Bibliometric Analysis in Scientific Research: Applications, Limitations, and Key Considerations for Authors

Namrata Dagli [1], Mainul Haque [2,3], Santosh Kumar [4]

Please Click on Photo



Keywords

Bibliometric Analysis, Bibliometrics, Application, Bias, Scientometric, Big data analysis, Bibliometric tools, Limitations

Bibliometric analysis, the quantitative analysis of scientific academic literature, has become an integral tool in evaluating research impact and scientific productivity. It offers valuable insights into co-authorship patterns, research trends, and scholarly collaborations [1]. The use of bibliometric analysis and other large-scale analytical approaches has increased in the scientific literature on mapping. This increase is driven by several factors reshaping the research landscape, such as technological advancements, digital academic database proliferation, and growing research paper publications [2]. Enhanced computational power and advanced data analysis tools allow for efficient processing of vast scientific literature. Additionally, the shift to digital publishing and the expansion of online academic databases have made research outputs more accessible, providing a rich data source for bibliometric analysis. Understanding this analytical approach is crucial for the scientific community in today's competitive academic environment.

Bibliometric analysis relies on academic databases such as Web of Science, Scopus, PubMed, Google Scholar, and IEEE Xplore for comprehensive literature and citation data [3]. Essential software tools include VOSviewer, CiteSpace, Gephi, Biblioshiny, HistCite, Pajek, and the Sci2 Tool [4-11]. The applications of these analytical approaches are diverse and wide-ranging. Bibliometric analyses can reveal emerging research trends [11,12] and interdisciplinary connections, guiding research investments. Institutions and funding bodies use bibliometrics to assess research performance and inform funding decisions. These tools can map collaboration networks, helping researchers identify potential partners and institutions for alliance opportunities. Policymakers can use bibliometric data to inform science policy, identify national research impact, and guide resource allocation. Researchers can use these metrics for self-evaluation and to support career

1. Center for Global Health Research, Saveetha Medical College, Saveetha Institute of Medical, Dental & Technical Sciences, Chennai, Tamil Nadu, India.
2. Unit of Pharmacology, Faculty of Medicine and Defence Health, Universiti Pertahanan Nasional Malaysia (National Defence University of Malaysia), Kuala Lumpur, Malaysia.
3. Department of Research, Karnavati Scientific Research Center (KSRC) Karnavati School of Dentistry, Karnavati University, Gandhinagar, Gujarat, India.
4. Department of Periodontology, Karnavati School of Dentistry, Karnavati University, Gandhinagar, Gujarat, India.

Correspondence

Mainul Haque. Unit of Pharmacology, Faculty of Medicine and Defence Health, Universiti Pertahanan Nasional Malaysia (National Defence University of Malaysia), Kem Perdana Sungai Besi, 57000 Kuala Lumpur, Malaysia.
Email: runurono@gmail.com, mainul@upnm.edu.my.
Cell Phone: +60109265543

DOI: https://doi.org/10.3329/bjms.v24i1.78602

advancement, while publishers can utilize bibliometrics to assess journal performance.

Despite these diverse applications, there are many limitations and biases, too. One crucial issue is the geographic and linguistic bias inherent in many bibliometric databases. Research has shown considerable citation indices underrepresent publications from non-English-speaking countries and developing nations [13]. This bias skews our understanding of global research output and potentially marginalizes important work from underrepresented regions. For authors from these regions, this presents an additional challenge in gaining recognition for their work. The rise of open-access publishing is another factor altering the research landscape. Studies have shown that open-access articles generally receive more citations due to their greater accessibility [14]. This trend highlights the importance of considering open-access options for authors while publishing.

