a class of drugs that are used mainly in the treatment of depression, anxiety, and some other disorders.

Results

Figures 1-3 outline the strategy used in this study and show the output of documents in each step of the analyses. The initial searches performed in SF, when CAPLUS and MEDLINE were searched at the same time (the default setting in SF), retrieved documents found in both databases. Such searches retrieved 4,769 documents on lisinopril, 10,327 on lovastatin, and 3,363 on SSRI antidepressants. Since there is some overlap of coverage between CAPLUS and MEDLINE, the document sets for lisinopril (1,037 duplicates), lovastatin (2,045 duplicates), and SSRI antidepressants (861 duplicates) were removed.

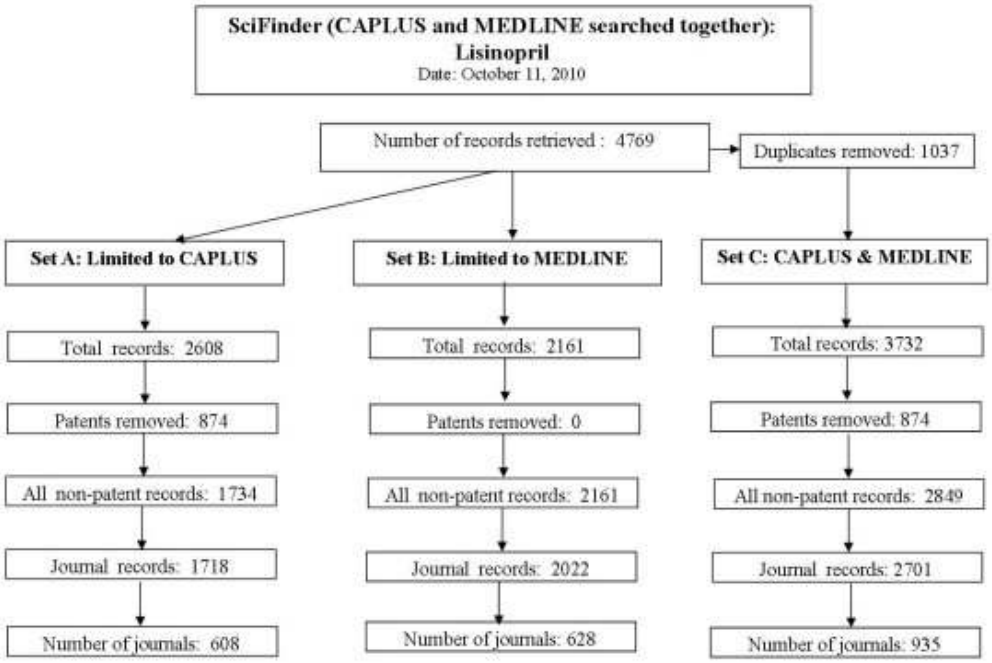


Figure 1: Strategy for retrieving literature on lisinopril and output of documents at individual steps of the process.

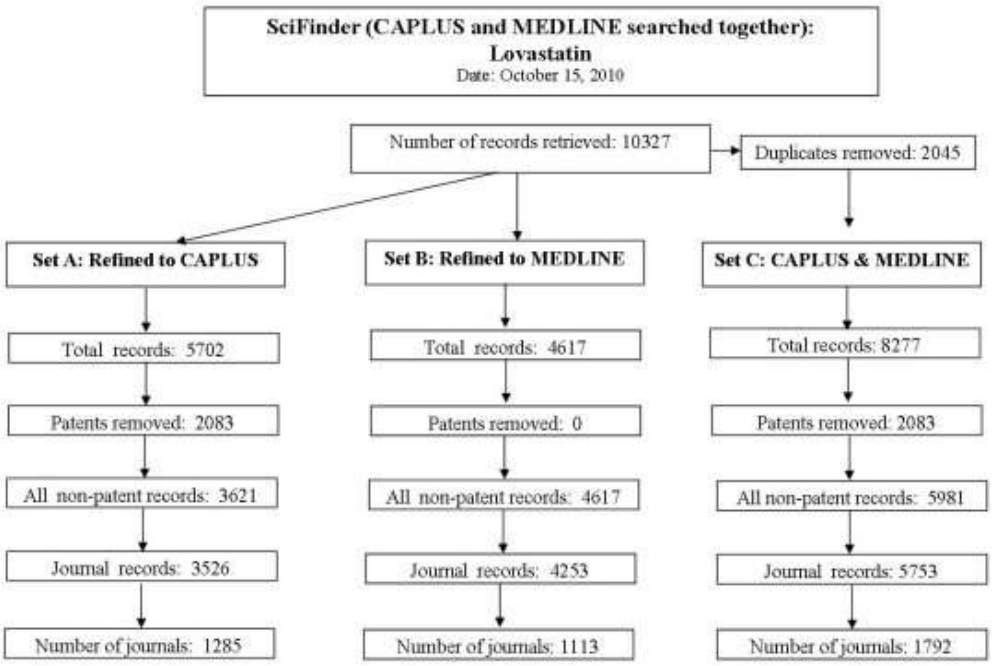


Figure 2: Strategy for retrieving literature on lovastatin and output of documents at individual steps of the process.

