https://doi.org/10.51408/issi2025_132

Implicit Reporting Standards in Bibliometric Research: What Can Reviewers' Comments Tell Us About Reporting Completeness?

Dimity Stephen[1], Alexander Schniedermann[2], Andrey Lovakov[3], Marion Schmidt[4], Matteo Ottaviani[5], Nikita Sorgatz[6], Roberto Cruz Romero[7], Torger Möller[8], Valeria Aman[9], Stephan Stahlschmidt[10]

[1]*stephen@dzhw.eu*, [2]*schniedermann@dzhw.eu*, [3]*lovakov@dzhw.eu*, [4]*schmidt@dzhw.eu*, [5]*ottaviani@dzhw.eu*, [7]*cruzromero@dzhw.eu*, [8]*moeller@dzhw.eu*, [9]*aman@dzhw.eu*
German Centre for Higher Education Research and Science Studies, Berlin (Germany)

[6]*nikita.sorgatz@hu-berlin.de*
German Centre for Higher Education Research and Science Studies, Berlin (Germany)
Robert K. Merton Center for Science Studies, Humboldt-Universität zu Berlin, Berlin (Germany)

[10]*stahlschmidt@dzhw.eu*
German Centre for Higher Education Research and Science Studies, Berlin (Germany)
Unit of Computational Humanities and Social Sciences (U-CHASS), EC3 Research Group, University of Granada, Granada (Spain)

Abstract

The rapid growth in the number of bibliometric studies in recent years has been accompanied by increasing diversity in the quality of the reporting of these studies' methodologies and results. This ongoing study explores and systematises the quality and completeness of reporting bibliometric research using a bottom-up approach based on open peer review. We first identified 89 bibliometric studies published in library and information science (LIS) journals and conference proceedings and non-LIS journals, and then retrieved the 194 corresponding first-round reviews. From these reviews we extracted 968 reviewer comments pertaining to aspects of reporting the details of these studies, and inductively classified these comments into 11 broad thematic categories and 68 sub-categories. Our preliminary results find that 77% of comments overall and the majority in each broad category were critical, which could be expected given the purpose of peer review to identify opportunities for improvement. In contrast, comments relating to the provision of study data and to the overall assessment of articles were more likely to be positive. The most common themes of reviewers' comments were critically appraising the details of the data, methods, visualisations and tables used, and the clarity of the research questions and text. The finalised results will provide a precise and practical outline of concrete items that should be reported in bibliometric research according to the implicit community standard. Our findings will highlight particular features of bibliometric reporting that could be strengthened, complementing existing initiatives to generate guidance for the complete and accurate reporting of bibliometric studies.

Introduction

Publication output in the field of bibliometrics is growing at an unchecked rate. Larivière (2012) and Jonkers and Derrick (2012) detected a sudden spurt in bibliometric studies in 2003 and growth has only accelerated since then: the number of publications increasing 12-fold from around 800 in 2000-04 to over 10,000 by 2015-19 (González-Alcaide, 2021). Notably, the share of these studies published in library and information science (LIS) journals – the field historically central to

bibliometrics – has steadily decreased over time from around 70% in the 1980s and 1990s to 40% in 2010 (Larivière, 2012) to around 25% in 2019 (González-Alcaide, 2021).

This rapid growth in bibliometric studies may be attributed to several diverse factors. For instance, the prominence of bibliometrics in international, national, and institutional research evaluation and management activities (Cabezas-Clavijo et al., 2023; González-Alcaide, 2021) has raised its profile amongst scholars in all fields. Further, the increasing availability of data sources and analytical software has made bibliometrics accessible to anyone (Cabezas-Clavijo et al., 2023; Boyack, Klavans & Smith, 2022). Viewed cynically, these advances have opened the field to "academic opportunists", who may perceive bibliometric analyses as a quick and easy approach to boosting their publication output (González-Alcaide, 2021). Viewed positively, the self-monitoring capacity in the diverse research fields has been empowered substantially. From either perspective, the prominence and accessibility of bibliometrics has thus generated a wave of interest in our field across disciplines.

While this widespread uptake should be celebrated as an acknowledgment of our field's relevance and potential to contribute broadly to academia, if unchecked, it may also negatively impact the quality, rigour, and development of our field. For instance, our central theories and principles are unlikely to be known to researchers dropping by from other fields to borrow methods and data. Consequently, the bibliometric corpus may be diluted with studies that make minimal contributions to the field or misuse methods and indicators (Jonkers & Derrick, 2012; González-Alcaide, 2021). Individually, such studies are unlikely to have a notable impact on the field. However, in large numbers, they can collectively produce misleading effects, which damages both the theoretical growth of our field and its reputation among academics and policy-makers (Boyack et al., 2022).

