Conference paper presented at SLA 2013 Annual Conference & Info-Expo, San Diego, California.
2016 update: Note that the *Scopus* interface changed since this paper was written. To extract the cited references of your results in *Scopus*, click on "More..." > "View references" and then select all the references and click on "Export."

Comparing the Indexing of Cited Journals to Identify the Premier Database for a Specific Discipline

Giovanna Badia, B.A., M.L.I.S.
Liaison Librarian, McGill University
Schulich Library of Science & Engineering
Montreal, Quebec, Canada

Introduction

Librarians continually seek to find better ways to connect people with the information they need. In research-intensive academic institutions, libraries often subscribe to more than one database to cover any given discipline. While databases overlap, they also contain unique content and features that differentiate them from their competitors. From these databases, can we identify the premier database in a field? Librarians need to know which database provides the most comprehensive coverage in a subject area, in order to conduct efficient searches and teach others how to develop their search strategies, as well as make sound fiscal decisions regarding database subscriptions.

This paper will employ a multi-step method to determine the most comprehensive database in the field of chemical engineering. Although the literature has reported use of citation analysis to identify top journals in chemical engineering for collection development purposes (Barrett and Barrett 1957, Burton 1959, Patterson 1945), these studies need to be updated since they were published in the 1940s and 1950s. The methodology used in these older studies will be taken a step further to find the database with the most comprehensive coverage in chemical engineering.

The research methodology involves: 1) selecting one to five titles from the core journals in chemical engineering; 2) examining the references cited in articles published in these core journals over a period of time to create a list of all cited journal titles; and 3) determining whether or not cited journal titles are indexed in a selection of databases, depending on the discipline. In this paper, four chemical engineering journal titles are selected, i.e., *AIChE Journal*, *Canadian Journal of Chemical Engineering*, *Chemical Engineering Science*, and *Industrial & Engineering Chemistry Research*, the references in these journals are examined for the year 2011, and the database that indexes the majority of cited journal titles, as well as provides references to the most items published within these cited titles, is identified as the premier database to search for chemical engineering topics.

Only a few studies, such as in the areas of agriculture, nursing and allied health, have analyzed the indexing of cited journal titles to compare databases (Kawasaki 2002, Delwiche, Schloman, and Allen 2010). This paper will apply the methodology to the field of chemical engineering and discuss how citation analysis can be used to determine the core database in any discipline. Additionally, it will describe distinctive features of the databases investigated, i.e., *Compendex*, *Inspec*, *Scopus*, and *Web of Science*, to aid the library professional in providing database recommendations for specific contexts.

Background

The citation analysis of articles, which were either published in a specific journal(s), authored by researchers in a particular discipline(s), or written by faculty from a particular institution(s), has been employed in a variety of ways, such as to determine core journals in a discipline for collection development and retention purposes in libraries, to identify developing trends in a research area, to rank universities, to characterize author collaboration networks, etc. (Bar-Ilan 2008) Librarians have analyzed citation data to create core journal lists by subject to guide their purchasing and weeding decisions. For example, Burton (1959) identified the 74 most cited journals in chemical engineering by examining the number of times that journals were cited by the articles published in three American chemical engineering journals (i.e., *Chemical Engineering*, *Chemical Engineering Progress*, and *Industrial & Engineering Chemistry*) during 1957. The most cited journals were ranked in descending order by the number of citations received, and tables were included that showed the countries of origin for the cited journals and the time periods in which they were cited. Burton stated in his article that he hoped this information would "prove useful for selecting general periodicals in the engineering fields involved, in formulating retention and binding policies, and in suggesting something of the information gathering habits of engineers in America" (Burton 1959, 70). Other studies have also applied a similar methodology to list the most cited journals in chemical engineering (Barrett and Barrett 1957, Patterson 1945).

