Objective

Citation analysis projects constitute a significant portion of the research requests submitted to the Taubman Health Sciences Library (THSIL). Researchers and administrators are increasingly interested in citation analysis as a way to describe research performance or impact, with particular interest on the cited-by and h-index metrics. Several commercial products provide access to citation analysis metrics, but users are often unaware of these features and generally do not understand what the numbers really mean or how they are used in which they can be appropriately applied. The purpose of this study is to compare the citation analysis functions of Collexis Research Profiles, ISI Web of Science, and Scopus in terms of accuracy, coverage, and overall functionality.

Methodology

A random sample of 10 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan Library System Health System and conducted by a THSIL librarian. The 50 author names were run against each of the three designated databases: Collexis Research Profiles, ISI Web of Science, and Scopus using the specific search strategies detailed below. Citation analysis metrics for total number of publications, h-index and total cited-by citations were recorded. For the graphs displayed below, a further random sample of 10 names was selected.

Search Strategies:

Collexis Research Profiles:
• Using ‘Full Name’ search feature, enter authors’ last name
• Select appropriate author from list
• Record total number of publications and h-index from Researcher Profile
• Manually calculate cited-by counts from citation counts from Researcher Profile

ISI Web of Science:
• Using ‘Routable’ search field, enter author’s last name and first initial with an asterisk
• Further disambiguation is necessary, enter first and middle initial with an asterisk
• Select relevant records and select Citation Tracker feature
• Use the Create Citation Report tool to generate a citation report
• Record total number of publications, h-index, cited-by citations by citation counts

Scopus:
• Using Author search feature, enter authors’ last name and first initial
• Further disambiguation is necessary, enter full author name
• Select relevant records and select Citation Tracker feature
• Record total number of publications, h-index cited-by citations by citation counts

Limitations

Search strategies were not intended to be comprehensive or capture a definitive list of each author’s research output. Rather, the intent was to utilize search features within each resource for author disambiguation and compare the utility and accuracy of the results.

The different data sources represented in this study each have their own limitations. While none of these limitations represent fatal flaws in the respective products, they are valid concerns or considerations that any end user needs to consider when evaluating the product and its results.

Collexis Research Profiles:
• UM’s current instantiation includes only those authors with a primary appointment to clinical or research units at the University of Michigan Medical School (UMMS). While all of the sampled authors met this criteria, the resource is currently not comprehensive for all UM or WMF faculty.
• UM’s current instantiation is manually curated to resolve author name disambiguation issues and validated against faculty卡尔目录Ia for minor modifications to increase publication count accuracy. However, both tools and cited-by counts are currently based on ISI Web of Science MEDLINE subset, which does not represent full UM 1ENG.EL.
• There is no way to manually refine a name search within the Research Profiles.
• Cited-by counts are not retrieved and need to be manually calculated.

ISI Web of Science:
• Citation info using the ‘District Author Set’ tool is markedly smaller than citation info using author last name and first initial. Many do not appear to have been imported into the provided author sets.
• Some common names/set are impossible to disambiguate using the ‘District Author Set’ tool. Full first name and last name are not interchangeable, making the disambiguation process for common names even more difficult.

Scopus:
• For name disambiguation issues, a user can select only up to 15 names. This resulted in the exclusion of many individual publications linked to an author’s name record.
• Unclear algorithm for calculating h-index using different database features. In several instances, a different h-index was obtained from the same set of articles due to an unexplained difference in the number of articles used to calculate h-index.
• A query to Scopus support was still open at the time the poster was prepared.
• During the period of data collection, the Scopus database presented the following error message: “Scopus is currently experiencing issues with the set of results and display of search numbers. We aim to correct this processing error as soon as possible. Please forgive the inconvenience.” Communication with the Scopus support indicated that the problem was fixed but the message provided throughout the data collection period.

Conclusion

The citation analysis functions for each of the three tools have unique strengths and weaknesses. Clearly, author name disambiguation continues to pose a problem, particularly in Scopus and ISI Web of Science. UM’s instantiation of the Collexis Research Profiles goes a long way toward addressing this issue through manual curation, but this is a time- and resource-intensive solution that may not be appropriate for all institutions. Collexis Research Profiles, however, uses only a subset of citations from ISI Web of Science to calculate h-index and citation metrics. While all three products provide readily available citation statistics clear through their reporting features, end users should be aware of the limitations of coverage and accuracy when using this information.

Acknowledgements

*Funding provided by the Michigan Institute for Clinical Health Research, NIH Grant UL1RR024986.
**Funding provided by the National Center for Integrative Biomedical Informatics, NIH Grant U54DA031519.