

Invited State-Of-The-Art Review

Five major challenges for medical bibliometrics

Jens F. Rehfeld

Department of Clinical Biochemistry, Copenhagen University Hospital – Rigshospitalet

Dan Med J 2026;73(1):A09250723. doi: 10.61409/A09250723

ABSTRACT

Bibliometrics, in terms of counting articles and books, has always existed. Since 1955, however, additional parameters such as article citations, impact factors and h-indexes have accompanied science – not least medical science. Today, bibliometrics is big business where private companies sell all kinds of bibliometric data. To handle these, it is essential to become aware of the limitations and pitfalls in bibliometrics. Accordingly, the present review describes five major challenges under the headings: Quality; Impact; Co-authorship; Databases, and Fraud.

KEY POINTS

Five challenges for bibliometrics

- Bibliometrics does not provide measures of scientific quality.
- There is no causal relationship between article citedness and journal impact factor.
- Some articles have too many authors because non-contributing authorships have become a trading commodity among scientists.
- Databases that deliver bibliometrics are grossly divergent and inaccurate in the registration of scientific articles.
- The number of predatory journals and “paper mills” that sell low-quality and fraudulent articles is rapidly growing.

Every research-minded physician knows the concept of bibliometrics, occasionally termed scientometrics. Basically, bibliometrics attempts to quantify scientific productivity by counting the number of the primary scientific products, the scientific publications. Such measurements are widely used to evaluate individual researchers, research groups, scientific institutes, entire universities and even entire nations.

Science and scientific publications, however, have grown exponentially over the past 70 years.

Considering the recent growth, especially in densely populated countries like China and India, the world now has many millions of scientists. Correspondingly, the number of scientific journals and publications has increased dramatically. Moreover, the social and political impact of science has become obvious and critically important for modern societies dealing with issues related to health, social welfare, economy, technology, industry, agriculture and the climate, among others. In line with this development, the need for reliable and distinctive bibliometrics has grown and now extends beyond counting scientific articles, books, PhD theses and patents. More sophisticated forms of citation analysis include the number of citations (with and without self-citations), citations only in certain high-impact journals, journal impact factor (JIF) scores, and h-index variations. These measures have become commodities in a vast industry of private companies that offer a range of individual, institutional, national and international rankings. Naturally, the rankings nourish questionable competitions.

A fundamental requirement for bibliometric measurements is, of course, their reliability, which is closely related to their value. As described below, however, bibliometrics now faces an increasing number of serious problems and pitfalls that question its value – and hence its usefulness and very existence.

Bibliometric challenges

I. Quality

Reliability and reproducibility are fundamental attributes of scientific results. Their quality depends both on the significance of the reported discovery, the accuracy of the methods employed and common sense and objectivity in the interpretation of the results. Finally, clarity and language in the mediation of the results, as well as the selection of relevant publication sites (journal or book), also matter in the evaluation of quality. So far, however, the only way of assessing quality has been “peer review” in its broadest sense. By definition, in bibliometrics, quantitative measures do not measure quality. Of course, citation analysis may provide information about breakthrough potential, scientific fashion and popularity of specific issues. And high h-indexes for individual scientists may indicate consistency and engagement, but they also vary across fields of research. Again, however, evaluation of scientific quality as such requires “peer review” [1-3].

II. Impact

Eugene Garfield is considered the father of bibliometrics. In 1955, he founded the Institute for Scientific Information in Philadelphia, which published the citation indexes which constitute the basis of citation analysis [4]. Subsequently, in his further refinement of citation analysis, he invented the concept of Journal Impact Factor (JIF) as a tool for editors and publishers of scientific journals. JIF provides information about how well articles in a given journal, on average, are cited. That is, of course, useful information for journal editors and publishers, because citations indicate the general impact of a given journal. However, unfortunately, JIF has also been widely misused

by individual scientists as an author impact factor rather than a journal impact factor. Even Garfield has regretted and warned against this development [5]. The misuse stems from the fact that JIF is based on average citations. In other words, many so-called high-impact journals achieve their ranking through relatively few groundbreaking articles whose high citation counts elevate the other, less significant articles in the journal. Therefore, there is no causal relationship between article citedness and journal impact, as previously pinpointed and quantified [1, 3, 6].

III. Co-authorship

A generation or two ago, leading medical scientists were single authors on most of their publications. Occasionally, they included one or two co-authors. Well-known local examples include Nobel Prize winner Jens Chr. Skou (1918-2018) and Nobel-nominated Jørgen Lehmann (1898-1989). Jørgen Lehmann, the discoverer of the drugs dicumarol and para-aminosalicylic acid (PAS) in tuberculosis therapy, had published 129 articles, of which 75 were proper original papers. Only 14 had a co-author, but Lehmann was always the first author [7]. The publications of Jens Chr. Skou showed a similar pattern. He single-authored the pioneering papers describing the discovery of the sodium pump (for which he won the Nobel Prize). In his later original articles, he occasionally had one co-author. Only one of his papers had three authors. Skou later regretted that co-authorship because he considered his own contribution to the study less significant [8].

