

“From Global Indicators to Local Applications”

7-9 September 2022 | Granada, Spain

#STI22GRX

Methodological shortcomings of bibliometric papers published in the journal Sustainability (2019-2020)

Álvaro Cabezas-Clavijo*, Yusnelkis Milanés-Guisado** and Ángel M. Delgado-Vázquez**

* alvaro.cabezas@unir.net

Universidad Internacional de La Rioja (UNIR), Avenida de la Paz, 137, Logroño, 26006 (Spain).

** ymilgui@upo.es; adelvaz@upo.es

Universidad Pablo de Olavide, Ctra. de Utrera, Km 1, Sevilla, 41013 (Spain).

Introduction

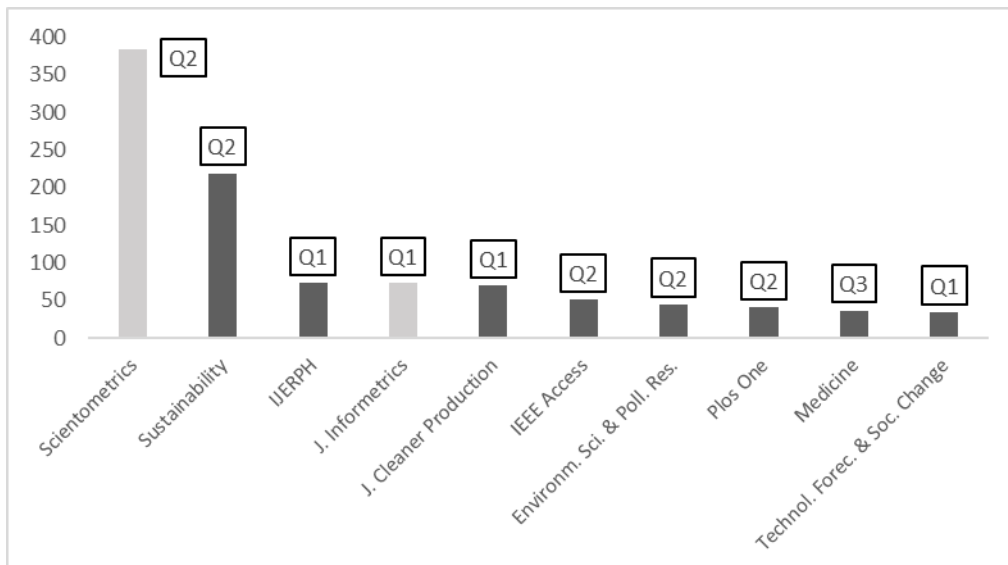
The number of papers using bibliometric methods has burgeoned in recent years, with many more bibliometric articles being published in journals outside the Information Science (IS) field than within it (González-Alcaide, 2021). Bibliometric techniques have become popular in recent years due to the intensive use of these types of indicators in institutional evaluations, academic reports and even the general press. As a result, they are used not only by bibliometrics experts but by academics and professionals from all fields (Petrovich, 2022). In addition, easily accessible and user-friendly tools can be found for both automated analysis and visualisation of results, in some cases free of charge.

One of the clearest manifestations of this phenomenon is the publication of bibliometric analyses on a specific subject or area, which tend to be published mainly in journals outside the field of IS (Ellegaard & Wallin, 2015). There is a perception among the bibliometric community that these publications in journals outside the discipline may lack the quality controls and standards required by journals in this field (Jonkers & Derrick, 2012), raising the need for greater rigour and critique prior to publication (González-Alcaide, 2021).

The emergence in recent years of the so-called ‘mega-journals’ has also opened up a new field for the publication of bibliometric studies. These journals are characterised by the large volume of papers published; a peer review process that assesses only the scientific solvency of the work as opposed to its complexity, usefulness, originality or interest (Björk & Catani, 2016); full open access to the publication through article processing charges (APC); and their multidisciplinary nature, with publication of papers in various areas of knowledge (Spezi et al., 2017).

In fact, in the period from 2019 to 2020, *Scientometrics*, the flagship journal in the field of Bibliometrics, published the most bibliometric articles, followed by the two mega-journals *Sustainability* and *International Journal of Environmental Research and Public Health* (IJERPH). Interestingly, of the ten journals with the highest bibliometric output, only two are from the IS area (see Figure 1).