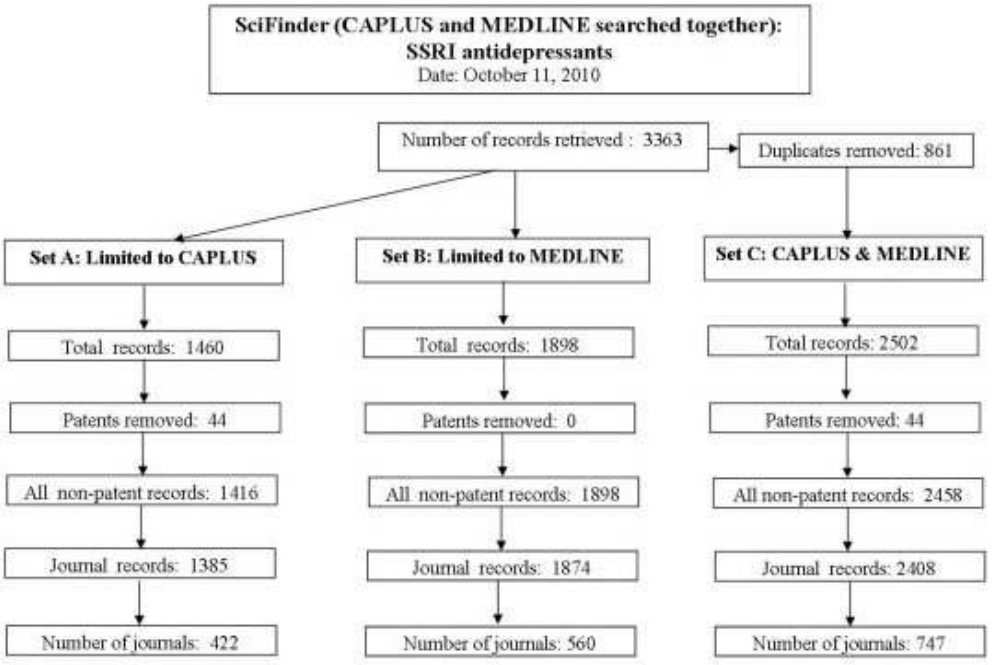


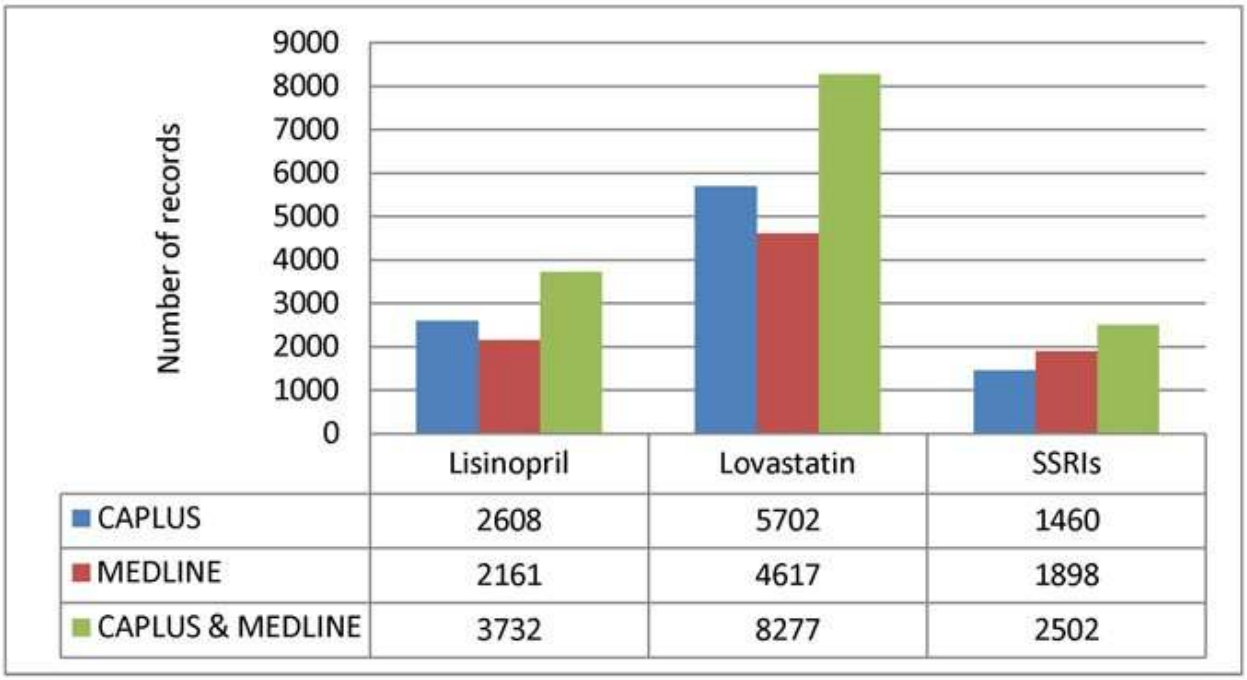
Figure 3: Strategy for retrieving literature on SSRI antidepressants and output of documents at individual steps of the process.

The document set obtained when lovastatin was used as a key word was processed differently than the document sets obtained when lisinopril and SSRI antidepressants were used as keywords, because the number of retrieved documents on lovastatin exceeded 10,000 -- the maximum number of documents from which duplicates can be removed in SF. In order to be able to remove the duplicates, the document set obtained for lovastatin was split into two sub-sets (using the "Refine by publication year" command) that contained documents published in two time periods: (1) 1980-2009 and (2) 2010. The subsets thus obtained were analyzed and the results obtained with them were later combined.

The results obtained for the sets of lisinopril, lovastatin, and SSRI antidepressants when SF searched

CAPLUS and MEDLINE together are presented as "CAPLUS & MEDLINE." These sets contained all documents that were unique to one or the other database, as well as one copy of the documents found in both databases. The results obtained for these sets showed that the overlap of content between CAPLUS and MEDLINE was found to be from 20-26%, depending on the drug literature studied.

The documents retrieved from the initial keyword searches (when CAPLUS and MEDLINE were searched together) were refined by database, to limit the content of the sets only to documents found in the individual databases. Figure 4 compares the total literature outputs obtained from CAPLUS (document set containing documents present only in CAPLUS), MEDLINE (document set containing documents present only in MEDLINE), and CAPLUS & MEDLINE (document set containing documents retrieved when CAPLUS and MEDLINE were searched in SF together, from which all duplicates were removed).



	Lisinopril	Lovastatin	SSRIs
CAPLUS	2608	5702	1460
MEDLINE	2161	4617	1898
CAPLUS & MEDLINE	3732	8277	2502

Figure 4: Total output of literature on lisinopril, lovastatin, and SSRI antidepressants retrieved from CAPLUS, MEDLINE, and CAPLUS & MEDLINE.

Tables 1-3 show the distribution of the retrieved documents by document type. In the document sets "CAPLUS" and "CAPLUS & MEDLINE" there were 874, 2,083, and 44 patents on lisinopril, lovastatin, and SSRI antidepressants, respectively. These patents were removed to obtain document sets that contained only non-patent literature.

