



RESEARCH ARTICLE

Scopus 1900–2020: Growth in articles, abstracts, countries, fields, and journals

Mike Thelwall  and Pardeep Sud 

University of Wolverhampton, Wolverhampton, UK

an open access  journal



Citation: Thelwall, M., & Sud, P. (2022). Scopus 1900–2020: Growth in articles, abstracts, countries, fields, and journals. *Quantitative Science Studies*, 3(1), 37–50. https://doi.org/10.1162/qss_a_00177

DOI:
https://doi.org/10.1162/qss_a_00177

Peer Review:
https://publons.com/publon/10.1162/qss_a_00177

Received: 19 October 2021
Accepted: 14 December 2021

Corresponding Author:
Mike Thelwall
m.thelwall@wlv.ac.uk

Handling Editor:
Ludo Waltman

Copyright: © 2022 Mike Thelwall and Pardeep Sud. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: academic publishing, academic publishing trends, scholarly databases, Scopus

ABSTRACT

Scientometric research often relies on large-scale bibliometric databases of academic journal articles. Long-term and longitudinal research can be affected if the composition of a database varies over time, and text processing research can be affected if the percentage of articles with abstracts changes. This article therefore assesses changes in the magnitude of the coverage of a major citation index, Scopus, over 121 years from 1900. The results show sustained exponential growth from 1900, except for dips during both world wars, and with increased growth after 2004. Over the same period, the percentage of articles with 500+ character abstracts increased from 1% to 95%. The number of different journals in Scopus also increased exponentially, but slowing down from 2010, with the number of articles per journal being approximately constant until 1980, then tripling due to megajournals and online-only publishing. The breadth of Scopus, in terms of the number of narrow fields with substantial numbers of articles, simultaneously increased from one field having 1,000 articles in 1945 to 308 fields in 2020. Scopus's international character also radically changed from 68% of first authors from Germany and the United States in 1900 to just 17% in 2020, with China dominating (25%).

1. INTRODUCTION

Science is not static, with the number of active journals increasing at a rate of 3.3%–4.7% per year between 1900 and 1996 (Gu & Blackmore, 2016; Mabe & Amin, 2001). Bibliometric studies covering a substantial period need to choose a start year and be aware of changes and any anomalies during the time covered. Citations over a long period are needed in bibliometric studies of the evolution of journal (Jayaratne & Zwahlen, 2015), field (Pilkington & Meredith, 2009), author (Maflahi & Thelwall, 2021), or national (Fu & Ho, 2013; Luna-Morales, Collazo-Reyes et al., 2009) research impact over time. Unless constrained by a research question, the logical start year for bibliometric studies covering many years might be either the most recent date when there was a change in the character of the bibliometric database used, or the earliest year when sufficient articles were indexed according to some criteria. It is therefore useful to assess the temporal characteristics of bibliometric databases to aid decisions by researchers about when to start, particularly as some facets, including narrow fields, average citations and the presence of abstracts, are not currently straightforward to obtain from the web interfaces of citation indexes. This article focuses on one of the major citation indexes, Scopus.

Little is known about the historical coverage of the major citation indexes, other than the information reported by their owners. This typically gives overall totals rather than yearly breakdowns (e.g., Clarivate, 2021; Dimensions, 2021; Elsevier, 2021). Scopus currently has wider coverage of the academic literature than the Web of Science (WoS) and CrossRef open DOI-to-DOI citations, similar coverage to Dimensions, but much lower coverage than Google Scholar and Microsoft Academic (Martín-Martín, Thelwall et al., 2021; Singh, Singh et al., 2021; Thelwall, 2018). Lower coverage than Google Scholar and Microsoft Academic is a logical outcome of the standards that journals must meet to be indexed by Scopus (e.g., Baas, Schotten et al., 2020; Gasparyan & Kitas, 2021; Prancutè, 2021; Schotten, Meester et al., 2017) and WoS (Birkle, Pendlebury et al., 2020). Nevertheless, non-English journals seem to be underrepresented in both Scopus and WoS (Mongeon & Paul-Hus, 2016). One source of difference between WoS and Scopus is that WoS aims to generate a balanced set of journals to support the quality of citation data used for impact evaluations (Birkle et al., 2020). While a larger set of journals would be better for information retrieval, a more balanced set would help citation data that is field normalized or norm-referenced within its field (e.g., adding many rarely cited journals to a single field would push existing journals into higher journal impact factor quartiles and increase the field normalized citation score of cited articles in the existing journals). Even if two databases cover the same journals they can index different numbers of articles from them, due to errors or different rules for categorizing a document as an article (Liu, Huang, & Wang, 2021). Overall, while Dimensions provides the most free support for researchers (Herzog, Hook, & Konkiel, 2020), Scopus seems to be the largest quality-controlled citation index and also covers substantially more years than Dimensions or the WoS Core Collection: It is therefore a logical choice for long-term investigations. No study seems to have analyzed the historical coverage of any citation index, however, with the partial exception of the WoS Century of Science specialist offering (Wallace, Larivière, & Gingras, 2009).

Some date-specific information is known about Scopus. It was developed by Elsevier from 2002, released in 2004 (Schotten et al., 2017), and has since incorporated many articles from before its start date. In the absence of systematic evidence of Scopus coverage changes over time, Scopus-based studies needing long-term data have often chosen 1996 as a starting point in the originally correct belief (Li, Burnham et al., 2010) that there was a change in Scopus in this year (e.g., Budimir, Rahimeh et al., 2021; Subbotin & Aref, 2021; many Thelwall papers). In 2015, Scopus recognized 1996 as a watershed year for coverage and added 4 million earlier articles and associated references into the system (Beatty, 2015). Because of this update, 1996 may no longer be a critical year. The current article explores whether 1996 or any other year represents a shift in Scopus coverage and reports a selection of more fine-grained information to help researchers using Scopus for historical data, by allowing them to pick a starting year with sufficient data for their study.

The indexing of abstracts is also important. Abstracts in academic articles typically summarise the parts of an article, usually reusing sentences from the main body (Atanassova, Bertin, & Larivière, 2016). Some journals require a structured format, ensuring that background, methods, results, and implications are all covered in a simple format (Nakayama, Hirai et al., 2005). They are needed for studies that attempt to predict future citation counts (Stegehuis, Litvak, & Waltman, 2015), or to map the development of fields or their evolution based on the terms in article titles, abstracts, and keywords (e.g., Anwar, Bibi, & Ahmad, 2021; Blatt, 2009; Kallens & Dale, 2018; Porturas & Taylor, 2021). The proportion of articles with abstracts is also relevant for the scope of keyword-based literature searches that cover many decades (e.g., Sweileh, Al-Jabi et al., 2019), since the searches will be less effective for articles without abstracts, if these are more common in some years. Abstracts have been mainly

studied for their informational role (e.g., Jimenez, Avila et al., 2020; Jin, Duan et al., 2021) or writing style (Abdollahpour & Gholami, 2018; Kim & Lee, 2020).