Well-documented data and methods are central to the reliability, reproducibility, and robustness of bibliometric studies (Boyack et al., 2022). Evidence of issues in the reporting of bibliometric studies remains currently rather anecdotal. However, a small number of studies that empirically examined reporting quality have found wide variation in the reporting of study characteristics, with good reporting of e.g., search terms, but poor reporting of database characteristics (Koo & Lin, 2023); that substantial numbers of studies lacked the sufficient detail necessary for replicating their findings (Boyack et al., 2022); and that under-reporting of methodological details was widespread in studies both within and outside the LIS field (Cabezas-Clavijo et al., 2023). These findings suggest that the broad community of scholars using bibliometrics could benefit from the guidance in the responsible and effective use of bibliometric data and methods that has long been called for (e.g., Glänzel & Schoepflin, 1994; Glänzel, 1996; González-Alcaide, 2021).

A first step toward providing this guidance is being made with the "Guidance List for repOrting Bibliometric AnaLyses" (GLOBAL) project, which seeks to implement reporting guidelines for bibliometric studies (Ng et al., 2023). GLOBAL comprises a scoping review for existing reporting recommendations and then harnesses the bibliometric community's expertise in developing guideline content.

Establishing and maintaining this continuously evolving shared set of concepts not only facilitates scientific communication, laying the groundwork for progress, but also has the potential to shape education and training in bibliometrics methods. The examination of current reporting standards therefore serves as a critical reflection of our methods, and a consequent broad discussion enables the professional community to agree upon claims for authority and legitimisation and to continue former work to develop the field (American Society for Cell Biology, 2012; Hicks et al., 2015).

Research aims and approach

The aim of this study is to explore and systematise problems in the quality and completeness of reporting bibliometric research. We do so by investigating the question, what reporting issues are identified by peer reviewers in their reviews of bibliometric studies? Our approach is to qualitatively examine peer reviews of bibliometric studies and identify aspects that reviewers raise as well- or poorly reported. For example, reviewers may ask for additional information regarding databases, sample sizes, search terms, filter criteria, or the indicators used, suggesting the provided details were insufficient for understanding or reproducing the study. Instead of pre-defining a set of subjectively ideal reporting criteria, our approach focuses on issues that have been identified by diverse academic peers in open peer review procedures at both central and peripheral bibliometric outlets. As such, our inductive and descriptive approach facilitates a discussion of what features of bibliometrics-based studies the community criticises (or compliments), complementing parallel efforts to jointly define reporting standards in a top-down approach.

Methods

Identification and sampling of bibliometric studies

The methods applied in the study are shown in Figure 1. We used a two-step process to identify bibliometric studies in journals and conference proceedings. We defined a bibliometric study as a study that used a bibliometric data source (e.g., WoS, Scopus) and one or more metadata fields (e.g., journal, discipline) to compare two or more entities or groups (e.g., authors, institutions, countries) to contribute knowledge to its field (e.g., one database covers more journals than another or one institution is more productive than another).

Our sampled studies included articles in LIS and non-LIS journals, and submissions to a LIS conference. As noted, a substantial number of bibliometric studies now appear in non-LIS journals and these may be reviewed by peers less familiar with the details necessary to sufficiently report a bibliometric study than reviewers of articles in LIS journals, which potentially increases the diversity of aspects raised. Similarly, conference papers are usually shorter in length than articles and may contain fewer methodological details and results than articles, and so reviewers may highlight particularly important features when these are missing. This sample of both articles and conference submissions may thus capture a wide array of issues raised by reviewers, aligning with the interdisciplinarity present in bibliometrics.

