4-2014

Chemical Literature Databases: Conflict of Interest?

Belinda L. Hurley
Ohio State University, hurley.50@osu.edu

Title: Chemical Literature Databases: Conflict of Interest?

Author: Belinda L. Hurley, Assistant Professor, Ohio State University Libraries, The Ohio State University, hurley.50@osu.edu

Abstract:

A publisher of a research database controls the search algorithms of its database and, at a minimum, partially controls the indexing metadata attached to journal articles indexed in its database. A publisher of a journal partially controls the indexing metadata attached to articles in that journal. A publisher who publishes both a research database and journals that are indexed in that database has significant control over two major aspects of the discovery process. This allows the possibility for a publisher to introduce a bias into its algorithm that could favor discovery of its own articles. This work looks at search results from three scientific databases and discusses this possible conflict of interest.

Keywords:

database, chemistry database, bias, SciFinder, Scopus, discovery

Introduction

Many articles have been written that compare and contrast the primary databases used by chemical researchers (Baykoucheva 2011; Gavel & Iselid 2007; Grey et al. 2012; Li et al. 2010). These and other articles have considered similarities, differences, advantages, and disadvantages in databases in areas such as citation analysis, h-factors, number of journals indexed, number of articles retrieved, user friendliness and much more. One aspect of chemical information retrieval for which there is a lack of discussion is the possible conflict of interest that arises when the publisher of scientific journals is also the publisher of a scientific database. Does such a publisher intentionally (or unintentionally) establish algorithms within its database that steer users toward its own publications? Even if one had access to the proprietary details of a database's algorithms, can such a question be answered with any degree of certainty? Discovery through a database depends on both the algorithms of the database and the metadata attached to each journal article. A publisher of a database and a journal indexed in that database obviously has significant control over both ends of this process and could introduce a bias that favors discovery of its own publications. This paper attempts to raise an awareness of this possible conflict of interest.

First and Foremost, Awareness of the Issue

Because this paper is unique in the idea it explores, a note on the underlying message is appropriate. Unlike much of the literature related to research databases, this paper is not meant to evaluate how a given research database works (its algorithms), nor is it about the options a database presents to users, nor about the results that could be obtained if the database were used in a proper (or even moderately proper) manner. The data presented

below look at the results that a typical user would obtain during a typical search and how those results might vary from database to database. The intent of this paper is not to reexamine other papers' studies on the mechanics of using databases. Search results and whether those results show any evidence of publisher bias is the sole focus. Certainly, the algorithms and options presented by a research database have a significant effect on the results obtained. However, it is no secret that users of databases display less than desirable habits when searching for literature and database providers are most definitely aware of user habits. In short, the results obtained by a typical user are pertinent to this work; how or why those results are obtained is not pertinent to this work, if one accepts that the majority of database users approach searches in a similar (less than ideal) manner. The foremost aim of this work is to bring attention to a possible conflict of interest and in doing so it looks at the results that a typical user *would* obtain, rather than the results a librarian would obtain or the results librarians believe users could be obtaining. There is ample library science literature indicating that the typical user puts forth as little effort as possible when performing a search in a research database. In fact, as Markey (2007) specifically noted in her review of literature related to end-user searching habits, "[m]ost end users accept the IR system's default values" (Bishop et al. 2000; Cooper 2001; Jones et al. 2000; Markey 2007). This holds not only for less experienced searchers, but also for more experienced searchers. Jones et al. (2000) gathered results from over 30,000 queries over a 61-week period using transaction log analysis of searches performed in the New Zealand Digital Library interface. Similar to most database interfaces, this user-friendly interface provides users with several query options in the default search screen and additional query options in the advanced search screen. Jones et al.

concentrated their analysis on searches of Computer Science Technical Reports with the assumption that the "computer science research community could be thought of as 'best case' users … given their familiarity with software and Boolean logic." Their analysis revealed that over two thirds of users used the default settings "as is," making no changes and taking no advantage of additional search options. In fact, approximately halfway through their study, the researchers changed the default options to see if such a change would have any effect on users' preferences. The change had no effect; users performed searches with the new set of default values at the same rate as they had with the previous set of default values (over two-thirds of searches). In addition to accepting the default values with which they are presented, the majority of users consistently take the shortest route when completing other search steps. Jones et al. (2000) also noted that "queries tend to be short and simple." Markey (2007) states that "less than 20% of queries in end-user searches bear Boolean operators," "[t]runcation occurs in less than 5% of end-user searches," and "[l]ess than 10% of user queries enlist relevance feedback." In their study of the information seeking behavior of engineering and philosophy graduate students, Korobili et al. (2011) noted that "students … did not invest time and effort in using complex tools in their research process." In fact, "[a] significant percentage (42.3%) … never or very seldom modified the initial statement if the results were not satisfactory."

It is, therefore, safe to assume that the majority of searches performed in databases are performed as noted above, using the shortest available route. This is not news to librarians and, more important with regard to this work, this is not news to the providers of research databases.

The above background information on typical user habits is presented as it is paramount to the fundamental purpose of this work. What algorithms a database provider chooses, how those choices determine outcome of results obtained, or why a provider chooses those algorithms is not at issue here. The issue being explored is whether those algorithms might produce results that are biased in favor of the publisher associated with the database provider. It should be noted that bias is not necessarily a pejorative concept. Indeed, those searching for chemical literature rely on a bias that provides a systematic distortion (versus a random distortion) and, therefore, a bias can exist for perfectly sound and valid reasons. In other words, there exists a reproducible difference in the results of an examination versus the results that theoretically should occur for that examination. Of course, the definition of "should occur" is open to a wide interpretation, thereby making it difficult to make a claim of the intentional introduction of a bias. Therefore, this work does not address whether a research database provider intentionally (or unintentionally) introduces a bias favoring related publications. It simply encourages a mindfulness of the possibility of a bias. Essentially, it is of value to know of the existence of a bias or the possibility that a bias could arise in a given situation, especially if a user (or librarian) had not considered the possibility.