While the use of citation analysis for collection development has been reported in the literature since the early twentieth century (Gross and Gross 1927), analyzing citation data to identify the most comprehensive database in a discipline is a relatively new idea that has not yet been applied to the field of chemical engineering. During the 1990s, the members of the Nursing and Allied Health Resources Section (NAHRS) of the Medical Library Association developed a protocol for conducting citation analysis studies (which became known as "mapping studies") of individual disciplines in nursing and allied health to increase librarians' understanding of the published literature, and how to access it, in these areas (Delwiche, Schloman, and Allen 2010). The protocol involves selecting one to five core journals in the discipline being investigated, examining the references in these journals for three years, and checking the coverage of these journals in a selection of databases appropriate to the specialty. One of the aims of these mapping studies is "to determine the extent to which coverage of the contents of the most highly cited journals in a given specialty/discipline is provided by major bibliographic indexes" (Delwiche, Schloman, and Allen 2010, 4). NAHRS members have utilized this protocol in over 35 published studies investigating the characteristics of the literature in nursing and allied health disciplines (e.g., emergency nursing, medical-surgical nursing, dietetics, occupational therapy, physical therapy, etc.) since 1997. In addition to the NAHRS group, Kawasaki (2002) updated an already published list of core journals in agricultural sciences, which was created “through citation analysis and other bibliometric techniques” (p. 33) to compare coverage of agricultural topics in the *Agricola*, *Biological Abstracts (BIOSIS), CAB Abstracts*, and *Biological & Agricultural Index Plus* databases.

Building on the research mentioned above, this paper makes use of citation analysis to create a list of cited journals in chemical engineering that is used, in turn, to compare the coverage of the chemical engineering literature across the following four databases: *Compendex*, *Inspec*, *Scopus*, and *Web of Science*.

Methodology

The author completed the following steps to find the most comprehensive database in chemical engineering (see also Figure 1):