We first required a sample of bibliometric studies in each category (NLIS article, LIS article, LIS conference submission) with open peer reviews. The 27th International Conference on Science, Technology and Innovation Indicators (STI 2023) used an open peer review process, so we used these submissions to represent conference proceedings (sample A). To identify bibliometric studies in these submissions, we extracted the title, abstract and keywords (TAK) of each submission from the Orvium website using the rvest (Wickham, 2022a), tidyverse (Wickham, et al. 2019), jsonlite (Ooms, 2014), and stringr (Wickham, 2022b) R packages. We then narrowed the submissions to those that used a bibliometric data source by searching for any of the following (case insensitive) terms in the TAK: Web of Science, WoS, Scopus, Dimensions, Openalex, Open Alex, Pubmed, Crossref, SciELO, Wikidata, Overton, altmetric, bibliometric data, DOAJ. We then manually screened the full-texts of these submissions to assess whether they fulfilled the aforementioned criteria of using a metadata field and comparing groups to make a knowledge claim, and retained those that fulfilled these criteria as bibliometric studies.

To identify N/LIS articles with open peer reviews, we performed two searches of the online WoS database on 29 February 2024, including the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts & Humanities Citation Index. First, to identify LIS articles (sample B), we searched for any of the aforementioned bibliometrics data sources in the Topic (TAK) field. In addition, we restricted the publication years to 2018-2024, the WoS Subject Category to "Information & Library Science", the document type to article, and filtered the results to those articles that were open access (OA) and had open peer reviews available. We performed the same search to identify non-LIS articles (sample C), with the following changes: Category was not "Information & Library Science", the title did not include “Protocol”, Dimensions and Pubmed were removed, and “scientometric” was added as a search term, as we observed authors to use bibliometric and scientometric interchangeably. “Protocol” was excluded to remove study protocols. Dimensions was removed because it is unlikely to refer to the database outside of LIS, and Pubmed was removed as its inclusion returned many out-of-scope systematic reviews. For both samples, we then manually screened the studies’ full-texts to retain those that fulfilled our criteria as bibliometric studies. We then downloaded the first-round peer reviews for all bibliometric studies via the “Open Peer Reviews” link in WoS.

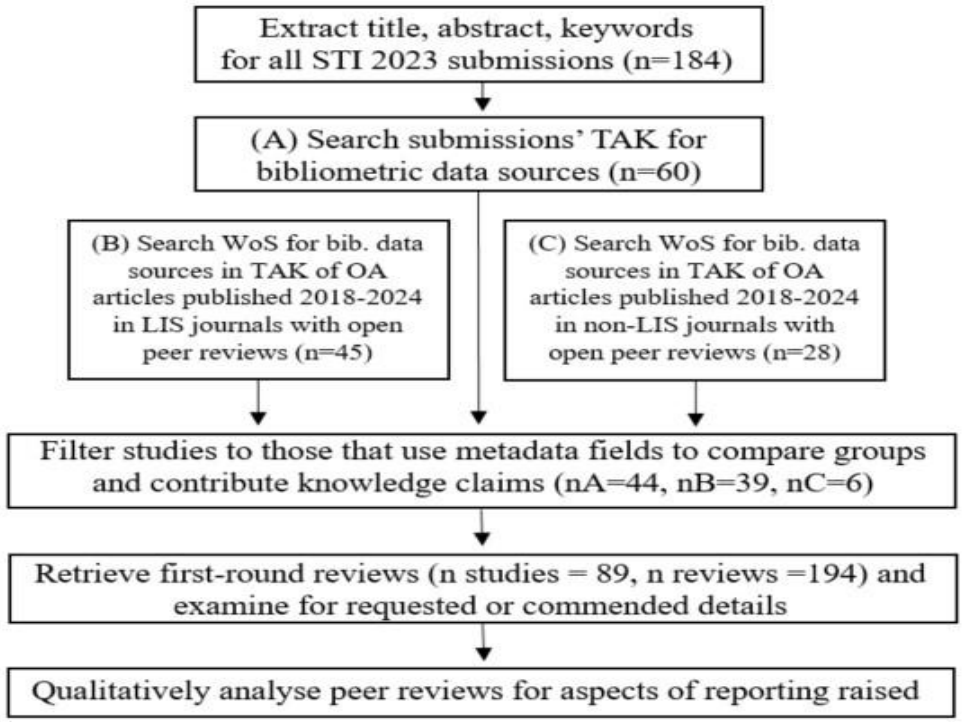

Figure 1. Flowchart of method.

Qualitative analysis of reviewer comments

To prepare for our qualitative analysis of the peer reviewers' comments, we first extracted all comments pertaining to reporting a bibliometric study from the reviewers' reports for all three samples. In this process, each team member examined and extracted comments from approximately 20 peer reviews. The comments could be positive, such as praise for the clear or detailed description of the methodology; negative, such as critiquing the study's limitations; or neutral, such as suggestions for additional references. At this stage, we aimed to collect as much information as possible and filter out irrelevant information later in the analysis.

