Scientometrics (2019) 119:543–547
https://doi.org/10.1007/s11192-019-03040-z

Missing documents in Scopus: the case of the journal *Enfermeria Nefrologica*

Erwin Krauskopf[1,2]

Received: 31 January 2019 / Published online: 9 March 2019

Abstract
In a recent paper, a group of researchers estimated various bibliometric indicators for the Spanish journal *Enfermeria Nefrologica* using the software "Publish or Perish", retrieving data exclusively from Google Scholar. Since their study revealed an unusual high number of citations for the documents published by the journal, we became interested in repeating the bibliometric analysis using data from Scopus. Surprisingly, our analysis revealed a high variability in the number of documents published each year. Therefore, the journal's website was accessed to confirm whether this irregularity was due to the journal's publication frequency. According to the data collected, only 50.2% of the documents published by the journal between 2006 and 2017 were registered by Scopus. Such omission-induced errors raise concerns about the validity of various indicators. This study shows that while Scopus needs to improve its quality control systems, editorial management teams need to routinely check the information being indexed by the databases.

Keywords Scopus · Missing issues · Missing documents · Indexing · Journal

Introduction

A recently published bibliometric study (Cobo Sanchez et al. 2018) analyzed the impact of manuscripts published by a Spanish journal entitled *Enfermería Nefrológica* between 1998 and 2017. In this study the authors estimated various bibliometric indicators using the software "Publish or Perish" by retrieving data exclusively from Google Scholar. While Google Scholar may present advantages over other databases such as the ability to recognize different forms of citation and the inclusion of material not covered by Scopus or Web of Science (thesis, dissertations and technical reports) (Martell 2009), caution should be exercised with the conclusions derived from their analysis. In fact, a study that examined the citation levels of articles published in medical journals indexed by Web of Science, Scopus and Google Scholar, established that the citations received differed quantitatively and qualitatively among the three databases (Kulkarni et al. 2009). In this context, it is

✉ Erwin Krauskopf
erwin.krauskopf@unab.cl

[1] Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile

[2] Fundacion Ciencia y Vida, Zanartu 1482, Santiago, Chile

valid to ask whether the citation of a webpage that has not been submitted to peer review is quantitatively equivalent to a citation made by a peer-reviewed article published in a journal indexed by Scopus or Web of Science. Subsequent studies showed that the lack of a quality control system for data collected by Google Scholar allows the number of recorded citations to be manipulated (Bohannon 2014; Lopez-Cozar et al. 2012).

Since the original study (Cobo Sanchez et al. 2018) revealed an unusual high number of citations for the documents published by this journal, I became interested in establishing how many derived from indexed journals. As the journal has been indexed by Scopus since 2006, the bibliometric analysis was repeated for the documents published between 2006 and 2017.

Methods

Bibliometric information was retrieved from Scopus on the week of January 23 2019 for the journal *Enfermeria Nefrologica* (and its previous title *Revista de la Sociedad Española de Enfermeria Nefrologica*). The downloaded data was sorted and processed using Microsoft Excel.

Results

According to Scopus, a total of 559 documents were published by the journal *Enfermeria Nefrologica* between 2006 and 2017. However, the number of documents published each year was highly variable (Fig. 1). It was unusual to observe more than 50% reduction in the number of published documents for only one year (15 documents in 2008) while five years later an increase of approximately 400% in the number of documents. Nonetheless, a closer look at the data from 2013 revealed that the journal published a special issue that contained 116 documents.

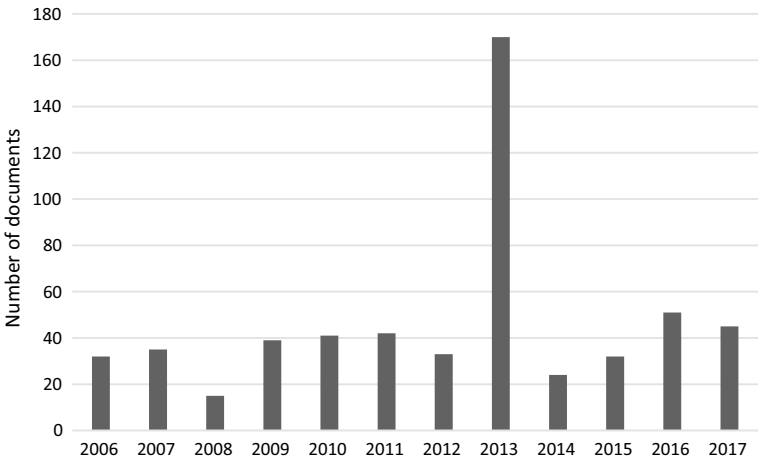

Fig. 1 Trends in the number of documents published annually by *Enfermeria Nefrologica*

Certainly, the irregular publishing frequency of the journal *Enfermeria Nefrologica* seemed anomalous as every year Scopus strictly analyzes the performance of each indexed journal through various metrics, such as number of articles published. Therefore, in order to verify the data collected from Scopus, the journals' webpage (www.revistaseden.org) was accessed in order to establish the publication frequency and the number of documents published each year. Surprisingly, the number of documents published each year was not uneven after all (Table 1). In fact, the journal published four regular issues every year, with an average of 42 documents each year (not including supplementary issues that were introduced in 2012). According to the webpage, the journal published a total of 1113 documents between 2006 and 2017. Thus, Scopus only registered 50.2% of these documents.