Leading scientific journals such as *Nature* and *Science* and medical journals like the *New England Journal of Medicine* and the *Lancet* still publish original articles with few authors, but far more articles today have many. There are several examples of articles with hundreds of co-authors, and some even with thousands. Such multi-authorships may be justified in highly complex, long-lasting studies in particle physics, climate studies, genetics, epidemiology, studies of rare diseases and perhaps in a few drug trials. Some studies in these fields happen to require several unique panels of specialised technology and may consequently have many authors. Some multiauthor articles, however, have addressed the co-author situation in confusing ways. For instance, with triple-divided author groups: First, a group of 10-15 scientists whose names appear as usual at the top of the article as authors. Additionally, an asterisk at the end of the article informs that the above-mentioned author group drafted the paper on behalf of, say, 150 contributors of data to the study. Their names and affiliations were also listed at the end of the paper. However, the readers were further informed that the paper was a product of a specific consortium counting 1,200 members, whose names were also printed (although in small print) in a supplement. Several of the consortium members also included the paper on their list of publications and scored the citations, because their names had been printed with the publication. Such examples and authorship practices push bibliometrics into the grey zone of misconduct. The described example is not unique. Besides, it is also quite common that articles with even modest numbers of authors include the names of persons with limited, if any, factual contribution to the studies, so-called “grace authorships”.

Multi-authorships may question the justification of bibliometric scores among authors. A radical

way of limiting the number of the “passive” or “grace” co-authors on publications could be a simple rule according to which the number of citations for a given paper is divided into equal fractions by the authors, irrespective of position in the author sequence. Hence, for a paper with 100 citations and 20 authors, each author should obtain only five citations. Such a rule would – in the hands of authentic first and last authors – probably limit the number of mid-sequence authors who have not contributed substantially. In the delicate process of excluding authors with very limited or no real contributions, referral to the acknowledgements section in the article may be an option.

IV. Database accuracy

There are several databases for bibliometric metrics (article counts, citations, h-indexes, etc.). For decades, it has been a significant problem that the data released from these bases deviates considerably for the same scientist and the same papers. Some explanations suggest that databases with small numbers (for instance, “Web of Science”) are more critical and include only original articles and review articles from journals, while excluding book chapters. Maybe they also exclude articles in low-impact journals. Conversely, databases with larger numbers (such as “Google Scholar”) are assumed to include non-scientific papers (e.g., newspaper articles and feature articles).

A recent experience with references in a review published in a respected medical journal exemplifies the problem [9]. In the proof version of the manuscript, the publisher of the journal (John Wiley & Sons Ltd.) had indicated in which of four well-known databases (“Chemical Abstract Services (CAS)”; “Google Scholar”; “PubMed”; Web of Science (WoS)), each of the 189 references appeared. CAS cited 151 references (80%); Google Scholar cited 189 (100%); PubMed cited 151 (80% – but with some references different to those included by CAS); and WoS cited 121 (63%). However, no systematics could explain the differences. This point is illustrated by the 11 references to original articles in the esteemed *Am. J. Physiol.*: Ten were cited by CAS (91%); all were cited by Google Scholar (100%); PubMed cited only four (36%); and WoS five (45%). Similar patterns were found for references from *J. Biol. Chem.* and, e.g., the *Lancet*. Thus, the discrepancies are gross and haphazard. The example shows that only Google Scholar citations were complete, whereas both PubMed and WoS citations were falsely low.

V. Fraud

The growth in science and scientific publications has both benefits and drawbacks. It is unfortunately accompanied by rapidly growing metastatic side effects such as severe misconduct (fabrication of data, falsification of data and plagiarism). This growth may have been facilitated by today’s easier communication via the internet and open-access publishing. Fraud in bibliometrics occurs because the number of publications, h-indexes and citations is believed by many – often younger scientists – to be decisive promoters of their career.

In a remarkable recent article by Proceedings of the National Academy of Sciences of the United

States of America (PNAS), Richardson et al. addressed a new reality of science [10]. They mention first that scientific contributions are “increasingly evaluated by potentially misleading proxies such as the h-index, journal impact factor, university rankings and scientific prizes” [2, 3]. Subsequently, they describe how these proxies have resulted “in increasing competition and inequality in how resources and rewards are distributed, which leaves the scientific enterprise more susceptible to defection” [11, 12]. This defection includes, e.g., 3.8% of 20,000 articles published between 1995 and 2014 containing duplicated images [13]; recent entities denoted “paper mills” that sell mass-produced low-quality and fabricated research articles” [11, 14]. Moreover, “Agents for some paper mills have tried to bribe journal editors and hijack entire editorial processes”. To add insult to injury, the Richardson article concludes that “the number of fraudulent publications is growing at a rate far exceeding the growth of legitimate publications”.