The following data present an initial look at some results related to this issue. If, in future discussions on and investigations into this matter, such a bias was established for a given database, a database publisher could (and should) defend its algorithms, however, there is no specific reason to believe that they could not reasonably justify the workings of their chosen algorithms. If there exists a bias in a research database, it could be that a provider is intentionally introducing a bias to steer users to its related publications, however, it also could

be that a database provider has found that the users of its database have a preference for a type of publication that is more likely to be aligned with one of their publications. The list of reasons why a bias might be preferable (or non-preferable), the reasons why a set of algorithms is chosen, or the reasons why a given algorithm might produce a set of results is a long list, but those issues are, appropriately, not discussed here. Instead, it is hoped that this work can help to start a discussion on whether database bias of this nature is worth considering.

The following takes a look at the possibility of the introduction of a bias in research databases by analyzing the results obtained from literature searches within Scopus, SciFinder ("SF"), and Thomson Reuters' Web of Science, Science Citation Index Expanded ("WOS").

A Look at Some Relevant Data

The data presented below were gathered in a rigorous, but necessarily limited fashion and provide an initial glimpse of data relevant to the question at hand. In fact, because of the limited amount of data collected, these overall results must be classified as non-rigorous and publishers of these databases and journals can rightfully state that these limited results cannot establish the existence or non-existence of bias. This work required that each of the journal articles in nine separate searches (over 5,000 journal articles) be assigned to a specific publisher. For the three databases analyzed, information provided by the database and related to the identification of the publisher associated with each journal article is restricted to journal titles. Although programming could be established to transform a list of publications into a corresponding list of the publishers of those publications, doing so is beyond the scope of this work. Therefore, in this work the number of searches and results that could be analyzed was limited by the amount of time and effort one could reasonably put forth in manually assigning a

publisher to each result. Unless and until database publishers provide user-friendly access to metadata that directly link the publisher of a journal title to each individual result or programming is established to complete this analysis, the process of establishing such data on the analysis of a large number of searches over a wide variety of topics will remain time prohibitive. Without doubt, any rigorous analysis intended to firmly establish the existence of some type of bias would require a much larger number of searches to be statistically sound and could still be open to doubt due to the many user-determined variables that occur during any search.

Methodology

Three distinctly different searches (three "search sets" or Search #1, Search #2 and Search #3) were performed in each of the three databases, for a total of nine searches. Although the different interfaces make it impossible to enter the search information in an identical manner, terms and limiters were chosen to replicate the process as identically as possible across the databases and care was taken to minimize the user input to best replicate the anticipated actions of a typical user, based on the assumptions listed above related to user habits. As few changes as possible were made to the default values and these changes were made only in an attempt to make searches in all three databases as similar as possible with as little input as possible (and also, in some cases, to minimize the number of results). The search terms were chosen from three distinctly different areas of chemical research to allow for analysis that takes into account bias that might be introduced due to the focus of each database. In their own words, Scopus focuses on "scientific, medical, technical, and social science" literature (Boyle & Sherman 2005), SF focuses on "chemistry and related science

information" (CAS 2014), and WOS focuses on "scientific and technical journals across 100 disciplines" (Thomson Reuters 2012). Search # 1 involved the terms "polymers" and "intrinsic porosity," Search #2 involved the terms "chiral catalysis" and "oxidation," and Search #3 involved the terms "cell death" and "nanoparticles." For easy reference, Table 1 presents the specific parameters used for each of the nine total searches. Each of the nine searches was limited to journal or review articles written in English over the entire date range available within each database (hereafter, journal articles and review articles are collectively referred to as "articles" or "results"). As noted in Table 1, searches performed in Scopus were restricted to the subject areas of life sciences, health sciences and physical sciences and searches performed in WOS were restricted to Science Citation Index Expanded. In all searches performed in SF, duplicates were removed and only results "containing both of the concepts" were analyzed. Although Scopus and WOS, by default, provide users with the AND operator, the unique nature of the SF interface does not do so. The word "and" was used in the search string entered in SF simply in an effort to mimic the most likely user method of entering the desired search terms. This choice was made based on the search habits of the majority of users (discussed above) combined with Jones et al.'s (2000) findings indicating that although the majority of searchers do not use Boolean operators, "and" is the most commonly occurring query term. Jones et al. explain this by indicating that in most cases, "and" constitutes a stopword rather than the intentional use of a Boolean operator. In the SF searches, "and" serves as a stopword in the initial search entry. When users are then presented with SF's mandatory "Research Topics Candidates" screen, the choice of "containing both of the concepts" effectively serves as the AND operator and brings the SF searches into reasonable

alignment with searches performed in Scopus and WOS. With the exception of the above-noted parameters, all other parameters were left in the default setting for each database.

The full set of results of each of the nine searches were sorted by journal title and each title was assigned to either its publisher (Elsevier, American Chemical Society ("ACS"), Wiley/Blackwell ("Wiley"), or Royal Society of Chemistry ("RSC")) or, for titles not published by any of those four publishers, a category denoted as "Other." In addition to analysis of the full set of results of each search, the first 100 results from Searches #2 and #3 were analyzed in the same manner as the full set of results. Search #1 produced fewer than 100 results in each of the databases, and, therefore, only the full sets of the results of Search #1 were analyzed. After each result was assigned to a publisher, the percentage of the results of each search attributed to a given publisher was plotted. (Note: The plots and tables indicate the percentage of results, not the percentage of titles.)

Data

The analyses of the full sets of results are represented graphically in Figures 1-3. The analyses of the limited sets of results are represented graphically in Figures 4 and 5. Table 2 presents the numerical data from which the plots were created.

Search #1. All three of the databases produced fewer than 100 results in Search #1 (polymers and intrinsic porosity) and all three databases produced the highest percentage of results from Elsevier publications, with Scopus producing 35.4%, SF producing 24.7%, and WOS producing 27.1% of their results from Elsevier publications. The percentage of results from ACS publications was only slightly less in the SF and WOS searches (21.2% and 22.9%, respectively), however, the percentage of results from ACS publications in the Scopus search was significantly

less at 13.5%. The percentage of results from Wiley and RSC publications were more similar across the three databases than those of Elsevier and ACS publications.

