

# Review of: "Comparison of self-citation patterns in WoS and Scopus databases based on national scientific production in Slovenia (1996-2020)"

Silvio Peroni<sup>1</sup>

<sup>1</sup> University of Bologna

**Potential competing interests:** The author(s) declared that no potential competing interests exist.

## Metadata

**Title:** Comparison of self-citation patterns in WoS and Scopus databases based on national scientific production in Slovenia (1996-2020)

**Author:** Gordana Budimir, Sophia Rahimeh, Sameh Tamimi, Primož Južnič

**Submitted to:** Scientometrics

## Review

In this article, the authors compare the presence of self-citations in Web of Science (WoS) and Scopus according to the scientific production in Slovenia. The intended goal is to check whether the two citation indexes are different in their self-citation counts and, even when different, whether they are comparable to some extent.

In general, these kinds of study are welcome, since they offer a view on a particular country that can help both researchers and policymakers in their analysis. However, some issues should be addressed, and some clarification should be provided before the article is acceptable for publication.

### 1. The goal of the article too late in the text

The authors did a great job of summarising prior works on comparing Scopus and WoS and on self-citations in general. However, the description of the actual goal of their study is introduced only at the end of the introduction. The introduction may cause confusion and misunderstanding to the reader, who may not understand the central claim of this article until page 3. It would be good to restructure the introduction by anticipating the goal of the study. Also, the authors may reformulate the points at the end of the introduction as research questions (RQs), to recall them explicitly when presenting and discussing the results.

### 2. Use of p-value as a measure of significance

The authors use p-values to stress the significance of the results they have obtained. However, several studies in the past

have demonstrated the possible drawbacks in using p-values to convey significance and correctness of the results, e.g. see [\[1\]](#) and [\[2\]](#).

The suggestion would be to abandon p-value (and the "statistical significance" way of thinking) to replace it with measures that estimate the size of an effect (e.g. how massive is the difference between two means), and to accompany such measurement with confidence intervals, that provides a degree of uncertainty and are very informative (in particular when drawn in a graph).

Traditional statistical significance tests give the impression that all kinds of problems that require subjective interpretation can be solved in a completely objective way (i.e. the result is either statistically significant or not). However, understanding whether a difference is practically important needs a subjective interpretation. In the estimation statistics approach, instead, the need for subjective interpretation is explicitly recognized.

### 3. Methodology to identify self-citations

The authors started from the authority control systems they have available in Slovenia (i.e. SICRIS and COBISS) to identify Slovenian authors. However, it is not clear how they retrieved the authors' documents from Scopus and WoS. Did SICRIS and COBISS have precise alignments (for documents and people) to Scopus and WoS? If not, how were the authors able to understand which authors and documents to consider in Scopus and WoS?

A more detailed description of which data are contained in SICRIS and COBISS may help to understand more clearly the method used by the authors to retrieve the documents considered in their study.

### 4. Related to the ageing of self-citations

Good to see how self-citations ageing in time. However, in order to have a clearer picture in comparing WoS and Scopus, it would be good to see if similar self-citations patterns can also be found when considering synchronous citations, i.e. those an article gave using the bibliographic references in its reference list, instead of those that received.

### 5. Self-citations affected by external events

At the end of the paper, the authors say "we believe that the self-citation behaviour of Slovenian researchers cannot be so different from other researchers and that our results might provide some broader information on both database". While from a perspective which looks at comparing WoS and Scopus this claim may be valid, indeed it does not hold in general. Self-citations may be affected by several external events, such as national research assessment exercises. To this end, please see [\[3\]](#) and [\[4\]](#) [disclosure: I am the author of the first article].

### 6. Typos

Fig. 1 caption: "dotted" -> "dashed"

## References

1. ^ Geoff Cumming. (2013). *The New Statistics*. *Psychol Sci*, vol. 25 (1), 7-29. doi:10.1177/0956797613504966.
2. ^ David Colquhoun. (2014). *An investigation of the false discovery rate and the misinterpretation of p -values*. *R. Soc. open sci.*, vol. 1 (3), 140216. doi:10.1098/rsos.140216.
3. ^ Silvio Peroni, Paolo Ciancarini, Aldo Gangemi, Andrea Giovanni Nuzzolese, Francesco Poggi, Valentina Presutti. (2020). *The practice of self-citations: a longitudinal study*. *Scientometrics*, vol. 123 (1), 253-282. doi:10.1007/s11192-020-03397-6.
4. ^ Alberto Baccini, Giuseppe De Nicolao, Eugenio Petrovich. (2019). *Citation gaming induced by bibliometric evaluation: A country-level comparative analysis*. *PLoS ONE*, vol. 14 (9), e0221212. doi:10.1371/journal.pone.0221212.