Figure 1. Journals with the highest bibliometric article production (2019-2020)



Source: prepared by the author based on Web of Science data (SCI and SSCI). Only articles and reviews are considered. Search equation: 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” (Topic) and 2019-2020 (Year Published). The quartile of the journal in Journal Citation Reports (JCR) 2020 is indicated. Journals in the IS area are shaded in grey.

The journals Sustainability and IJERPH are owned by MDPI, a Swiss publisher with 386 ‘gold’ open-access publications¹. MDPI has come under fire for its aggressive editorial practices, mass publication of special issues and lax peer review processes (Repiso, Merino-Arribas, & Cabezas-Clavijo, 2021; Siler, 2020), which has led some authors to describe it as a predatory publisher (Oviedo-García, 2021).

Using Sustainability (the journal outside the IS area that publishes the most bibliometric articles) as a case study, this study uses content analysis to determine various parameters relating to the methodological rigour and reproducibility of the papers published in this journal in 2019 and 2020. In particular, analysis has been performed of the samples and time periods used in the analyses, and whether the authors adequately report the search strategy and the data sources used. Finally, studies are carried out to confirm whether they break down the document types and languages considered in their papers differentiating between articles published in ordinary issues and those published in special issues.

In this sense, we are following the recommendations of the articles that suggest carrying out content analysis to verify the methodological quality of articles published in journals outside the IS area (González-Alcaide, 2021; Jonkers & Derrick, 2012) or in journals published by MDPI (Oviedo-García, 2021).

Materials and methods

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

This search equation is: bibliometric* OR scientometric* OR webometric* OR altmetric* OR informetrics* OR “citation analysis” OR “citation study” OR “scholarly productivity” OR

¹ Data as of 24/04/2022, taken from <https://www.mdpi.com/about/journals>

“publication analysis” OR “scholarly impact” OR “patent citation” (Topic) and 2019-2020 (Year Published) and SUSTAINABILITY (Publication Titles).

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 12/04/2021.

The criteria for inclusion of papers in the final analysis were:

- It was a bibliometric analysis of a geographical, institutional, or theme-based nature.
- The document was an article or a review.
- Availability of the full text of the document.

After reviewing the papers, 16 of the 220 documents retrieved were excluded, either because they were not one of the document types defined (n=1) or they did not perform a domain analysis (n=15). Accordingly, the final sample consisted of 204 papers.

Analysis variables

The variables analysed were:

- Sample: number of papers included in the analysis of each article.
- Period: number of years analysed in each bibliometric study.
- Data sources: the data sources for the search upon which the bibliometric analysis is based.
- Search equation: whether the authors specify the exact search equation applied, indicating both the keywords used and the fields in which the search is made.
- Document type: The document types taken into account in the bibliometric analysis.
- Language: the languages taken into account in the bibliometric analysis.
- Special or regular issue: whether the paper is published in a regular issue or a special issue of the journal.

Statistical analysis

The results were expressed as medians and IQRs for continuous variables, and numbers and percentages for categorical variables. The χ^2 test was used to compare categorical variables, while the Mann-Whitney U test was used for quantitative variables with non-parametric distribution.

Results

Of the 204 documents analysed, 133 (65.2%) were published in special issues, and 71 (34.8%) in regular issues ($p < 0.001$). The average number of papers analysed per study was 804 [222-2,611] over an average duration of 23.0 [15-33.8] years. The sample was smaller for papers published in special issues (647.5 [221.5_2,061.8]) than for papers published in regular issues (1,021 [225.5_3,567.8], $p = 0.239$). Interestingly, in 18 of the papers the samples consisted of less than 100 documents and between 100 and 200 documents in the case of 23 others. The number of years analysed was somewhat lower for papers published in special issues compared to those published in regular issues (22.5 [14_33.0] vs 24 [18_35], $p = 0.393$). In six of the studies it was impossible to determine the sample analysed, and the years of the period analysed were not specified in four of the 204 papers reviewed.