Search #2, Full Results. Search #2 (chiral catalysis and oxidation) produced the highest number of total results in all three databases. Scopus produced 1023 results, SF produced 2161 results, and WOS produced 496 results. The distributions across Scopus and SF are remarkably similar. The percentages of publications from Elsevier are 29.3% in Scopus and 29.7% in SF, with WOS presenting a similar percentage of 29.3%. Likewise, the percentage of publications from ACS only ranges from 28.4% in SF to 28.5% in Scopus, however, WOS presented a notable difference with only 15.5% of its results arising from ACS publications. Another notable difference is the percentage of results from WOS attributed to Wiley publications. At 25.6%, this percentage is significantly larger than both Scopus (17.1%) and SF (14.6%).

Search #3, Full Results. The distribution of results in Search 3 (cell death and nanoparticles) was markedly different than that in Searches #1 and #2, with the "Other" category containing the largest percentage of results from all three databases. Likewise, the second largest percentage arose from Elsevier publications and a very small percentage of results arose from RSC publications in all three databases. Results from ACS and Wiley publications were similar to each other both within each database and across the three databases with a very limited range of 11.4% to 13.8%. The total number of results in this search fell in between those of Searches 1 and 2 for all three databases.

Search #2, Top 100 Results. The distribution of the top 100 results for Search #2 (chiral catalysis and oxidation) was notably altered from that of the distribution of the full set of results for the same search. Additionally, each of the databases displayed unique distributions

for these results. The percentage of results from Elsevier publications from all three databases was over 10% less for the top 100 results versus the full set of results. In contrast, the percentage of results from Wiley publications from all three databases was substantially greater for the top 100 results versus the full set of results. Interestingly, the percentage of results from Scopus and WOS for ACS publications was roughly the same for the top 100 results and the full set, but the percentage of results from SF for ACS publications was more than 5% less for the top 100 results.

Search #3, Top 100 Results. The distribution of the top 100 results versus the full set of results for Search #3 was somewhat the same, although it is perhaps noteworthy that the percentage of Elsevier publications dropped in both SF and WOS, but climbed over 12% in Scopus in moving from the full set of results to the top 100 results.

Discussion

Excluding even the specific nature of a research topic, there are still many factors that can influence the results returned in a science literature search. Optimizing a database for breadth can often lead to less than optimal results for specificity and vice versa. Finding the right balance while providing a user-friendly interface, that does not require excessive input from users is, no doubt, a challenge for any database vendor. Users and librarians have many aspects to take into consideration when choosing a research database. The consequences of making an appropriate or an inappropriate choice as they pertain to research undertaken in a typical chemistry department at a research university and specifically the possibility of the presence of a bias favoring a given publisher contained within the search algorithms of a database are, therefore, of importance.

As noted above, the results of a literature search are highly dependent on both the algorithms of the chosen database and how those algorithms interact with the metadata attached to a journal article. Obviously, a publisher of academic journals in given areas of research who also owns a major database used for research in those same areas possesses significant control over both ends of the literature discovery process. This is the case for both Chemical Abstract Services' SciFinder and Elsevier's Scopus. As a division of the American Chemical Society (ACS), Chemical Abstract Services has an inseparable relationship with the publications of the ACS indexed in SF. Likewise, there exists a similar relationship between Scopus and Elsevier publications. On the other hand, Thomas Reuters does not directly publish scientific academic journals, and therefore, its Web of Science has no direct relationship with the journals it indexes.

It would, however, be inaccurate to use the term "control" to imply that WOS serves as a control database in the data presented herein. Likewise, one cannot imply that journals other than ACS and Elsevier journals (RSC, Wiley and others) can serve as "control" journals. Making these assumptions implies that there is some type of measurable standard in WOS that is common to both SF and Scopus and a similar assumption with regard to journals. The inclusion of WOS, however, can serve as a quasi-marker of the results that might be expected from a database provider that has no associated publications. Although there are many commonalities across the databases, the different date ranges, different foci, different indexing of titles and many other factors will certainly have a complex effect on the results. Therefore, is it virtually impossible to establish a control database and, likewise, impossible to directly compare and contrast results from various databases as they relate to the possible existence of the bias

described above. The above data, however, can be used as examples to show some differences and similarities between science databases (as used for chemistry research) and to establish an awareness of possible database bias.

The subjects chosen for each of the three searches involved relatively recent research and, in fact, although the date range for each of the databases varied (at the time of this work the ranges were: SF 1905-present, Scopus 1960-present, and WOS 1900-present), all the results obtained in all the searches were dated post 1962 as noted in Table 3.

Although some journal titles are indexed for all years of publication in each of the three databases, many journal titles have different indexing start dates (and possibly end dates) in each of the three databases (due to either the choice of indexing dates by a database or the given range of all titles in the database). Overlap and the presence of unique titles within the three databases is difficult to establish and is typically dependent on the database provider's information, which is often less than explicit (Gavel & Iselid 2007; Gluck 1990). It can be said with certainty though that the different providers are likely to place a difference emphasis on various subfields, as indicated by their own descriptions of coverage. As previously noted, coverage is just one of many factors that could lead to different results even if all other parameters could somehow be held constant. In many ways coverage can be categorized with the various algorithms of the three databases. The details of the algorithms and coverage are not at issue for this work, however--only the results as they pertain to the question at hand. Coverage, however, is arguably one of the most likely factors to influence results in a manner that might imply bias and, therefore, it is mentioned here.

It should also be noted that the results from each of the three searches revealed that in not a few cases Scopus retrieved articles from ACS publications that were not retrieved by SF and, vice versa, in not a few cases SF retrieved articles from Elsevier publications that were not retrieved by Scopus.

The immediately above discussion about different factors that can affect search results is presented to place an emphasis on the valid complexity of the specific results obtained from a given database and to avoid the false misinterpretation of the following discussion as a claim that all or the majority of differences in results arise from artificial database bias originating from intentional manipulation by the database providers and publishers.

