Ganti L, Persaud NA, Stead TS. Bibliometric analysis methods for the medical literature. *Academic Medicine & Surgery*. Published online January 30, 2025.
doi:10.62186/001c.129134

Review

Bibliometric analysis methods for the medical literature

Latha Ganti, MD, MS, MBA[1], Nadiya Amanda Persaud, B.S.[1], Thor Shiva Stead, ScB[2]

[1] Research, Orlando College of Osteopathic Medicine, [2] Brown University

Keywords: bibliometric, citation analysis, VosViewer

https://doi.org/10.62186/001c.129134

Academic Medicine & Surgery

A bibliometric analysis of medical literature provides valuable insights into the structure, trends, and impact of research in the medical field. It can be a critical tool for advancing knowledge, improving healthcare outcomes, and supporting strategic decisions in the medical and scientific communities. In this article, the authors describe a step-by-step approach on how to perform a bibliometric analysis of the medical literature, based on review of many different styles of bibliometric analyses in various fields of medicine and surgery.

INTRODUCTION

DEFINITION AND PURPOSE

A bibliometric analysis is a systematic method for quantitatively evaluating scientific literature to identify patterns, trends, and key contributions within a specific field of study.[1] This approach relies on mathematical and statistical techniques to analyze bibliographic data, such as publication records, citation metrics, and authorship details, typically sourced from databases such as PubMed, Scopus, Web of Science, or Google Scholar.[2] By examining the relationships and impacts of published works, bibliometric analysis provides a comprehensive understanding of research activity and its development over time.

The primary objective of bibliometric analysis is to provide insights into the evolution and structure of a research domain.[3] Researchers use it to uncover historical trends, measure the impact of specific studies or authors, and identify influential journals or institutions. It also helps in discovering emerging topics, collaborative networks, and areas requiring further investigation. This analysis is particularly useful for guiding new researchers or organizations in identifying key contributors, highly cited articles, and research gaps, especially important for global threats such as the COVID-19 pandemic.[4]

Bibliometric analysis has wide-ranging applications across disciplines. In academia, it supports systematic literature reviews by identifying foundational works and influential frameworks. For policymakers, it provides evidence-based insights to allocate funding and prioritize research initiatives.[5] Institutions use bibliometric studies to benchmark their performance, assess the productivity of faculty members, and identify collaboration opportunities.[6] This approach also plays a critical role in interdisciplinary studies, helping to trace how knowledge flows between fields and facilitating the integration of ideas. In medicine, bibliometric studies can highlight emerging therapies or innovative methodologies by tracking citations and keywords in leading journals.[7,8]

HISTORY AND EVOLUTION OF BIBLIOMETRICS

Bibliometric analysis has its roots in the intersection of information science, statistics, and scholarly publishing. Its development was driven by the need to evaluate and organize the ever-growing body of scientific literature systematically. The origins of bibliometric analysis can be traced to the early 20th century, with foundational ideas emerging from librarianship, citation analysis, and the sociology of science.

The conceptual foundation for bibliometric analysis was created by Paul Otlet and Henri La Fontaine, whose efforts to organize global knowledge culminated in the development of the Universal Decimal Classification.[9] This system sought to categorize and systematize information, serving as a precursor to bibliometric methods. Librarians and early information scientists began developing methods to classify and manage large volumes of scholarly materials. The emergence of library cataloging systems provided the groundwork for later bibliometric practices.[10]

One of the earliest examples of citation-based analysis came from Gross and Gross,[11] who analyzed chemistry journal citations to identify influential journals. This marked the start of systematic efforts to evaluate scientific literature quantitatively. Samuel C. Bradford[12] introduced Bradford's Law of Scattering, which demonstrated how articles on a specific topic are unevenly distributed across journals. His work became instrumental in identifying core journals in various disciplines. Another foundational contribution came from George Zipf[13] whose statistical analysis of word frequencies, known as Zipf's Law, was later adapted to analyze patterns in scientific publications and citations.

