Research Paper

The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics

Alvaro Cabezas-Clavijo[1], Yusnelkis Milanés-Guisado[2], Ruben Alba-Ruiz[3†], Ángel M. Delgado-Vázquez[2]

[1]Universidad Internacional de La Rioja (UNIR), Avenida de la Paz, 137, Logroño, 26006, Spain
[2]Universidad Pablo de Olavide, Ctra. de Utrera, Km 1, Sevilla, 41013, Spain
[3]Hospital Universitario Virgen de las Nieves, Avenida de las Fuerzas Armadas, 2, Granada 18014, Spain

Abstract

Purpose: The aim of this article is to explore up to seven parameters related to the methodological quality and reproducibility of thematic bibliometric research published in the two most productive journals in bibliometrics, Sustainability (a journal outside the discipline) and Scientometrics, the flagship journal in the field.

Design/methodology/approach: The study identifies the need for developing tailored tools for improving the quality of thematic bibliometric analyses, and presents a framework that can guide the development of such tools. A total of 508 papers are analysed, 77% of Sustainability, and 23% published in Scientometrics, for the 2019-2021 period.

Findings: An average of 2.6 shortcomings per paper was found for the whole sample, with an almost identical number of flaws in both journals. Sustainability has more flaws than Scientometrics in four of the seven parameters studied, while Scientometrics has more shortcomings in the remaining three variables.

Research limitations: The first limitation of this work is that it is a study of two scientific journals, so the results cannot be directly extrapolated to the set of thematic bibliometric analyses published in journals from all fields.

Practical implications: We propose the adoption of protocols, guidelines, and other similar

† Corresponding author: Ruben Alba-Ruiz (ruben.alba.ruiz@gmail.com; ORCID: 0000-0002-8194-2264).

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

tools, adapted to bibliometric practice, which could increase the thoroughness, transparency, and reproducibility of this type of research.

Originality/value: These results show considerable room for improvement in terms of the adequate use and breakdown of methodological procedures in thematic bibliometric research, both in journals in the Information Science area and journals outside the discipline.

Keywords: Thematic bibliometric analyses; Sustainability; Scientometrics; Reproducibility; Methodological quality

1 Introduction

The number of papers using bibliometric methods has skyrocketed in recent years, with many more bibliometric articles published in journals outside the field of Information Science (IS) than within the area (González-Alcaide, 2021). Scientometric techniques have become popular in recent years because of the intensive use of this type of indicators in institutional evaluations, academic reports, and even in the general press (Petrovich, 2022); thus, they are used not only by experts in bibliometrics, but also by all types of academics and professionals. To this must be added the improvement of coverage and relevant data available through major providers such as Scopus or Web of Science (Pranckutė, 2021), and the flourishing of tools that are easy to access and use, in many cases free of charge (Bibliometrix or VosViewer, among others), both for the automated analysis of the results and for their visualisation (Moral-Muñoz et al., 2020). The fact that bibliometric studies do not require the involvement of human research subjects may be also a relevant factor in the decision of performing this type of analyses. Likewise, early-career scholars and PhD students may be advised by supervisors to carry out bibliometric studies to get an overall perspective of a particular research topic.

One of the clearest manifestations of this phenomenon is the performance of bibliometric analyses on a particular subject or subject domain, which are usually published primarily in journals outside the IS area (Ellegaard & Wallin, 2015). From the bibliometric community, it is perceived that these works in journals outside the discipline may lack the quality controls and standards required in the journals of the area (Jonkers & Derrick, 2012); therefore, some authors demand greater rigour and critical sense prior to publication (González-Alcaide, 2021).

If we consider the period 2019-2021, of the ten most productive journals in bibliometrics, only two belong to the IS area. Scientometrics, the flagship journal in the field of bibliometrics, published by Springer, is the outlet with the highest number of articles on bibliometrics, followed by two mega-journals: Sustainability, and

International Journal of Environmental Research and Public Health (IJERPH), both published by Multidisciplinary Digital Publishing Institute (MDPI). Nine of the ten most productive journals in this field are in the first or second journal impact factor quartile of their discipline, according to data from Journal Citation Reports (JCR) 2021 (Figure 1).
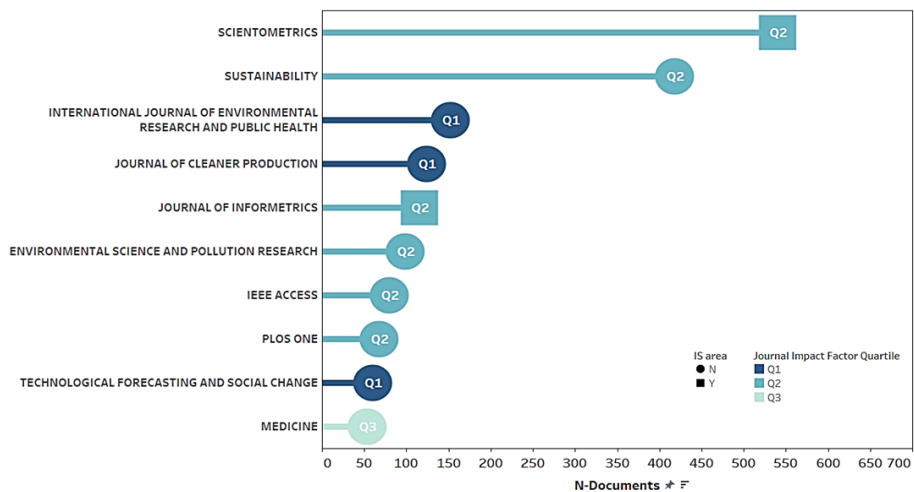

Figure 1.   Journals with the highest bibliometric production (2019-2021).
Source: Own elaboration based on Web ofScience data (SCI and SSCI indexes). All languages are considered. Search equation: TS=(bibliometric* OR scientometric* OR webometric* OR altmetric* OR informetrics* OR "citation analysis" OR "citation study" OR "scholarly productivity" OR "publication analysis" OR "scholarly impact" OR "patent citation") AND PY=(2019-2021) AND DT=(article OR review)[1]. Early Access articles were excluded. The highest quartile of the journal in Journal Citation Reports (JCR) 2021 is indicated.

1.1   Methodological quality and reproducibility of research

In the context of the replication crisis in science (Saltelli & Funtowicz, 2017), measuring the methodological quality of studies becomes particularly important. Thus, Lindner et al. (2018) urged conducting assessments of the quality of the methods used in research and of the reproducibility of the results. They also advocate for replication/confirmation studies. Along the same lines, Moher et al. (2020) proposed as one of their five Hong Kong principles for evaluating researchers to "value the accurated and transparent reporting of all research", and therefore advised the use of reporting guidelines to ensure complete and adequate reporting of the methods used. They also recommended the public availability of raw data, code, and

[1]TS=Topic; PY=Publication Year; DT=Document Type.

The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

materials that allow research to be reproduced. However, studies in sectors such as Business and Management found that only 4.7% of journals were expressly open to accepting replications of previous works (Tipu & Ryan, 2021), and that only between 0.1% and 1.5% of papers published in top Economics and Business and Management journals were replication studies (Mueller-Langer et al., 2019; Ryan & Tipu, 2022). Very low levels of replication are also reported in other Social Science fields, around 1% in Psychology (Makel et al., 2012), and slightly above 0.1% in Education (Makel & Plucker, 2014), albeit with an upward trend in recent years (Perry et al., 2022).