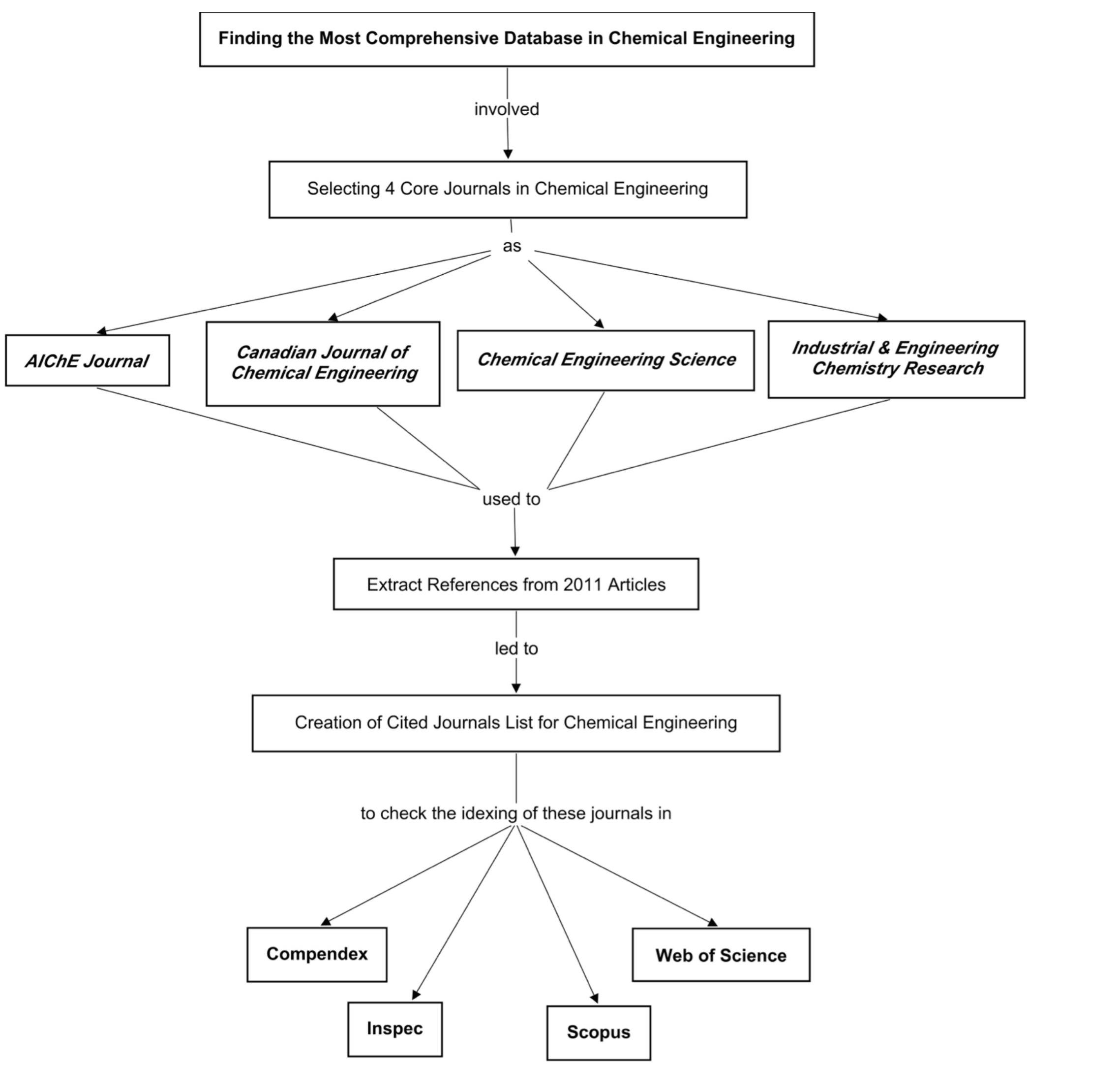


Figure 1. Concept map of methodology

1) Select four core journals in chemical engineering as the source of your data.

The NAHRS protocol suggests selecting one to five core journals, depending on the discipline, that provide a comprehensive coverage of the field. Choosing specialty journals or too few journals may lead to an incorrect or very narrow representation of citation patterns, whereas picking too many journals may cause the results to be scattered in various directions (Delwiche, Schloman, and Allen 2010). To guide the selection of the core journals for this study, the author conducted a literature search on serials in chemical engineering (Burton 1959, Peters, Hartmann, and Van Raan 1988, Modak and Madras 2008, Prathap 2011, Yin 2011), examined journal articles by McGill University's Chemical Engineering faculty to identify the most popular titles in which to publish, and obtained informal feedback from a chemical engineering professor at McGill. From the common information collected by these three methods, the following long-standing, core, chemical engineering journals were chosen:

- *AIChE Journal (*published by the American Institute of Chemical Engineers and Wiley-Blackwell; started in 1955)
- *Canadian Journal of Chemical Engineering* (published by Wiley-Blackwell on behalf of the Canadian Society for Chemical Engineering; started as the *Canadian Journal of Research* in 1929)
- *Chemical Engineering Science* (published by Elsevier; started in 1951)
- *Industrial & Engineering Chemistry Research* (published by the American Chemical Society; started 1909 under a different title; underwent multiple variations in title over the years)

2) Extract the references that were cited in the articles published in these core journals over a one year period.

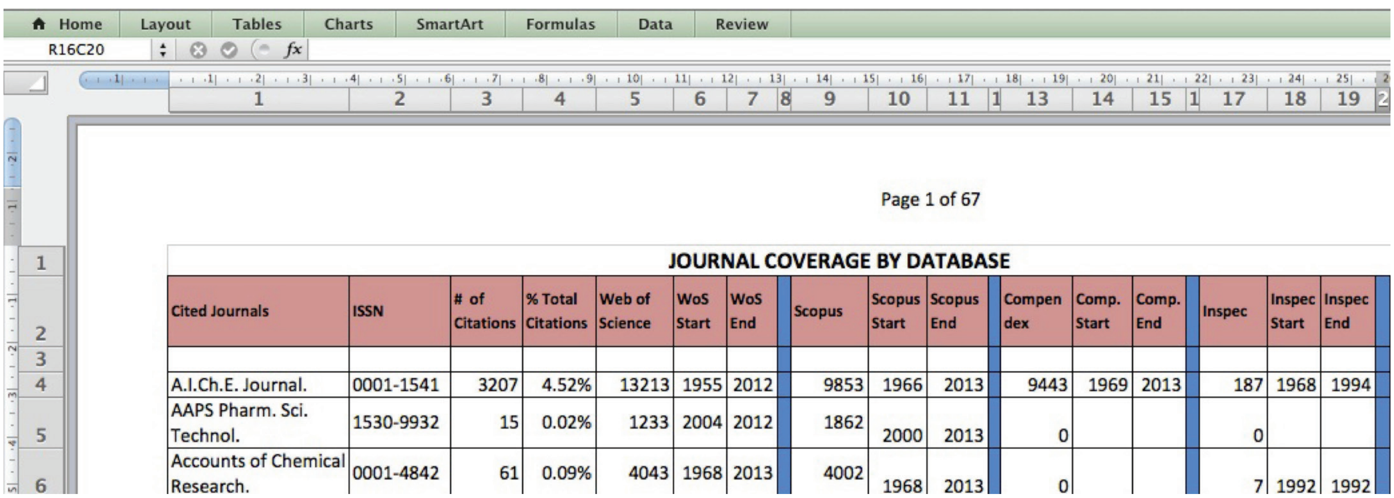
The cited references of 2011 articles published in the journal, *Industrial & Engineering Chemistry Research*, were exported directly from the journal publisher's website to EndNote. *Scopus* was used to extract the cited references for the other three journals, since they did not have this exporting capability on their websites. A different search was conducted in *Scopus* for each journal name and was limited to articles published in 2011. The author used the "View Cited By…" option in *Scopus* to obtain a list of cited references for the search results, and then exported these cited references to EndNote. A total of 90,945 cited references were obtained.