Bibliometric analysis relies on several major academic databases and bibliometric tools, each with strengths and limitations. PubMed offers free access to biomedical literature but lacks citation data. Web of Science and Scopus provide comprehensive citation analysis but are subscription-based. Google Scholar offers broad coverage but lacks quality control. Dimensions give a more extensive range of research outputs but are relatively new. The bibliometric tools vary in focus, too. The choice of database and tool significantly impacts bibliometric results, and no single option is without bias or limitations [11]. Researchers should carefully select their resources based on specific needs and be aware of potential biases in coverage across languages, geographic regions, and disciplines. Ideally, using multiple databases and tools, combined with expert knowledge [15] and qualitative assessment, provides the most comprehensive and balanced view of the research landscape. Also, bibliometric analysis varies across disciplines, necessitating context-specific interpretation of data, such as humanities prioritizing monographs and books often underrepresented in citation databases [16,17]. Another critical limitation of bibliometric analysis is its heavy reliance on keywords, particularly for keyword-based trend and thematic analyses. The effectiveness of keyword-based analysis depends on the choice and consistency of keywords used by authors, indexers, and searchers. Variations in terminology across disciplines, evolving scientific language, and inconsistent keyword selection can lead to the omission of relevant research or the inclusion of tangentially related work. This issue is particularly significant in interdisciplinary research, where terminology may differ substantially between fields, potentially obscuring essential connections or contributions to the field. Authors should strategically choose trending keywords to overcome the limitation. Language barriers further intensify the limitations of keyword-based analysis, as research published in non-dominant languages may use different terms for similar concepts. As the scientific community continues to rely on bibliometric tools for thematic and trend analyses, it is crucial to develop more sophisticated approaches to capture the semantic content of research beyond simple keyword matching. This indicates the need for guidelines for systematic and consistent keyword selection in scientific manuscripts. Integrating context-aware algorithms and multi-lingual analysis tools might provide a more comprehensive view of the scientific landscape. Authors should be aware that while citation counts and impact factors offer valuable information, they don't tell the whole story of a paper's significance or a researcher's contributions [18]. Authors should also consider alternative metrics that capture broader societal impact, such as mentions in policy documents or media coverage [19,20]. Most importantly, authors should focus on conducting rigorous, innovative research rather than chasing metrics.

Future bibliometric analyses can be enhanced by data science, artificial intelligence, and machine learning and integrated with Altmetrics for more accurate and deeper insights [19]. However, the increasing reliance on bibliometrics advances might spark ongoing debates about the ethics of quantifying scientific impact and challenge the diversity and creativity of research. The scientific community must thoughtfully integrate quantitative tools with human expertise and qualitative assessments to support curiosity, creativity, and knowledge pursuit. A balanced approach to evaluation is required to ensure that metrics enhance rather than constrain scientific progress [21]. Despite its limitations, bibliometric analysis is a crucial tool for mapping the scientific landscape and guiding policy decisions, and its significance in the scientific community cannot be underestimated. As authors, being aware of the strengths and limitations allows for obtaining more accurate results and interpretation of bibliometric analyses. The key findings of this editorial are depicted in Figure 1.