Table 1. Distribution of documents retrieved from CAPLUS, by document type.

Document Type	Lisinopril	Lovastatin	SSRIs*
Book	1		1
Computer Optical Disk	2	13	9
Conference	21	65	27
Dissertation	2	29	3
Editorial	1		1
General Review	123	359	495
Journal	1718	3526	1385
Letter		2	
Meeting Abstract	1	16	11
Online Computer File	35	62	31
Patent	874	2083	44
Report		1	

*SSRIs, SSRI antidepressants

Table 2. Distribution of documents retrieved from MEDLINE, by document type.

Document Type	Lisinopril	Lovastatin	SSRIs*
Article	2022	4254	1879
Bibliography	2	5	
Biography	2	8	
Commentary	58	167	20
Conference	5	6	2
Conference Article	2	5	
Editorial	25	50	6
General Review	235	501	652
Historical	3	10	6
Journal	2022	4253	1874
Letter	100	279	15

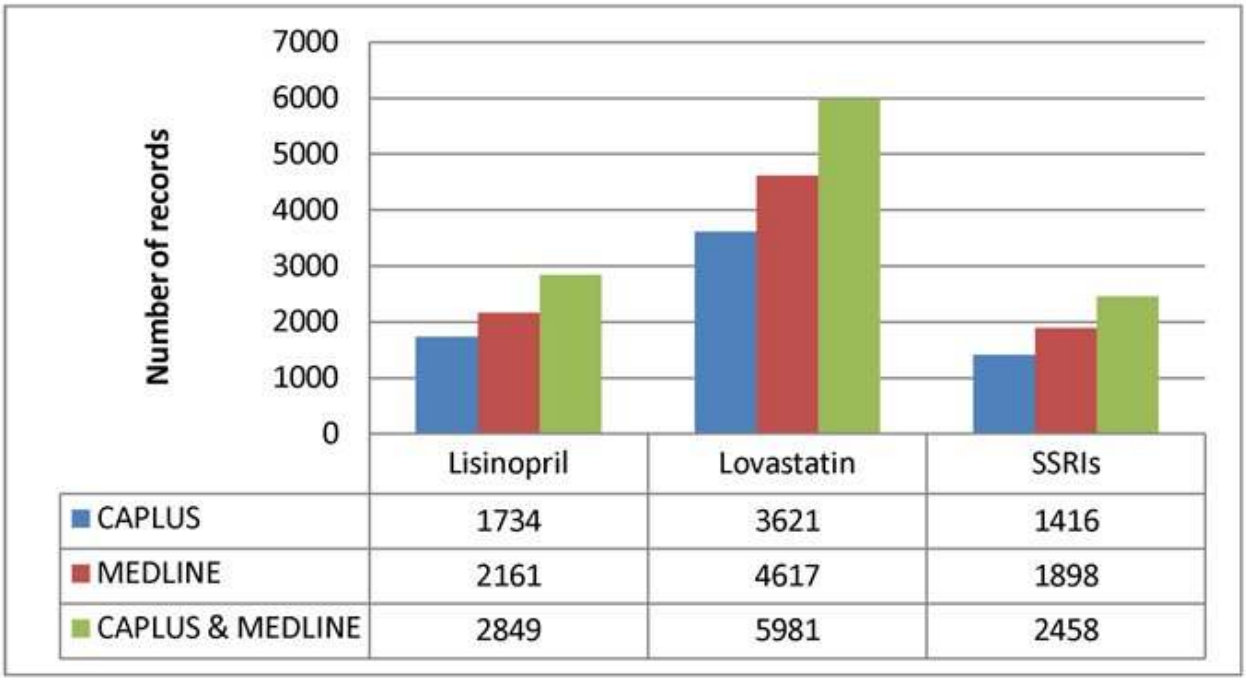
*SSRIs, SSRI antidepressants

Table 3. Distribution of documents retrieved from set CAPLUS & MEDLINE, by document type.

Document Type	Lisinopril	Lovastatin	SSRIs*
Article	996	2231	1028
Bibliography	2	5	
Biography	2	5	
Book	1		1
Commentary	55	162	17
Computer Optical Disk	2	13	9
Conference	25	70	29
Conference Article	1	4	
Dissertation	2	29	3
Editorial	23	48	4
General Review	284	180	873
Historical	3	7	5
Journal	2701	5753	2408
Letter	99	270	14
Meeting Abstract	1	16	11
News Announcement	7	21	
Online Computer File	34	62	31
Patent	874	2081	44

*SSRIs, SSRI antidepressants

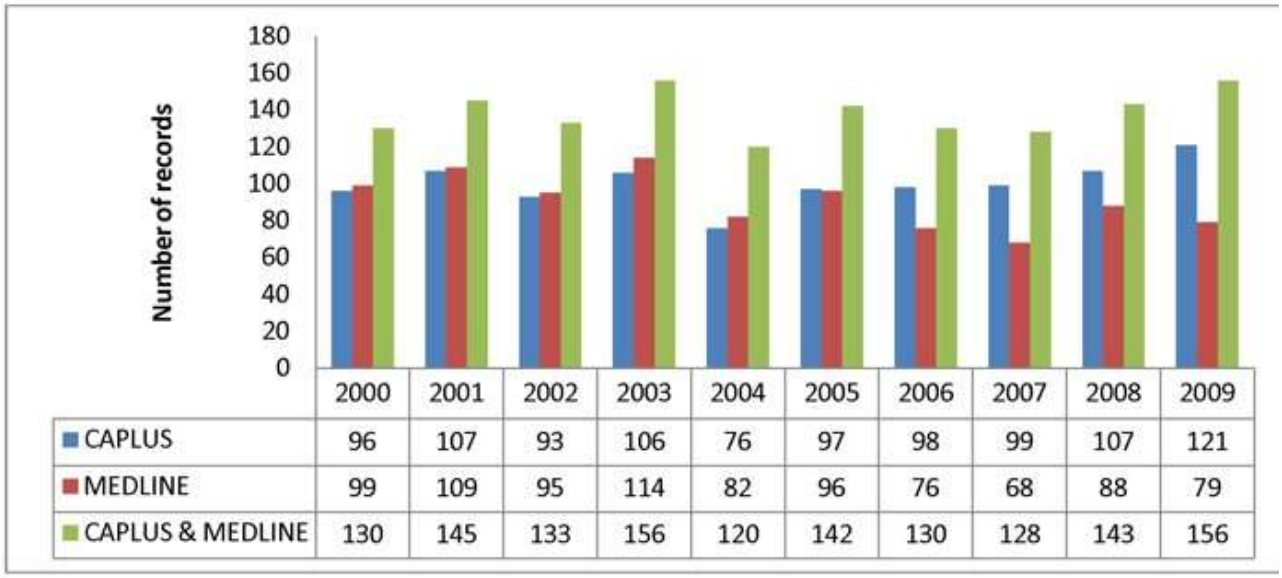
Figure 5 shows the output of the non-patent literature retrieved from the databases.



