**FREQUENTLY ASKED QUESTIONS**

Please check carefully the answers to these frequently asked questions and the provided references. We will not answer any e-mails with questions that are covered by material presented in the Frequently Asked Questions or in the introductory documentation to the databases.

1. Where can I see the detailed **Methods** used to calculate the composite citation indicator and to construct the citation databases?

Reply: Please read the 3 associated PLoS Biology papers that explain the development, validation and use of the composite citation indicator metrics and databases

Ioannidis JPA, Boyack KW, Baas J. Updated science-wide author databases of standardized citation indicators. PLoS Biol. 2020 Oct 16;18(10):e3000918. <https://doi.org/10.1371/journal.pbio.3000918>

Ioannidis JPA, Baas J, Klavans R, Boyack KW. A standardized citation metrics author database annotated for scientific field. PLoS Biol. 2019 Aug 12;17(8):e3000384. <https://doi.org/10.1371/journal.pbio.3000384>

Ioannidis JP, Klavans R, Boyack KW. Multiple Citation Indicators and Their Composite across Scientific Disciplines. PLoS Biol. 2016 Jul 1;14(7):e1002501. <https://doi.org/10.1371/journal.pbio.1002501>

1. What do the listed **retraction data** mean?

Reply: Since 2024, the annual editions of the databases include also information for each top-cited scientist on the number of retractions, the total number of citations received to their retracted papers, and the total number of citations they have received from any retracted papers. The Methods and some analyses of the retraction data can be found in: Ioannidis JPA, Pezzullo AM, Cristiano A, Boccia S, Baas J. Linking citation and retraction data reveals the demographics of scientific retractions among highly cited authors. PLoS Biology 2025, 23(1):e3002999. <https://doi.org/10.1371/journal.pbio.3002999> .

1. How can I correct **errors** in the databases?

Reply: The databases use Scopus data without further alteration. Only Scopus can correct their data and, if corrected, the correct versions will be used in the next iteration of our databases which are published annually in an archival form and will not changed until the next annual update. The published version reflects Scopus author profiles at the time of calculation. We thus advise authors to ensure that their Scopus profiles are accurate. REQUESTS FOR CORRECIONS OF THE SCOPUS DATA (INCLUDING CORRECTIONS IN AFFILIATIONS) SHOULD NOT BE SENT TO US. They should be sent directly to Scopus, preferably by use of the Scopus to ORCID feedback wizard (https://orcid.scopusfeedback.com/) so that the correct data can be used in any future annual updates of the citation indicator databases. Similarly, if any data on retractions are deemed to be incorrect, requests for correction should be addressed to the Retraction Watch database.

1. Are the databases stratified by **gender** and **age**?

Reply: The main databases do not contain information on gender and age on the listed scientists. Information on publication age can be inferred from the listed year of starting to publish. However, one may peruse another database that we have created and which provides information also on gender as well as stratified top-cited scientists according to publication age in 4 strata (first publication before 1992, 1992-2001, 2002-2011, 2012 or later). The database is in doi: 10.17632/wwykk8d48g See also: Ioannidis JPA, Boyack KW, Collins TA, Baas J. Gender imbalances among top-cited scientists across scientific disciplines over time through the analysis of nearly 5.8 million authors. PLoS Biol. 2023 Nov 21;21(11):e3002385. <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002385>

1. Does **number of publications** contribute to the composite citation indicator and respective ranking?

Reply: No, number of publications does not count at all. Impact may be achieved with very few or many publications. If a scientist with very few papers (even one or two papers) can achieve very high impact, they are not excluded from the database simply because they published few papers. Conversely, some scientists publish extremely high numbers of papers and this may reflect either amazing genuine productivity or manipulative practices. We have created a database that includes all scientists who have published more than 60 full papers (full articles, reviews, or conference papers, excluding from the count all other types of published items) in any single calendar year. The database is in: https://elsevier.digitalcommonsdata.com/datasets/kmyvjk3xmd/2 See also: Ioannidis JPA, Klavans R, Boyack KW. Thousands of scientists publish a paper every five days. Nature. 2018 Sep;561(7722):167-169 <https://www.nature.com/articles/d41586-018-06185-8> and Ioannidis, J.P.A., Collins, T.A. & Baas, J. Evolving patterns of extreme publishing behavior across science. Scientometrics (2024). <https://doi.org/10.1007/s11192-024-05117-w>

1. How could I correct **affiliations**?

Reply: To correct affiliations, please use the same exact process outlined under “How can I correct errors in the databases” sending the corrections directly to Scopus, not to us. We have no access or jurisdiction over the Scopus data. If corrected in Scopus, the corrected/preferred affiliation will appear in the next annual iteration of our databases. Scopus uses a machine learning approach to select only one affiliation from each author, based on the most recently published papers. This means that only one of possibly many affiliations is selected and this may not be the one that a scientist might prefer for themselves. It is impossible to know for millions of scientists which affiliation they prefer.