The term "bibliometrics" was coined by Alan Pritchard,[14] who defined it as "the application of mathematical and statistical methods to books and other media of communication." This provided a formal identity to the field. One of the most transformative contributions came from Eugene Garfield in the mid-20th century. In 1955, Garfield introduced the concept of citation indexing[15]

which enabled researchers to trace connections between papers. His creation of the Science Citation Index (SCI) in 1964 revolutionized the study of scholarly communication, making citation analysis widely accessible. Lotka's Law described the frequency distribution of authorship in scholarly publications, showing that a small number of authors produce the majority of articles in a given field.[16]

With the advent of computers and digital databases in the 1980s, bibliometric analysis became more sophisticated. Digital tools like Web of Science and Scopus enabled researchers to perform large-scale analyses of citations, keywords, and collaborations. Network visualization tools such as VOSviewer and CiteSpace further enhanced the ability to explore relationships within bibliographic data.[17]

Derek J. de Solla Price (1963), often called the father of scientometrics, significantly influenced bibliometrics by studying the exponential growth of scientific literature. His concept of "Big Science" highlighted the increasing scale and complexity of research activities.[18]

In the realm of sociology, Robert K. Merton (1968) provided a theoretical foundation for citation practices. His "Matthew Effect" described how prominent scientists often receive disproportionate credit, influencing patterns of recognition in academia.[19]

REVIEW

KEY COMPONENTS OF BIBLIOMETRIC ANALYSIS[20-22]

1. Citation Analysis: This technique examines the frequency and patterns of citations a publication, author, or journal receives. It provides an indicator of the work's impact within the scientific community.
2. Co-Citation Analysis: This method explores the relationships between articles that are frequently cited together, revealing connections and clustering within the research field.
3. Keyword Analysis: By analyzing recurring keywords, researchers can track dominant themes and emerging areas of interest within a discipline.
4. Authorship and Collaboration Analysis: This aspect focuses on the productivity of individual authors or institutions and the extent of collaboration within the research community.
5. Bibliographic Coupling: This approach identifies links between articles based on shared references, which can indicate intellectual similarities or influences.
6. Network Visualization: Software tools are often used to map and visualize complex relationships between authors, papers, or keywords, making it easier to understand research dynamics.

STEP BY STEP APPROACH TO PERFORMING A BIBLIOMETRIC ANALYSIS

Writing a robust bibliometric review requires careful planning, rigorous methodology, and clear presentation of results.[23] Below is a step-by-step approach:

Define Your Research Scope and Objectives

1. Identify Research Goals: Decide what you aim to achieve—e.g., analyzing publication trends, and identifying key authors, institutions, journals, or topics in a medical field.
2. Set a time frame: Choose the publication period you wish to analyze (e.g., last 10 years).
3. Select a Topic/Field: Narrow down the focus to a specific topic, such as the use of tenecteplase in ischemic stroke,[24] post-stroke depression,[25] advances in autism,[26] the doctor-patient relationship,[27] or access to healthcare in rural communities.[28]The scope of the focus (whether narrow or broad) depends on the purpose of the review (what information is hoped to be gleaned from citation analysis).

Search and Collect Data

1. Choose a Bibliographic Database: Use reputable databases like PubMed, Scopus, Web of Science, or Embase. Each have their pros and cons [**Table 1**].
2. Create a Search Strategy:
   - Use relevant keywords (e.g., "cardiovascular diseases") and Boolean operators (AND, OR, NOT).
   - Apply filters for language, article type, and timeframe.
3. Download Records:
   - Export citation data, including titles, authors, abstracts, keywords, publication years, journals, and DOI.
   - Preferred formats include .csv, .xls, or .ris for further analysis.

Step 3: Data Cleaning and Preparation

1. Remove Duplicates: Use reference management software like EndNote, Mendeley, or Zotero to clean duplicates.
2. Standardize Data: Ensure consistent formatting for author names, affiliations, and journal names.
3. Filter Relevant Records: Exclude irrelevant articles based on abstracts and titles.