3) Create a cited journals list from your extracted references.

The author sorted the references in EndNote by reference type and excluded all items that were not journal articles, as well as manually checked the remaining references to remove any books or other items that were mislabeled as a journal article. An existing output style in EndNote was then modified to display only the journal names for any given reference. The "Copy Formatted" command in EndNote was employed to copy the journal names from EndNote and paste them into an Excel document. Finally, the "PivotTable" command in Excel was used to automatically count the occurrences of a journal name in the document and create a list with the journal names and the number of times each journal was cited. The author had to manually find and merge the full and abbreviated versions of a journal title to manage duplicate entries in the list. Journals cited less than five times were deleted from the list. The final cited journals list consisted of 970 journal titles.

4) Check the indexing of cited journals across databases that cover the chemical engineering literature.

The *Compendex*, *Inspec*, *Scopus*, and *Web of Science* databases were picked for comparison due to their inclusion of references to the chemical engineering literature and their accessibility to the author through institutional subscriptions. All the cited journals were searched by title or ISSN in each of the four databases. The journal ISSNs were found by using the “Sources” index in *Scopus* and the *CAS Source Index (CASSI) Search Tool*. Both the title or ISSN of the journal were used to search if no record was found using only one search strategy. The number of records, date of the earliest record, and date of the most recent record in each database was recorded in an Excel spreadsheet for each journal title (see Figure 2).



JOURNAL COVERAGE BY DATABASE

Cited Journals	ISSN	# of Citations	% Total Citations	Web of Science	WoS Start	WoS End	Scopus	Scopus Start	Scopus End	Compen dex	Comp. Start	Comp. End	Inspec	Inspec Start	Inspec End
A.I.Ch.E. Journal.	0001-1541	3207	4.52%	13213	1955	2012	9853	1966	2013	9443	1969	2013	187	1968	1994
AAPS Pharm. Sci. Technol.	1530-9932	15	0.02%	1233	2004	2012	1862	2000	2013	0			0		
Accounts of Chemical Research.	0001-4842	61	0.09%	4043	1968	2013	4002	1968	2013	0			7	1992	1992

Figure 2. Screenshot of a section of the author’s cited journals list in Excel

Note that this step is a very time-consuming process if there are a large number of journals to be checked which, in turn, may limit the number of databases that can be compared at a given time. In this instance, the author took approximately 100 hours to search 970 journals across four databases and record the pertinent data in Excel. The author plans to compare additional subject-specific databases that search the chemical engineering literature (e.g., *SciFinder*) in future studies.

5) Sort and compare the information found.

Columns of Excel data were sorted to compare the database coverage of these cited journals in chemical engineering. For example, the column that contained the number of search results for each journal in *Web of Science* (see the fourth column in Figure 2) was sorted by its smallest to largest values in order to quickly determine how many journals were not indexed in

Web of Science. Non-indexing was indicated with a value of zero. Excel's "MAX" formula function was also used to contrast the number of records found in the different database columns in order to quickly identify the database with the highest number of records for an individual journal title.

Results

All together, the articles published in the year 2011 in the *AIChE Journal*, *Canadian Journal of Chemical Engineering*, *Chemical Engineering Science*, and *Industrial & Engineering Chemistry Research* cited a total of 90,945 publications, out of which 90.5% were references to journal articles. 970 journal titles (from 86.3% of the cited journal article references) constituted the list of cited journals that was used to check database coverage. Journal titles that were cited four times or less were excluded from the list. The average number of times cited per journal was 73. The top three frequently cited journals were three of the four core chemical engineering titles chosen as the source of data for this study, i.e., *Industrial & Engineering Chemistry Research* (cited 5873 times, 8.3% of the total times cited for all the journals on the list), C*hemical Engineering Science* (cited 5068 times, 7.1% of the total), and the *AIChE Journal* (cited 3207 times, 4.5% of the total). The fourth core title, *Canadian Journal of Chemical Engineering*, ranked 26th on the list (cited 480 times, 0.7% of the total). The selection of the *Canadian Journal of Chemical Engineering* as a core journal was more strongly influenced by chemical engineering faculty at McGill University, located in Canada. Its lower ranking may indicate that this title has more local, rather than international, influence.