Abstracts are known to have changed in format over time and individual journal policies have evolved. For instance, although Scopus has indexed *Landscape History* since 1979, the first abstract from this journal was in 1989 for the article, “Cairns and ‘cairn fields’; evidence of early agriculture on Cefn Bryn, Gower, West Glamorgan,” although this seemed to be an author innovation, starting their article with a short section entitled “Summary” rather than a journal-required or optional abstract. From browsing the journal, 1997 seems to be the year when abstracts were first mandatory, representing a policy change. In some fields, abstracts were published separately to articles in dedicated abstracting periodicals (e.g., *Biological Abstracts*) so that potential readers would have a single paper source to help them quickly scan the contents of multiple journals (Manzer, 1977). For example, early mathematics papers tended not to have abstracts, but very short summaries were instead posted by independent reviewers in publications such as *Zentralblatt MATH* (Teschke, Wegner, & Werner, 2011) and *Mathematical Reviews* (Price, 2017). Some journals also had sections dedicated to summarizing abstracts of other journals’ contents (e.g., Hollander, 1954). Despite the value and different uses of abstracts, no study seems to have assessed the historical prevalence or length of abstracts associated with articles in any major database.

The coverage of bibliometric databases is a separate issue to their citations, although the two are connected. Nothing is known about trends in average citation counts for Scopus, but a study of references in the WoS Century of Science 1900–2006 found an increasing number of citations per document, from less than 1 in 1900 to an arithmetic mean of 8 (Social Sciences), 10 (Natural Sciences and Engineering), and 22 (Medicine) in 2006, based on a 10-year citation window (Wallace et al., 2009). Changes over time in the types of journals cited by articles in the WoS also been investigated, showing reduced concentration (Larivière, Gingras, & Archambault, 2009).

Driven by the above issues, the goal of the current paper is to present a descriptive analysis of Scopus 1900–2020 in terms of the annual numbers of articles published as well as its field coverage, citation counts, and abstracts.

2. METHODS

2.1. Data

Documents in Scopus are assigned a type, such as book, trade journal article, or academic journal article. Of these, academic journal articles are the most relevant to research evaluation and bibliometrics and so other types were ignored. Documents are also usually assigned narrow and broad fields in Scopus, with 337 narrow fields being declared (Elsevier, 2021), although some are not used (e.g., 3699 Sports Science, 3323 Social Work, 2509 Nanotechnology). The records for all journal articles in Scopus were downloaded through its Applications Programming Interface (API) with narrow field queries because this is the easiest way to identify which narrow field articles belong to via the API. Queries were submitted in the following form, where 1213 is the narrow field code for Visual Arts and Performing Arts.

SUBJMAIN(1213) AND DOCTYPE(ar) AND SRCTYPE(j)

The query was also submitted with every other narrow field code and every year between 1900 and 2020 (sent as the API *year* query parameter). The queries were submitted at

substantially different time periods as part of an ongoing updating exercise to keep within Scopus usage restrictions and to avoid overloading the Scopus servers. The three batches used for this article were downloaded as follows.

- Article records for 1900–1995 were downloaded in September 2021.
- Article records for 1996–2013 were downloaded in November–December 2018.
- Article records for 2014–2020 were downloaded in January 2021.

The differences between download years should only influence the citation data, unless Scopus has substantially changed its 1996–2013 coverage after 2018, which seems unlikely. There may be minor changes due to new de-duplication algorithms or other improvements, however.

The data was checked for consistency by generating time series for the number of articles per year, per narrow field. Some gaps were identified due to software errors, and these were filled by redownloading the missing data for the narrow field and year within 2 months of the original download date.

2.2. Analysis

Purely descriptive data is presented, matching the purpose of this article. Since some bibliometric studies use article abstracts, statistics are reported for articles containing abstracts as well as for all articles.

It is not straightforward to identify whether an article has an abstract, so a rule was generated to estimate this. Some articles in Scopus have abstracts indexed as part of their record, although they may not always be called “abstract” in the published article (e.g., “Summary”). These abstracts typically include copyright statements and sometimes only a copyright statement is present in the abstract field. As in previous papers from the authors’ research group (e.g., Fairclough & Thelwall, 2021), a heuristically chosen 500 character minimum (about 80 words) was set as indicative of a reasonably substantial abstract that is unlikely to be purely a copyright statement.

As an indicator of field breadth, data is reported for the number of narrow fields containing a given number of articles. Since a field may be large due to a single journal, data is also reported for the number of fields containing a given number of journals, as a rough indicator of diversity of content (although individual megajournals can also have diverse content: Siler, Larivière, & Sugimoto, 2020).

Average citations per year are reported with both the traditional arithmetic mean and the more precise geometric mean for typical highly skewed citation count data (Fairclough & Thelwall, 2015; Thelwall, 2016). The citation count data is not symmetrical (e.g., equally distributed on either side of the mean) but is highly skewed: While most articles have zero or few citations, so that their citation counts are slightly less than the mean, the citation counts of a small number of highly cited articles are far greater than the mean (Price, 1976; Seglen, 1992). For example, the skewness is enormous at 107 for the 2004 citation counts and even larger for recent years (387 in 2020), whereas the skewness of the normal distribution is 0.

3. RESULTS AND DISCUSSION

The main results are introduced and discussed below. Additional graphs and brief discussion are in the online supplement and the full data behind all graphs is also online, both on FigShare at <https://doi.org/10.6084/m9.figshare.16834198>.