Step 4: Data Visualization

1. Use Software like VOSviewer, Gephi, or CiteSpace can create co-authorship, keyword co-occurrence, and citation network maps [**Table 2**]
2. Generate Charts and Graphs:
   - Use tools like Excel or Tableau for simple trends and distribution graphs.

Figures 1 and 2 summarize in detail how to conduct a bibliometric analysis using VOSViewer software and data from Web of Science.

LIMITATIONS

Bibliometric analysis offers an objective, data-driven approach to evaluating scientific progress, enabling the identification of impactful research and influential contributors. It also facilitates a macro-level understanding of

Table 1. Databases for bibliometric data in medicine

	Pros	Best Use	Cons
PubMed	• Free • Includes articles from highly reputable journals. • Offers MeSH (Medical Subject Headings) for robust keyword analysis.	Thematic analysis Life science focus	• Lack of detailed citation analysis features, inability to calculate author metrics like h-index directly
Web of Science	• Citation tracking • Multidisciplinary	Best citation and network analysis	• Subscription required. • Potential biases in journal selection, language limitations (mainly English), exclusion of non-journal publications like books and conference papers.
Scopus	• Broad coverage • robust author metrics (h-index, author impact)	Multidisciplinary analysis	• Subscription required. • Potential biases in journal selection (Elsevier) • Over-reliance on impact factor-based metrics, potentially neglecting high-quality research in less prominent journals.
Embase	• Drug research • clinical trials • Includes conference abstracts	Pharmacological studies	• Subscription required • Potential biases in journal selection (Elsevier)
Google Scholar	• Free • Broad coverage	Initial search, gray literature	• lack of comprehensive coverage of scholarly literature • difficulty in filtering out non-peer-reviewed content • potential for bias in search results due to its algorithm, • lack of expert curation compared to dedicated academic databases
Cochrane Library	Evidence-based content	Systematic review focus	• Narrow scope- only Cochrane reviews included, although quality of data overall is excellent

research trends, which is invaluable in a rapidly expanding scientific landscape. However, the technique has limitations.[29] Over-reliance on citation metrics can skew interpretations, as high citation counts do not always equate to high-quality research. For example, bibliometric methods seldom account for the nature of citations—whether they are positive, negative, or purely perfunctory—and thus may inflate the perceived relevance of certain studies.[30] Furthermore, the time-lag inherent in accruing citations often fails to reflect the current impact of newer work.[31] Additionally, bibliometric tools generally neglect non-journal outputs, such as patents or policy documents, undervaluing important sources of practical influence.[32] Lastly, biases inherent in bibliographic databases—such as limited coverage of certain languages or regions—can affect results.

CONCLUSION

Bibliometric analysis emerged from efforts to organize and evaluate scientific literature systematically, drawing on contributions from library science, documentation, and statistical analysis. Bibliometric analysis is a robust method for exploring and interpreting academic literature. By quantifying and visualizing research dynamics, it provides valuable insights into the development of scientific fields, supporting decision-making in academia, industry, and policy. Despite its limitations, bibliometric analysis remains a powerful tool for exploring the evolution of knowledge.

Submitted: January 28, 2025 EST. Accepted: January 28, 2025 EST. Published: January 29, 2025 EST.

Table 2. Software for bibliometric analysis

	Pros	Cons
VOSviewer	• User-friendly interface for beginners • Excellent for network visualization • Free	• Limited metrics customization • Focused on visualizations, lacks in-depth statistics
Bibliometrix (R)	• Open-source and free • Highly customizable and flexible • Advanced statistical analysis	• Requires R programming knowledge • Steeper learning curve for non-technical users
CiteSpace	• Specializes in trend and temporal analysis • Good for citation bursts and co-citation • Free	• Complex interface • Requires substantial time to master advanced features
Gephi	• Advanced network analysis • High degree of customization • Free	• Requires technical expertise • Not tailored specifically for bibliometrics
SciVal	• Robust metrics and benchmarking tools • Comprehensive institutional data • Subscription-based	• Expensive for individual researchers or small institutions
InCites	• Excellent for institutional benchmarking • Integrated with Web of Science • Comprehensive analytics	• Requires subscription to Web of Science • High cost
Publish or Perish	• Free and lightweight • Pulls data from multiple sources • Simple interface for basic metrics	• Limited visualization capabilities • Relies on Google Scholar, prone to data inconsistency
HistCite	• Simple interface • Focuses on citation analysis • Free	• Outdated design • Limited visualization and compatibility with modern databases
Dimensions	• Combines citations, funding, and altmetrics • Freemium model available • Easy-to-use	• Limited older data compared to Scopus/Web of Science • Advanced features require payment
Pajek	• Excellent for large datasets • Specialized in social network analysis • Free	• Very technical interface • Limited to network analysis, not general bibliometrics