	Lisinopril	Lovastatin	SSRIs
CAPLUS	1734	3621	1416
MEDLINE	2161	4617	1898
CAPLUS & MEDLINE	2849	5981	2458

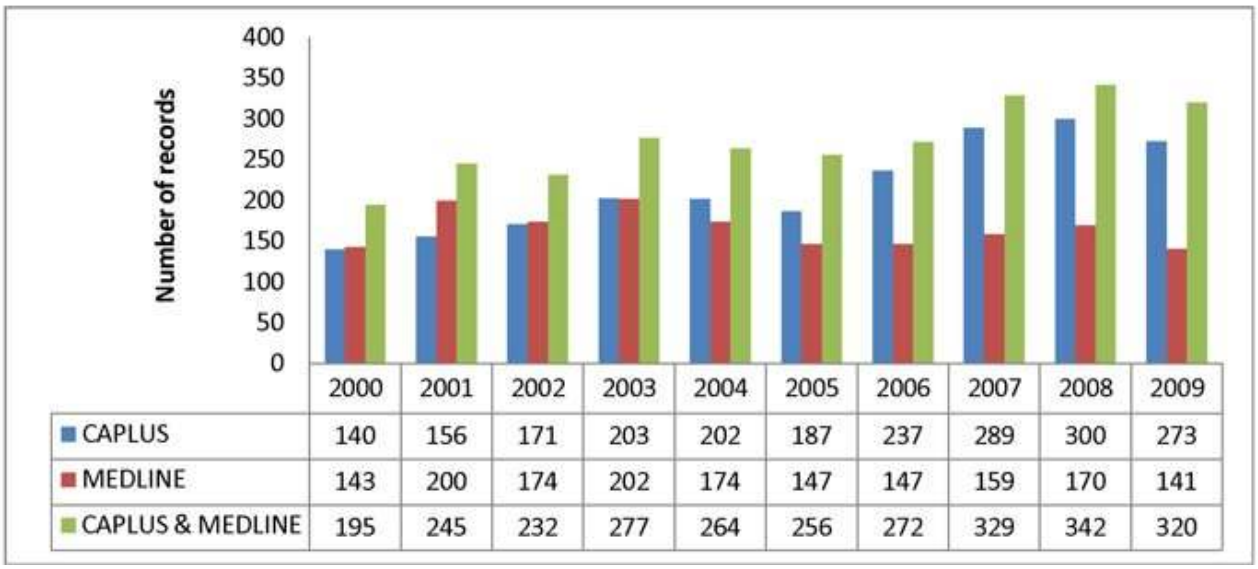
Figure 5: Output of non-patent literature on lisinopril, lovastatin, and SSRI antidepressants retrieved from CAPLUS, MEDLINE, and CAPLUS & MEDLINE.

The sets containing non-patent records were further refined by document type to limit them to journal records, which were further analyzed by publication year. The annual output of journal records was determined for each year throughout the whole publication history of the drugs, until October 11 or October 15, 2010, when the searches were performed. The first documents on lisinopril, lovastatin, and the SSRI antidepressants retrieved from the databases were published in 1981, 1976, and 1991, respectively. Figures 6-8 illustrate the annual output of documents from the databases during a 10-year period of time -- from 2000 to 2009 (the year 2009 was the last complete year of this study).



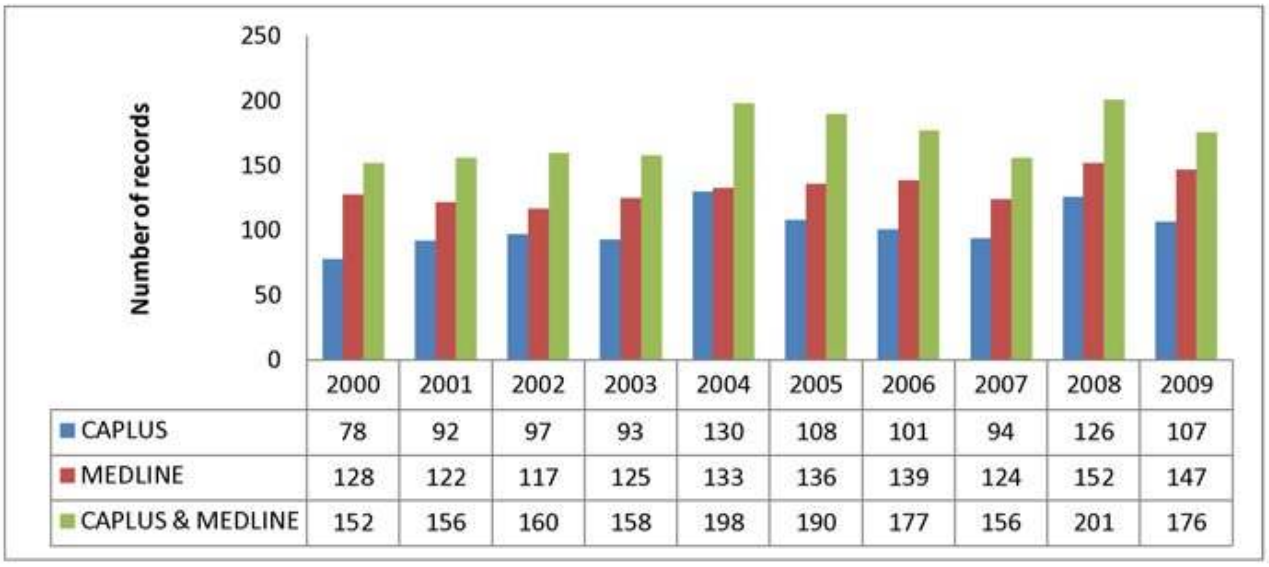
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
CAPLUS	96	107	93	106	76	97	98	99	107	121
MEDLINE	99	109	95	114	82	96	76	68	88	79
CAPLUS & MEDLINE	130	145	133	156	120	142	130	128	143	156

Figure 6: Annual output of journal articles on lisinopril retrieved from CAPLUS, MEDLINE, and CAPLUS & MEDLINE (2000-2009).