With the above information in mind, an analysis of the data is presented here.

Of all three search sets, the results of Search #1 (polymers and intrinsic porosity) presented the least consistent distribution of Elsevier and ACS publications between the Scopus results and the SF results. If one wanted to make a case for database bias these results might add to that argument. Although the results from SF and WOS are remarkably similar, the results from Scopus consist of a significantly higher percentage of Elsevier publications than the percentage of Elsevier publications resulting from the SF search and the WOS search (35.4% in Scopus versus 24.7% in SF and 27.1% in WOS). Additionally, the results from Scopus show only 13.5% from ACS publications whereas SF and WOS show 21.2% and 22.9%, respectively, of their results from ACS publications. Two things should be taken into consideration with these data, however. First, all three databases returned less than 100 results from this search. Care should be exercised in drawing strong conclusions due to the low number of results in this search set. Not only did this search set return the least number of results, the results from this set covered

the most expansive time range of all three search sets. Therefore, these data were fewer and spread out over a longer time. These two factors strongly imply a scarcity of results relative to the other search sets and, therefore, less statistical reliability for this search set. On the other hand, looking only at the results within this search set, the similarity in the percent of "Other" publications from all three databases (28.1%, 29.4%, 29.2%) suggests a consistency that is not present in the results from Elsevier and ACS publications.

Search Set #2 (chiral catalysis and oxidation) produced remarkably similar distributions from Scopus and SF, but a somewhat different distribution from WOS. The fact that all three databases produced the highest number of results from Elsevier publications with almost identical percentages (29.3%, 29.7%, and 29.2% respectively) supports the validity of Elsevier being the top publisher (quantitatively) for research related to these specific search terms. With the exception of the Elsevier results and, to a lesser extent, the RSC results, the distribution produced by WOS, however, is quite different than the similar distributions produced by Scopus and SF. This difference could, in fact, be the result of a different search algorithm within WOS or it could originate in the fewer number of results (and thus, statistically weaker data) from WOS versus Scopus and SF (n=496, 1023, and 2161, respectively). Overall, a claim of publisher bias by either Scopus or SciFinder, based strictly on the distributions in this search set, is not supported and, in fact, these data point toward the lack of any bias.

Of note within this search set is the large discrepancy in the number of results returned by SF (2161) versus Scopus (1023). As noted above, the results from Scopus for Search Set #2 range from 1977-present and those from SciFinder range from 1970-present. It should be noted that only 10 of the 2161 results from SciFinder are dated between 1970 and 1977. Over

their respective full time ranges, Scopus returned 292 results from 15 ACS publications and 300 results from 51 Elsevier publications while SF returned 610 results from 27 ACS publications and 641 results from 54 Elsevier publications. In other words, SF returned slightly more than twice as many results as Scopus and these extra results were fairly evenly distributed over the various publishers. This suggests that for whatever reason(s), SF's algorithm is picking up more papers with these search terms from Elsevier journals than Scopus is picking up from Elsevier journals (and, likewise, more from ACS journals, too). In fact, SF produced 90 results from the title *Tetrahedron* (Elsevier) with a date range of 1984-present and Scopus produced only 36 results from *Tetrahedron* with a date range of 1991-present. Only eight of the SF results were dated pre-1991, thereby precluding the assumption that the extra results are due to the longer date range. This also suggests a fairly even distribution of extra *Tetrahedron* articles throughout the SF results. Similar results were returned from the two databases from *Tetrahedron Letters* (also Elsevier). Therefore, whatever difference(s) exist in SF's algorithms in this search versus Scopus' algorithms, those differences do not appear to produce biased results.

Conversely, one can consider the ACS results of Search Set #2. As noted above, although the percentage of ACS journal articles retrieved from SF and Scopus are virtually the same, SF returned slightly over twice as many articles from ACS publications as Scopus. As in the case of the Elsevier titles, this doubling was fairly consistent across the ACS titles and can be seen in Table 4. Overall no distinct bias can be attributed to the SF results because although SF did return more than twice the total number of results as Scopus, the "extra" results appear to be distributed evenly among the various publishers.

Of course, no user of a database is likely to peruse 2161 results or even 1023 results. Mansourian and Ford's (2007) work on searchers' perceptions of missing important information quotes a user as saying, "there are only a certain number of articles you can look at when you go through a PubMed search." Holman's (2011) study on search strategies found "students skimmed search results quickly, rarely looking beyond the first two pages." Jansen and Spink (2003) report that in a typical web search over 80% of end users view three or fewer results from three or fewer pages of results. It is not unreasonable to assume that a user of a scientific research database might be willing to spend a bit more time and effort perusing a larger number of results. However, these data regarding a typical web search support the idea that the user of a scientific database is likely to peruse only a limited number of results. Arguably, the articles most likely to be viewed by a user engaged in the above searches would be the articles that appear at the top of the results screen. Therefore, the top 100 results from each of the three databases were analyzed from Search #2 and from Search #3. There are a variety of ways a user could limit his search results, however, in keeping with the simplest route, no additional limiters were employed; the top 100 results were simply taken from the full sets of results. The top 100 results, by default, are the most recent (chronological) 100 articles in the case of Scopus and WOS. SF returns results based on accession number which roughly translates to the most recent 100 articles (with very limited exceptions for results indexed in other than chronological order).

Unlike the remarkably similar distributions of results in Scopus and SF for the full sets of results in Search #2, the top 100 results in Search #2 in these two databases displayed some notable differences in their distribution over the various publishers. When pared down to 100,

the largest percentage of results was no longer attributable to Elsevier publications as had been the case in all three databases for the full sets. In fact, Scopus returned the most results from ACS journals while SF returned the most from Wiley journals and WOS returned the most from "other" journals. Notably, however, the *ratio* of Elsevier results to ACS results is virtually identical for Scopus and SF (0.70 for both). Considering the limited number of results analyzed, it is to be expected that these comparisons will not be as statistically rigorous; however, the similarity in these ratios strongly suggests the absence of a bias.