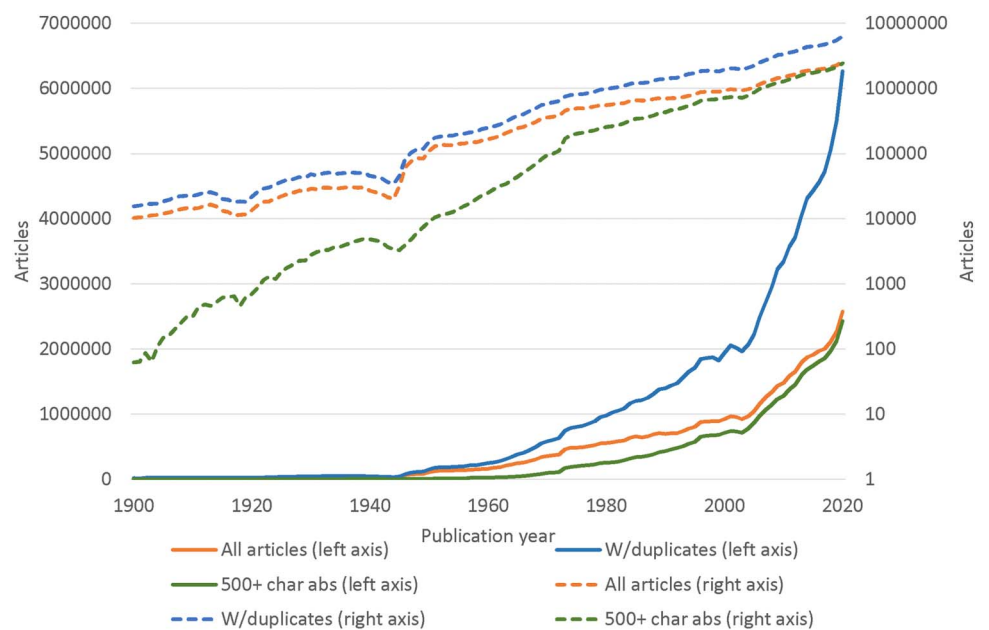


Figure 1. Number of documents of type journal article indexed in Scopus, as retrieved by its API. There are separate lines for all articles, all articles in each field (i.e., counting an article n times if it is in n narrow fields) and the number of articles with 500+ word abstracts.

3.1. Total Number of Articles

The number of articles in Scopus shows exponential growth from 1900 to at least 2020 (Figure 1). The extent to which the trend reflects the technical limitations of Scopus and its indexing policy rather than the amount of scholarly publishing is unclear because not all journals qualify for indexing (e.g., Mabe & Amin, 2001). The kink in the logarithmic line in the year that Scopus launched, 2004, suggests that its expansion accelerated more quickly after then. More specifically, the initial release in 2004 and subsequent backfilling projects were surpassed by subsequent expansions of additional journals. The graph from 1970 can be compared to the equivalent WoS volume of coverage in response to the DT=(Article) query. WoS does not have a kink in 2004, suggesting that this is a Scopus phenomenon (WoS has a similar exponentially increasing shape, with sudden increases in 1996, 2015, and 2019: See the online supplement for a graph).

Both world wars resulted in decreases in coverage, presumably due to many scientists and journal staff switching to unpublished military research or service (e.g., Hyland, 2017). Conditions were described as “extremely difficult” for journal publishing in the second world war (Anonymous, 1944) and there would also have been problems with international transport for printed journals. For example, the number of *Nature* articles per year indexed in Scopus decreased temporarily during both world wars. At the same time, war created the need for new types of research, leading to the emergence of new fields, such as occupational medicine (Smith, 2009) and operational research (Molinero, 1992), but this did not immediately translate into expanded academic publishing overall.

3.2. The Proportion of Articles with Abstracts

The proportion of articles with a substantial abstract of at least 500 characters has increased from 1% in 1900 to 95% in 2020 (Figure 2). A 500-character abstract contains about 80 words,

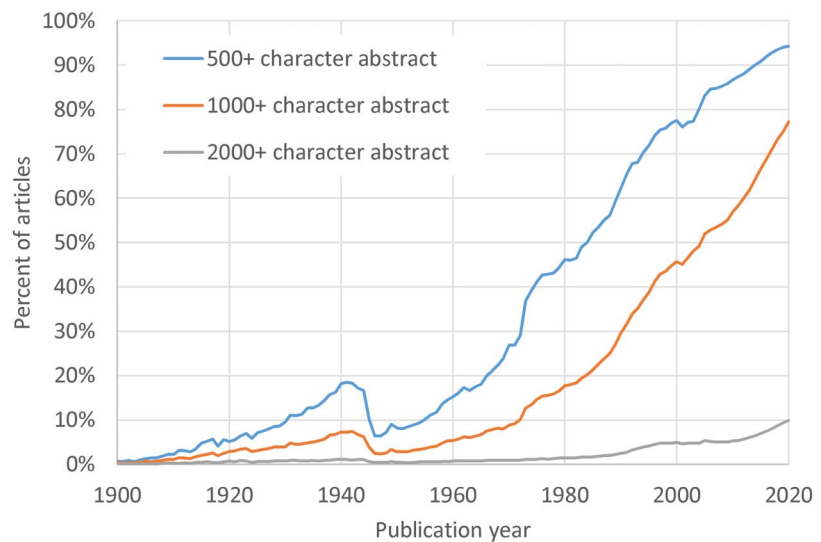


Figure 2. The percentage of Scopus journal articles with abstract length at least 500, 1,000, or 2,000 characters.

so is short but nontrivial even if the copyright statement is included, as the example below illustrates.

“© 2019 Brill Academic Publishers. All rights reserved. This paper presents the new and actually the first diplomatic publication of the unique 16th-century copy of the Church Slavonic Song of Songs translated from a Jewish original, most likely not the proper Masoretic Text but apparently its Old Yiddish translation. This Slavonic translation is extremely important for Judaic-Slavic relations in the context of literature and language contacts between Jews and Slavs in medieval Slavia Orthodoxa.” (Grishchenko, 2019).

A 1,000-character abstract has about 160 words and these longer abstracts have become increasingly common. In contrast, long abstracts with at least 2,000 characters and about 320 words are still rare, accounting for only 10% of articles in 2020.

The increasing percentage of articles with nontrivial abstracts presumably reflects their increasing necessity in scientific research due to their role in attracting readers (and hence citations for the publishing journal). The trend found here may also partly reflect Scopus ingesting early sources that omitted abstracts, although no evidence was found for this as a cause. In contrast, some of the few early abstracts indexed by Scopus were not part of the original article. For example, some early psychology articles (e.g., Pressey, 1917) had abstracts attached to them in Scopus that apparently originated from APA Psycnet (e.g., <https://doi.org/10.1037/h0070284>) and may have been extracted by PsycInfo from early psychology abstracting journals (e.g., *Psychological Abstracts*). Thus, the early results may partly reflect retrospective attempts to add abstracts. One early journal with genuine abstracts was the *Journal of the American Chemical Society*, which allowed articles to have a separate section at the end entitled *Summary*. While this could be interpreted as part of the article, it has a different heading format and could reasonably be classed as an abstract. At least one author conceived the summary as being separate from the article, stating, “The foregoing article may be summarized as follows:” (Clark, 1918).

