RECRUS
Res. Newsl.

November 2021
Vol. 1 Issues 9
Page 69

What Are the World Top Scientists?

By: Boon-How Chew

In its latest version 3 "Updated science-wide author databases of standardized citation indicators", the Stanford-based authors have again made available the ranking list of the world top 2% scientists (100,000 scientists) according citation index derived from an online publication database [1,2]. News may be coming from institutions and individuals scientists of how many of them and who or what they are in the list, respectively. A few glances on the lists on Excel files accessible [here] show that names change on the lists in the different versions, and quite sizable of scientists are coming from local institutions. However, it is of first importance to understand the 'strange' bibliometric ranking indicators.

The raw data are number of citations of individual publication from the Scopus and specified into six indicators and one composite score (see below). These are considered to indicate the citation impact of the published research and validated against Nobel laureates taken as the world leading scientists [3]. The higher the scores indicate the higher impact of the scientists. Separate data are shown for career-long and single-year impact. Career-long data are updated from 1st January 1996 to end-of-2020. The single-year impact is based on citations accumulated in the single year 2020. Metrics with and without self-citations and ratio of citations to citing papers are given. Scientists are classified into 22 scientific fields and 176 subfields. Field- and subfield-specific percentiles are also provided for all scientists who have published at least five papers.

1. Total citations [**NC**]
2. Hirsch index (h-index) [**H**],
3. Schreiber coauthorship-adjusted hm-index [**HM**],
4. Total citations to papers in different authorship positions either as single [**NS**], single or first [**NSF**], or single or first or last author [**NSFL**], since these positions are usually pivotal to the published work)
5. A composite indicator (**C**).

Scopus contains complete information including reference lists for documents published since 1996 [3]. The Scopus identifiers have also been shown to have very high precision for matching to specific single authors when comparison was made against funding databases. The initiation year of 1996 will limits the metrics's ability in assessing the full citation profile of a scientist prior to 1996. This limitation on the career-long impact could be minimised by supplementing it with the single-year impact. The recent citation impact of a scientist such as that is shown by the single-year impact will be more reflective of the current visibility and future citation rate e.g., the number of citations in a given year is the strongest predictor of citations in future years and perhaps also the potential for continuing to perform influential work in the near future.

The composite scores of the multiple indicators were first assessed in 2013 evaluated the top 30,000 scientists (out of 84,116) based on each of six citation indicators and also developed a composite score that combines the six indicators [3]. Correlation analyses showed that NC and H correlated strongly strong (r = 0.88), no correlation between the NC and Hm (r = 0.00), but after excluding physics (strongest negative correlation), the correlation of NC with the Hm across other disciplines varied between r = 0.21 to 0.76, and always smaller than the correlation between NC and H. There was a negative correlation between NC and each of the three author-order-based citation indicators NS (r = -0.43), NSF (r = -0.22), and NSFL (r = -0.04). These weak and even negative correlations were generally seen in disciplines that collaborate extensively such as in physics, biology and medicine. This means selection of top scientists based on total citations only may include many scientists (>50%) without a leading, key contribution to the published work (as the single, first or last authors of influential papers). See the box below an excerpt from the paper that illustrate the significance of authorship order.

This was the selection approach without considering the order of authorship by the Thomson Reuters of the most influential scientists in the world who had not been the first or last author of any highly cited paper. Whereas Clarivate Analytics included less than 0.1% of the world scientists in its evaluation [3]. Google Scholar reports citation index too but not all researchers have created a profile.

For example, of the 100 most-cited authors according to total citations (NC = 7,457–19,245 Scopus citations received in 2013), almost half of them (n = 48) had received only 0–20 citations to papers or other scholarly documents for which they were the single author, and 15 of them did not have even a single cited single-authored paper or other scholarly document. Moreover, of the 100 most-cited authors according to total citations, 12 had received fewer than 100 citations in papers for which they were the single or first author, and 11 had received fewer than 1,000 citations to all papers for which they were single, first, or last author.

The composite index C is the summed of standardized scores of each of the six log-transformed citation indicators (NC, H, Hm, NS, NSF, NSFL). It is standardized on the maximum score achieved for that particular indicator with value ranged from 0 to 1, where 1 is given to the scientist with the highest raw value for the respective indicator. Of the 1,000 top-ranked scientists according to the composite score, only 322 were ranked such based on total citations [3].

Assuming Nobel laureates are leading scientists in the world, further analyses with the indicators showed that different scientists populated the top ranks when different indicators were used. Many Nobel laureates and other influential scientists rank among the top 1,000 with the composite indicator, but rank much lower based on total citations [3]. Conversely, many of the top 1,000 authors on total citations had no single/first/last-authored cited paper. More Nobel laureates are among the top authors when the ranking is according to the composite score (31 of the 47 laureates of year 2011–2015) than by single indicators of the total citations (15 of 47), h-index (18 of 47) or Hm index (26 of 47). Multiple indicators and their composite give a more comprehensive picture of impact as they are less influenced by extreme values in only one indicator. Nevertheless, no method can pick all the best scientists as evidenced by that 40/47 of the Nobel laureates are among the top 30,000 by at least one of the six indicators. Scientists with earlier citation peaks and poor coverage of certain fields in Scopus could be misclassified based on these indicators.

Interesting results from the paper [3] also showed that there was tremendous variation in the number of papers and the citations per paper across the top 1,000 scientists with the highest composite score, the number of papers varied from 26 to 2,533 and the citations per paper varied from 1.6 to 288.9. These two indicators are not included in the calculation of the composite index C because they are believed to be better captured by the already defined indicators, and to discourage rewarding scientists for publishing more papers of sloppy and over-claimed science. This conveys a strong message that quality is different from quantity, both can co-exist but quality should take the precedence, and without quality as the prime characteristic of a research there will not be good impact we want to see. Nonetheless, the 'quality' here has to be qualified qualitatively of each of the scientist and of each piece of the published work because the citation impact can be due to different reasons such as early work (case reports), preliminary research, different site, study samples of another population, atypical results, questionable methods, updated results (systematic reviews), or innovative studies of gap-filling studies, definitive evidence (both positives and negatives) from experimental studies at routine practices or life and living of the target users of the research.

Recommendation

Use the composite indicator as it has been proven to more informative than any of the single indicator. Examine the multiple indicators in parallel may be useful for certain quality of the scientist. All indicators need to be used with knowledge of both their strengths and weaknesses. Bear in mind that it is difficult to differentiate the ranking of scientists because small differences in their indicators. Although the impact indicators across the different disciplines have been standardised and levelled but comparisons between them should remain cautious because the ideal proportional representation of scientists from different fields among those claimed to be top scientists should be is not known. Similarly, the impact indicators do not tell about the full potential of recent published works and young scientists who have just published their first few papers. If insights are needed for these impact, one can limit the evaluation of indicators to papers published in the last few years.

References

1. Baas J Boyack K, Ioannidis JPA. "August 2021 data-update for "Updated science-wide author databases of standardized citation indicators"", Mendeley Data, V3, 2021, doi: 10.17632/btchxktzyw.3
2. Ioannidis JPA, Baas J, Klavans R, Boyack KW. A standardized citation metrics author database annotated for scientific field. PLoS Biol. 2019;17(8): e3000384. https://doi.org/10.1371/journal. pbio.3000384
3. Ioannidis JPA, Klavans R, Boyack KW. Multiple Citation Indicators and Their Composite across Scientific Disciplines. PLoS Biol. 2016;14(7): e1002501. doi:10.1371/journal.pbio.1002501