The author compared the indexing of the 970 cited journals in *Compendex*, *Inspec*, *Scopus*, and *Web of Science* by answering the questions below.

Which database indexes the maximum number of cited journal titles? *Scopus* and *Web of Science* index 934 titles (96.3%) and 927 (95.6%) respectively (see Figure 3). Note that for a journal to be counted as being indexed by a database, the database needed to provide references to more than 30 items published in the journal for over more than a one year period.

Database Name	Number of cited journals indexed	Percentage of 970 cited journals
Compendex	788	81.2%
Inspec	518	53.4%
Scopus	934	96.3%
Web of Science	927	95.6%

Figure 3. Database coverage of cited journals

Which database indexes the largest number of items from these journals? *Web of Scienc*e indexes the largest number of items for the most number of titles on the cited journals list (see Figure 4). Compared to the other databases being investigated, *Web of Science* provides equal or more coverage of 603 titles (62.2%).

Database Name	**Number of cited journals with the max. records**	**Percentage of 970 cited journals**
Compendex	152	15.7%
Inspec	24	2.5%
Scopus	517	53.3%
Web of Science	603	62.2%

Figure 4. Database coverage by largest number of items indexed within the most number of cited journals

Which database goes back the furthest in time in the indexing of these journals? *Scopus* provides references to articles published in the earliest year for the most journals (662 titles, 68.2%) on the list (see Figure 5). However, *Scopus* has gaps in the coverage of many of its journals for the period before 1996, which have been reported in the published literature (for example, see Goodman and Deis 2007). The indexing of a journal may start in 1951, stop, then continue from 1960 to the present. The author recorded the periods of coverage of the cited journals found in *Scopus* from its "Sources" index, and found that 101 journals out of the 662 titles, going back the furthest in time, had gaps in their time coverage. This means that *Scopus* provides uninterrupted time coverage for only 561 journals going back to the earliest year (57.8%), which is comparable to the coverage in *Web of Science* (559 journals, 57.6%).

Database Name	**Number of cited journals going back furthest**	**Percentage of 970 cited journals**
Compendex	309	31.9%
Inspec	203	20.9%
Scopus	662 561 (uninterrupted time coverage)	68.2% 57.8% (uninterrupted time coverage)
Web of Science	559	57.6%

Figure 5. Database coverage by cited journals going back the furthest

Which database currently indexes the greatest number of cited journals? *Scopus* currently indexes 818 of these journals (84.3%), closely followed by *Web of Science* (795 journals, 82%).

Database Name	**Number of cited journals currently indexed**	**Percentage of 970 cited journals**
Compendex	593	61.1%
Inspec	389	40.1%
Scopus	818	84.3%
Web of Science	795	82.0%

Figure 6. Current database coverage of cited journals

The Z Test Calculator for 2 Population Proportions (Stangroom 2013), applying a two-tailed hypothesis, was employed to analyze the differences between the two leading databases in answer to the questions listed above. The difference in the amount of coverage between *Scopus*, the leader, and *Web of Science*, the runner-up, was shown to be statistically insignificant for indexing:

- the most cited journal titles
  (0.7% difference, z-score = 0.8041, p-value = 0.42372, result is not significant at $p < 0.01$)
- the oldest time period
  (0.2% difference, z-score = 0.0919, p-value = 0.92828, result is not significant at $p < 0.01$)
- the greatest number of cited journals up to the present time
  (2.3% difference, z-score = 1.3949, p-value = 0.16452, result is not significant at $p < 0.01$)

Scopus and *Web of Science* can be considered tied in these categories.

However, *Web of Science* has a statistically significant lead over the runner-up, *Scopus* (8.9% difference, z-score = 3.9526, p-value = 8E-05, result is significant at $p < 0.01$), for indexing the largest number of items for the most journal titles. Due to the latter tie-breaker, *Web of Science* would be the most comprehensive database to use for searching chemical engineering topics compared to *Compendex*, *Inspec,* and *Scopus*.