The field of bibliometrics has not been characterised by promoting replication studies, despite the fact that, as Glänzel (1996) rightly pointed out, "reproducibility is one of the basic criteria for any science. Under identical conditions research results should be reproducible in bibliometrics, too. The reproducibility of results can only be guaranteed, if all sources, procedures and techniques are properly documented in scientific publications". Authors such as Harzing (2016), also highlighted that replication studies are essential for the advancement of scientometrics, and even proposed the inclusion of a Replication Studies section in scientific journals. Among the few cases found in the scientometric field, we can cite replication studies on nepotism in peer reviews (Sandström & Hällsten, 2008), or on publications of excellence in Nursing (Nicoll et al., 2020). Due to its similarity to our research, we should point out the replication carried out by Liu (2022) of a thematic bibliometric study, in which she identifies different methodological gaps that significantly affect the results and their interpretation.

In early 2023, the first protocol aimed to provide guidelines for the reporting of bibliometric analyses was launched. The Guidance List for the repOrting of Bibliometric AnaLyses (GLOBAL) "will help promote transparency and completeness in reporting bibliometric and related analyses and provide a framework for authors to report methods and results" (Ng et al., 2023). As GLOBAL has just been released at the time of this study, it is not feasible to measure the impact that this tool may have on the bibliometric community.

Until the GLOBAL protocol was released, there were no specific protocols or guidelines that indicated how a study should be conducted in the field of bibliometrics, as there are for systematic reviews or other types of evidence synthesis. However, some existing guidelines may be applicable, at least in part, to identifying the different items that should be reported and how to report them in a bibliometric study. Thus, the Search (S) extension derived from the PRISMA protocol - Preferred Reporting Items for Systematic reviews and Meta-Analyses- (Rethlefsen et al., 2021) identifies up to 16 variables that should be reported in a completely transparent and replicable systematic review, including aspects such as the breakdown of the

Research Paper

databases and platforms used, the "copied and pasted exactly as run" search equation, the limits and restrictions used, the search filters, or the date of the search.

Many of these variables can also be used to properly report the designed search equation in a bibliometric analysis, as well as to accurately document the procedures and choices made in the search process and determination of the final analysis sample. Adherence to the PRISMA-S items is linked to the accuracy in reporting the search strategies and procedures employed (Sadeghi-Ghyassi et al., 2022) and is therefore related to a higher methodological quality of the research, as has been exposed in several works assessing the rigour of systematic reviews, mainly in the health field (Biocic et al., 2019; Koffel, 2015; O'Donohoe et al., 2021). The only study located that analyses the quality of search strategies in systematic reviews in the area of Information Science (Salvador-Oliván et al., 2018) concludes that the search methods are poorly reported, and that it is necessary to improve this information in order to replicate such systematic studies. With this objective in mind, a group of specialists in information and library science and evidence synthesis methodology has proposed a standardised data structure to report all the necessary details that allow the reproducibility of database searches, and enable their reuse and interoperability between information systems (Haddaway et al., 2022).

We need to mention that methodological flaws are not exclusive to bibliometrics. Other research techniques, such as clinical interventions (Chou et al., 2007), expert panels (Evans, 1997), or interviews (Alsaawi, 2014) also present shortcomings, which limit the reproducibility and reliability of results. Notwithstanding, the focus of this work is on methodological limitations in bibliometric research.

2　Objectives

Following in the footsteps of the studies by González-Alcaide (2021) and Jonkers and Derrick (2012), which call for an in-depth study of the quality and thoroughness of bibliometric content published mainly in journals outside the field of Information Science, and that of Oviedo-García (2021), which specifically requests the same to be done with journals published by MDPI, this article explores different aspects related to the methodological quality and reproducibility of bibliometric research published in the two most productive journals in bibliometrics, Sustainability and Scientometrics. Specifically, the following objectives are proposed:

● To identify the characteristics of the thematic bibliometric articles published in the journals Scientometrics and Sustainability in the period 2019-2021.

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

- To determine the number of methodological flaws identified for each of the seven variables studied, which are broken down in the subsection *Analysis of variables and determination of shortcomings*.
- To compare the results of both subsamples, in order to determine which of the two journals publishes work of greater rigour from the point of view of the methodological quality and reproducibility of the research.
- To propose measures to improve the reporting of thematic bibliometric research, thus increasing its methodological quality.

3 Materials and methods

3.1 Journals sampled: Sustainability and Scientometrics

Sustainability is a journal published since 2009 by the Swiss publisher MDPI, which publishes 418 open-access titles through a golden route. This publisher is questioned for its aggressive editorial practices, for the massive publication of special issues and for the laxity of its peer review processes (Repiso et al., 2021; Siler, 2020), which, together with its high rates of self-citation and intra-citation, has led some authors to describe it as a predatory publisher (Oviedo-García, 2021). Sustainability was listed as a questionable journal in products released by national authorities, such as the Norwegian Register of Scientific Journals, Series and Publishers. It was also part of the initial version of the China Early Warning Journal List, which was created in 2020 by the Chinese Academy of Sciences to combat academic misconduct. Among the 22 MDPI journals included in the list, Sustainability was the only one classified as "medium risk" (Early Warning Journal List, 2020). Nevertheless, it was subsequently removed from the list in 2021 and 2023 releases.

However, MDPI is also a publisher valued by authors, especially for its high bibliometric indexes, and for its very fast review and publication times, with an average of approximately 40 days from manuscript submission to publication date (Csomós & Farkas, 2022; Repiso et al., 2021). These reasons are probably behind the increase in the market share of this publisher, which in some countries, such as Romania or Poland, is above 30% (Csomós & Farkas, 2022). In Spain, MDPI was the publisher with the highest number of works published in 40 out of its 70 universities in 2021 (Delgado López-Cózar & Martín-Martín, 2022).

Scientometrics, started in 1978, is considered the flagship journal of bibliometrics (Chen et al., 2002), and since its inception has been the main publication forum for academics and professionals in this field. Although there are currently numerous options for the dissemination of works in the field of metric studies of science, such as the Journal of Informetrics, Quantitative Science Studies, or Research Evaluation,

this outlet is the oldest and most productive in the field of knowledge, and is considered in the academic community to be a prestigious and relevant journal that publishes quality bibliometric studies.

3.2 Data retrieval and inclusion criteria

In order to identify bibliometric publications in both journals, we used the search equation designed by González-Alcaide (2021), which, in turn, is based on previous studies. This search equation was:

TS=(bibliometric* OR scientometric* OR webometric* OR altmetric* OR informetrics* OR "citation analysis" OR "citation study" OR "scholarly productivity" OR "publication analysis" OR "scholarly impact" OR "patent citation") AND PY=(2019-2021) AND SO=(Sustainability OR Scientometrics)[2].

The data source searched was the Web of Science Core Collection, selecting the Science Citation Index (SCI) and Social Science Citation Index (SSCI) databases. The search was carried out on 04/12/2021 for the 2019-2020 period and updated on 09/14/2022 for including the year 2021.

This search yielded a total of 1005 papers, 419 from Sustainability and 586 from Scientometrics. On this set of results, the following inclusion criteria were applied to determine the final sample to be analysed (Figure 2). These criteria were:

- C1. It is an article or a review (document types Article or Review, according to the classification made by Web of Science).
- C2. It is an article published in a volume in 2019, 2020 or 2021. "Early access" papers but not definitively published in any of these three years are excluded.
- C3. Access to the full text of the paper is available to the authors.
- C4. It is a thematic bibliometric analysis, i.e., it studies a discipline or a specific subject. It therefore excludes research on a specific characteristic of the sample (e.g., language, OA, gender, etc.). This criteria also excludes analysis of specific journals or groups of journals within a particular field. Additionally, studies that solely focus on the work of a specific researcher or institution, or analyses that exclusively examine an institutional or geographic domain are excluded.