Hence, the lower number of documents registered by Scopus is due to missing issues? Or perhaps missing documents? In the case of the Scopus data from 2008, is due to both reasons. Our analysis revealed that two complete issues (3 and 4) were missing from the Scopus database, accounting for 22 documents. Additionally, four documents from 2008 were also missing (two from issue 1 and two from issue 2). In fact, for all the years analyzed in this study we discovered documents published on regular issues that were not registered by Scopus. Another question that arises regards supplementary issues. Why does Scopus only contain data for the year 2013 if these issues have been published non-stop since 2012?

Discussion and conclusions

One of the key aspects of subscription-based indexing databases such as Scopus are its accuracy and consistency. While several studies have reported data quality problems (Krauskopf 2017; Van Eck and Waltman 2017; Liu et al. 2018; Zhu et al. 2018), only a single editorial warned about the consequences of missing issues from such databases (Kellner and Azevedo 2013). In this particular case, the editorial team detected that an

Table 1 Number of documents published by *Enfermeria Nefrologica*, sorted by year and issue number

Year	Issue 1	Issue 2	Issue 3	Issue 4	Supp.
2006	8	9	10	9	
2007	9	10	11	8	
2008	9	10	11	11	
2009	11	12	12	12	
2010	11	10	12	11	
2011	13	10	9	11	
2012	11	10	11	11	118
2013	10	11	10	10	116
2014	12	10	10	9	117
2015	10	10	10	11	111
2016	11	10	11	14	138
2017	11	12	11	12	6
2018	11	10	12	10	14

Data was collected from the journal's website (www.revistaseden.org) on January 25th 2019

Supp supplementary issue

issue had not appeared in Web of Science, even though a latter issue was already available through Web of Science. Unfortunately, by the time the Journal Citation Reports 2012 was released, the problem had not been fixed so it affected the journal's impact factor.

Such omission-induced errors raise concerns about the validity of various indicators. For instance, the journal's Citescore for 2017 is based on the citations count from 2017 for all the documents published between 2014 and 2016. In the case of the journal *Enfermeria Nefrologica*, 12 citations were received by 107 documents published, according to Scopus. However, a total of 494 documents (128 from regular issues 366 from supplementary issues) were actually published by the journal, therefore, the Citescore provided is erroneous.

This also extends to other indicators such as the Scimago Journal Rank (SJR). This size-independent indicator establishes the scientific influence of journals considering the number of citations received and the prestige of the journals where these citations originate. Thus, how many documents were considered to determine the journal's 2017 SJR? According to the Scimago web page (www.scimagojr.com), *Enfermeria Nefrologica* published 555 documents (476 citable items and 79 non-citable items) between 2015 and 2017, whereas Scopus accounts for 128 documents and the journal's website for 388 documents.

So, who should be accountable for such differences? It seems that Scopus' quality control system is failing to verify whether all the manuscripts from a journal are being registered. A previous study also reported problems with quality control of duplicate records in Scopus (Valderrama-Zurian et al. 2015). Nevertheless, the journal's editorial management team also need to assume responsibility for their lack of awareness. It is their responsibility to routinely check if all the issues and manuscripts have been registered by Scopus. In fact, the editorial management team should be concerned as to why only 10 documents from 2018 have been registered by Scopus.

Fifty years ago, Pritchard (1969) defined bibliometrics as the application of mathematical and statistical methods to books and other media of communication. Since then, bibliometrics has progressed to become a powerful tool for research management and for the development of scientific policies. However, if the data being collected for bibliometric assessments is flawed, this will severely affect scientific evaluation.

Acknowledgements This work was supported by the Grant Proyecto de Financiamiento Basal AFB170004.

References

Bohannon, J. (2014). Google Scholar wins raves-but can it be trusted. *Science, 343,* 14.

Cobo Sanchez, J. L., Ochando Garcia, A., Blanco Mavillard, I., Cirera Segura, F., Crespo Monteo, R., & Casas Cuesta, R. (2018). Análisis del impacto del a producción científica de la revista Enfermería Nefrológica entre 1998–2017. *Enfermeria Nefrologica, 21*(4), 349–358.

Kellner, A. W. A., & Azevedo, R. A. (2013). A closer look at the impact factor (JCR 2012): problems, concerns and actions needed. *Anais da Academia Brasileira de Ciencias, 85*(3), 859–862.

Krauskopf, E. (2017). Call for caution in the use of bibliometric data. *Journal of the Association for Information Science and Technology, 68*(8), 2029–2032.

Kulkarni, A. V., Aziz, B., & Shams, I. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA, 302*(10), 1092–1096.

Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in Web of Science—An explorative study. *Journal of Informetrics, 12*(3), 985–997.

Lopez-Cozar, E.D., Robinson-Garcia, N., & Torres-Salinas, D. (2012) *Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting*. ArXiv Preprint. 2012. Retrieved from http://arxiv.org/labs/1212.0638. Accessed January 27, 2019.

Martell, C. (2009). A citation analysis of college & research libraries comparing Yahoo, Google, Google Scholar, and ISI web of knowledge with implications for promotion and tenure. *College & Research Libraries, 70*(5), 460–472.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 24,* 348–349.

Valderrama-Zurian, J. C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics, 9*(3), 570–576.

Van Eck, N. J., & Waltman, L. (2017). Accuracy of citation data in Web of Science and Scopus. In *ISSI 2017—16th international conference on scientometrics and informetrics, conference proceedings* (pp. 1087–1092).

Zhu, J., Hu, G., & Liu, W. (2018). DOI errors and possible solutions for Web of Science. *Scientometrics*. https://doi.org/10.1007/s11192-018-2980-7.