Using Citation Analysis to Determine the Premier Database in a Specific Discipline

The superior coverage of the chemical engineering literature in the multidisciplinary databases, *Web of Science* and *Scopus*, over the subject-specific databases, *Compendex* and *Inspec*, is logical when you consider that chemical engineering is a multidisciplinary field. The references that were cited in the articles published in the year 2011 in the four core chemical engineering journals were to publications in various subject areas, i.e., chemical engineering, chemistry, math, physics, the health sciences, etc.

The author was initially concerned that selecting only four core chemical engineering journals as the source of this study's data would not accurately capture the multidisciplinary nature of the field. These concerns proved to be unfounded. The cited references extracted were multidisciplinary in nature since the core journals selected are broad in scope. To revalidate this in a different way, the author also extracted the cited article references found in papers published by McGill University's chemical engineering faculty in 2011. These 20 chemical engineering professors altogether published 90 articles in 60 different journals that produced 1,984 references to 129 journals. Journals cited less than five times were not counted. 115 out of the 129 journals (89.1%) were the same titles that were on the author's cited journals list.

As demonstrated thus far, comparing the indexing of cited journal titles can be successfully used to determine the premier database for a specific discipline, including multidisciplinary fields. In addition to the indexing of the chemical engineering journals that is presented here, this method of citation analysis has also been applied to database coverage of the literature in the fields of agriculture, and nursing and allied health (Kawasaki 2002, Delwiche, Schloman, and

Allen 2010). Following the steps in the "Methodology" section of this paper should help you to identify the most comprehensive database in any given discipline. While the use of citation analysis is a relatively straightforward process, it can be labor-intensive if the list of cited journal titles used for searching in the multiple databases is long. To be efficient, the cited journals list can be sorted in Excel from the most times cited to the least cited so that you can check database coverage of the most frequently cited journals first. If you see a statistically significant difference in coverage between the databases when searching the top 20% of frequently cited journals on the list, which according to Bradford's law of distribution (Bradford 1985) will account for most of the total number of citations on your list, you may decide not to search the remaining titles on the list and, thus, eliminate many hours from the process.

Searching the top 200 frequently cited journals (20.6 %) on the author's list included 82.2% of the total number of citations on the list. The results reflected the general trend previously seen in Figure 3. *Scopus* indexes all 200 top journals (100% coverage), *Web of Science* covers 198 titles (99% coverage), *Compendex* indexes 190 journals (95% coverage), and *Inspec* provides references to 122 journal titles (61% coverage). The difference between *Compendex* and *Web of Science*/*Scopus* decreases when examining the most frequently cited journals. *Compendex* is smaller in size and scope than *Web of Science*/*Scopus*, but its coverage of the vast majority of frequently cited journal titles (95%) indicates that it is a powerful, subject-specific database in engineering that is not to be discounted.

Discussion and Conclusions

This paper employed citation analysis to compare the coverage of chemical engineering journals across four databases. *Web of Science* was identified as providing more comprehensive coverage of the chemical engineering literature than *Compendex*, *Inspec*, and *Scopus*. When confronted with searching only one of these four databases, the study's results indicate that *Web of Science* is the best source to choose for a one-stop search. This finding may interest librarians who are searching, and/or teaching others how to search, for information on chemical engineering topics. The description of this study's methodology may also be helpful to librarians who seek to identify the premier database in a discipline, whether for efficient searching purposes and/or to make the best financial decisions regarding the purchase/cancellation of database subscriptions.

Further research in this area would include: 1) comparing the indexing of cited journals for other subject-specific databases that cover the chemical engineering literature, such as *SciFinder*; 2) exploring more closely the depth of journal coverage in these databases by examining whether there are gaps in time for the cited journals indexed in all these databases, not just *Scopus*; and 3) performing the same test search across all of these databases to compare the precision and recall of search results.

While the quantity of literature coverage is important for determining the comprehensiveness of a database, it is just one criteria that individuals may be mindful of when choosing where to search first. Ease of use and specific database features may also play an important role in the person's decision. Ease of use is comparable across the databases investigated in this study since *Compendex* and *Inspec* (both available via the Engineering Village search platform at the author's institution), *Web of Science*, and *Scopus* have very similar search interfaces, with multiple search boxes available and the ability to do a keyword search using Boolean operators, proximity operators, exact phrases, and truncation.