In 2020 the median abstract length was 1,367 characters or 200 words. This median is presumably partly due to some journals having a 200-word abstract length limit in their

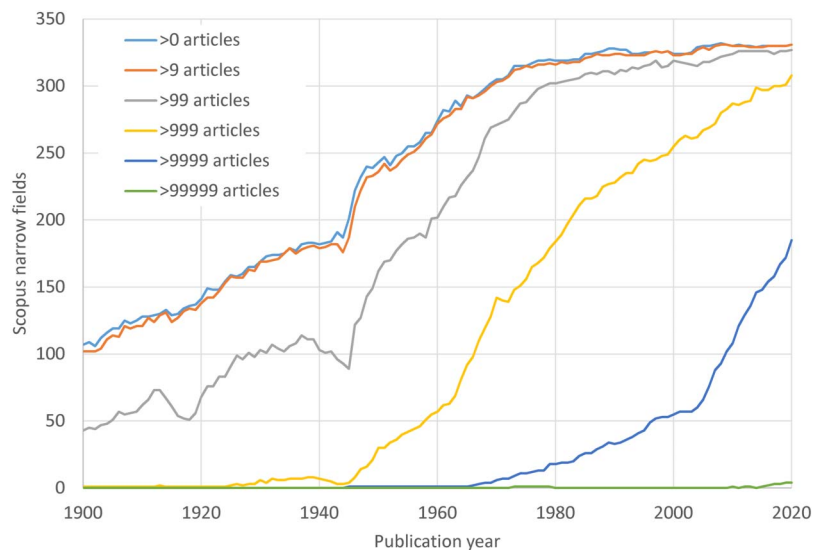


Figure 3. Number of Scopus narrow fields with specified minimum numbers of articles.

guidelines for authors (e.g., *Quantitative Science Studies*, *Nature Scientific Reports*, most Royal Society journals, many Wiley journals).

3.3. Narrow Field Coverage

Scopus has over 100 narrow fields with some articles for 1900, with the number of narrow fields increasing over time (Figure 3). The increasing shapes of the lines reflect Scopus narrow fields having uneven sizes, with most growing in size as the database grows overall. The number of narrow fields in Scopus is relevant for studies that attempt to present a broad picture of science. It is not clear whether the increasing number of substantial narrow fields reflects the greater coverage of Scopus or increased specialization in science, however. Thus, the analysis of long-term cross-science trends is a particularly difficult issue.

Almost all Scopus narrow fields include few journals (<10) until after the Second World War, when the number of narrow fields with at least 10 different journals increases from 25 (Figure 4). By 2020, most narrow fields included at least 100 different journals.

3.4. Number of Journals and Average Journal Size

Scopus indexed few journals in 1900, with growth starting after the Second World War or the end of the 1960s, if only articles with 500+ character abstracts are included (Figure 5). Surprisingly, the growth in the number of journals slowed and then stopped by 2020, perhaps due to the increasing number of general or somewhat general megajournals (Siler et al., 2020) adequately filling spaces that new niche journals might previously have occupied. The journal count for 2020 may also increase as back issues of new journals are added in 2021 and afterwards.

The number of articles per journal fluctuated considerably between 1900 and 1980, with apparently thinner journals during both world wars (Figure 6). From 1980, journals seemed to grow in average size, perhaps aided by online-only journals without print journal limits on the annual numbers of articles. The apparent accelerated growth after 2010 is presumably due to

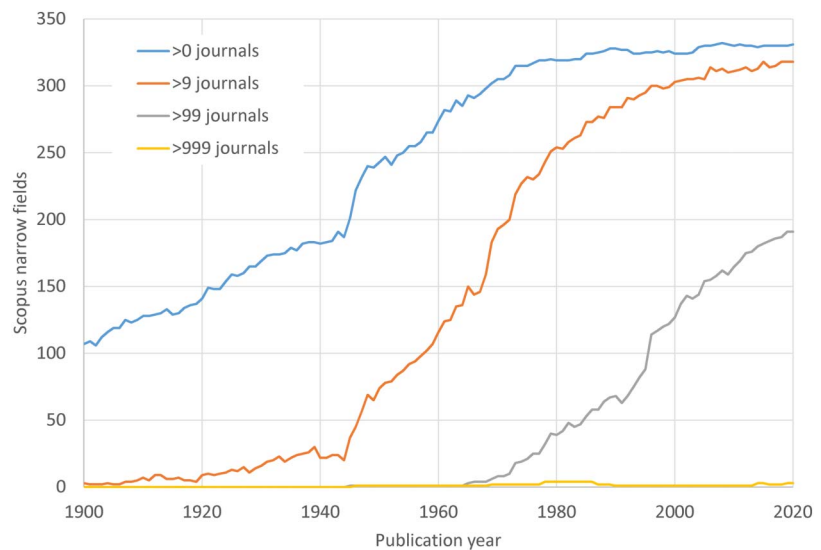


Figure 4. Number of Scopus narrow fields with specified minimum numbers of different journals.

increases in the number and size of online-only megajournals, starting in 2006 with *PLOS ONE* (Domnina, 2016), which had 230,518 articles in Scopus by 2020. The 10 largest journals in Scopus in 2020 were all arguably megajournals (*Scientific Reports*, *IEEE Access*, *PLOS ONE*, *Sustainability*, *International Journal of Environmental Research and Public Health*, *Applied Sciences*, *International Journal of Molecular Sciences*, *Science of the Total Environment*, *Sensors*, *Energies*), with only *Science of the Total Environment* existing before *PLOS ONE*. Megajournals have expanded to cover multiple more specialist roles, impinging on multiple fields (Siler et al., 2020). These combined factors seem likely to be the cause of tripling the average number of articles per journal between 1980 and 2020.

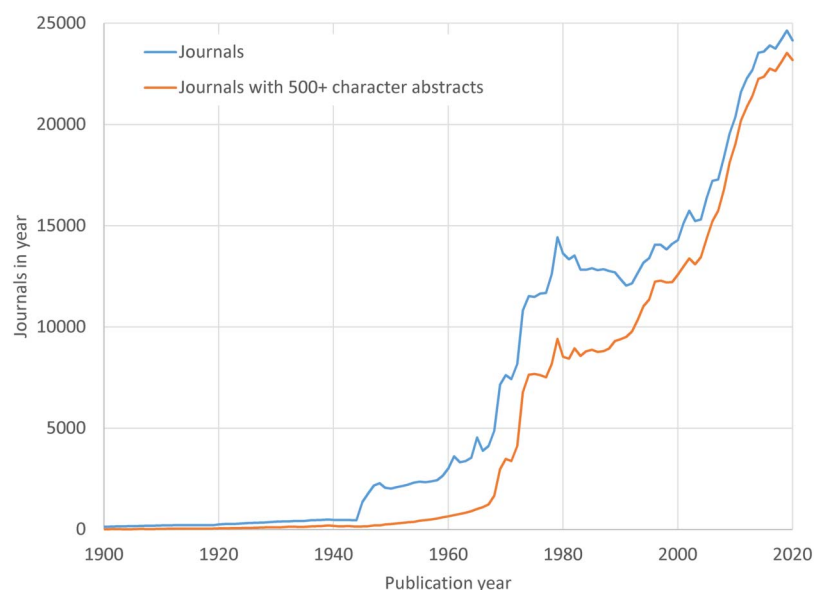


Figure 5. Number of different journals in Scopus by year.