Papers that did not meet criteria C1 or C2 were excluded from the final sample without the need to review their content. Criterion C3 was met in all cases. To determine the suitability for criterion C4, each of the papers was reviewed manually by the authors. To this end, the title, abstract and, when necessary, the full text of the papers were analysed. In the case of Sustainability, most of the papers reviewed were thematic bibliometric analyses (93.5% of the sample after the exclusion of C1-C2

[2]TS= Topic; PY=Publication Year; SO=Sources

The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

papers). In contrast, the thematic bibliometric analyses published in Scientometrics constituted 21.7% of the results retrieved in the initial search (after the exclusion of C1-C2 papers). Thus, the final sample was composed of 508 papers, 77% from Sustainability (391 papers) and 23% published in Scientometrics (117 papers).
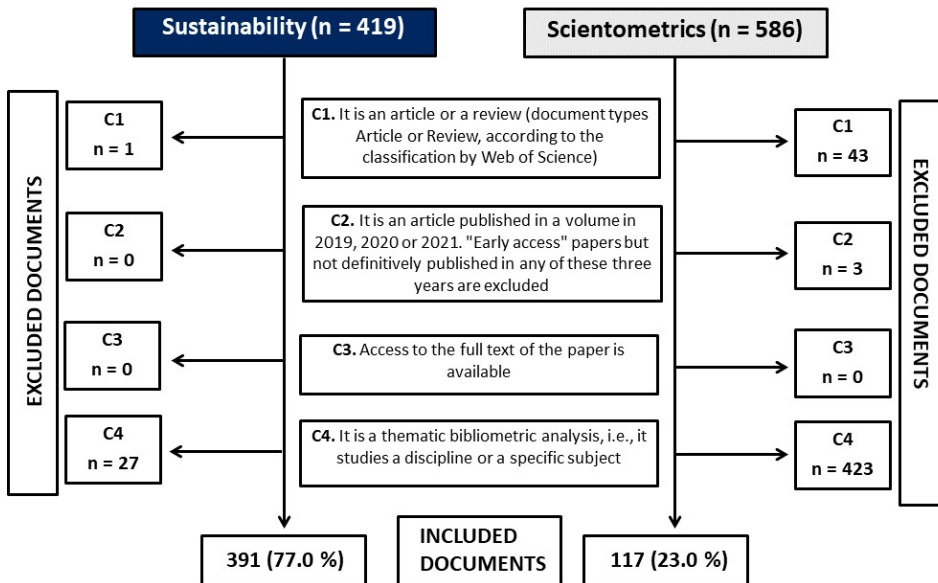

Figure 2.   Inclusion and exclusion criteria used and determination of the final sample analysed.

3.3   Analysis of variables and determination of shortcomings

The following analysis variables were defined.

1.   Search strategy: The search carried out by the authors, with an indication of the terms used, the field where the search is carried out, and the filters or limitations indicated. In the case of different searches being indicated, the most general one is taken into account, in line with what is indicated in the Sample variable.

2.   Sample: The number of papers analysed in each bibliometric analysis. If several searches are carried out, the most comprehensive search conducted is taken into account. The final number of papers analysed is recorded, not the number of papers retrieved in the initial searches.

3.   Date of search: The date of the search that serves as the basis for the analysis. If this date is not specifically specified, but it is indicated that the search period is, for example, from 2000 to March 2021, it is understood that the

Research Paper

search date is March 2021.

4. Period: The number of years analysed in each bibliometric analysis, including the start and last year of the study. It is recorded whether the last year of the sample is analysed in its entirety, or only in some months.

5. Data sources: The data source(s) used for the retrieval of the records analysed in the bibliometric study.

6. Document types: The document types analysed in the bibliometric study.

7. Language: The languages analysed in the bibliometric study.

Based on these variables, seven different types of shortcomings were identified for each of the articles.

1. Shortcoming search string. It is marked as a flaw if the paper does not break down the keywords used or the search fields used. It is not necessary for the paper to indicate the exact search string used as long as it indicates the keywords, how they are combined (Boolean operators) and the search field.

2. Shortcoming sample. It is marked as a flaw if the total number of documents analysed is not indicated, if sources are analysed separately without integrating the results (e.g., Web of Science and Scopus), or if the sample is less than 200 documents, since in these cases, bibliometric studies are not recommended and it is advisable to use other techniques that are more suitable (Donthu et al., 2021).

3. Shortcoming Search date. It is classified as a flaw if the search date is not indicated. Both the indication of the exact date and the month of search, or some similar formula (e.g., "at the beginning of 2020") are considered valid.

4. Shortcoming period. It is considered a flaw when the initial or final year of the analysed sample is not specified. Likewise, papers that analyse some of the years partially (for example, from 2000 to May 2021) are considered in this category. This circumstance has been derived from the information of the search date, or, if this information does not exist, from the date of submission of the manuscript to the journal (for example, for a paper that analyses the period 2000-2020 and that is submitted to the journal or published in the same year 2020, the last year is considered incomplete).

5. Shortcoming sources. This limitation applies to papers where the source used is not indicated, or which, following the recommendations of Liu (2019), indicate the use of Web of Science or Web of Science Core Collection, but without specifying the exact indexes where the search has been performed (Science Citation Index, Emerging Sources Citation Index, etc.).

6. Shortcoming document type. It is considered a limitation when it is not

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

indicated whether all the document types have been analysed, or only a selection of them. That is, when there is a lack of information on this variable.

7. Shortcoming languages. It is considered a limitation when it is not indicated whether the documents have been analysed in all the languages of the sample, or only in a selection of languages. That is, when there is no information on this variable.

Therefore, each article has a number of shortcomings ranging from 0 (no flaws) to 7 (all flaws).

3.4 Statistical analysis

The results were expressed as medians and interquartile ranges (IQRs) for continuous variables, and numbers and percentages for categorical variables. The comparison of subgroups was done by the Chi-square test or 2 x 2 tables, and the Fisher exact test when the number of events was less than 5. The Mann-Whitney U test was used for quantitative variables with non-parametric distribution. We used the Welch's t-test to measure equality of variances (Welch, 1947) and the Hedges' g to compare the magnitude of effects across different groups (Krzywinski & Altman, 2014). R Statistical Software (Foundation for Statistical Computing, Vienna, Austria), using the ggstatsplot approach (Patil, 2021) was employed.

4 Results

4.1 Descriptive data of the sample

The sample size was n=508, with a median of 846 papers analysed per study (average: 11,875.2). The minimum number of papers analysed was 14, and the maximum was 2,685,356. Notably, the number of papers studied in articles published in Scientometrics was much higher, with 2,009 papers (average: 40,615.4) compared to Sustainability, with 678 articles (average: 4,016).

The periods analysed were also longer in Scientometrics (median: 25 years) compared to a median of 22 years for studies published in Sustainability. In terms of the number of sources used, the median was one in all cases, although the mean was 1.9 for Scientometrics and 1.4 for Sustainability. In other words, in general, the studies published in Scientometrics were broader in their samples and in the years analysed, as well as in the number of sources used (table 1).

Research Paper

Table 1.   Descriptive data relating to the parameters Sample, Period and Number of Sources.

	Sustainability	Scientometrics	Total	p^*
N (%)	391 (77%)	117 (23%)	508 (100 %)	0.740
Sample analysed	678 [193-2206]	2,009 [607.5-8394.75]	846 [229.5-2932.5]	0.001
Period analysed (years)	22 [14-32]	25 [12-37]	23 [14-33.5]	0.426
Number of sources analysed	1 [1-1]	1 [1-2]	1 [1-1]	0.231

Data are expressed as median (interquartile range) or n (%), p^* is the p value associated to the likelihood that there is no difference between the groups being compared. P-value < 0.05 is considered statistically significant.

Table 2 shows the information on the sources used in the primary studies, as well as the types of documents and languages considered.

In terms of sources, 46.5% of the sample used only Web of Science (WoS) as a data source, and 26.8% used only Scopus. Only 4.1% of the papers analysed did without both data sources, with the use of alternative sources to WoS or Scopus being more frequent in articles published in Scientometrics (9.4%) than in those published in Sustainability (2.6%).