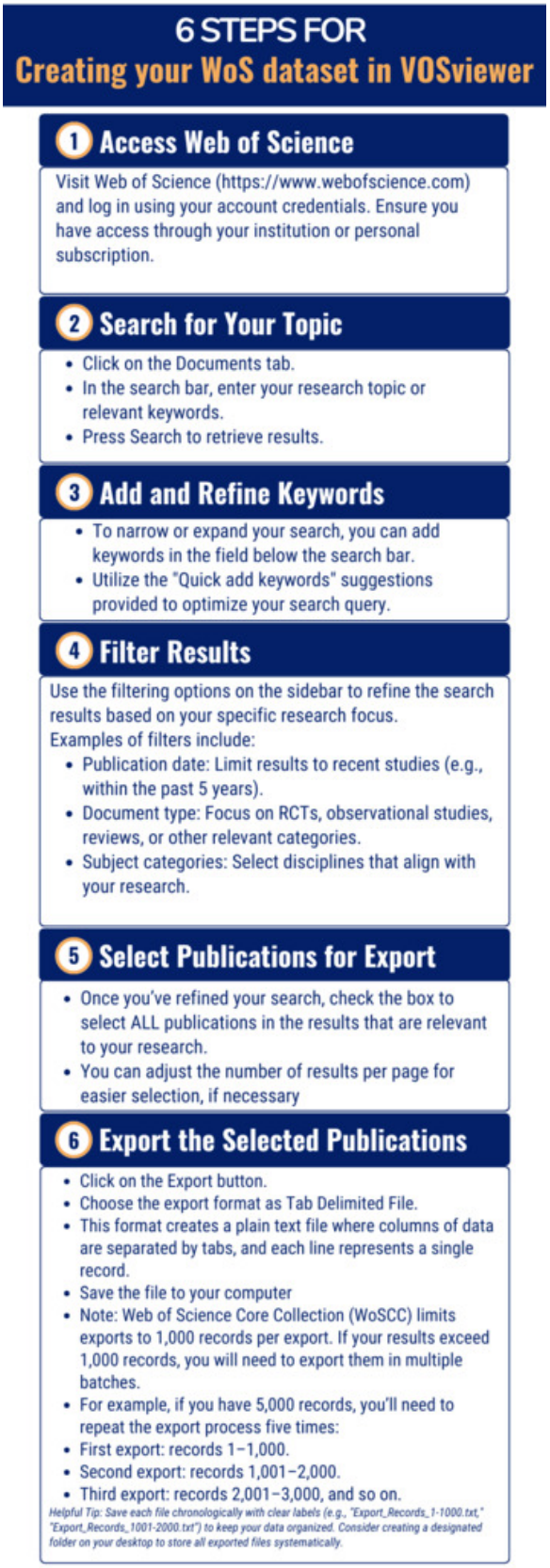

Figure 1. 6 Steps for Creating Your WoS Dataset in VOSviewer

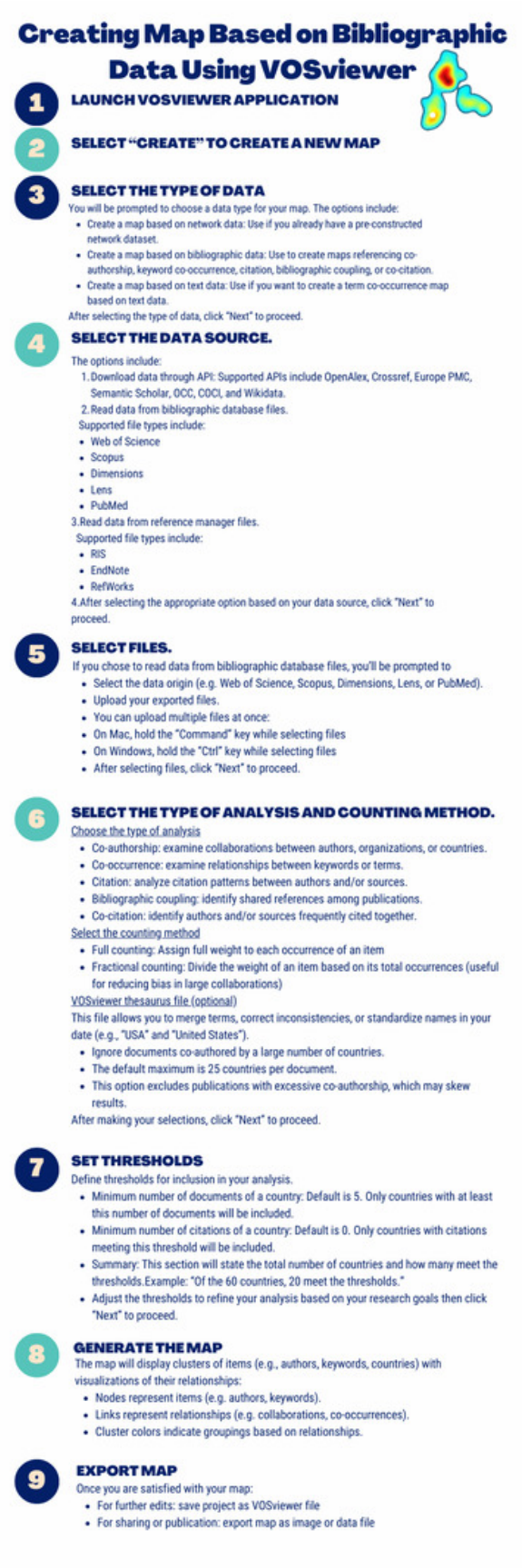

Figure 2. Creating Map Based on Bibliographic Data Using VOSviewer

REFERENCES

1. L NMK, George RJ, PS A. Bibliometric analysis for medical research. *Indian Journal of Psychological Medicine*. 2022;45(3):277-282. doi:10.1177/02537176221103617

2. NOAA Library: Bibliometrics and Research Evaluation: Home. https://library.noaa.gov/bibliometrics

3. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*. 2021;133:285-296. doi:10.1016/j.jbusres.2021.04.070

4. Mohadab ME, Bouikhalene B, Safi S. Bibliometric method for mapping the state of the art of scientific production in Covid-19. *Chaos Solitons & Fractals*. 2020;139:110052. doi:10.1016/j.chaos.2020.110052

5. Agarwal A, Durairajanayagam D, Tatagari S, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. *Asian Journal of Andrology*. 2015;18(2):296. doi:10.4103/1008-682x.171582

6. Ninkov A, Frank JR, Maggio LA. Bibliometrics: Methods for studying academic publishing. *Perspectives on Medical Education*. 2021;11(3):173-176. doi:10.1007/s40037-021-00695-4

7. Hu Y, Yu Z, Cheng X, Luo Y, Wen C. A bibliometric analysis and visualization of medical data mining research. *Medicine*. 2020;99(22):e20338. doi:10.1097/md.0000000000020338

8. Kokol P, Vošner HB, Završnik J. Application of bibliometrics in medicine: a historical bibliometrics analysis. *Health Information & Libraries Journal*. 2020;38(2):125-138. doi:10.1111/hir.12295

9. The Universal Decimal Classification for Libraries. *Science*. 1944;100(2588):94. doi:10.1126/science.100.2588.94

10. Mann T. *Library Research Models: A Guide to Classification, Cataloging, and Computers.*; 1993. http://ci.nii.ac.jp/ncid/BA20445717