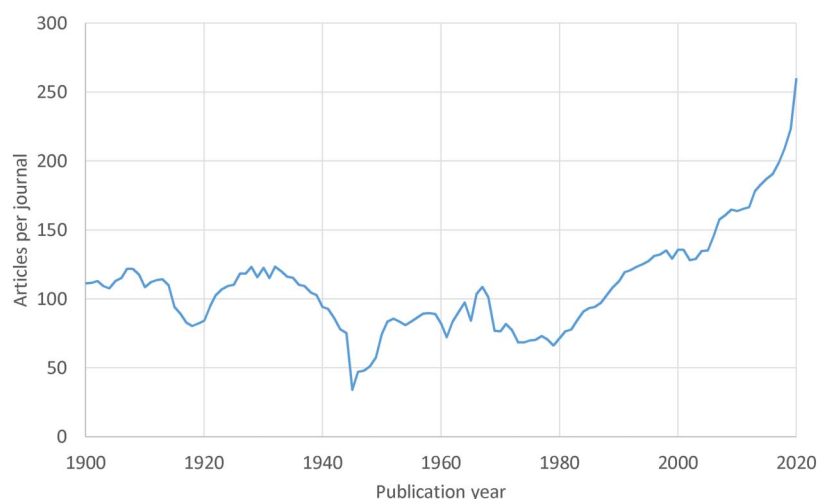


Figure 6. The average (mean) number of articles per journal in Scopus.

3.5. International Coverage (Authorship)

The national character of Scopus has changed dramatically over the 121 years covered (Figure 7). Initially, over two-thirds of first authors with known country affiliations were from the United States and Germany (Figure 8), but by 2020 China had substantially more articles than these two combined, and India had the third most articles (Figure 8). The number of articles with country affiliations dropped substantially during the Second World War, although the cause is unknown (e.g., Scopus indexing discrepancies, journal policy changes, or

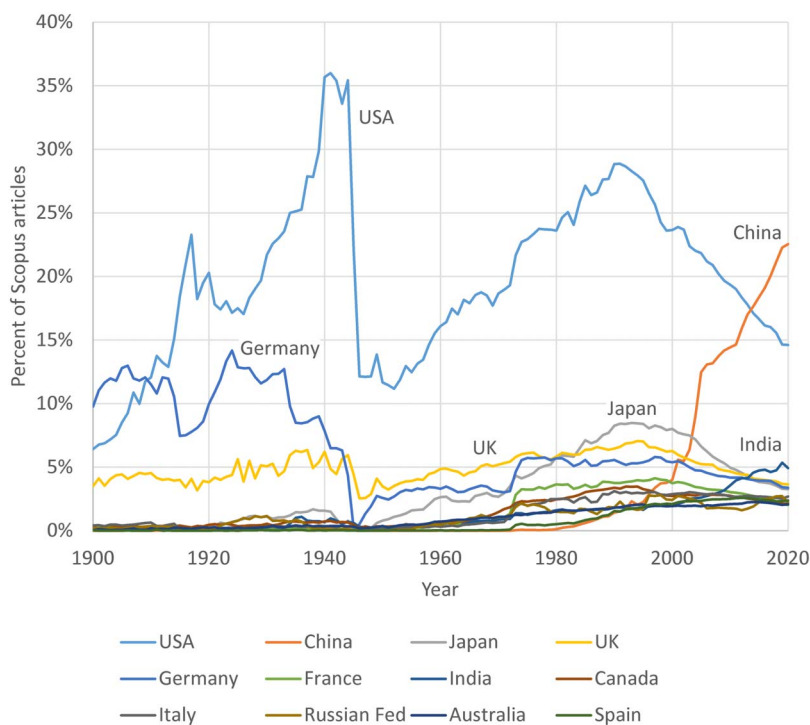


Figure 7. The percentage of Scopus articles with first author from the 12 countries with the most articles. Articles in multiple narrow fields are counted once for each narrow field.

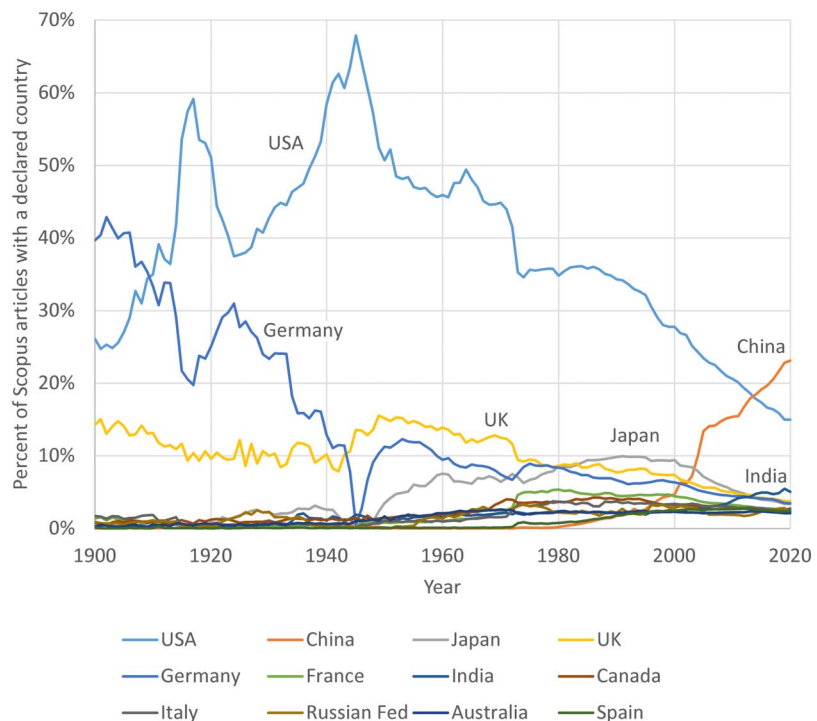


Figure 8. The percentage of Scopus articles with first author from the 12 countries with the most articles, excluding articles where the first author country is unknown (i.e., changing only the denominator from the previous graph). Articles in multiple narrow fields are counted once for each narrow field.

scientists omitting affiliations). Germany's contribution to the international literature dropped dramatically during both world wars, presumably by cutting it off from the publishing houses of the United Kingdom and United States. Germany's decline in the 1930s may have been partly due to the anti-Semitic policies of the Nazi party disrupting scholarship and causing a mass exodus of skilled researchers (e.g., in maths: Siegmund-Schultze, 1994).