The full set of results from Search Set #3 (cell death and nanoparticles) presented database distributions very different from those of Search Sets #1 and #2. This was not unexpected as the search terms, although still related to chemical research, are also closely related to medical literature. As such, a search related to these terms would be expected to tap into more traditional medical journals in addition to traditional chemistry journals. Indeed, each of the three databases returned the most results from "other" publications. The distribution of all results across publishers, however, was fairly similar across all three databases. The largest deviation from these similarities was the number of articles from Elsevier publications returned by Scopus (33.6%), which was notably higher than the percentage of articles from Elsevier publications returned by SF and WOS (26.9% and 26.7%, respectively). With Scopus returning 742 total results, SF returning 603 results and WOS returning 490 results, it is reasonably fair to say that the number of results were high enough to provide a reliable comparison.

The top 100 (chronological) results for Search Set #3, were also analyzed. The higher percentage of articles from Elsevier publications returned by Scopus in the full sets of results

became significantly higher than the percentage of articles from Elsevier publications returned by SF and WOS when looking at the sets of 100 results. Scopus returned over twice as many Elsevier articles as SF (46 versus 19) and almost twice as many as WOS (46 versus 24). All of the top 100 results returned by Scopus were from 2011, however, only 66 of the results returned by SF were from 2011 (31 were from 2010, 2 from 2009, and 1 from 2008). These numbers from the top 100 results combined with the slightly higher percentage of Elsevier publications in the full set somewhat support that Scopus seems to be providing a higher percentage of results from Elsevier publications on medical literature searches than other databases.

Conclusions

The following conclusions must be interpreted with an understanding that the limited data from these three different sets of search terms cannot rigorously establish the existence or non-existence of publisher bias.

1. In general, these data do not indicate that the publishers of Scopus and SF provide results with a bias that favors their respective related publications.

2. Because different database publishers have different foci , users should be aware of the alignment (or misalignment) of their research focus with a given database focus, and especially be aware of the possibility that such alignment might tend more toward one publisher than another, which could, in turn, bias their expected results.

3. When perusing a small set of results from a scientific database search, users should be especially aware of the possibility that the results might not be representative of the true distribution of relevant literature across publishers.

4. An additional source of possible misrepresentation of the true distribution of relevant literature across publishers on a given topic is the date range of the database and/or the date range of coverage of various titles. The importance of this factor can vary depending on the time-frame during which research on the chosen topic has been undertaken by the research community and should be taken into consideration.

References

Baykoucheva, S. 2011. Comparison of the contributions of CAPLUS and MEDLINE to the performance of SciFinder in retrieving the drug literature. *Issues in Science and Technology Librarianship* 66. [Internet]. [Cited April 8, 2014]. Available from: http://www.istl.org/11-summer/refereed1.html

Bishop A.P., Neumann L.J., Star S.L., Merkel C., Ignacio E. & Sandusky R.J. 2000. Digital libraries: Situating use in changing information infrastructure. *Journal of the American Society for Information Science* 51(4):394-413. DOI: 10.1002/(SICI)1097-4571(2000)51:4<394::AID-ASI8>3.0.CO;2-Q

Boyle F. & Sherman D. 2005. Scopus™: The product and its development. *The Serials Librarian* 49(3):147-153. DOI:10.1300/J123v49n03_12

CAS, a Division of the American Chemical Society. 2014. SciFinder - The choice for chemical research.™ [Internet]. [Cited March 20, 2014]. Available from: http://www.cas.org/products/scifinder.

Cooper, M.D. 2001. Usage patterns of a web-based library catalog. *Journal of the American Society for Information Science and Technology* 52(2):137-148. http://editlib.org/p/90564/

Gavel, Y. & Iselid, L. 2007. Web of Science and Scopus: a journal title overlap study. *Online Information Review* 23(1):8-21. DOI 10.1108/14684520810865958

Gluck, M. 1990. A review of journal coverage overlap with an extension to the definition of overlap. *Journal of the American Society for Information Science* 41(1):43-60. DOI: 10.1002/(SICI)1097-4571(199001)41:1<43::AID-ASI4>3.0.CO;2-P

Grey J.E., Hamilton M.C., Hauser A., Janz M.M., Peters J.P. & Taggart F. 2012. Scholarish: Google Scholar and its value to the sciences. *Issues in Science and Technology Librarianship* 70. [Internet]. [Cited November 28, 2012]. Available from: http://www.istl.org/12-summer/article1.html

Jones S., Cunningham S.J., McNab R. & Boddie S. 2000. A transaction log analysis of a digital library. *International Journal on Digital Libraries* 3(2):152. DOI: 10.1007/s007999900022

Li J., Burnham J.F., Lemley T. & Birtton R.M. 2010. Citation Analysis: Comparison of Web of Science, Scopus, SciFinder, and Google Scholar. *Journal of Electronic Resources in Medical Libraries* 7:196-217. DOI:10.1080/15424065.2010.505518

Markey, K. 2007. Twenty-five years of end-user searching, part 1: Research findings. *Journal of the American Society for Information Science and Technology* 58(8):1071-1081. DOI: 10.1002/asi.20462

Thomson Reuters. Science Citation Index. 2012. [Internet]. [Cited December 27, 2012]. Available from: http://thomsonreuters.com/products_services/science/science_products/a-z/science_citation_index/.