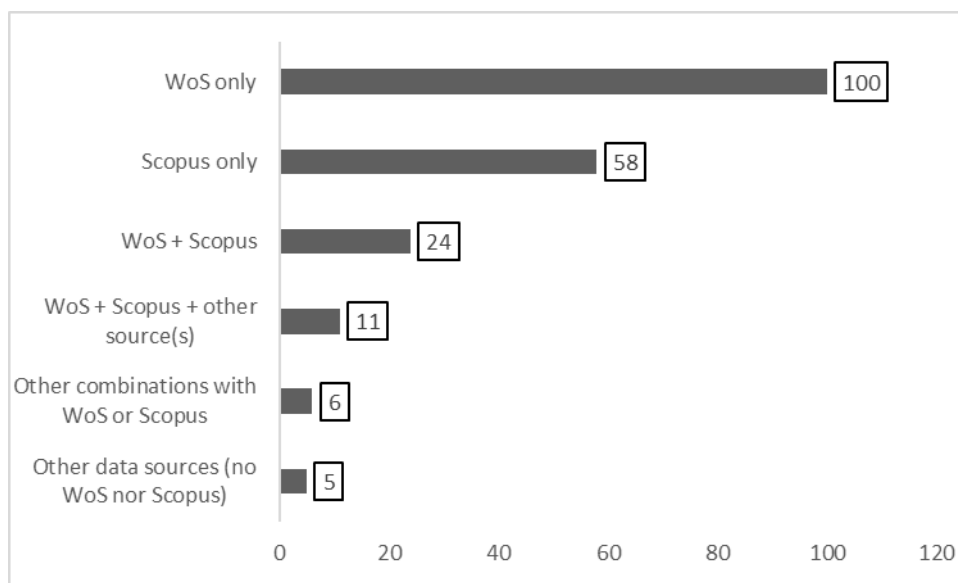
Table 1. Descriptive statistics of the Sample and Period variables

	n	Median [IQR]	p value
Sample	198	804.5 [221.5_2,611.3]	0.239
Special Issue	130	647.5 [221.5_2,061.8]	
Regular Issue	68	1,021 [225.5_3,567.8]	
Period (years)	200	23.0 [15_33.8]	0.393
Special Issue	130	22.5 [14_33.0]	
Regular Issue	70	24 [18_35]	

Data are expressed as medians (interquartile range) or n (%)

The most common approach was to carry out bibliometric analyses based solely on Web of Science (WoS) (100 papers, 49%) or Scopus (58 papers, 28.4%), although 11.8% of the papers used a combination of both sources. The remaining papers also used other additional sources in combination with WoS, Scopus or both simultaneously, while only five papers (2.5%) did not use WoS or Scopus (see Figure 2). Of the 139 papers that made single or combined use of WoS, 90 of them (64.7%) did not report the specific indexes used for data collection on this platform owned by Clarivate Analytics ($p < 0.001$).

Figure 2. Data sources used



Half of the papers analysed reported the exact search strategy used, with a slightly higher percentage of papers published in regular issues giving an adequate breakdown (53.5%) compared to those published in special issues (48.1%) ($p = 1.000$). As for the document types included in the analyses, 23% of the papers included all of them and 62.7% selected some of them ($p < 0.001$). Finally, 14.2% of the papers did not indicate the types of documents analysed in the bibliometric study.

In terms of languages, 21.6% of the sample analysed papers in all the languages available in the selected data sources, while 31.9% only considered papers in English. Meanwhile, 91 of these studies (44.6%) did not specify whether the papers were analysed in all the languages of the database or only in one of them. Papers published in regular issues tend to analyse only English publications (42.3%) while in the case of papers published in special issues, the

number of papers analysing only English documents and those analysing documents in all languages is similar (Table 2).

Table 2. Accuracy of the search equation, document types and languages analysed

		N	%	p value*
Exact Search String	Yes	102	50.0	0.462
	Special Issue	64	48.1	
	Regular Issue	38	53.5	
	No	102	50.0	
	Special Issue	69	51.9	
	Regular Issue	33	46.5	
Document Type	All	47	23.0	0.692
	Special Issue	32	24.1	
	Regular Issue	15	21.1	
	Some types	128	62.7	
	Special Issue	84	63.2	
	Regular Issue	44	62.0	
	No data	29	14.2	
	Special Issue	17	12.8	
	Regular Issue	12	16.9	
Language	All	44	21.6	0.026*
	Special Issue	36	27.1	
	Regular Issue	8	11.3	
	Only English	65	31.9	
	Special Issue	35	26.3	
	Regular Issue	30	42.3	
	Other combinations	4	2.0	
	Special Issue	3	2.3	
	Regular Issue	1	1.4	
	No data	91	44.6	
	Special Issue	59	44.4	
	Regular Issue	32	45.1	