With regard to the document types, in the sample as a whole, the most frequent typology analysed was "article", together with other specifically indicated types (review, letters, proceeding papers, book chapters, etc.), as this situation occurred in one-third of the works analysed. The next most common situation was, in the case of Sustainability, the analysis exclusively of the "article" typology (31.2%), while in Scientometrics it was the analysis of all document types (19.7%), although with a very similar percentage to that of the "articles only" category.

With regard to the language variable, if we exclude the papers for which this data is not provided, the most frequent was the analysis of papers only in English (33.7%). This is the case for Sustainability, while Scientometrics detects the same number of papers analysing documents only in English as in all languages (17.9%). A tiny number of papers are found that analyse the results of papers in English and some other specific language(s) (2.6%), or that study papers in a language other than English (specifically, Chinese in one case and Ukrainian in the other) (0.4%).

4.2   Shortcomings

In the sample as a whole, a total of 1,304 shortcomings are found, 1,000 in the case of Sustainability, and 304 in the journal Scientometrics, with an average of 2.6 shortcomings per article in both journals, and a median of 2 flaws in both cases (Figure 3).

The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

Table 2.   Descriptive data relating to the parameters Sources, Document types and Languages.

	Sustainability n= 391 (77%)	Scientometrics n= 117 (23%)	Total n= 508 (100%)	p*
Sources, no. (%)				0.007
WoS only	181 (46.3%)	55 (47.0%)	236 (46.5%)	
Scopus only	113 (28.9%)	23 (19.7%)	136 (26.8%)	
WoS + Scopus only	52 (13.3%)	11 (9.4%)	63 (12.4%)	
WoS or/and Scopus + other source(s)	34 (8.7%)	17 (14.5%)	51 (10.0%)	
Other sources (not WoS nor Scopus)	10 (2.6%)	11 (9.4%)	21 (4.1%)	
No data	1 (0.3%)	0 (0.0%)	1 (0.2%)	
Document type, no. (%)				0.036
Articles + Other doc types	126 (32.2%)	43 (36.8%)	169 (33.3%)	
Articles only	122 (31.2%)	22 (18.8%)	144 (28.3%)	
All	79 (20.2%)	23 (19.7%)	102 (20.1%)	
Other doc types	1 (0.3%)	1 (0.9%)	2 (0.4%)	
No data	63 (16.1%)	28 (23.9%)	91 (17.9%)	
Language, no. (%)				<0.001
English only	150 (38.4%)	21 (17.9%)	171 (33.7%)	
All	68 (17.4%)	21 (17.9%)	89 (17.5%)	
English + others	11 (2.8%)	2 (1.7%)	13 (2.6%)	
Not English	1 (0.3%)	1 (0.9%)	2 (0.4%)	
No data	161 (41.2%)	72 (61.5%)	233 (45.9%)	

Data are expressed as n (%), $p*$ is the p value associated to the likelihood that there is no difference between the groups being compared. P-value < 0.05 is considered statistically significant.

$t_{Welch}(186.37) = -0.27, p = 0.79, \hat{g}_{Hedges} = -0.03, CI_{95\%} [-0.24, 0.18], n_{obs} = 508$
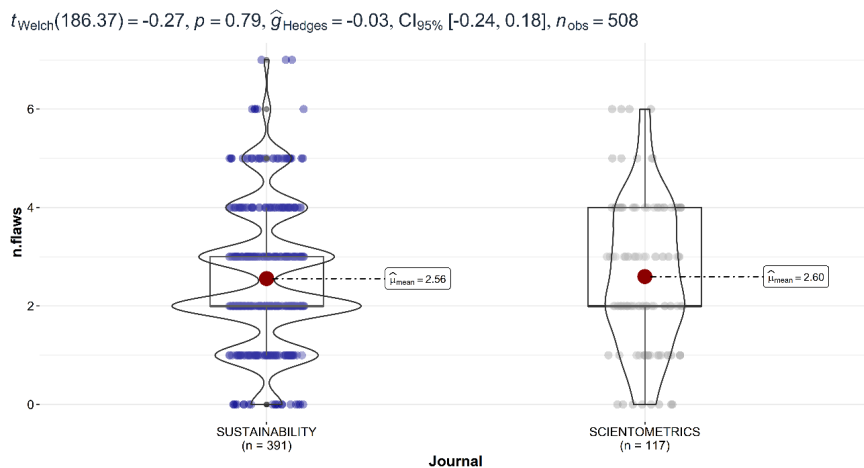

Figure 3.   Distribution of number of flaws by journal.

t Welch: test to measure equality of variances; g Hedges: test to compare the magnitude of effects across different groups; CI: confidence intervals; n obs: number of observations.

Research Paper

Table 3 and Figure 4 show for each of the seven variables the percentage of papers with shortcomings for both journals. Sustainability has a slightly higher percentage of flaws than Scientometrics in the variable Search String (38.4% vs. 35%, p = 0.295), which is more pronounced in the parameters Sample (8.7 points difference, p = 0.039), Period (10.1 points difference, p = 0.024) and Sources (7.4 points difference, p = 0.102).

Table 3.   Number and percentage of shortcomings.

	Sustainability n= 391 (77%)	Scientometrics n= 117 (23%)	Total n= 508 (100 %)	$p*$
Search string, n (%)	150 (38.4)	41 (35.0)	191 (37,6)	0.295
Sample, n (%)	114 (29.2)	24 (20.5)	138 (27.2)	0.039
Search date, n (%)	164 (41.9)	54 (46.2)	218 (42.9)	0.241
Period, n (%)	176 (45.0)	40 (34.9)	216 (42.5)	0.024
Sources, n (%) *	172 (65.6)	45 (57.0)	217 (63.6)	0.102
Doc_types, n (%)	63 (16.1)	38 (23.9)	91 (17.9)	0.039
Language, n (%)	161 (41.2)	72 (61.5)	233 (45.9)	<0.001

Data are expressed as n (%). * For Sources category, only applicable data was included (Total n = 341; Sustainability (n = 262), Scientometrics (n = 79). $P*$ is the p value associated to the likelihood that there is no difference between the groups being compared. P-value < 0.05 is considered statistically significant.

The bibliometric analyses published in Scientometrics showed a greater number of shortcomings than those published in Sustainability in the criteria Search date (4.3 points difference, p = 0.241), Document types (7.8 points difference, p = 0.039), and Language, with the latter parameter showing the greatest difference between subsamples, with more than 20 points difference between Scientometrics (61.5%), and Sustainability (41.2%) (p < 0.001).
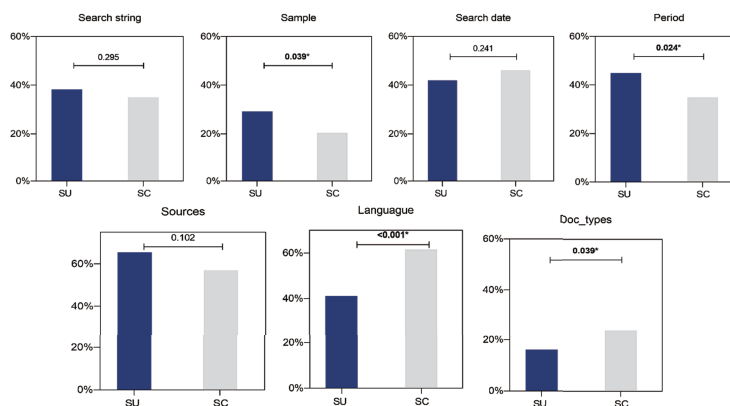

Figure 4.   Comparison of shortcomings among journals for each of the parameters. SU= Sustainability; SC= Scientometrics.