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
CAPLUS	140	156	171	203	202	187	237	289	300	273
MEDLINE	143	200	174	202	174	147	147	159	170	141
CAPLUS & MEDLINE	195	245	232	277	264	256	272	329	342	320

Figure 7: Annual output of journal articles on lovastatin retrieved from CAPLUS, MEDLINE, and CAPLUS & MEDLINE (2000-2009).



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
CAPLUS	78	92	97	93	130	108	101	94	126	107
MEDLINE	128	122	117	125	133	136	139	124	152	147
CAPLUS & MEDLINE	152	156	160	158	198	190	177	156	201	176

Figure 8: Annual output of journal articles on SSRI antidepressants retrieved from CAPLUS, MEDLINE, and CAPLUS & MEDLINE (2000-2009).

Figure 9 shows the number of journals covered by each database.

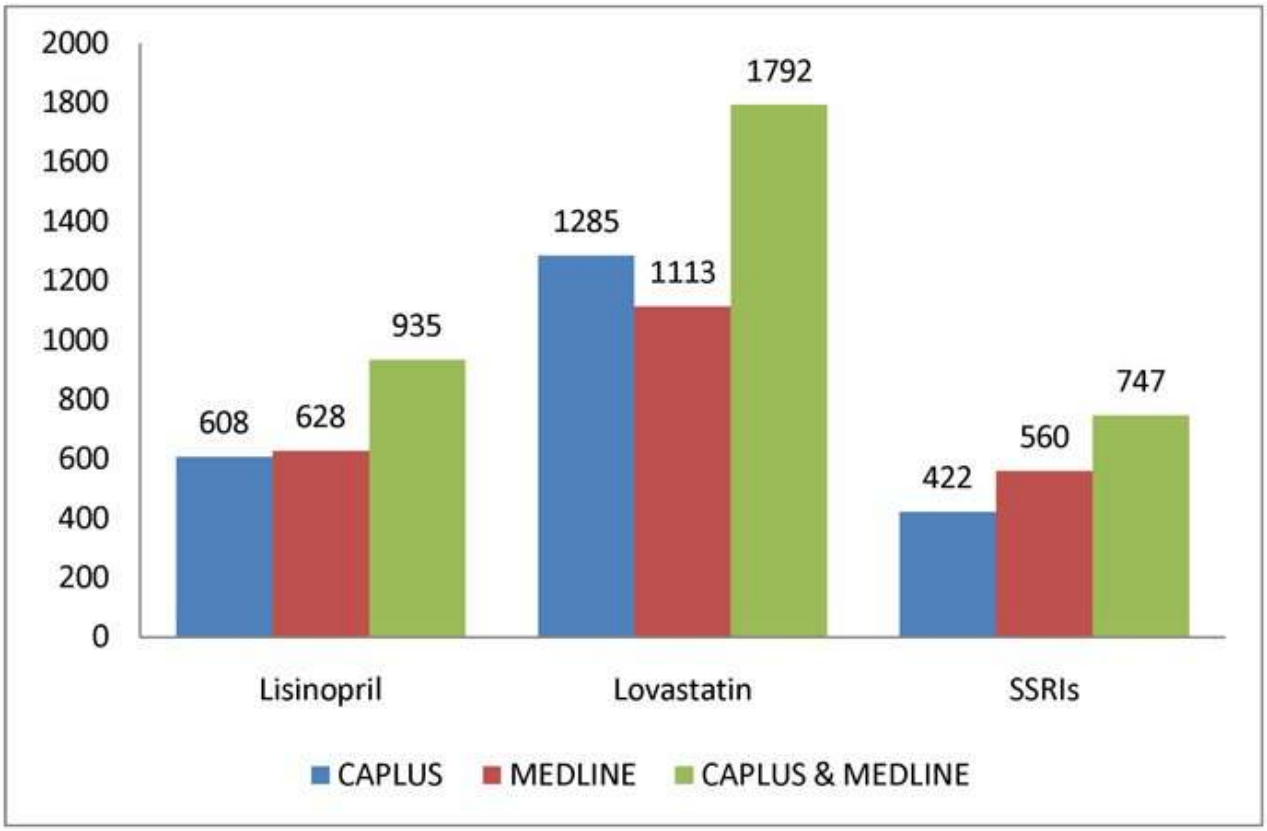


Figure 9: Number of journals containing documents on lisinopril, lovastatin, and SSRI antidepressants that were covered by CAPLUS, MEDLINE, and CAPLUS & MEDLINE.

Additional evaluation of the journal coverage was performed by comparing the lists of the top 20 journal titles from which the databases contained the highest number of articles (Tables 4-6).

Table 4. The top 20 journals with the highest number of records on lisinopril found in CAPLUS and MEDLINE (journal titles in boldface are shared by both databases).