11. Gross PLK, Gross EM. College libraries and chemical education. *Science*. 1927;66(1713):385-389. doi:10.1126/science.66.1713.385

12. Bradford SC. Sources of Information on Scientific Subjects. *Engineering: An Illustrated Weekly Journal*. 1934;137:85-86. https://www.scirp.org/reference/referencespapers?referenceid=3383890

13. Zipf GK. *Human Behavior and the Principle of Least Effort.*; 1949. http://psycnet.apa.org/psycinfo/1950-00412-000

14. Pritchard J. Statistical Bibliography or Bibliometrics. *Journal of Documentation*. 1969;25:348-349.

15. Garfield E. Citation indexes for science. *Science*. 1955;122(3159):108-111. doi:10.1126/science.122.3159.108

16. Lotka A. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*. 1926;16(12):317-323. https://www.jstor.org/stable/24529203

17. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2009;84(2):523-538. doi:10.1007/s11192-009-0146-3

18. De Solla Price DJ. *Little Science, Big Science.*; 1963. doi:10.7312/pric91844

19. Merton RK. The Matthew effect in science. *Science*. 1968;159(3810):56-63. doi:10.1126/science.159.3810.56

20. Gan YN, Li DD, Robinson N, Liu JP. Practical guidance on bibliometric analysis and mapping knowledge domains methodology – A summary. *European Journal of Integrative Medicine*. 2022;56:102203. doi:10.1016/j.eujim.2022.102203

21. Krüger AK, Petersohn S. 'I want to be able to do what I know the tools will allow us to do': Practicing evaluative bibliometrics through digital infrastructure. *Research Evaluation*. 2022;31(4):475-485. doi:10.1093/reseval/rvac009

22. Öztürk O, Kocaman R, Kanbach DK. How to design bibliometric research: an overview and a framework proposal. *Review of Managerial Science*. 2024;18(11):3333-3361. doi:10.1007/s11846-024-00738-0

23. Klarin A. How to conduct a bibliometric content analysis: Guidelines and contributions of content co-occurrence or co-word literature reviews. *International Journal of Consumer Studies*. 2024;48(2). doi:10.1111/ijcs.13031

24. Bhasin G, Ganti L. Bibliometric analysis of the usage of tenecteplase for stroke. *International Journal of Emergency Medicine*. 2024;17(1). doi:10.1186/s12245-024-00738-7

25. Zheng D, Vaughn S, Gould M, Ganti L. Bibliometric analysis of depression in post-stroke patients. *International Journal of Emergency Medicine*. 2024;17(1). doi:10.1186/s12245-024-00725-y

26. Chowdhury M, Gould M, Ganti L. Advances in Autism: a bibliometric analysis. *Health Psychology Research*. 2024;12. doi:10.52965/001c.125844

27. Konda M, Gould M, Mangal R, Daniel A, Stead T, Ganti L. The Doctor-Patient relationship: A bibliometric analysis. *Health Psychology Research*. 2023;11. doi:10.52965/001c.90429

28. Natarajan A, Gould M, Daniel A, Mangal R, Ganti L. Access to healthcare in rural communities: A bibliometric analysis. *Health Psychology Research*. 2023;11. doi:10.52965/001c.90615

29. Passas I. Bibliometric Analysis: the main steps. *Encyclopedia*. 2024;4(2):1014-1025. doi:10.3390/encyclopedia4020065

30. Aksnes DW, Langfeldt L, Wouters P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *Sage Open*. 2019;9(1). doi:10.1177/2158244019829575

31. Teplitskiy M, Duede E, Menietti M, Lakhani KR. How status of research papers affects the way they are read and cited. *Research Policy*. 2022;51(4):104484. doi:10.1016/j.respol.2022.104484

32. Wallin JA. Bibliometric methods: Pitfalls and possibilities. *Basic & Clinical Pharmacology & Toxicology*. 2005;97(5):261-275. doi:10.1111/j.1742-7843.2005.pto_139.x