The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

Finally, the analysis of the frequency of articles according to the number of flaws shows similar profiles for both journals, with 2 being the most common number of shortcomings for each of the subsamples. Up to 28 papers were detected without flaws, which represents a percentage of 5.1% in Sustainability and 6.8% in Scientometrics (p = 0.404). At the other extreme, there are 127 papers with four or more shortcomings, with a percentage of 23.9% for Sustainability and 29.1% for Scientometrics (p = 0.267).
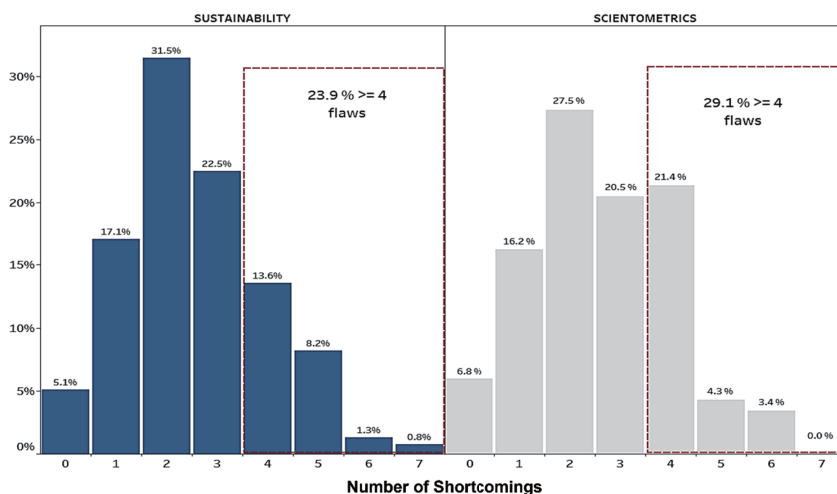

Figure 5.   Percentage of articles by number of shortcomings for each journal.

5   Discussion

In this paper, we have studied several characteristics associated with the methodological quality of the bibliometric thematic analyses disseminated in the two journals that published the largest amount of scientometric research in the period 2019-2021. A practically identical number of flaws were found in both journals. Specifically, of the seven parameters studied, an average of 2.6 shortcomings were detected in the sample as a whole, with no significant differences between the journals analysed.

The flagship bibliometrics journal Scientometrics has more flaws than Sustainability in three of the seven parameters studied, namely in the reporting of the period, in the breakdown of the types of documents used, and in the reporting of the language. Sustainability showed a higher number of shortcomings in the breakdown of the search strategy, the sample of papers, the sources used, and the search date. It also showed a lower number and diversity of sources used.

It should be emphasised that although all the identified shortcomings affect the quality of the studies, their significance may vary. In this investigation we did not study the degree of harm caused by each type of flaw. While some deficiencies, such as an erroneous search strategy, may impact the paper's outcomes, others, like the reporting of document types or languages used, are related to the level of transparency and replicability of the research, but may not affect the results obtained.

The variables which, in the sample as a whole, showed the lowest degree of compliance were the report of the sources used (63.6%), and the breakdown of the languages analysed (45.9%), while those relating to the sample (27.2%) and the document types (17.9%) were those which indicated the highest degree of compliance.

Although we would like to share Lund's (2022) statement that "there are not many bibliometric studies that are of poor quality", these data show that there is considerable room for improvement in terms of the proper use and breakdown of methodological procedures in thematic bibliometric research.

5.1   Search strategy / Samples analysed

Of particular concern is the lack of transparency in the reporting of executed searches. It should be borne in mind that in this variable we have not taken a criterion as demanding as the one proposed by PRISMA-S (Rethlefsen et al., 2021), such as "Include the search strategies for each database and information source, copied and pasted exactly as run", but that a search was considered reproducible when the keywords, Boolean operators and the search field were specified. The most common reason for this parameter to be qualified as a shortcoming was that the search field was not reported, but only the keywords. This is relevant because the number of results retrieved in the databases differs significantly depending on the field in which the search is executed. In some cases, however, not even the specific terms used were mentioned. It was not the purpose of this paper to determine the quality of the search equation, but a brief review reveals that in many cases the terms used are not the most appropriate for the objectives of the specific study, sometimes because they are too broad, sometimes because they are too specific. In other cases, the use of inverted commas in the transcription of searches is not rigorously employed, or it is not clearly indicated how the terms have been combined using Boolean operators. Finally, we have also found in some of the studies that several searches are broken down, without it being clear which is the one taken into consideration in the work, or how they have been combined.

The development of the search strategy should be a well-documented process, clearly specifying the terms used, the search fields where the query was executed, the operators used to connect the search(s), as well as special characters (such as

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

asterisks or inverted commas), as any small change in the search string may generate different results (Romanelli et al., 2021). The use of incorrect search strategies affects the precision and completeness of results, leading to unreliable conclusions (Salvador-Oliván et al., 2019). Examples of such strategies include utilizing a single keyword to define a field of study, improperly utilizing Boolean operators to combine search terms, using ambiguous terms, or selecting inadequate search fields. In this regard, knowledge of the database syntax, coverage, and changes over time is essential to avoid errors in interpreting results. The cases of thematic analyses in Artificial Intelligence (Liu, 2021) and Climate Change (Liu, 2022) are paradigmatic of the challenges involved in performing a rigorous and error-free bibliometric analysis. In this sense, we recommend the development of more precise search strategies, for which various techniques can be employed, such as using search equations based on the previous literature, deriving terms from artificial intelligence tools, or having committees of experts in the field that may propose/refine strategies (Rethlefsen et al., 2021). Finally, it is highly advisable to include documentalists or librarians to help generate more effective search strategies, as this is associated with higher quality and reproducibility of the research (Koffel, 2015).

In this research, we have set a threshold of 200 papers to consider an analysis as using a small sample, so that the bibliometric approach is probably not the most appropriate for studying the selected topic. While this threshold is derived from the recommendations of Donthu et al. (2021) and has been considered as a minimum requirement in some thematic bibliometric analyses (Gao & Wang, 2022; Kumar et al., 2021; Tamala et al., 2022), it must be acknowledged that the 200-paper threshold is somewhat arbitrary. However, we believe it to be a reasonable threshold, considering that the papers that used a sample of less than 200 articles in our research analysed an average of only 8.2 papers per year. It seems difficult to identify trends or derive robust conclusions on topics with such small numbers.

The fact that as many as 12 of the papers analysed samples of less than 50 papers, or that 57 papers used samples of less than 100 papers, shows that either the topics are so specific that a bibliometric analysis does not make sense, or that the searches are poorly designed. Furthermore, in some cases, the initial search retrieved tens of thousands of papers, which after review by the authors and the exclusion of irrelevant articles remained in the hundreds or tens of papers, which also points to a significant dissonance between the search strategies designed and the final results, impeding the reproducibility of such analyses (Boyack et al., 2022).

5.2 Search dates and periods

It is also striking that up to 42.5% of the papers studied committed some kind of

Research Paper

shortcoming in relation to the period analysed, the most common being that incomplete years were studied (the last of the period analysed), which prevents a strict comparison with previous years. Given that one of the main objectives of thematic analyses is usually to verify the quantitative evolution in the number of papers generated this methodological limitation prevents us from reaching solid conclusions about the strength of a given research topic. We are not talking about a few days to complete the year, but rather that in up to 70 of the papers, in the last year of the sample, a period of less than six months was analysed. It is probably the authors' desire to make their papers more up-to-date and relevant that leads them to include the current year in the analysis.