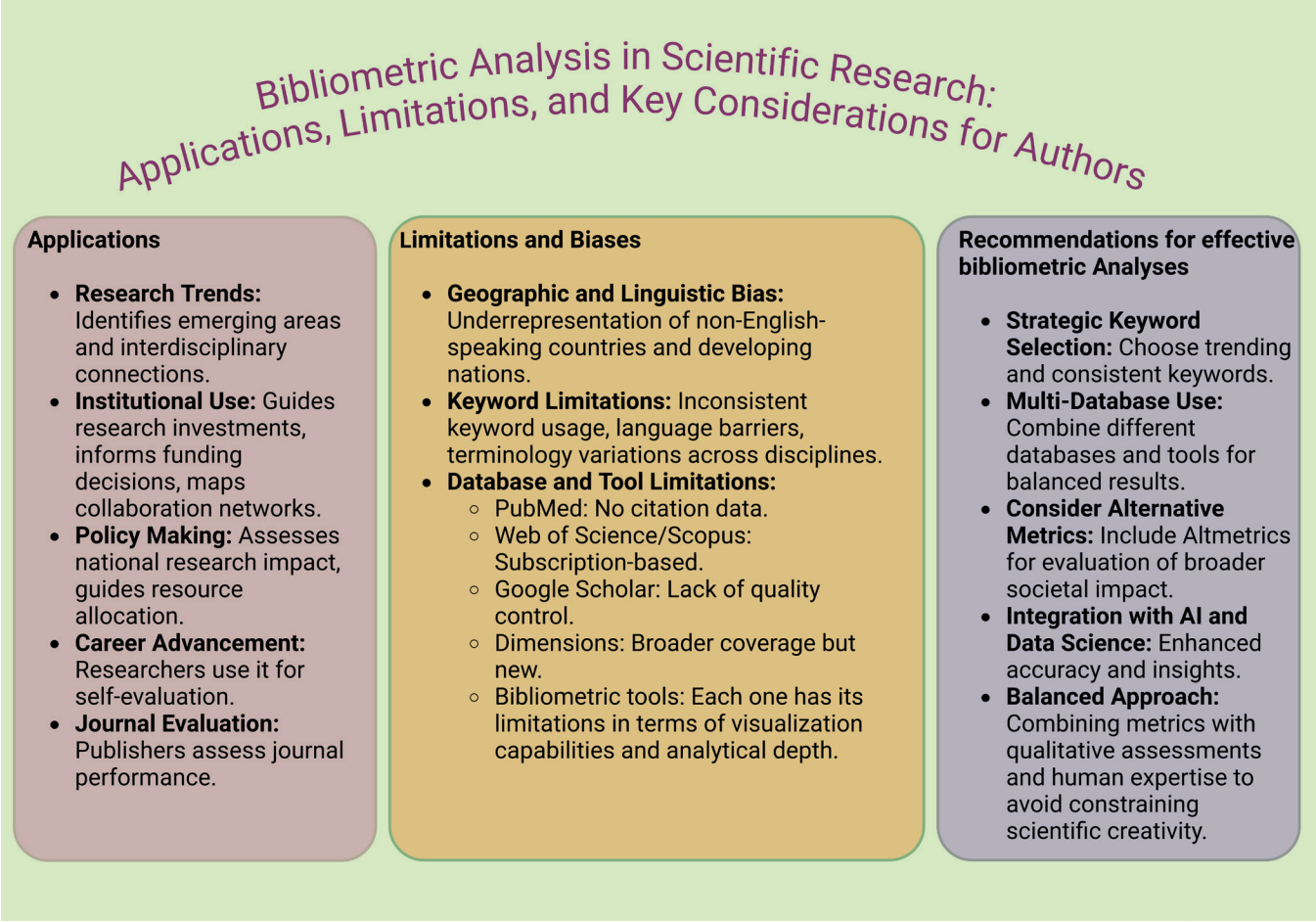

Figure 1: The Principal Findings of This Paper.

Notes: This figure has been drawn using the premium version of BioRender [22] [(https://biorender.com/) Accessed on August 29th, 2024) with the agreement license number LB278OKBTV.

Image Credit: Namrata Dagli.

CONSENT FOR PUBLICATION

The author reviewed and approved the final version and has agreed to be accountable for all aspects of the work, including any accuracy or integrity issues.

DISCLOSURE

The author declares that they do not have any financial involvement or affiliations with any organization, association, or entity directly or indirectly related to the subject matter or materials presented in this editorial. This includes honoraria, expert testimony, employment, ownership of stocks or options, patents, or grants received or pending royalties.

DATA AVAILABILITY

Information is taken from freely available sources for this editorial.

Authorship Contribution

All authors contributed significantly to the work, whether in the conception, design, utilization, collection, analysis, and interpretation of data or all these areas. They also participated in the paper's drafting, revision, or critical review, gave their final approval for the version that would be published, decided on the journal to which the article would be submitted, and made the responsible decision to be held accountable for all aspects of the work.