It has been suggested that systematic production of fake papers mainly occurs outside Europe and North America, i.e., in large Asian countries such as China, India, Indonesia and Pakistan. However, Richardson et al. show that major “production of low-quality and fraudulent science can occur anywhere” [15-17]. That is a lesson to be learned also in Danish medical science.

Conclusion

The five shortcomings described above cast doubt on the value of bibliometric parameters. On the other hand, the bibliometric industry has grown economically worldwide, to what appears to be a billion-dollar business. It is therefore naïve to believe that such a golden enterprise should collapse in the near future. Hopefully, however, the misuse and plain criminal fraud activities in the enterprise will diminish. For genuine science, it is important to minimise reliance on bibliometrics as much as possible and to approach what cannot be avoided with a great deal of scepticism. Moreover, careful “peer review” should be encouraged at all levels of scientific evaluations (manuscripts, research positions, funding applications, support for research institutes (public and private), university research etc.). Along this line, the San Francisco Declaration on Research Assessment (DORA) from 2012 should be implemented as much and far as possible [18, 19].

Correspondence *Jens F. Rehfeld*. E-mail: jens.frederik.rehfeld@regionh.dk

Accepted 26 November 2025

Published 12 December 2025

Conflicts of interest none. The author has submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. This is available together with the article at ugeskriftet.dk/DMJ.

References can be found with the article at ugeskriftet.dk/DMJ

Cite this as Dan Med J 2026;73(1): A09250723

doi 10.61409/A09250723

Open Access under Creative Commons License [CC BY-NC-ND 4.0](#)

REFERENCES

1. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ*. 1997;314:498-502 <https://doi.org/10.1136/bmj.314.7079.497>
2. Belter CW. Bibliometric indicators: opportunities and limits. *J Med Libr Assoc*. 2015;103:219-221 <https://doi.org/10.3163/1536-5050.103.4.014>
3. Siler K, Larivière V. Who games metrics and rankings? Institutional niches and journal impact factor inflation. *Res Policy*. 2022;81:104608 <https://doi.org/10.1016/j.respol.2022.104608>
4. Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*. 1955;122:108-111 <https://doi.org/10.1126/science.122.3159.108>
5. Garfield E. The history and meaning of journal impact factor. *JAMA*. 2000;295:90-93 <https://doi.org/10.1001/jama.295.1.90>
6. Seglen PO. Causal relationship between article citedness and journal impact. *J Am Soc Inf Sci*. 1994;45:1-11 [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<1::AID-ASI1>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<1::AID-ASI1>3.0.CO;2-Y)
7. Rehfeld JF. Jørgen Lehmann – den kliniske biokemiker. *Klin Biokemi Norden*. 2005;17:9-17 (ISSN 1101-2013)
8. Skou JC. Om heldige valg. Eller hvad frøer, krabber og hajer også kan bruges til. Aarhus University Publ. 2013 (ISBN: 9788779344761)s
9. Rehfeld JF. Cholecystokinin: Clinical aspects of the new biology. *J Intern Med*. 2025;298:251-267 <https://doi.org/10.1111/joim.20110>
10. Richardson RAK, Hong SS, Byrne JA, Stoeger T, Nunes Amaral LA. The entities enabling scientific fraud at scale are large, resilient and growing rapidly. *Proc Natl Acad Sci U S A*. 2025;122(32):e2420092122. <https://doi.org/10.1073/pnas.2420092122>
11. Byrne JA, Abalkina A, Akinduro-Aje O, et al. A call for research to address the threat of paper mills. *PLoS Biol*. 2024;22:e3002931. <https://doi.org/10.1371/journal.pbio.3002931>
12. Abalkina A, Aquarius R, Bik E, et al. “Stamp out paper mills” – science sleuths on how to fight fake research. *Nature*. 2025;637:1047-1050 <https://doi.org/10.1038/d41586-025-00212-1>
13. Bik EM, Casadevall A, Fang FC. The prevalence of inappropriate image duplication in biomedical research publications. *mBio*. 2016;7:e00809-16. <https://doi.org/10.1128/mBio.00809-16>
14. COPE, STM. Paper mills – research report from COPE & STM. 2022. <https://doi.org/10.24318/jtbG8IHL>
15. Piller C. Picture imperfect. *Science*. 2024;385:1406-1412. <https://doi.org/10.1126/science.adt3535>
16. O'Grady C. The reckoning. *Science*. 2024;383:1046-1051. <https://doi.org/10.1126/science.adp0437>
17. Kupferschmidt K. Researcher at the center of an epic fraud remains an enigma to those who exposed him. *Science*. 2018;80:9281-9294. <https://doi.org/10.1126/science.aav1079>
18. Weston K. Kicking the impact factor habit: how institutes and funders are turning the DORA theory into practice. *EMBO Rep*. 2018;39:11
19. Rehfeld JF, Gøtze JP. DORA versus JIF: et spørgsmål om kvalitetsvurdering af forskning. *Ugeskr Laeger*. 2018;180:2312-2313