For its part, the indication of the search date is another essential element to allow the reproducibility of the research, given the dynamic nature of data sources and their changing character. Unless there is an error or omission on our behalf, we have not located other studies that study this variable in the bibliometric analyses. In the only study detected that analyses this variable in systematic reviews in the field of Library and Information Science, the search date was an element that was omitted in 56.5% of the papers (Salvador-Oliván et al., 2018). On the other hand, it is a parameter that has been studied for systematic reviews in different medical fields, detecting between 0% and 12% of papers that did not indicate the search date (Beller et al., 2013; Biocic et al., 2019; Franco et al., 2018; O'Donohoe et al., 2021; Sadeghi-Ghyassi et al., 2022). The specific date of search is an element that needs to be reported in a systematic review, according to PRISMA-S (Rethlefsen et al., 2021) and which we believe should be extended to bibliometric studies, in order to allow reproducibility of research.

5.3   Sources

In this study, we have verified the limited diversity of the sources used in the bibliometric studies analysed, given the overwhelming dominance of Web of Science and Scopus. Only in a residual 4.1% of the papers was neither of the two sources used. The use of a greater number of local sources is lacking, which would provide a new scenario for comparisons, for example, between the literature circulating at the international level and that which is recorded in local or regional sources.

Another significant conclusion is that 75.2% of the analysed articles used a single data source (WoS or Scopus) in their bibliometric analysis. Although we do not consider this as a shortcoming, current recommendations in a responsible metrics scenario suggest the use of a diverse range of databases, avoiding the use of single sources that may show biased results (Cabezas-Clavijo & Torres-Salinas, 2021). Web of Science and Scopus are high-quality databases, designed, among other

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

purposes, for bibliometric analyses, providing extensive possibilities for data querying and downloading. Their characteristics, strengths, and weaknesses have been repeatedly highlighted in the literature (Aksnes & Sivertsen, 2019; Mongeon & Paul-Hus, 2016; Visser et al., 2021), enabling authors to acknowledge their limitations in their bibliometric studies. It contrasts with the difficulty of extracting structured information from other data sources which are not originally designed for performing bibliometric analyses. We consider that a greater use of alternative sources, such as DOAJ, Redalyc or Dialnet, is directly linked to the fact that the producers of these resources allow a more appropriate use for bibliometric research, for example by facilitating the search by a greater number of fields, support for combined searches, or the downloading of records in standardised formats. This low use of alternative sources may be overshadowing research topics and trends that are locally relevant but not visible in mainstream journals.

On the other hand, the low level of disaggregation of the specific indexes used within Web of Science is certainly surprising. We have identified that, considering only papers using WoS, 65.6% of the studies in Sustainability and 57% of those published in Scientometrics fail to provide this information. In the case of Scientometrics, this figure is slightly higher than reported previously (Liu, 2019). It is advisable to indicate the subset used within WoS, given that institutional subscriptions may include a greater or lesser number of indexes, or differ in the period covered.

Authors should be aware that results obtained from Web of Science depend on the type of access granted to each institution for the database. This implies that the particular indices that an organization can search and the coverage of a certain database might vary from one institution to another (Gusenbauer & Haddaway, 2020). However, these variations are sometimes unknown for researchers, which may affect the accuracy of their reporting and hinder the reproducibility of the research. Likewise, wide-used databases may also present mistakes in their references. These errors could be due to the quality of metadata provided by publishers, such as mistakes in the original paper, or to errors introduced in the data processing stage by the database operators (Pranckutė, 2021). Although the rate of errors for Scopus and Web of Science is relatively low, it may affect outcomes, specially when working with small samples (Franceschini et al., 2016). Therefore, researchers should be familiar with the databases characteristics and limitations before choosing any of them for bibliometric research purposes.

5.4 Document types and languages

As for the types of documents, 17.9% of the papers did not specify this information.

Research Paper

When it is detailed, it is found that almost a quarter of the papers (24.4%) analyse all the document typologies retrieved in their search strategy, while the remaining papers analyse the typology "article", either exclusively or combined with other typologies. Given that the bibliographic databases include not only research documents, but also other types of materials (editorial material, proceedings papers, meeting abstracts, corrections, book reviews, book chapters, etc.), it is advisable to indicate the exact document types analysed, as this may condition the final results. It is also important to bear in mind that some typologies, such as proceeding papers, can have delays of up to three years in their indexing in Web of Science (Maddi & Baudoin, 2022), which can have a notable effect on bibliometric analyses in certain areas.

Almost half of the papers (45.9%) did not indicate the languages of the documents analysed. Although most of the papers use Web of Science or Scopus, sources that mainly compile documents in English (Vera-Baceta et al., 2019), it should not be forgotten that in up to 18 Web of Science categories, mainly in the Humanities, the percentage of documents in languages other than English exceeds 20%, and in five of these 18 it exceeds 50% (Liu, 2017). The fact that there are almost twice as many analyses of papers in English alone (171) than in all languages (89) highlights the need to study not only the scientific production visible in English in the major databases, but also that produced in other languages, given that the subject matters, methodologies, approaches and findings may differ significantly. In this sense, it would be desirable that the more inclusive perspective advocated by statements such as the Helsinki Initiative on Multilingualism in Scholarly Communication (Federation of Finnish Learned Societies et al., 2019) be more clearly incorporated into thematic bibliometric studies, in order to endow them with greater academic and social value.

As it cannot be assumed that there is a default mode of performing a bibliometric analysis (e.g., considering all document types, or all languages), the reporting of these variables is key to assure the transparency and reproducibility of results (Rethlefsen et al., 2021). Conceptually, it is important to specify which document types are selected, as not all of them may convey relevant scientific information (Glänzel & Moed, 2002). Likewise, the choosing of only one language (e.g., English) can introduce biases through certain topics or research lines, so researchers in the field should always make clear the methodological decisions taken into account.

6   Limitations and future research directions

The first limitation of this work is that it is a study of two scientific journals, so

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

the results cannot be directly extrapolated to the set of thematic bibliometric analyses published in journals from all fields. However, the fact that we have taken the two journals with the highest number of bibliometric studies, the flagship journal of scientometric research, on the one hand, and the journal that publishes the most thematic bibliometric analyses, on the other hand, suggests that the level of shortcomings in other journals may be similar to that found in our study. In any case, an analysis taking a larger sample would allow us to confirm this scenario.

Secondly, our study adopts a perspective that focuses almost exclusively on aspects related to search reporting, which is only one aspect of a bibliometric analysis. A study that aimed to comprehensively measure the quality of thematic bibliometric research should also consider aspects such as the relevance of the search conducted, the data cleaning techniques, the indicators selected, the analysis software, the statistical techniques employed, as well as aspects related to data interpretation, and the robustness and transparency of the research. Likewise, it is important to underline that not all types of shortcomings analysed have the same impact on the methodological quality of a bibliometric research. Therefore, it would be necessary to delve into these aspects to determine the degree of harm caused by each kind of flaw.

Thirdly, although our work analyses bibliometric studies, at no time does it use indicators of this nature, so another line of work could focus on determining the relationship between the flaws detected and extrinsic characteristics of the papers such as the number of authors, number of institutions, type of collaboration, or citations received by the contributions. Finally, the authors' expertise in bibliometric matters could be related to the number of shortcomings in the papers.

7　Conclusion - The need to develop tailored tools for improving the quality of thematic bibliometric analyses

Protocols such as PRISMA have become very relevant in recent years for the detailed reporting of systematic reviews. In the data collection of this work, we have found that many of the studies analysed used these guidelines to show the phases of the study, the databases used, or the inclusion and exclusion criteria. Although PRISMA has been extended, in practice, on many occasions to the reporting of the bibliometric analysis process, this model is insufficient to rigorously show the steps and phases of a scientometric study.

The Search extension of PRISMA, on which we have based ourselves to construct some of the variables in this study, although also insufficient to capture all the key aspects of a bibliometric analysis, can be a good starting point to improve the transparency and rigour of bibliometric studies. While a thematic bibliometric

analysis is not a systematic review, it is clear that there are common elements, and that models such as this one, suitably adapted to the reality of scientometrics, could contribute to improving the reporting and reproducibility of this kind of research.