REFERENCES

1. Pritchard A. Statistical bibliography or bibliometrics. *J Doc*. 1969;**25**:348. Available from: https://cir.nii.ac.jp/crid/1570009750342049664
2. Dagli N, Haque M, Kumar S. The Explosion of Scientific Publications: Navigating the Evolving Landscape of Research Dissemination. *Bang J Med Sci*. 2024;**23**(04). (In Press)
3. Zhu J, Liu W. A tale of two databases: the use of Web of Science and Scopus in academic papers. *Scientometrics*. 2020;**123**:321–335. Doi:10.1007/s11192-020-03387-8
4. Van Eck NJ, Waltman L. VOSviewer Manual. Leiden, Netherlands: Leiden University; 2023. [Cited 2024 March 26]. Available from: http://ttps://www.VOSviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf.
5. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci Technol*. 2006;**57**(3):359-377. doi:10.1002/asi.20317
6. Bastian, M., Heymann, S., & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*. 2009: 3(1), 361-362. doi.:10.1609/icwsm.v3i1.13937 Ba
7. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr*. 2017;**11**:959-75. doi:10.1016/j.joi.2017.08.007
8. Garfield E, Paris S, Stock WG. HistCite: A software tool for informetric analysis of citation linkage. *Inf-Wiss Prax*. 2006;**57**(8):391-400. doi:10.1515/iwp-2006-057
9. Batagelj V, Mrvar A. Pajek-program for large network analysis. *Connections*. 1998; **21**(2):47-57. Available at: https://vlado.fmf.uni-lj.si/pub/networks/doc/pajek.pdf [Accessed August 28, 2024]_
10. Börner K, Herr II BW, Combs J, Fortin JP, Paley J, Theriault T, et al. The Science of Science (Sci2) Tool. Cyberinfrastructure for Network Science Center, Indiana University. 2009. Available at https://sci2.cns.iu.edu/docs/Sci2_Handout.pdf [Accessed August 28, 2024]_
11. Osińska V, Klimas R. Mapping science: tools for bibliometric and altmetric studies. *Quart Uni Borås* . 2021;**26**(4):909. Available at https://informationr.net/ir/26-4/paper909.html [Accessed August 28, 2024]
12. Xu H, Winnink J, Yue Z, Zhang H, Pang H. Multidimensional Scientometric indicators for the detection of emerging research topics. *Technol Forecast Soc*. 2021;**163**:120490. doi:10.1016/j.techfore.2020.120490
13. Olechnicka A, Ploszaj A, Celińska-Janowicz D. The geography of scientific collaboration. 1st Edition. 2018 London, Routledge. https://doi.org/10.4324/9781315471938 [Accessed August 28, 2024]
14. Tang A, Li KK, Han S et al. Amplifying research influence through the social network, open access publishing, and international collaboration: A mediation analysis on nursing studies literature. *J Nurs Scholarsh*. 2023;**55**(2):477-483. doi: 10.1111/jnu.12827.
15. Kasaraneni H, Rosaline S. Automatic merging of Scopus and Web of Science data for simplified and effective bibliometric analysis. *Ann Data Sci*. 2024;**11**(3):785-802.
16. Green R. Locating sources in humanities scholarship: The efficacy of following bibliographic references. *Libr Quart*. 2000;**70**(2):201-29. doi: 10.1086/630018
17. Szomszor M, Adams J, Fry R, Gebert C, Pendlebury DA, Potter RWK, Rogers G. Interpreting Bibliometric Data. *Front Res Metr Anal*. 2021;5:628703. doi: 10.3389/frma.2020.628703.
18. Thelwall M, Kousha K, Stuart E, Makita M, Abdoli M, Wilson P, Levitt J. Do bibliometrics introduce gender, institutional or interdisciplinary biases into research evaluations? Research Policy. 2023;**52**(8):104829. doi:10.1016/j.respol.2023.104829
19. Gumpenberger C, Glänzel W, Gorraiz J. The ecstasy and the agony of the altmetric score. *Scientometrics*. 2016;**108**(2):977-982. doi: 10.1007/s11192-016-1991-5
20. Sugimoto CR, Work S, Larivière V, Haustein S. Scholarly use of social media and altmetrics: A review of the literature. *J Assoc Inf Sci Technol*. 2017;**68**(9):2037-62.
21. Torres-Salinas D, Orduña-Malea E, Delgado-Vázquez Á, Gorraiz J, Arroyo-Machado W. Foundations of Narrative Bibliometrics. *J Informetr*. 2024;**18**(3):101546. doi: 10.1016/j.joi.2024.101546
22. BioRender. Available at https://www.biorender.com/ [Accessed August 29th, 2024], 2024.