We then categorised the comments into broad themes based on the overarching concept of the comment. Here, in a group process, we discussed the comments' focus and identified and allocated comments to one or more high-level categories. To enhance the specificity of the concepts addressed, we as a group then further assessed the comments in each category and identified a set of more specific sub-categories. For instance, the reviewer comment "What is the unit for y-axis in Figure 7?" was first assigned to the broad category of *Visualisations and Tables* and then sub-categorised to *(Un)clear presentation*. Sub-categories were named neutrally as comments could be positive, neutral, or negative. Once classified, all comments in each sub-category were reviewed for consistency and reclassified to other or new sub-categories as required. In this way, we inductively classified all comments to both broad categories and more specific sub-categories based on the concept addressed in the comment.

Results

The total sample examined consisted of 194 reviews of 89 bibliometric studies: 11 reviews of 6 studies published in NLIS journals, 79 reviews of 39 studies published in LIS journals, and 104 reviews of 44 LIS conference papers. The LIS articles were all published in Quantitative Science Studies, as the only WoS-indexed LIS journal with open peer review. The NLIS studies were published in six journals: Ecological Solutions and Evidence, Engineering Reports, Environmental Research Letters, Internet Technology Letters, Journal of Oral Rehabilitation, and Royal Society Open Science. The low number of NLIS studies occurs as the open peer review restriction severely limited the sample. On average, reviews of conference papers were 287 words in length (range = 31-1,091 words), which was – as could be expected – shorter than article reviews. Reviews of bibliometric studies in the LIS journals were notably longer (mean = 710 words, range = 76-2,605) than articles in NLIS journals (mean = 536 words, range = 35-2,062).

The initial coding of the reviews identified 1,030 relevant comments. Sixty-two comments were later deemed to be out of scope of the analysis and removed, leaving 968 comments in scope. The first classification process identified 11 broad themes: *Clarity and validity of concepts*; *Clarity of presentation*; *Description of data/methods*; *Description of results*; *Visualisations and tables*; *Limitations*; *Conclusions*; *Open Science/Reproducibility*; *Declarations*; *Links to literature/references*; and *Overall assessment*. The second classification process identified 68 sub-categories of these themes. Table 1 shows the number and percentage of comments in the 11 broad categories and the number and percentage of each category's comments that were negative (i.e. critical of the manuscript), neutral, or positive. As comments could be classified to more than one category, the total count of comments exceeds 968.

Table 1. The number and percentage of comments in the 11 broad categories and the number and percentage of comments in each category that were positive, neutral, or negative, ordered by the total number of comments.

Category	*No. (%) comments*	*No. (%) negative*	*No. (%) neutral*	*No. (%) positive*
Description of data / methods	329 (29.1)	287 (87.2)	5 (1.5)	37 (11.2)
Clarity of presentation	139 (12.3)	89 (64.3)	0 (0.0)	50 (35.7)
Visualisations and tables	136 (12.0)	118 (86.8)	6 (4.4)	12 (8.8)
Description of results	131 (11.6)	111 (84.7)	11 (8.4)	9 (6.9)
Overall assessment	118 (10.5)	51 (42.9)	0 (0.0)	67 (56.8)
Links to literature / references	83 (7.4)	69 (84.1)	2 (2.4)	12 (14.5)
Clarity and validity of concepts	62 (5.5)	57 (91.9)	0 (0.0)	5 (8.1)
Conclusions	59 (5.2)	54 (91.5)	0 (0.0)	5 (8.5)
Open science / reproducibility	42 (3.7)	20 (47.6)	0 (0.0)	22 (52.4)
Limitations	29 (2.6)	20 (69.0)	2 (6.9)	7 (24.1)
Declarations	1 (0,1)	1 (100.0)	0 (0.0)	0 (0.0)
Total	1,129 (100)	877 (77.7)	26 (2.3)	226 (20.0)

Nearly a third of reviewers' comments pertained to the authors' description of the data and or methods used in the study (329, 29.1%), the majority of which (87.2%)

were critical, while 11.2% of comments praised the methodological information presented. The next most common comments regarded the clarity of the information presented (12.3%), and the visualisations and tables used (12.0%). In the former category around two-thirds of comments were critical of, for instance, the clarity of research questions and the structure of the text, while one-third of comments regarded these features positively. Comments regarding the content and presentation of visualisations and tables, however, were largely critical (86.8%). Similarly, overall, 77.7% of comments were critical of the manuscripts' reporting, which aligns with the aim of peer review to identify potential issues and suggest improvements to the authors. In contrast, comments relating to open science/reproducibility (e.g., the provision of the data and or scripts used in the study) and the overall assessment of the study (e.g., its contextualisation in the existing literature, appropriateness of its design to address the research question, and its originality, utility, and relevance) were more often positive than negative. However, this latter instance may have been influenced by the fact that all articles examined were eventually accepted for publication.