We therefore call, along the lines suggested by Boyack et al. (2022), for greater collaboration between the actors involved in the publication of bibliometric research (authors, editors, reviewers), allowing the development of protocols, guidelines, frameworks, or checklists, which contribute to enhancing the transparency and reproducibility of thematic bibliometric analyses. Likewise, all these actors must be aware that it is necessary to apply more rigorous criteria in their respective roles in order to raise the quality of research in this field. In this regard, the launch of the GLOBAL (Guidance List for the repOrting of Bibliometric AnaLyses) initiative (Ng et al., 2023) may represent a significant step towards this aim.

Finally, we believe it is essential to continue studying the characteristics of thematic bibliometric studies in order to detect areas for improvement and to promote a more thorough and informed use of this powerful quantitative methodology.

Conflict of interest

No funding was received for conducting this study. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

Data availability statement

The associated data for this paper (DOI 10.5281/zenodo.8262602) can be accessed in the Zenodo database (https://doi.org/10.5281/zenodo.8262602).

Author contributions

Alvaro Cabezas-Clavijo (alvaro.cabezas@unir.net): Conceptualization (Lead), Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Equal), Project administration (Equal), Resources (Equal), Supervision (Equal), Validation (Equal), Visualization (Equal), Writing – original draft (Equal), Writing

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

– review & editing (Equal); Yusnelkis Milanés-Guisado (ymilgui@upo.es): Data curation (Equal), Formal analysis (Lead), Methodology (Equal), Software (Lead), Supervision (Equal), Validation (Equal), Visualization (Equal); Ruben Alba-Ruiz (ruben.alba.ruiz@gmail.com): Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Equal), Project administration (Equal), Resources (Equal), Validation (Equal), Visualization (Equal), Writing – review & editing (Equal); Ángel M. Delgado-Vázquez (adelvaz@upo.es): Conceptualization (Equal), Data curation (Equal), Investigation (Equal), Methodology (Equal), Resources (Equal), Supervision (Equal), Validation (Equal), Writing – review & editing (Equal).

References

Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. *Journal of Data and Information Science*, 4/1: 1-21. DOI: 10.2478/jdis-2019-0001

Alsaawi, A. (2014). A Critical Review of Qualitative Interviews. *European Journal of Business and Social Sciences*, 3/4: 149-56. DOI: 10.2139/ssrn.2819536

Beller, E. M., Chen, J. K.-H., Wang, U. L.-H., & Glasziou, P. P. (2013). Are systematic reviews up-to-date at the time of publication?. *Systematic Reviews*, 2/1: 1–6.

Biocic, M., Fidahic, M., & Puljak, L. (2019). Reproducibility of search strategies of non-Cochrane systematic reviews published in anaesthesiology journals is suboptimal: primary methodological study. *British Journal of Anaesthesia*, 122/6: e79–81. DOI: 10.1016/j.bja.2019.02.014

Boyack, K. W., Klavans, R., & Smith, C. (2022). Raising the bar for bibliometric analysis'. *STI 2022 Conference Proceedings*. Presented at the 26th International Conference on Science and Technology Indicators, September 7, Granada. DOI: 10.5281/zenodo.6975632

Cabezas-Clavijo, Á., & Torres-Salinas, D. (2021). Bibliometric reports for institutions: best practices in a responsible metrics scenario'. *Frontiers in Research Metrics and Analytics*, 6: 696470. DOI: 10.3389/frma.2021.696470

Chen, C. M., McCain, K., White, H., & Lin, X. (2002). Mapping Scientometrics (1981–2001). *Proceedings of the American Society for Information Science and Technology*, 39/1: 25–34.

Chou, R., Fu, R. W., Carson, S., Saha, S., & Helfand, M. (2007). Methodological shortcomings predicted lower harm estimates in one of two sets of studies of clinical interventions. *Journal of Clinical Epidemiology*, 60/1: 18-28. DOI: 10.1016/j.jclinepi.2006.02.021

Csomós, G., & Farkas, J. Z. (2022). Understanding the increasing market share of the academic publisher "Multidisciplinary Digital Publishing Institute" in the publication output of Central and Eastern European countries: a case study of Hungary. *Scientometrics*. DOI: 10.1007/s11192-022-04586-1

Delgado López-Cózar, E., & Martín-Martín, A. (2022). Detectando patrones anómalos de publicación científica en España: Más sobre el impacto del sistema de evaluación científica'.

Research Paper

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133: 285–96. DOI: 10.1016/j.jbusres.2021.04.070

Early Warning Journal List. (2020). Retrieved 31 August 2023, from https://earlywarning. fenqubiao.com/#/en/

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact?. *Scientometrics*, 105/3: 1809–31. DOI: 10.1007/s11192-015-1645-z

Evans, C. (1997). The Use of Consensus Methods and Expert Panels in Pharmacoeconomic Studies. *PharmacoEconomics*, 12/2: 121-9. DOI: 10.2165/00019053-199712020-00003

Federation of Finnish Learned Societies, Committee for Public Information, Finnish Association for Scholarly Publishing, Universities Norway, & European Network for Research Evaluation in the Social Sciences and the Humanities. (2019). *Helsinki Initiative on Multilingualism in Scholarly Communication*., pp. 621757-Bytes. Retrieved October 28, 2022, from <https://figshare.com/articles/Helsinki_Initiative_on_Multilingualism_in_Scholarly_Communication/7887059>

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). Empirical analysis and classification of database errors in Scopus and Web of Science. *Journal of Informetrics*, 10/4: 933–53. DOI: 10.1016/j.joi.2016.07.003

Franco, J. V. A., Garrote, V. L., Escobar Liquitay, C. M., & Vietto, V. (2018). Identification of problems in search strategies in Cochrane Reviews. *Research Synthesis Methods*, 9/3: 408–16. DOI: 10.1002/jrsm.1302

Gao, H. C., & Wang, S. J. (2022). The Intellectual Structure of Research on Rural-to-Urban Migrants: A Bibliometric Analysis. *International Journal of Environmental Research and Public Health*, 19/15: 9729. DOI: 10.3390/ijerph19159729

Glänzel, W. (1996). The need for standards in bibliometric research and technology. *Scientometrics*, 35/2: 167–76. DOI: 10.1007/BF02018475

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research'. *Scientometrics*, 53/2: 171-93.

González-Alcaide, G. (2021). Bibliometric studies outside the information science and library science field: uncontainable or uncontrollable?. *Scientometrics*, 126/8: 6837–70. DOI: 10.1007/s11192-021-04061-3

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11/2: 181–217. DOI: 10.1002/jrsm.1378

Haddaway, N. R., Rethlefsen, M. L., Davies, M., Glanville, J., McGowan, B., Nyhan, K., & Young, S. (2022). A suggested data structure for transparent and repeatable reporting of bibliographic searching. *Campbell Systematic Reviews*, 18/4: e1288. DOI: 10.1002/cl2.1288

Harzing, A.-W. (2016). Why replication studies are essential: learning from failure and success. *Cross Cultural & Strategic Management*, 23/4: 563–8. DOI: 10.1108/CCSM-07-2016-0133

Jonkers, K., & Derrick, G. E. (2012). The bibliometric bandwagon: Characteristics of bibliometric articles outside the field literature. *Journal of the American Society for Information Science and Technology*, 63/4: 829–36.