TABLES AND FIGURES

Table 1 Search Parameters

Table 2 Numerical Results

Table 3 Date Ranges of Results

Table 4 ACS Title Results, Search #2

Figure 1 Search #1, Full Results

Figure 2 Search #2, Full Results

Figure 3 Search #3, Full Results

Figure 4 Search #2, Top 100 Results

Figure 5 Search #3, Top 100 Results

Table 1 Search Parameters

		Scopus	SciFinder	Web of Science
Search 1	**Search Term(s)**	***polymers*** in Article Title, Abstract, Keywords AND ***intrinsic porosity*** in Article Title, Abstract, Keywords	Research Topic = ***polymers and intrinsic porosity***	***polymers*** in Topic AND ***intrinsic porosity*** in Topic
	# of Results	96	92†	48
Search 2	**Search Term(s)**	***chiral catalysis*** in Article Title, Abstract, Keywords AND ***oxidation*** in Article Title, Abstract, Keywords	Research Topic = ***chiral catalysis and oxidation***	***chiral catalysis*** in Topic AND ***oxidation*** in Topic
	# of Results	1023	2161†	496
Search 3	**Search Term(s)**	***cell death*** in Article Title, Abstract, Keywords AND ***nanoparticles*** in Article Title, Abstract, Keywords	Research Topic = ***cell death and nanoparticles***	***cell death*** in Topic AND ***nanoparticles*** in Topic
	# of Results	742	603†	490
	Limiters/Refiners*	Document Type = Article or Review Date Range = All years (1960-present) Language = English Subject Areas = Life Sciences, Health Sciences, Physical Sciences	Document Type = Journal, Review Publication Years = 1800-present Language = English	Document Type = Article, Review Timespan = All years (1899-present) Language = English Citation Database = Science Citation Index Expanded

***** These parameters were the same for all three searches.

† Duplicates Removed, References containing "both of the concepts"

Table 2 Numerical Results

Full Results

		% Elsevier	% ACS	% Wiley	% RSC	% Other
Search 1	Scopus (n=96)	35.4	13.5	14.6	8.3	28.1
	SciFinder (n=92)	24.7	21.2	18.8	5.9	29.4
	WOS (n=48)	27.1	22.9	14.6	6.2	29.2
Search 2	Scopus (n=1023)	29.3	28.5	17.1	7.7	17.3
	SciFinder (n=2161)	29.7	28.4	14.6	7.4	19.8
	WOS (n=496)	29.2	15.5	25.6	6.9	22.8
Search 3	Scopus (n=742)	33.6	12.0	13.5	1.3	39.6
	SciFinder (n=603)	26.9	11.6	11.4	1.5	48.6
	WOS (n=490)	26.7	13.8	12.2	2.4	44.8

Top 100 Results

	% Elsevier	% ACS	% Wiley	% RSC	% Other
Scopus (n=100)	19	27	24	9	21
SciFinder (n=100)	16	23	27	11	23
WOS (n=100)	18	14	29	8	31
Scopus (n=100)	46	5	13	4	32
SciFinder (n=100)	19	11	11	4	55
WOS (n=100)	24	11	16	8	41

Table 3 Date Ranges of Results

	Date Range of Results for Indicated Database		
Search Subjects	**Scopus**	**SF**	**WOS**
Polymers, Intrinsic Porosity	1962-	1968-	1992-
Chiral Catalysis, Oxidation	1977-	1970-	1992-
Cell Death, Nanoparticles	1998-	1993-	1998-

Table 4 ACS Title Results, Search #2

Sampling of Results by ACS Journal Title for Search Set #2		
ACS Journal Title	**Results from SF**	**Results from Scopus**
J of the American Chemical Society	225	113
J of Organic Chemistry	165	95
Organic Letters	93	39
Analytical Chemistry	2	1