CAPLUS		MEDLINE		CAPLUS & MEDLINE	
Source Title	Recs	Source Title	Recs	Source Title	Recs
J Hypertens	50	**J Cardiovasc Pharmacol**	74	J Cardiovasc Pharmacol	79
J Cardiovasc Pharmacol	45	**J Hum Hypertens**	70	J Hum Hypertens	70
Hypertension	42	**J Hypertens**	53	J Hypertens	63
A J P[a]	38	**Hypertension**	52	Hypertension	59
Kidney Int	28	**Am J Cardiol**	46	Am J Cardiol	53
Am J Hypertens	27	**Am J Hypertens**	41	Am J Hypertens	46
Circulation	26	**Kidney Int**	37	Kidney Int	40
Am J Cardiol	22	Lancet	31	A J P[a]	38
Drugs	21	J Hypertens Suppl	29	Circulation	33
J Pharm Biomed Anal	21	**Drugs**	28	Drugs	33
J Hum Hypertens	20	**Circulation**	27	J Hypertens Suppl	29
JASN[b]	20	**Br J Clin Pharmacol**	25	Br J Clin Pharmacol	28
Br J Pharmacol	19	Ann Pharmacother	25	Am J Kidney Dis	25
J Pharmacol Exp Ther	19	J Clin Hypertens	21	Br J Pharmacol	21
Am J Kidney Dis	18	**Am J Kidney Dis**	20	J Pharm Biomed Anal	21
J Am Coll Cardiol	18	**A J P[a]**	20	Am Heart J	20
Br J Clin Pharmacol	17	**Br J Pharmacol**	19	Eur J Clin Pharmacol	20
Eur J Pharmacol	17	**J Pharmacol Exp Ther**	19	J Am Coll Cardiol	20
Nephrol Dial Transplant	16	**Eur J Pharmacol**	18	JASN[b]	20
Am Heart J	15	**JASN[b]**	18	J Pharmacol Exp Ther	19

[a]American Journal of Physiology

[b]Journal of the American Society of Nephrology

Table 5. The top 20 journals with the highest number of records on lovastatin found in CAPLUS and MEDLINE (journal titles in boldface are shared by both databases).

CAPLUS		MEDLINE		CAPLUS & MEDLINE	
Source Title	Recs	Source Title	Recs	Source Title	Recs
J Biol Chem	99	**Am J Cardiol**	148	Am J Cardiol	168
Am J Cardiol	72	**J Biol Chem**	145	J Biol Chem	159
Atherosclerosis[a]	57	**Atherosclerosis**[a]	108	Atherosclerosis[a]	117
J Lipid Res	54	**J Lipid Res**	76	J Lipid Res	79
Biochem Biophys Res[b]	41	**Biochim Biophys Acta**[c]	68	Biochim Biophys Acta[c]	71
Biochim Biophys Acta[c]	35	**Biochem Biophys Res**[b]	64	Biochem Biophys Res[b]	69
Proc Natl Acad Sci U S A	34	**Proc Natl Acad Sci U S A**	45	Circulation	53
Drug Metab Dispos	31	**Circulation**	42	Proc Natl Acad Sci U S A	52
Biochem Pharmacol	28	Clin Ther	37	Clin Ther	39
Circulation	26	Metabolism[d]	36	Metabolism[d]	37
J Med Chem	24	**Biochem J**	36	Biochem J	36
Ann Pharmacother	21	**Biochem Pharmacol**	32	Biochem Pharmacol	36
Arch Biochem Biophys	21	Eur J Clin Pharmacol	29	J Med Chem	35
Biochem J	21	Arterioscler Thromb Vasc[e]	28	Arch Intern Med	34
FEBS Lett	21	Arch Intern Med	27	Am J Med	31
J Pharmacol Exp Ther	21	**FEBS Lett**	26	Arterioscler Thromb Vasc[e]	31
Curr Med Res Opin	20	J Neurochem	26	Drug Metab Dispos	31
Int J Cancer	20	Am J Med	26	Eur J Clin Pharmacol	31
J Pharm Biomed Anal	20	**Cancer Res**	25	Drugs	29
Cancer Res	19	Drugs	25	FEBS Lett	29

[a]Includes documents published in all Atherosclerosis titles

[b]Biochem Biophys Res Commun

[c]Includes documents published in all Biochimica et Biophysica Acta titles

[d]Metabolism: clinical and experimental

[e]Arteriosclerosis, Thrombosis, and Vascular Biology

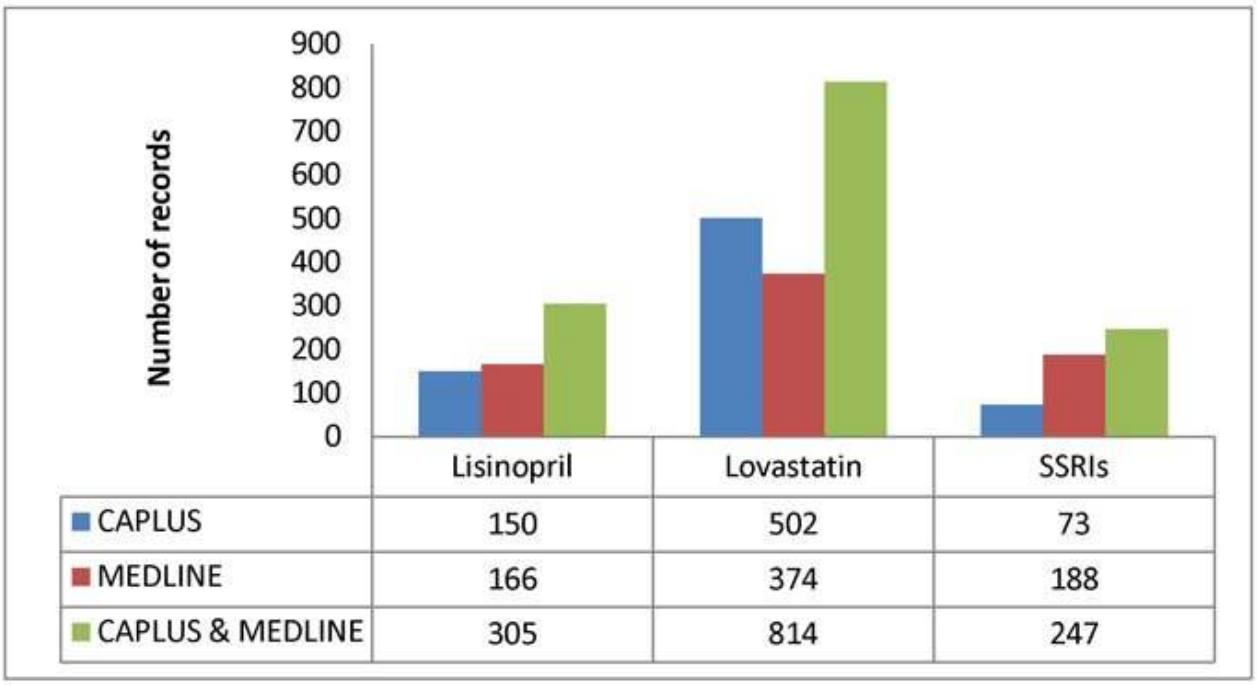
Table 6. The top 20 journals with the highest number of records on SSRI antidepressants found in CAPLUS and MEDLINE (journal titles in boldface are shared by both databases).