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

Koffel, J. B., & Rethlefsen, M. L. (2016). Reproducibility of Search Strategies Is Poor in Systematic Reviews Published in High-Impact Pediatrics, Cardiology and Surgery Journals: A Cross-Sectional Study. *PLOS ONE*, 11/9: e0163309. DOI: 10.1371/journal.pone.0163309

Krzywinski, M., & Altman, N. (2014). Comparing samples—part II. *Nature Methods*, 11/4: 355–6. DOI: 10.1038/nmeth.2900

Kumar, S., Lim, W. M., Pandey, N., & Christopher Westland, J. (2021). 20 years of Electronic Commerce Research. *Electronic Commerce Research*, 21/1: 1–40. DOI: 10.1007/s10660-021-09464-1

Lindner, M. D., Torralba, K. D., & Khan, N. A. (2018). Scientific productivity: An exploratory study of metrics and incentives. *PLOS ONE*, 13/4: e0195321. DOI: 10.1371/journal.pone.0195321

Liu, F. (2022). Retrieval strategy and possible explanations for the abnormal growth of research publications: re-evaluating a bibliometric analysis of climate change. *Scientometrics*. DOI: 10.1007/s11192-022-04540-1

Liu, W. S. (2017). The changing role of non-English papers in scholarly communication: Evidence from Web of Science's three journal citation indexes. *Learned Publishing*, 30/2: 115–23. DOI: 10.1002/leap.1089

Liu, W. S. (2019). The data source of this study is Web of Science Core Collection? Not enough. *Scientometrics*, 121/3: 1815–24. DOI: 10.1007/s11192-019-03238-1

Liu, W. S. (2021). Caveats for the use of Web of Science Core Collection in old literature retrieval and historical bibliometric analysis. *Technological Forecasting and Social Change*, 172: 121023. DOI: 10.1016/j.techfore.2021.121023

Lund, B. (2022). Bibliometrics Is Valuable Science. Why Do Some Journals Seem to Oppose It?. *Journal of Data and Information Science*, 7/3: 1–4. DOI: 10.2478/jdis-2022-0012

Maddi, A., & Baudoin, L. (2022). The quality of the web of science data: a longitudinal study on the completeness of authors-addresses links. *Scientometrics*. DOI: 10.1007/s11192-022-04525-0

Makel, M. C., & Plucker, J. A. (2014). Facts Are More Important Than Novelty: Replication in the Education Sciences. *Educational Researcher*, 43/6: 304–16. DOI: 10.3102/0013189X14545513

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research: How Often Do They Really Occur?. *Perspectives on Psychological Science*, 7/6: 537–42. DOI: 10.1177/1745691612460688

Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., Coriat, A.-M., et al. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, 18/7: e3000737. DOI: 10.1371/journal.pbio.3000737

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106: 213-28. DOI: 10.1007/s11192-015-1765-5

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la información*, 29/1. DOI: 10.3145/epi.2020.ene.03

Mueller-Langer, F., Fecher, B., Harhoff, D., & Wagner, G. G. (2019). Replication studies in economics—How many and which papers are chosen for replication, and why?. *Research Policy*, 48/1: 62–83. DOI: 10.1016/j.respol.2018.07.019

Research Paper

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*.

Nicoll, L. H., Oermann, M. H., Carter-Templeton, H., Owens, J. K., & Edie, A. H. (2020). A bibliometric analysis of articles identified by editors as representing excellence in nursing publication: Replication and extension. *Journal of Advanced Nursing*, 76/5: 1247–54. DOI: 10.1111/jan.14316

O'Donohoe, T. J., Bridson, T. L., Shafik, C. G., Wynne, D., Dhillon, R. S., & Tee, J. W. (2021). Quality of Literature Searches Published in Leading Neurosurgical Journals: A Review of Reviews. *Neurosurgery*, 88/5: 891–9. DOI: 10.1093/neuros/nyaa573

Oviedo-García, M. A. (2021). Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). *Research Evaluation*, 30/3: 405–19.

Patil, I. (2021). Visualizations with statistical details: The "ggstatsplot" approach. *Journal of Open Source Software*, 6/61: 3167. DOI: 10.21105/joss.03167

Perry, T., Morris, R., & Lea, R. (2022). A decade of replication study in education? A mapping review (2011–2020). *Educational Research and Evaluation*, 27/1–2: 12–34. DOI: 10.1080/13803611.2021.2022315

Petrovich, E. (2022). Bibliometrics in Press. Representations and uses of bibliometric indicators in the Italian daily newspapers. *Scientometrics*, 127/5: 2195–233. DOI: 10.1007/s11192-022-04341-6

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*, 9/1: 12. DOI: 10.3390/publications9010012

Repiso, R., Merino-Arribas, A., & Cabezas-Clavijo, Á. (2021). El año que nos volvimos insostenibles: Análisis de la producción española en Sustainability (2020). *El Profesional de la información*, 30/4: e300409. DOI: 10.3145/epi.2021.jul.09

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, 10/1: 39. DOI: 10.1186/s13643-020-01542-z

Romanelli, J. P., Pereira Gonçalves, M. C., de Abreu Pestana, L. F., Hitaka Soares, J. A., Stucchi Boschi, R., & Fernandes Andrade, D. (2021). Four challenges when conducting bibliometric reviews and how to deal with them. *Environmental Science and Pollution Research*, 28/43: 60448–58. DOI: 10.1007/s11356-021-16420-x

Ryan, J. C., & A Tipu, S. A. (2022). Business and management research: Low instances of replication studies and a lack of author independence in replications. *Research Policy*, 51/1: 104408. DOI: 10.1016/j.respol.2021.104408

Sadeghi-Ghyassi, F., Zarea Gavgani, V., Fathifar, Z., Makani, N., Vaez, R., & Montazeri, M. (2022). Quality of reporting of literature search strategies in systematic reviews published on the role of telehealth during COVID-19. *Journal of Information Science*, 01655515221087649. DOI: 10.1177/01655515221087649

Saltelli, A., & Funtowicz, S. (2017). What is science's crisis really about?. *Futures*, Post-Normal science in practice, 91: 5–11. DOI: 10.1016/j.futures.2017.05.010

The need to develop tailored tools for improving the quality of thematic bibliometric
analyses: Evidence from papers published in Sustainability and Scientometrics

Cabezas-Clavijo, A., et al.

Research Paper

Salvador-Oliván, J. A., Marco-Cuenca, G., & Arquero-Avilés, R. (2018). Las revisiones sistemáticas en Biblioteconomía y Documentación: análisis y evaluación del proceso de búsqueda. *Revista Española de Documentación Científica*, 41/2: e207–e207. DOI: 10.3989/redc.2018.2.1491

Salvador-Oliván, J. A., Marco-Cuenca, G., & Arquero-Avilés, R. (2019). Errors in search strategies used in systematic reviews and their effects on information retrieval. *Journal of the Medical Library Association : JMLA*, 107/2: 210–21. DOI: 10.5195/jmla.2019.567

Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, 74/2: 175–89. DOI: 10.1007/s11192-008-0211-3

Siler, K. (2020). Demarcating spectrums of predatory publishing: Economic and institutional sources of academic legitimacy. *Journal of the Association for Information Science and Technology*, 71/11: 1386–401. DOI: 10.1002/asi.24339

Tamala, J. K., Maramag, E. I., Simeon, K. A., & Ignacio, J. J. (2022). A bibliometric analysis of sustainable oil and gas production research using VOSviewer. *Cleaner Engineering and Technology*, 7: 100437. DOI: 10.1016/j.clet.2022.100437

Tipu, S. A. A., & Ryan, J. C. (2021). Are business and management journals anti-replication? An analysis of editorial policies. *Management Research Review*, 45/1: 101–17. DOI: 10.1108/MRR-01-2021-0050

Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121/3: 1803–13. DOI: 10.1007/s11192-019-03264-z

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2/1: 20-41. DOI: 10.1162/qss_a_00112

Welch, B. L. (1947). The generalisation of student's problems when several different population variances are involved. *Biometrika*, 34/1–2: 28–35. DOI: 10.1093/biomet/34.1-2.28