PLOS BIOLOGY

FORMAL COMMENT

Updated science-wide author databases of standardized citation indicators

John P. A. Ioannidis[1,2,3,4]*, **Kevin W. Boyack**[5], **Jeroen Baas**[6]

1 Department of Medicine, Stanford University, Stanford, California, United States of America, **2** Department of Epidemiology and Population Health, Stanford University, Stanford, California, United States of America, **3** Department of Biomedical Data Science, Stanford University, Stanford, California, United States of America, **4** Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, United States of America, **5** SciTech Strategies, Inc., Albuquerque, New Mexico, United States of America, **6** Research Intelligence, Elsevier B.V., Amsterdam, the Netherlands

* jioannid@stanford.edu

Check for updates

OPEN ACCESS

Citation: Ioannidis JPA, Boyack KW, Baas J (2020) Updated science-wide author databases of standardized citation indicators. PLoS Biol 18(10): e3000918. https://doi.org/10.1371/journal.pbio.3000918

Received: August 3, 2020

Accepted: September 18, 2020

Published: October 16, 2020

Copyright: © 2020 Ioannidis et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received no specific funding for this work.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests. JPAI is a member of the editorial board of *PLOS Biology*. JB is an Elsevier employee. Elsevier runs Scopus and ICSR Lab, which is the source of this data, and also runs Mendeley Data, where the database is now stored.

There was great interest in the databases of standardized citation metrics across all scientists and scientific disciplines [1], and many scientists urged us to provide updates of the databases. Accordingly, we have provided updated analyses that use citations from Scopus with data freeze as of May 6, 2020, assessing scientists for career-long citation impact up until the end of 2019 (Table-S6-career-2019) and for citation impact during the single calendar year 2019 (Table-S7-singleyr-2019). Updated databases and code are freely available in Mendeley (https://dx.doi.org/10.17632/btchxktzyw). The original database (version 1) can also be found in https://data.mendeley.com/datasets/btchxktzyw/1, the updated (version 2) can also be found in https://data.mendeley.com/datasets/btchxktzyw/2, and any subsequent updates that might appear in the future will be generally accessible in https://dx.doi.org/10.17632/btchxktzyw.

S6 and S7 tabulated data include all scientists who are among the top 100,000 across all fields according to the composite citation index [2] when self-citations are included and/or when self-citations are not included. Furthermore, in the current update, Tables S6 and S7 include also scientists who are not in the top 100,000 according to the composite index but are nevertheless within the top 2% of scientists of their main subfield discipline, across those that have published at least five papers. Another new feature in this update is that Tables S6 and S7 include new columns showing for each scientist the rank of their composite citation index within their subfield discipline (with and without self-citations) and the total number of authors within the subfield discipline. For example, for Kevin W. Boyack, rank is 50 and 52 for the composite citation index with and without self-citations, respectively, among the total of 10,391 scientists whose main subfield discipline is "Information and Library Sciences." This extension allows the inclusion of more comprehensive samples of top-cited scientists for fields that have low citation densities and therefore would be less likely to be found in the top 100,000 when all scientific fields are examined together. Comparisons of citation metrics are more meaningful when done within the same subdiscipline. Of course, even within the same subdiscipline, different areas may still possess different citation densities, and assessing citation indicators always require caution.

Field and subfield discipline categories use the Science-Metrix classification as in our previous work [1], but multidisciplinary journals that were previously not assigned to a Science-Metrix field or subfield [3] have now been assigned to a specific field and subfield using a character-based convolutional deep neural network. This machine learning approach was trained with a set consisting of over a million entries was found to be outperforming other approaches