CAPLUS		MEDLINE		CAPLUS & MEDLINE	
Source Title	Recs	Source Title	Recs	Source Title	Recs
J Clin Psychiatry	86	**J Clin Psychiatry**	104	J Clin Psychiatry	117
Eur Neuropsychopharmacol	48	**J Affect Disord**	52	J Affect Disord	57
J Clin Psychopharmacol	44	Int Clin Psychopharmacol	44	J Clin Psychopharmacol	54
Neuropsychopharmacology	44	**Eur Neuropsychopharmacol**	41	Eur Neuropsychopharmacol	48
Psychopharmacology	40	L'Encephale	33	Int Clin Psychopharmacol	44
CNS Drugs	37	**Neuropsychopharmacology**	33	Neuropsychopharmacology	44
Hum Psychopharmacol	31	**Psychopharmacology**	32	CNS Drugs	41
Biol Psychiatry	30	**J Clin Psychopharmacol**	28	Hum Psychopharmacol	36
J Affect Disord	27	**CNS drugs**	25	Biol Psychiatry	34
Int J Neuropsychopharmacol	25	Pharmacoepidemiol Drug[b]	23	L'Encephale	33
Neuropharmacology	21	J Psychopharmacol	22	Pharmacoepidemiol Drug[b]	26
Progress in Neuro Psycho[a]	20	**Biol Psychiatry**	21	Psychopharmacology	40
Drug Safety	17	**Br J Clin Pharmacol**	19	Int J Neuropsychopharmacol	25
Br J Clin Pharmacol	15	Depression and anxiety	19	Neuropharmacology	21
Behav Brain Res	14	PharmacoEconomics	19	Pharmacopsychiatry	21
Acta Psychiatr Scand Suppl	13	Pharmacopsychiatry	19	Br J Clin Pharmacol	20
Expert Opin Pharmacother	13	Cochrane Database Syst Rev	18	Progress in Neuro Psycho[a]	20
Brain Res	12	**Drug Saf**	17	Br J Psychiatry	20
Br J Pharmacol	12	Psychopharmacol Bull	17	Depression and anxiety	19
Clin Ther	12	Br J Psychiatry	17	PharmacoEconomics	19

[a]Progress in Neuro Psychopharmacology & Biological Psychiatry

[b]Pharmacoepidemiol Drug Saf

The non-patent documents retrieved from the databases were also analyzed by language of publication (Figure 10). The top five languages covered by each database are shown in Table 7.



	Lisinopril	Lovastatin	SSRIs
CAPLUS	150	502	73
MEDLINE	166	374	188
CAPLUS & MEDLINE	305	814	247

Figure 10: Number of non-patent documents published in languages other than English retrieved from CAPLUS, MEDLINE, and CAPLUS & MEDLINE.

Table 7. Number of records published in non-English languages and found in CAPLUS and MEDLINE searched individually or together in SciFinder.

CAPLUS		MEDLINE		CAPLUS & MEDLINE	
Language	Records	Language	Records	Language	Records
Lisinopril					
Chinese	66	Russian	44	Chinese	69
Japanese	36	German	31	Russian	45
German	5	Italian	24	Japanese	41
Portuguese	5	French	23	German	34
Russian	3	Spanish	16	Italian	24
Lovastatin					
Chinese	387	German	50	Chinese	406
Korean	25	French	44	German	61
Japanese	22	Spanish	43	Spanish	50
Russian	22	Italian	41	Russian	48
German	12	Chinese	40	French	46
SSRI Antidepressants					
Japanese	38	French	55	French	55
Chinese	20	German	46	Japanese	55
French	9	Japanese	22	German	46
German	2	Spanish	18	Chinese	20
Afrikaans	1	Polish	7	Spanish	18

Discussion

The selection of lisinopril, lovastatin, and SSRI antidepressants as models in this study was based on the fact that these drugs significantly differed in properties and that they had long clinical and publication histories. Although CAS and MEDLINE have different indexing approaches, CAS performs systematic indexing for SF, which allows precision searches to be accomplished even when only one of the possible terms is used to perform the searches (Ridley 2009). The option of retrieving references that include "the concept" of the search term has to be selected to benefit from this strategy.

CAPLUS and MEDLINE covered almost the same publication periods for the drugs included in this study, but the peak years in which these databases had the highest number of articles on a particular drug occurred at different times. While for CAPLUS the peak of articles on lisinopril occurred in 2009, the peak for the literature on this drug happened for MEDLINE six years earlier and coincided with the peak observed for CAPLUS & MEDLINE (figure 6). The peak of articles on lovastatin for CAPLUS, MEDLINE, and CAPLUS & MEDLINE occurred in 2008, 2003, and 2008, respectively (figure 7). As shown in figure 8, the year for which CAPLUS, MEDLINE, and CAPLUS & MEDLINE contained the highest number of documents on SSRI antidepressants was 2004, 2008, and 2008, respectively. Another study, which compared results from searches performed in CAPLUS and MEDLINE, also showed that the peak years for articles happened for these two databases in different years (Brown 2003).

As shown in tables 1-3, there were more document types in MEDLINE than in CAPLUS. The document types in CAPLUS & MEDLINE were a combination of the document types of these two databases. CAPLUS contained patents, a document type not covered by MEDLINE, but the latter outperformed CAPLUS in the number of non-patent documents. Analysis of the annual journal literature output showed that MEDLINE consistently retrieved more journal articles than CAPLUS until 2004 (for lisinopril) (figure 6) and until 2002 (for lovastatin) (figure 7), but this trend changed in the more recent years, when CAPLUS started retrieving more documents on these drugs than MEDLINE. Throughout the whole history of publication on the SSRI antidepressants, MEDLINE retrieved more journal articles than CAPLUS (figure 8).