3.6. Average Citation Counts

Articles accrue citations over time, so older articles have longer to be cited and should have more citations than newer articles, other factors being equal. This pattern is only partly evident in Scopus, however, since there is a peak in the year 2000 (Figure 9). This peak remains if the geometric mean is used (Fairclough & Thelwall, 2015), so it is not due to a few highly cited articles. The relatively few citations for articles published before 2000 could be due to a combination of factors, but the most likely seem to be

- shorter reference lists in older papers;
- a tendency to cite newer research in the digital age due to electronic searching, online first, and preprint archives;
- fewer references in older papers mentioning journal articles; and
- greater technical difficulty in matching citations to articles in older journals.

A similar lack of citations for older articles has been found for the WoS (Wallace et al., 2009). The size of the index does not seem to be an important consideration because, while there are

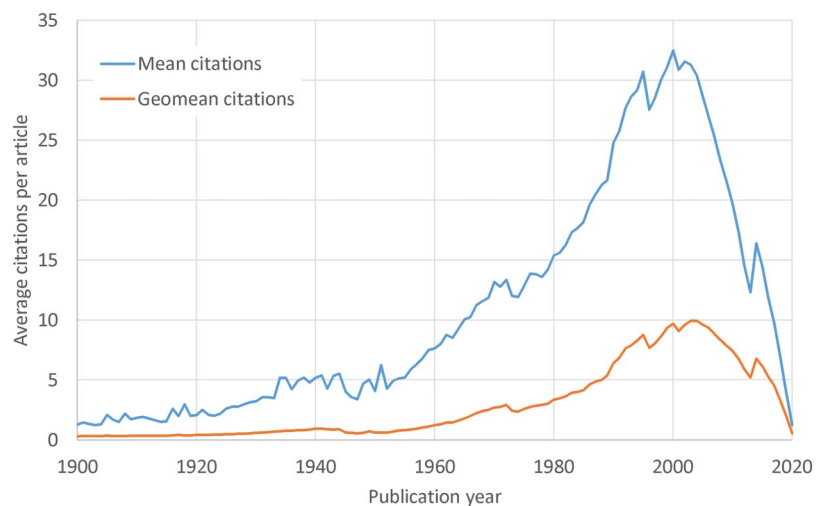


Figure 9. Average (arithmetic and geometric mean) citation counts for Scopus journal articles, by year. Citation data for 1900–1995 is from September 2021, for 1996–2013 is from December 2018, and for 2013–2020 is from January 2021.

fewer contemporary articles to cite old articles, there are also fewer articles to be cited, so the two factors largely cancel out under a scenario of constant growth. It is also possible that Scopus indexed a smaller fraction of the early scientific literature, therefore losing more old citations than contemporary citations. If true, this may again partly cancel out with Scopus presumably tending to preferentially index the most prestigious journals, therefore increasing the average number of citations per indexed article.

4. LIMITATIONS

The results are limited by the dates of the searches conducted and will be changed by any Scopus retrospective coverage increase. There is a small discrepancy between the total number of journal articles analyzed here (56,029,494) and the 56,391,519 reported by the Scopus web interface for the corresponding query, DOCTYPE(ar) AND SRCTYPE(j) AND PUBYEAR>1899 AND PUBYEAR<2021. The missing 362,025 journal articles seem too few (0.6%) to influence the analysis. The difference may derive partly from minor expansions of Scopus 1996–2013 after 2018, such as by adding the back catalogues of journals first indexed after 2018, especially megajournals, or by fixing indexing inconsistencies, such as reclassifying some documents as journal articles. There may also be technical issues with the API availability or processing that the consistency checks did not find.

An interpretation limitation for the analysis of abstracts is that it is not clear whether Scopus is comprehensive in its indexing of article abstracts, when they exist. No tests were performed to check whether articles without abstracts in Scopus had abstracts elsewhere, so this is unknown. One case of the opposite was accidentally found: an abstract in Scopus for the correct article that appeared to have been written afterwards and attached to it by a service that presumably informed Scopus, PsycInfo. Other sources of abstracts that could be compared with Scopus to check for this include Crossref (only publisher-supplied information, not always including abstracts; Waltman, Kramer et al., 2020), PubMed (biomedical science; e.g., Frandsen, Eriksen et al., 2019), and Microsoft Academic (soon to be discontinued; Tay, Martín-Martín, & Hug, 2021).

5. CONCLUSIONS

Overall, the results show that 1996 is no longer a watershed year (cf. Li et al., 2010) for Scopus coverage and that the three watersheds are the two world wars (dips in coverage) and 2004 (start of more rapid expansion and the Scopus launch year). This is true in terms of fields and citations, whereas for abstracts, the key date is the end of the Second World War. For journals, 1980 is another watershed for expanding average journal size and possibly also 2019 for a peak in the number of journals. The results therefore suggest 1946 as a logical earliest starting point for scientometric studies that require the longest reasonably consistent coverage. Nevertheless, this seems too long for most practical purposes (e.g., tracking the evolution of a journal over time) and the following practical suggestions are made to help decide on a suitable start year.

- Choose a starting year that is a watershed for the field(s) investigated, if relevant, and report any anomalies identified above during the period that might influence the results. All data is online for this <https://doi.org/10.6084/m9.figshare.16834198>.
- Set thresholds for the minimum number of articles, articles with abstracts, or average citation counts for the purposes of the study and use the graphs above to select the earliest year above the thresholds.
- If conducting a science-wide or international study, set thresholds for internationality or national field coverage and use the graphs above to select the earliest year above the thresholds. Also carefully consider the implications of the increasingly wide coverage of Scopus for more recent years.
- Explicitly acknowledge that the nature of the journal literature has changed during the years of the study in ways that cannot fully be considered, such as constantly expanding numbers and (for most periods) sizes of journals, and the international composition of authors.
- If using citation counts from before 2004, acknowledge that long-term trends will be influenced by lower average citations for earlier years, whether using a fixed citation window or counting citations to date. Lower level biases may also influence other years, however, as the publishing process evolves (e.g., speed, indexing).

The findings about abstracts are important for all articles that use long-term literature searches, because a Scopus search spanning from 1900 will be weak due to a lack of abstracts in the early years (e.g., Nsuala, Enslin, & Viljoen, 2015; Sweileh et al., 2019), but more powerful in recent years with near-universal abstract inclusion. Thus, variability in coverage should be considered when reporting or analyzing such results. This issue is particularly important for studies making claims such as “x increased in prevalence over time” based on searches of titles, abstracts, and keywords. Such increases could be artifacts of the more powerful searches in recent years because of the greater prevalence of abstracts.

AUTHOR CONTRIBUTIONS

Mike Thelwall: Methodology, Writing—Original draft, Writing—Review & editing. Pardeep Sud: Writing—Review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

This research was not funded.