Discussion and conclusions

This paper has explored various issues relating to the methodology used in bibliometric analyses published in the journal *Sustainability* in 2019 and 2020, showing that many of the papers published lack the methodological rigour that would normally be required. In particular, the following limitations have been identified:

- Small samples: bibliometric analyses do not make sense “when the scope of review is specific” or “when the dataset is small and manageable enough that its content can be manually reviewed” (Donthu, Kumar, Mukherjee, Pandey, & Lim, 2021). In this regard, a total of 41 papers (20.1%) were detected that analysed samples of less than 200 documents, suggesting that such studies could have been approached using other more appropriate techniques.

- Inadequate description of the search equation applied: not all the terms used in the search are broken down in detail, nor the fields in which the search is performed. 102 of the papers reviewed (50%) had shortcomings of this nature, making it difficult or impossible to reproduce the original bibliometric analysis.
- Lack of information on the indexes used within the Web of Science platform: 90 of the papers (44.1% of the total sample and 64.7% of the studies using WoS) did not indicate the specific sources from which the information had been extracted. Some of these papers did list the Web of Science Core Collection as the data source, but without specifying the specific databases searched. As highlighted by Liu (2019), this is an inadequate practice given that “many institutions may choose to subscribe to a customised subset of the whole core collection,” which significantly affects the reproducibility of the work.
- Lack of information on the document types and languages covered by the bibliometric analysis: up to 103 of the papers analysed (50.5%) did not break down either (or both) of these variables.

Considering these shortcomings, 181 of the 204 studies analysed (88.7%) have one or more methodological limitations which hinder or prevent their reproducibility. This shows that there is considerable room for improvement in the methodological quality of the bibliometric papers published in Sustainability, although there are no significant differences between the characteristics of the papers published in special issues and those published in regular issues. Accordingly, our results further underline the concerns raised by authors such as Jonkers & Derrick (2012) and González-Alcaide (2021) regarding the quality of bibliometric papers published in journals outside the IS field, or by Oviedo-García (2021) regarding the rigour of the peer review processes applied by MDPI.

One limitation of this study is that it is a case study of a single journal and it does not use a control group to determine whether the methodological shortcomings detected also occur in IS journals, and more specifically in the most productive journal in the field, *Scientometrics*. In terms of future studies, it would be interesting to extend the content analysis to other variables, such as the techniques used to analyse the results of these studies, or other aspects relating to their reproducibility or transparency. Finally, we suggest that these variables could be studied in other journals with extensive bibliometric production outside the IS field, as well as measuring their relationship with the formal characteristics of the papers, such as the number of authors, collaborations or citations received.

In short, there is a need to further study the characteristics of bibliometric analyses published in journals of all disciplines in order to promote increasingly rigorous use of this powerful quantitative methodology.

References

- Björk, B.-C., & Catani, P. (2016). Peer review in megajournals compared with traditional scholarly journals: Does it make a difference? *Learned Publishing*, 29(1), 9–12.
- 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, 285–296.

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, *105*(3), 1809–1831.

González-Alcaide, G. (2021). Bibliometric studies outside the information science and library science field: Uncontainable or uncontrollable? *Scientometrics*, *126*(8), 6837-6870.

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), 829–836.

Liu, W. (2019). The data source of this study is Web of Science Core Collection? Not enough. *Scientometrics*, *121*(3), 1815–1824.

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), 405–419.

Petrovich, E. (2022). Bibliometrics in Press. Representations and uses of bibliometric indicators in the Italian daily newspapers. *Scientometrics*. Retrieved April 26, 2022, from <https://doi.org/10.1007/s11192-022-04341-6>

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.

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), 1386–1401.

Spezi, V., Wakeling, S., Pinfield, S., Creaser, C., Fry, J., & Willett, P. (2017). Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review. *Journal of Documentation*, *73*(2), 263–283.