Evaluation of the retrieved documents by journal title showed significant differences in the number of journals covered by the databases. Figure 9 shows that, while CAPLUS and MEDLINE covered almost equal number of journals on lisinopril, CAPLUS covered 172 more journals than MEDLINE on lovastatin, and MEDLINE covered 138 more journals than CAPLUS on SSRI antidepressants. When the two databases were searched together (CAPLUS & MEDLINE), the number of journals covered was significantly higher than the number of journals covered by the individual databases.

Additional evaluation of the journal coverage consisted in analyzing and comparing the lists of the top 20 journal titles from which the databases have retrieved the highest number of articles. The results from these analyses showed that the lists of CAPLUS and MEDLINE shared 16, 12, and 10 journal titles that have published articles on lisinopril, lovastatin, and SSRI antidepressants, respectively (tables 4-6). The data presented in these tables also show that the number of articles from the same journals differed significantly between the databases. For example, from all shared journals MEDLINE had more documents on lisinopril and lovastatin than CAPLUS. From the 10 shared journal titles that had articles on SSRI antidepressants, CAPLUS had more documents from six and MEDLINE from four of these journals. The number of articles from the shared journals was significantly increased when the two databases were searched together (data presented as CAPLUS & MEDLINE).

Analysis of the documents by language of publication showed significant differences between the databases. While MEDLINE contained more non-patent documents on lisinopril and SSRI antidepressants in languages other than English, CAPLUS covered more such documents than MEDLINE on lovastatin (figure 10). Chinese and Japanese were the predominant languages for non-patent literature found in CAPLUS, while MEDLINE covered more documents in some of the European languages (table 7). When the databases were searched together, the output of documents in languages other than English was significantly higher than the output from the individual databases.

The total number of documents retrieved from SF was much larger when CAPLUS and MEDLINE were searched together than when they were searched separately. The original content found in the individual databases was 70-76%, depending on the drug literature. Previous studies have shown that, while the biomedical literature retrieved with PubMed significantly overlapped in content with the literature retrieved with Google Scholar (Anders and Evans 2010; Shultz 2007), CAPLUS covered a large volume of literature that was unique and could not be retrieved with Google Scholar (Levine-Clark and Kraus 2007).

Conclusions

MEDLINE is the most widely used database for retrieving the biomedical literature, but drug research is an interdisciplinary area that also requires monitoring of the chemical literature. The results reported in this article indicate that searching CAPLUS and MEDLINE together through SF significantly expands the capabilities of these two databases in retrieving literature on such interdisciplinary topic as drugs. For those who want to retrieve drug literature, using SF is a much better option than searching CAPLUS or MEDLINE alone, as these two databases complement very well each other with respect to journal coverage, document types, and languages in which the documents are published.

References

Anders, M.E. and Evans, D.P. 2010. Comparison of PubMed and Google Scholar literature searches. *Respiratory Care* 55(5):578-83.

Bandyopadhyay, Aditi. 2010. Examining Biological Abstracts on two platforms: what do end users need to know? *Science & Technology Libraries* 29(1):34-52.

Baykoucheva, Svetla. 2007. A new era in chemical information: PubChem, DiscoveryGate, and Chemistry Central. *Online* 31(5):16-20.

Baykoucheva, Svetla. 2008. Finding drug information in integrated chemistry and life sciences databases: PubChem and DiscoveryGate. *Abstracts of Papers, 236th ACS National Meeting, Philadelphia, PA, United States, August 17-21, 2008*: CINF-041.

Baykoucheva, Svetla. 2010. Selecting a database for drug literature retrieval: a comparison of MEDLINE, Scopus, and Web of Science. *Science & Technology Libraries* 29:276-288.

Bianchi, Stephanie. 2002. PubMed: for more than medicine this is one of the world's greatest databases. *Issues in Science & Technology Librarianship* 34 (Spring). [Internet]. [Cited July 11, 2011]. Available from: http://www.istl.org/02-spring/databases3.html

Bolek, Ann D. 2000. SciFinder Scholar. *Issues in Science & Technology Librarianship* 28 (Fall). [Internet]. [Cited July 11, 2011]. Available from http://www.istl.org/00-fall/databases.html.

Brown, Cecelia. 2003. The changing face of scientific discourse: analysis of genomic and proteomic database usage and acceptance. *Journal of the American Society for Information Science & Technology* 54(10):926-938.

Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., and Pappas, G. 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *Faseb Journal* 22(2):338-342.

Haldeman, Margaret, Vieira, Barbara, Winer, Fred, and Knutsen, Lars J.S. 2005. Exploration tools for drug discovery and beyond: applying SciFinder to interdisciplinary research. *Current Drug Discovery Technologies* 2(2):69-74.

Hightower, Christy and Caldwell, Christy. 2010. Shifting sands: science researchers on Google Scholar, Web of Science, and PubMed, with implications for library collections budgets. *Issues in Science and Technology Librarianship*. [Internet]. [Cited July 11, 2011]. Available from: http://www.istl.org/10-fall /refereed3.html

Levine-Clark, Michael, and Kraus, Joseph. 2007. Finding chemistry information using Google Scholar: a comparison with Chemical Abstracts Service. *Science & Technology Libraries* 27(4):3-17.

Oprea, Tudor I. and Tropsha, Alexander. 2006. Target, chemical and bioactivity databases - integration is key. *Drug Discovery Today: Technologies* 3 (4):357-365.

Ridley, Damon D. 2009. *Information Retrieval: SciFinder*. 2nd ed. Chichester, UK: John Wiley & Sons, Ltd.

Shultz, Mary. 2007. Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association* 95(4):442-445.

Suarez-Almazor, Maria E., Belseck, Elaine, Homik, Joanne, Dorgan, Marlene, and Ramos-Remus, Cesar. 2000. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Controlled Clinical Trials* 21 (5):476-487.

Weiner, Sharon A. 2009. Tale of two databases: the history of federally funded information systems for education and medicine. *Government Information Quarterly* 26 (3):450-8.

Previous Contents Next