There are additional features available in some of these databases that make them more advantageous to search in certain situations. For instance, *Web of Science* and *Scopus* list the references included in the bibliography of an article and link to references that cite the article (note that these are only for articles published since 1996 in *Scopus*), as well as allow you to perform a search within the cited references field. Citation analysis options are limited in *Compendex* and not available in *Inspec*. Engineering Village lists the number of times cited in *Scopus* for those articles in *Compendex* that are also indexed in *Scopus* and links to the citing references in *Scopus*. While all four databases have set their default search options to include a search for the user's terms in the author keywords field, *Compendex*, *Inspec*, and *Scopus* also search for the user's terms in the controlled vocabulary field. Scopus contains subject headings in its records that were obtained from various thesauri that Elsevier owns or licenses (e.g., the Ei thesaurus from *Compendex*), and there are subject headings for 80% of indexed publications in *Scopus* (Sciverse Scopus 2012). *Compendex* and *Inspec* contain their own thesauri and allow subject searching in addition to keyword searching in their databases.

Based on the maximum quantity of literature coverage, ease of use, and additional features in this database (such as cited reference searching), the author would suggest using *Web of Science* as the starting point when searching for chemical engineering topics, whether looking for a few articles or undertaking an in-depth review of the literature. If too many search results are retrieved, then a subject heading search could be performed in *Compendex* to obtain the most relevant articles on a topic and reduce the number of false hits. If not enough search results are obtained, either a thesaurus search in *Compendex* or a keyword search in *Scopus* (which will include searching within the controlled vocabulary field) may help to expand an individual's search strategy by providing additional search terms that the person may use. For a exhaustive search or systematic review of the literature, multiple databases (including but not limited to these four databases) would need to be searched since no single database covers all of the chemical engineering literature.

Endnotes

Bar-Ilan, J. 2008. "Informetrics at the beginning of the 21st century - A review." *Journal of Informetrics* no. 2 (1):1-52.

Barrett, R.L., and M.A. Barrett. 1957. "Journals most cited by chemists and chemical engineers." *Journal of Chemical Education* no. 34:35-38.

Bradford, S.C. 1985. "Sources of information on specific subjects (Reprinted from Engineering an Illustrated Weekly Journal, Vol 137, Pg 85-86, 1934)." *Journal of Information Science* no. 10 (4):176-180.

Burton, R.E. 1959. "Citations in American engineering journals I. Chemical engineering." *American Documentation* no. 10 (1):70-73.

Delwiche, F.A., B.F. Schloman, and M.P. Allen. 2010. Mapping the literature of nursing and allied health professions: Project protocol. Nursing and Allied Health Resources Section of the Medical Library Association.

Goodman, D., and L. Deis. 2007. "Update on Scopus and Web of Science." *The Charleston Advisor* no. 8 (3):15-18.

Gross, P.L.K., and E.M. Gross. 1927. "College libraries and chemical education." *Science* no. 66 (1713):385-389.

Kawasaki, J.L. 2002. "Indexing of core agriculture serials." *Quarterly Bulletin of the International Association of Agricultural Information Specialists* no. 47 (2):33-37.

Modak, J.M., and G. Madras. 2008. "Scientometric analysis of chemical engineering publications." *Current Science* no. 94 (10):1265-1272.

Patterson, A.M. 1945. "Literature references in *Industrial and Engineering Chemistry* for 1939." *Journal of Chemical Education* no. 22 (10):514-515.

Peters, H.P.F., D. Hartmann, and A.F.J. Van Raan. 1988. "Monitoring advances in chemical engineering." *Informetrics* no. 87/88:175-195.

Prathap, G. 2011. "Correlation between h-index, Eigenfactor™ and Article Influence™ of chemical engineering journals." *Current Science* no. 100 (9):1276.

Sciverse Scopus. 2013. *Content coverage guide*. Elsevier B.V. 2012 [cited May 3 2013]. Available from http://files.sciverse.com/documents/pdf/ContentCoverageGuide-jan-2013.pdf

Stangroom, J. 2013. *Z Test Calculator for 2 Population Proportions* 2013 [cited May 3 2013]. Available from http://www.socscistatistics.com/tests/ztest/Default.aspx

Yin, C.Y. 2011. "Do impact factor, h-index and Eigenfactor™ of chemical engineering journals correlate well with each other and indicate the journals' influence and prestige?" *Current Science* no. 100 (5):648-653.