DATA AVAILABILITY

The counts underlying the graphs are in the Supplementary material: <https://doi.org/10.6084/m9.figshare.16834198>.

REFERENCES

- Abdollahpour, Z., & Gholami, J. (2018). Building blocks of medical abstracts: Frequency, functions and structures of lexical bundles. *Asian ESP Journal*, 14(1), 83–111.
- Anonymous. (1944). Editorial. *British Journal of Industrial Medicine*, 1, 66. <https://doi.org/10.1136/oem.1.1.66>
- Atanassova, I., Bertin, M., & Larivière, V. (2016). On the composition of scientific abstracts. *Journal of Documentation*, 72(4), 636–647. <https://doi.org/10.1108/JDOC-09-2015-0111>
- Anwar, J., Bibi, A., & Ahmad, N. (2021). Behavioral strategy: Mapping the trends, sources and intellectual evolution. *Journal of Strategy and Management*. <https://doi.org/10.1108/JSMA-01-2021-0002>
- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), 377–386. https://doi.org/10.1162/qss_a_00019
- Beatty, S. (2015). Breaking the 1996 barrier: Scopus adds nearly 4 million pre-1996 articles and more than 83 million references. *Scopus Blog*. <https://blog.scopus.com/posts/breaking-the-1996-barrier-scopus-adds-nearly-4-million-pre-1996-articles-and-more-than-83>
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. https://doi.org/10.1162/qss_a_00018
- Blatt, E. (2009). Differentiating, describing, and visualizing scientific space: A novel approach to the analysis of published scientific abstracts. *Scientometrics*, 80(2), 385–406. <https://doi.org/10.1007/s11192-008-2070-3>
- Budimir, G., Rahimeh, S., Tamimi, S., & Južnič, P. (2021). Comparison of self-citation patterns in WoS and Scopus databases based on national scientific production in Slovenia (1996–2020). *Scientometrics*, 126(3), 2249–2267. <https://doi.org/10.1007/s11192-021-03862-w>
- Clarivate. (2021). *Web of Science platform: Web of Science: Summary of Coverage*. <https://clarivate.libguides.com/webofscienceplatform/coverage>
- Clark, W. B. (1918). Volumetric determination of reducing sugars. A simplification of scales' method for titrating the reduced copper without removing it from the residual copper solution. *Journal of the American Chemical Society*, 40(12), 1759–1772. <https://doi.org/10.1021/ja02245a002>
- Dimensions. (2021). *The data in Dimensions*. <https://www.dimensions.ai/dimensions-data/>
- Domnina, T. N. (2016). A megajournal as a new type of scientific publication. *Scientific and Technical Information Processing*, 43(4), 241–250. <https://doi.org/10.3103/S0147688216040079>
- Elsevier. (2021). *How Scopus works > Content*. <https://www.elsevier.com/solutions/scopus/how-scopus-works/content>
- Fairclough, R., & Thelwall, M. (2015). More precise methods for national research citation impact comparisons. *Journal of Informetrics*, 9(4), 895–906. <https://doi.org/10.1016/j.joi.2015.09.005>
- Fairclough, R., & Thelwall, M. (2021). Questionnaires mentioned in academic research 1996–2019: Rapid increase but declining citation impact. *Learned Publishing*. <https://doi.org/10.1002/leap.1417>
- Frandsen, T. F., Eriksen, M. B., Hammer, D. M. G., & Christensen, J. B. (2019). PubMed coverage varied across specialties and over time: A large-scale study of included studies in Cochrane reviews. *Journal of Clinical Epidemiology*, 112, 59–66. <https://doi.org/10.1016/j.jclinepi.2019.04.015>, PubMed: 31051247
- Fu, H. Z., & Ho, Y. S. (2013). Independent research of China in science citation index expanded during 1980–2011. *Journal of Informetrics*, 7(1), 210–222. <https://doi.org/10.1016/j.joi.2012.11.005>, PubMed: 32288781
- Gasparyan, A. Y., & Kitas, G. D. (2021). Editorial strategy to get a scholarly journal indexed by Scopus. *Mediterranean Journal of Rheumatology*, 32(1), 1–2. <https://doi.org/10.31138/mjr.32.1.1>, PubMed: 34386695
- Grishchenko, A. I. (2019). The Church Slavonic Song of Songs translated from a Jewish source in the Ruthenian Codex from the 1550s (RSL Mus. 8222): A new revised diplomatic edition. *Scrinium*, 15(1), 111–131. <https://doi.org/10.1163/18177565-00151P08>
- Gu, X., & Blackmore, K. L. (2016). Recent trends in academic journal growth. *Scientometrics*, 108(2), 693–716. <https://doi.org/10.1007/s11192-016-1985-3>
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), 387–395. https://doi.org/10.1162/qss_a_00020
- Hollander, F. (1954). The abstract section. *Gastroenterology*, 26(2), 319. [https://doi.org/10.1016/S0016-5085\(54\)80132-6](https://doi.org/10.1016/S0016-5085(54)80132-6)
- Hyland, C. J. (2017). English-Canadian social science, humanities, and law academic secondments during the Second World War and their contributions to Canadian external affairs. *International Journal of Canadian Studies*, 56, 81–114. <https://doi.org/10.3138/ijcs.56.2017-0004>
- Jayarathne, Y. S. N., & Zwahlen, R. A. (2015). The evolution of dental journals from 2003 to 2012: A bibliometric analysis. *PLOS ONE*, 10(3), e0119503. <https://doi.org/10.1371/journal.pone.0119503>, PubMed: 25781486
- Jimenez, S., Avila, Y., Dueñas, G., & Gelbukh, A. (2020). Automatic prediction of citability of scientific articles by stylometry of their titles and abstracts. *Scientometrics*, 125(3), 3187–3232. <https://doi.org/10.1007/s11192-020-03526-1>
- Jin, T., Duan, H., Lu, X., Ni, J., & Guo, K. (2021). Do research articles with more readable abstracts receive higher online attention? Evidence from Science. *Scientometrics*, 126(10), 8471–8490. <https://doi.org/10.1007/s11192-021-04112-9>
- Kallens, P. C., & Dale, R. (2018). Exploratory mapping of theoretical landscapes through word use in abstracts. *Scientometrics*, 116(3), 1641–1674. <https://doi.org/10.1007/s11192-018-2811-x>
- Kim, E.-S., & Lee, E.-J. (2020). A corpus-based study of lexical bundles and moves by English L1 and L2 writers in medical journal abstracts. *Korean Journal of English Language and Linguistics*, 20, 768–800. <https://doi.org/10.18823/asiateli.2021.18.1.9.142>
- Larivière, V., Gingras, Y., & Archambault, É. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the American Society for Information Science and Technology*, 60(4), 858–862. <https://doi.org/10.1002/asi.21011>
- Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation analysis: Comparison of Web of Science®, Scopus™, SciFinder®,