These preliminary results provide initial insights into the issues raised and details praised by reviewers of bibliometric studies. This study is ongoing and we intend to finalise the qualitative analysis of the reviewers' comments, particularly the sub-category level, which will provide greater granularity of the themes discussed in the comments and highlight specific aspects of the reporting of bibliometric studies that should be addressed by authors. Further, we plan to compare the theme and prevalence of comments between articles and conference submissions and between NLIS and LIS articles to investigate potential differences in reviewers' focus or authors' reporting between groups. Finally, we plan to distill the results into a precise and practical list of concrete items that should be reported in bibliometric research according to the implicit community standard, and present this for discussion at the conference.

We anticipate that our results will provide a descriptive and inductive perspective of the aspects of reporting bibliometric studies raised by peer reviewers. This will highlight particular features of bibliometric reporting that could be strengthened and complement initiatives such as GLOBAL, which take an expert-based top-down approach to generating guidance in complete and accurate reporting of bibliometric studies. The availability and up-take of such guidance could enhance the reliability, reproducibility, and robustness of bibliometric studies.

References

American Society for Cell Biology. (2012). San Francisco declaration on research assessment (DORA). https://sfdora.org/

Boyack, K. W., Klavans, R., & Smith, C. (2022). Raising the bar for bibliometric analysis. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), 26th International Conference on Science and Technology Indicators, STI 2022. sti22143. DOI: 10.5281/zenodo.6975632.

Cabezas-Clavijo, A., Milanés-Guisado, Y., Alba-Ruiz, R., & Delgado-Vázquez, A. M. (2023). The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and

Scientometrics. Journal of Data and Information Science, 8(4), 10–35. DOI: 10.2478/jdis-2023-0021.
Glänzel, W. (1996). The need for standards in bibliometric research and technology. Scientometrics, 35(2), 167–176. DOI: 10.1007/BF02018475.
Glänzel, W., & Schoepflin, U. (1994). Little scientometrics, big scientometrics... and beyond? Scientometrics, 30(2–3), 375–384. DOI: 10.1007/BF02018107.
González-Alcaide, G. (2021). Bibliometric studies outside the information science and library science field: uncontainable or uncontrollable? Scientometrics, 126(8), 6837–6870. DOI: 10.1007/s11192-021-04061-3.
Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Raflos, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. Nature, 520, 429–431. https://www.nature.com/articles/520429a.
Jonkers, K. & Derrick, G. (2012). The bibliometric bandwagon: Characteristics of bibliometric articles outside the field literature. Journal of the American Society for Information Science and Technology, 63(4), 829-836. DOI: 10.1002/asi.22620.
Koo, M., & Lin, S-C. (2023). An analysis of reporting practices in the top 100 cited health and medicine-related bibliometric studies from 2019 to 2021 based on a proposed guidelines. Heliyon, 9(6), e16780. DOI: 10.1016/j.heliyon.2023.e16780.
Larivière, V. (2012). The decade of metrics? Examining the evolution of metrics within and outside LIS. Bulletin of the American Society for Information Science and Technology, 38(6), 12-17. DOI: 10.1002/bult.2012.1720380605.
Ng, J. Y., Haustein, S., Ebrahimzadeh, S., Chen, C., Sabe, M., Solmi, M., & Moher, D. (2023). Guidance List for reporting bibliometric analyses (GLOBAL): A research protocol. Open Science Framework. DOI: 10.17605/OSF.IO/MTXBF.
Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between JSON data and R objects. arXiv. DOI: 10.48550/arXiv.1403.2805.
Wickham, H. (2022a). rvest: Easily Harvest (Scrape) Web Pages. (R package version 1.0.3). https://CRAN.R-project.org/package=rvest.
Wickham, H. (2022b). stringr: Simple, Consistent Wrappers for Common String Operations. (R package version 1.4.1). https://CRAN.R-project.org/package=stringr.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R, et al. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686.