Figure 1 Search #1, Full Results

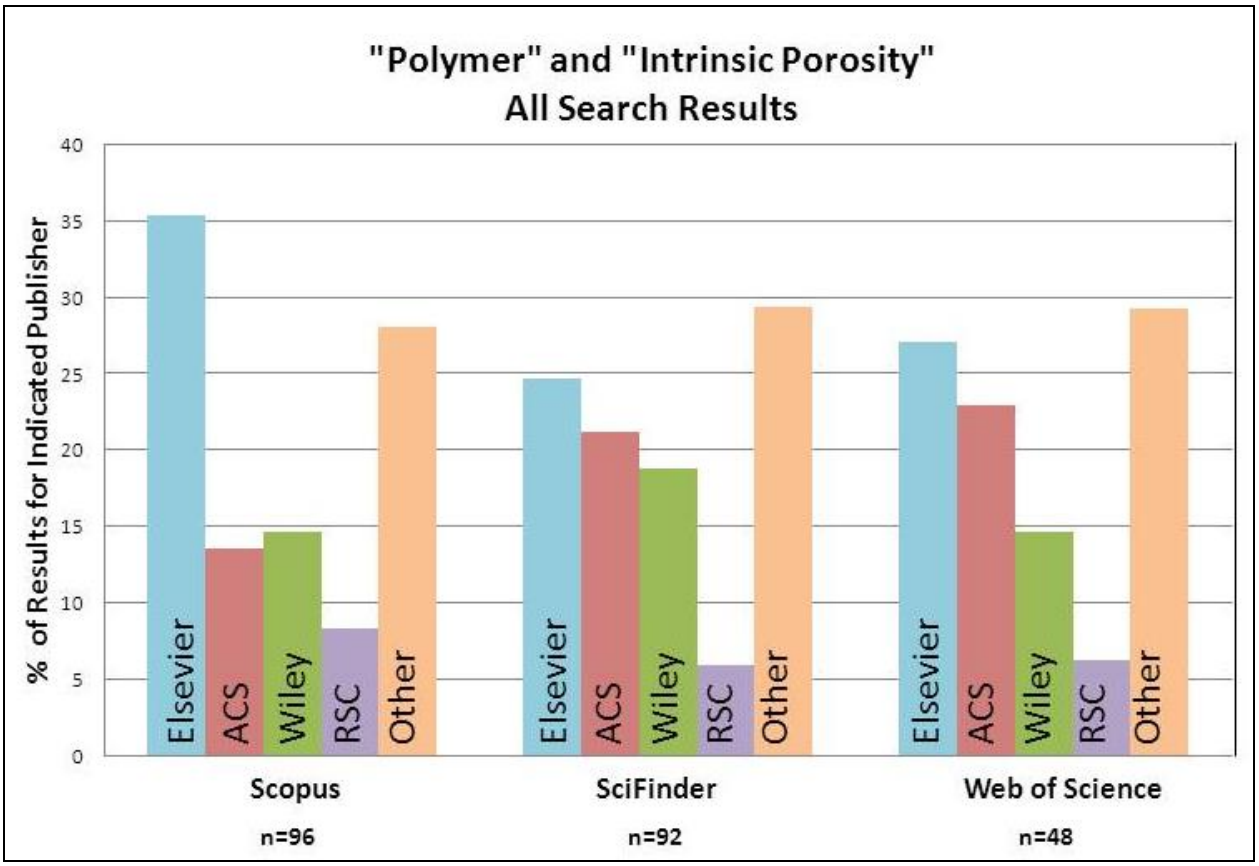

Figure 2 Search #2, Full Results

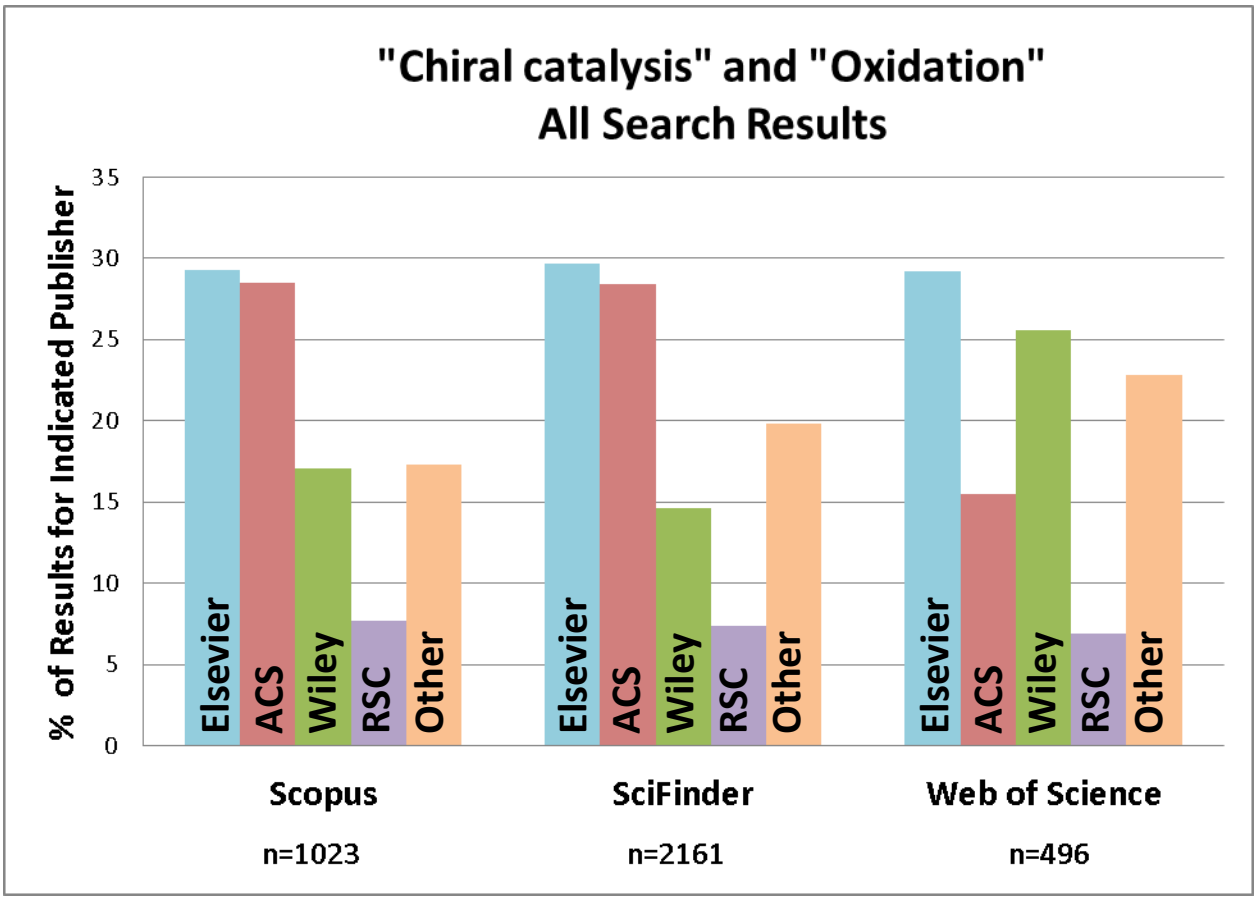

Figure 3 Search #3, Full Results