- and Google Scholar. *Journal of Electronic Resources in Medical Libraries*, 7(3), 196–217. <https://doi.org/10.1080/15424065.2010.505518>
- Liu, W., Huang, M., & Wang, H. (2021). Same journal but different numbers of published records indexed in Scopus and Web of Science Core Collection: Causes, consequences, and solutions. *Scientometrics*, 126(5), 4541–4550. <https://doi.org/10.1007/s11192-021-03934-x>
- Luna-Morales, M. E., Collazo-Reyes, F., Russell, J. M., & Pérez-Angón, M. Á. (2009). Early patterns of scientific production by Mexican researchers in mainstream journals, 1900–1950. *Journal of the American Society for Information Science and Technology*, 60(7), 1337–1348. <https://doi.org/10.1002/asi.21065>
- Mabe, M., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51(1), 147–162. <https://doi.org/10.1023/A:1010520913124>
- Maflahi, N., & Thelwall, M. (2021). Domestic researchers with longer careers generate higher average citation impact but it does not increase over time. *Quantitative Science Studies*, 2(2), 560–587. https://doi.org/10.1162/qss_a_00132
- Manzer, B. M. (1977). *The abstract journal, 1790–1920. Origin, development and diffusion*. Metuchen, NJ: The Scarecrow Press.
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871–906. <https://doi.org/10.1007/s11192-020-03690-4>, PubMed: 32981987
- Molinero, C. M. (1992). Operational research: From war to community. *Socio-Economic Planning Sciences*, 26(3), 203–212. [https://doi.org/10.1016/0038-0121\(92\)90011-S](https://doi.org/10.1016/0038-0121(92)90011-S)
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Nakayama, T., Hirai, N., Yamazaki, S., & Naito, M. (2005). Adoption of structured abstracts by general medical journals and format for a structured abstract. *Journal of the Medical Library Association*, 93(2), 237–242. PubMed: 15858627
- Nsuala, B. N., Enslin, G., & Viljoen, A. (2015). “Wild cannabis”: A review of the traditional use and phytochemistry of *Leonotis leonurus*. *Journal of Ethnopharmacology*, 174, 520–539. <https://doi.org/10.1016/j.jep.2015.08.013>, PubMed: 26292023
- Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management 1980–2006: A citation/co-citation analysis. *Journal of Operations Management*, 27(3), 185–202. <https://doi.org/10.1016/j.jom.2008.08.001>
- Porturas, T., & Taylor, R. A. (2021). Forty years of emergency medicine research: Uncovering research themes and trends through topic modeling. *American Journal of Emergency Medicine*, 45, 213–220. <https://doi.org/10.1016/j.ajem.2020.08.036>, PubMed: 33059985
- Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1), 12. <https://doi.org/10.3390/publications9010012>
- Pressey, S. L. (1917). Distinctive features in psychological test measurements made upon dementia praecox and chronic alcoholic patients. *Journal of Abnormal Psychology*, 12(2), 130. <https://doi.org/10.1037/h0070284>
- Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292–306. <https://doi.org/10.1002/asi.4630270505>
- Price, G. B. (2017). The founding of *Mathematical Reviews*. <https://www.ams.org/publications/math-reviews/GBaleyPrice.pdf>
- Schotten, M., Meester, W. J., Steiginga, S., & Ross, C. A. (2017). A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In F. J. Cantú-Ortiz (Ed.), *Research Analytics* (pp. 31–58). New York: Auerbach Publications. <https://doi.org/10.1201/9781315155890-3>
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628–638. [https://doi.org/10.1002/\(SICI\)1097-4571\(199210\)43:9<628::AID-ASIS>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASIS>3.0.CO;2-0)
- Siegmund-Schultze, R. (1994). “Scientific control” in mathematical reviewing and German-US-American relations between the two World Wars. *Historia Mathematica*, 21(3), 306–329. <https://doi.org/10.1006/hmat.1994.1027>
- Siler, K., Larivière, V., & Sugimoto, C. R. (2020). The diverse niches of megajournals: Specialism within generalism. *Journal of the Association for Information Science and Technology*, 71(7), 800–816. <https://doi.org/10.1002/asi.24299>
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113–5142. <https://doi.org/10.1007/s11192-021-03948-5>
- Smith, D. R. (2009). The historical development of academic journals in occupational medicine, 1901–2009. *Archives of Environmental & Occupational Health*, 64(Suppl. 1), 8–17. <https://doi.org/10.1080/19338240903284672>, PubMed: 20007113
- Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of Informetrics*, 9(3), 642–657. <https://doi.org/10.1016/j.joi.2015.06.005>
- Subbotin, A., & Aref, S. (2021). Brain drain and brain gain in Russia: Analyzing international migration of researchers by discipline using Scopus bibliometric data 1996–2020. *Scientometrics*, 126(9), 7875–7900. <https://doi.org/10.1007/s11192-021-04091-x>
- Sweileh, W. M., Al-Jabi, S. W., Zyoud, S. E. H., Shraim, N. Y., Anayah, F. M., ... AbuTaha, A. S. (2019). Bibliometric analysis of global publications in medication adherence (1900–2017). *International Journal of Pharmacy Practice*, 27(2), 112–120. <https://doi.org/10.1111/ijpp.12471>, PubMed: 30044514
- Tay, A., Martín-Martín, A., & Hug, S. E. (2021). Goodbye, Microsoft Academic—hello, open research infrastructure? *Impact of Social Sciences Blog*. https://eprints.lse.ac.uk/111325/1/impactofsocialsciences_2021_05_27_goodbye_microsoft_academic_hello.pdf
- Teschke, O., Wegner, B., & Werner, D. (2011). Glimpses into the history of *Zentralblatt MATH*. In O. Teschke, B. Wegner, & D. Werner (Eds.), *80 Years of Zentralblatt MATH* (pp. 1–16). Berlin: Springer. https://doi.org/10.1007/978-3-642-21172-0_1
- Thelwall, M. (2016). The precision of the arithmetic mean, geometric mean and percentiles for citation data: An experimental simulation modelling approach. *Journal of Informetrics*, 10(1), 110–123. <https://doi.org/10.1016/j.joi.2015.12.001>
- Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2), 430–435. <https://doi.org/10.1016/j.joi.2018.03.006>
- Wallace, M. L., Larivière, V., & Gingras, Y. (2009). Modeling a century of citation distributions. *Journal of Informetrics*, 3(4), 296–303. <https://doi.org/10.1016/j.joi.2009.03.010>
- Waltman, L., Kramer, B., Hendricks, G., & Vickery, B. (2020). *Open abstracts: Where are we?* <https://www.crossref.org/blog/open-abstracts-where-are-we/>