

Document downloaded from:

<http://hdl.handle.net/10251/84569>

This paper must be cited as:

Martín-Martín, A.; Ayllón, JM.; Delgado López-Cózar, E.; Orduña Malea, E. (2015). Nature's Top 100 Re-Revisited. *Journal of the American Society for Information Science and Technology*. 66(12):2714-2714. doi:10.1002/asi.23570.



The final publication is available at

<http://doi.org/10.1002/asi.23570>

Copyright Association for Information Science and Technology (ASIS&T)

Additional Information

"This is the peer reviewed version of the following article:
Martín-Martín, A., Ayllon, J. M., López-Cózar, E. D., & Orduna-Malea, E. (2015). Nature's top 100 Re-revisited. *JASIST*, 66(12), 2714., which has been published in final form at <http://doi.org/10.1002/asi.23570>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

Nature's Top 100 Re-Revisited

Alberto Martín-Martín

EC3 Research Group. Facultad de Comunicación y Documentación, Colegio Máximo de Cartuja s/n, Universidad de Granada, 18071, Granada, Spain. E-mail: albertomartin@ugr.es

Enrique Orduna-Malea

EC3 Research Group. Instituto de Diseño y Fabricación, Camino de Vera s/n, Polytechnic University of Valencia (UPV), 46022 Valencia, Spain. E-mail: enorma@upv.es

Juan M. Ayllón

EC3 Research Group. Facultad de Comunicación y Documentación, Colegio Máximo de Cartuja s/n, Universidad de Granada, 18071, Granada, Spain. E-mail: jayllon@ugr.es

Emilio Delgado López-Cózar*

EC3 Research Group. Facultad de Comunicación y Documentación, Colegio Máximo de Cartuja s/n, Universidad de Granada, 18071, Granada, Spain. E-mail: edelgado@ugr.es, Tel: 958243821.

Abstract:

To mark the 50th anniversary of the Science Citation Index, Nature published a list of the 100 most-cited papers of all time. It also included an alternative ranking from data provided by Google Scholar, which, as this letter illustrates, contains certain inconsistencies. This does not, however, diminish the usefulness of Google Scholar, not only in identifying the most-cited articles of all time, but also in reflecting the impact of other document types (especially books), thus redefining the concept of academic impact.

Keywords: Google Scholar, Web of Science, Highly cited documents, Rankings, Inconsistencies.

Dear Sir,

In a recent letter published in the Journal of the Association for Information Science and Technology, Bornmann (in press) criticizes Nature's top 100 ranking. Van Noorden, Maher and Nuzo (2014) requested this list of the most-cited research of all time from Web of Science to mark the 50th anniversary of the Science Citation Index. Bornmann expresses concern about the methods used to generate the list (based on raw citation counts and not normalized bibliometric indicators).

The Nature article also provides an alternative list of most-cited research contributions according to Google Scholar (available in the online version of the article as complementary material). Although we acknowledge that the main focus of the article is the data extracted from Web of Science, we believe it necessary to point out some discrepancies in the Google Scholar list.

For example, it lists “Protein measurement with the folin phenol reagent” as the second most-cited article (192,710 citations), contradicting the Web of Science ranking, which shows it to be the most cited by far (Garfield, 2005), a fact that merits a thorough discussion.

Apart from this issue (anecdotal, perhaps, but worth noting), there are certain inconsistencies that do not seem to have been considered in the Google Scholar list published by Nature. We discovered these inconsistencies when researching highly cited documents in Google Scholar (Martin-Martin, Orduna-Malea, Ayllon & Delgado Lopez-Cozar, 2014), and they relate, in particular, to the allocation of citations and the identification and linkage of different versions of the same documents.

According to our empirical data, the aforementioned article on protein measurement had attracted, as of May 2014, a total of 253,671 citations, whereas Nature’s ranking (extracted from Google Scholar on 17 October 2014) records only 192,710 citations. How can an article lose 60,961 citations in five months? Conversely, the “Diagnostic and statistical manual of mental disorders”, not included in the top ten despite having 185,000 citations in Google Scholar (as of October 2014), and almost 220,000 if we merge its various versions, would seem to have attracted a remarkable 55,170 citations from May to October.

Moreover, two different editions of “Molecular cloning” appear on the list. Adding up the two versions (and other unmerged records), the citations amount to 268,834, which would

promote this work to first place in the ranking. Likewise, we found 164 additional unmerged records for “A mathematical theory of communication”, where citations to the article and the subsequent book are mixed.

To what extent, therefore, can we trust this Google Scholar list?

Apart from these errors in the preparation of the list, mainly due to a lack of professional filtering (necessary if we wish to compare citations on Google Scholar with Web of Science), we wish to note two important findings: a) even when dirty (unfiltered), Google Scholar is capable of identifying the most-cited papers; and b) Google Scholar is capable of providing a complementary academic landscape (including books, heavily cited in certain fields, and traditionally discriminated against).

And this is what should be retained, regardless of positions or exact figures. Let us not fall into the classic trap of not seeing the wood for the trees.

References

Bornmann, L. (in press). Nature's top 100 revisited. *Journal of the Association for Information Science and Technology*. Retrieved April 20, 2015, from http://www.lutz-bornmann.de/icons/top_100.pdf

Garfield, E. (2005). The agony and the ecstasy- the history and meaning of the Journal Impact Factor. In *International Congress on Peer Review And Biomedical Publication*. Retrieved April 20, 2015, from <http://www.garfield.library.upenn.edu/papers/jifchicago2005.pdf>

Martin-Martin, A., Orduna-Malea, E., Ayllon, JM., & Delgado Lopez-Cozar, E. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? Retrieved April 20, 2015, from <http://arxiv.org/abs/1410.8464>

Van Noorden, R, Maher, B. & Nuzzo, R. (2014). The top hundred papers. *Nature* 514, 550-553.