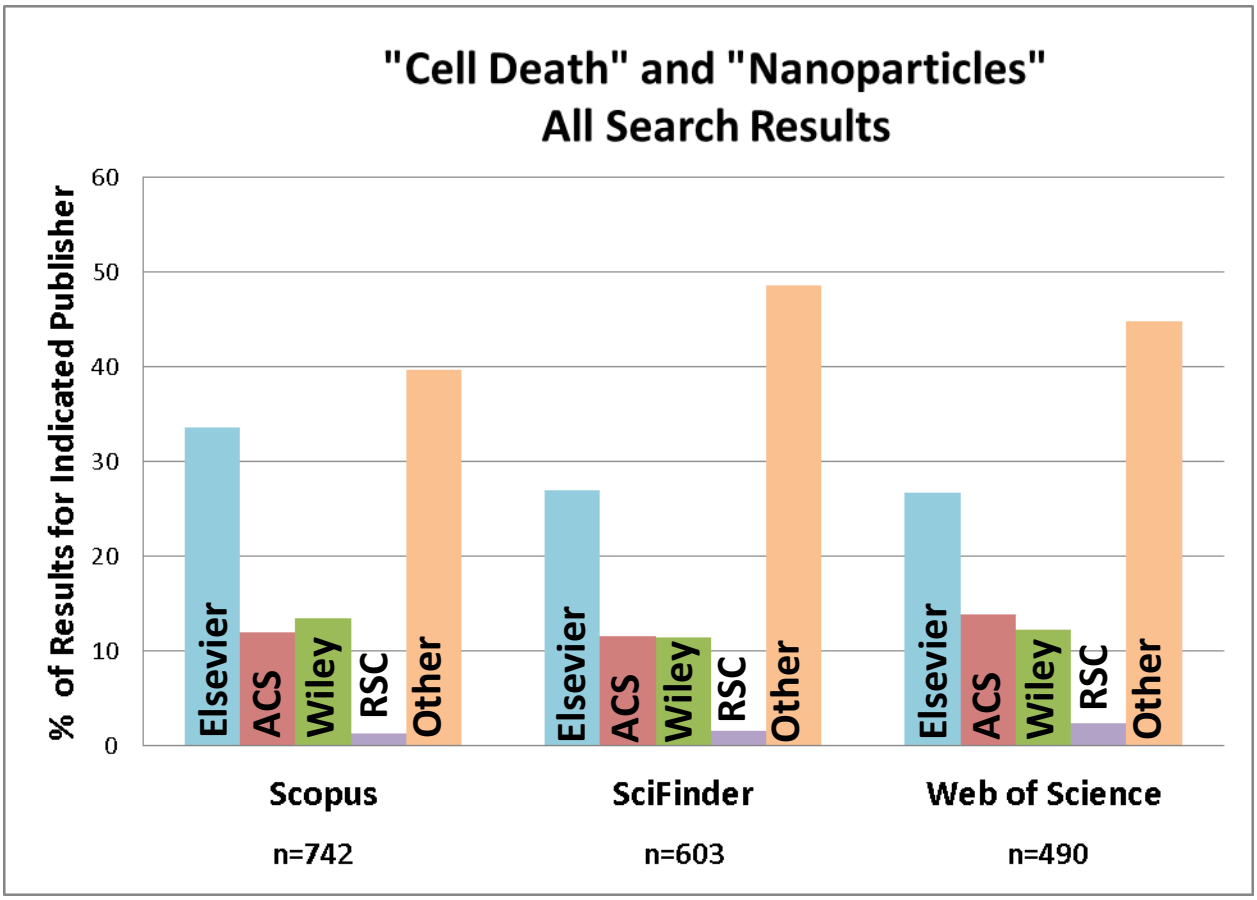

Figure 4 Search #2, Top 100 Results

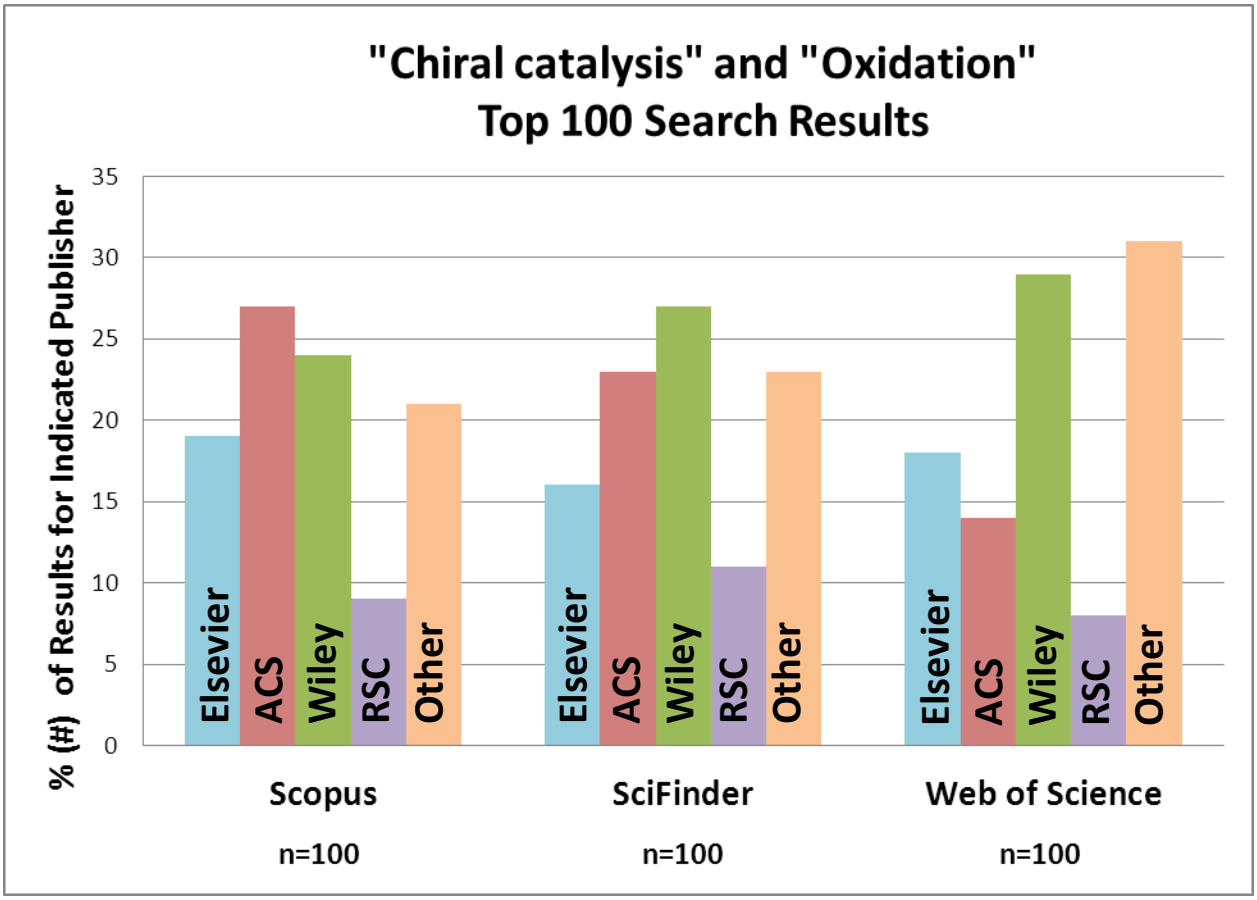

Figure 5 Search #3, Top 100 Results