1. Do the databases include only living persons or also **dead scientists**?

Reply: Both living and dead scientists are included in the datasets. Scientists with very old publication start year are likely to be dead, but also some young scientists may have died. A recent/current year of latest publication does not offer complete proof that a scientist is alive, since for some deceased scientists some of their classic works may be republished currently and some scientists continue having their name listed in work they initiated when they were alive but is published posthumously.

1. How are **self-citations** calculated and what are considered to be too many self-citations?

Reply: Self-citations are counted from all authors of a given paper. E.g. if a paper X has 10 authors, all the papers that cite X and have one of the 10 original authors appear in their authors’ list are counted as self-citations. Percentiles of citations for each scientific discipline are provided in the databases so that one can place numbers in context of the discipline. For more details see: Ioannidis JPA, Boyack KW, Baas J. Updated science-wide author databases of standardized citation indicators. PLoS Biol. 2020 Oct 16;18(10):e3000918. <https://doi.org/10.1371/journal.pbio.3000918>

1. Can one infer evidence for **gaming of citations** from the database?

Reply: The presented data may offer some hints of citation orchestration and similar manipulations, but judgment needs to be cautious and linked with additional evidence. For potential metrics of citation orchestration, see: A. arXiv:2406.19219 [cs.DL] <https://doi.org/10.48550/arXiv.2406.19219> B. Ioannidis JPA, Maniadis Z. In defense of quantitative metrics in researcher assessments. PLoS Biol. 2023;21(12):e3002408 C. Ioannidis JPA, Maniadis Z. Quantitative research assessment: using metrics against gamed metrics. Intern Emerg Med. 2024 Jan;19(1):39–47. D. Ioannidis JP. Features and impact in precocious citation impact. PLoS One. 2025 Aug 6;20(8):e0328531.

1. Do the data include **only citations from original papers?**

Reply: No, data are provided from all published items included in Scopus and these include large numbers of not only traditional articles, but also reviews, conference papers, editorials, notes, letters, news items, and more. Classification of article type may have substantial error in Scopus and the boundaries of what is “original” can be contested. E.g. some editorials may have some extremely original ideas, while some “original articles” may show zero originality. While most editorials get minimal or no citations, some may be highly cited and this demonstrates impact for the writer(s). If a published work can be cited, citations should be counted. Nevertheless, for those interested in authors with extreme editorializing work, an analysis is provided in: Ioannidis JPA. Prolific non-research authors in high impact scientific journals: meta-research study. Scientometrics. 2023;128(5):3171-3184. doi: 10.1007/s11192-023-04687-5.

1. Why do some data look so different from **Google Scholar and other citation databases**?

Reply: Each citation database has its own inclusion and eligibility criteria and therefore different citations are counted. Our databases use Scopus data. In most disciplines, GoogleScholar tends to count more citations than Scopus, as it is far less selective regarding the citing sources that it considers eligible. Many works have been published around issues with indexation, see for instance: Sauvayre R. Types of Errors Hiding in Google Scholar Data. J Med Internet Res. 2022 May 27;24(5):e28354. doi: 10.2196/28354. PMID: 35622395; PMCID: PMC9187964.

1. Why is my listed **primary subfield of research** not what I would have described for myself.

Reply: The subfields are named using the Science-Metrix classification and nomenclature and the 174 subfield names may not precisely correspond to what specific scientists think about themselves. One should nevertheless examine in Science-Metrix what areas are covered under each subfield – a perfect name representing equally all of them would be impossible. Moreover, the primary subfield is the one with the largest proportion of papers for a given scientist and some scientists are split across many diverse subfields. The databases provide also the secondary (second more frequent) subfield for each scientist.

1. Are all scientists listed in the databases **excellent scientists** and those not listed are not?

Reply: NO! If an author is not on the list it is simply because the composite indicator value was not high enough to appear on the list. It does not mean that the author does not do good work. Similarly, some of the listed scientists may not be excellent. Citation metrics have limitations.

1. Do you offer certificates or other tokens of **recognition** for the included scientists or other sorts of **endorsement** for the use of these data?

Reply: No, this follows from the answer to question 13 above. We recognize that since we have made our datasets publicly available, many other scientists and non-scientists, administrators, institutions, organizations, or other entities may have used them. E.g., they may have summarized the information that we have provided, recast the datasets, disseminated them, or interpreted them in various ways. We trust that open public availability is worthwhile for transparency, but we should not be asked to endorse any of the multifarious ways that these datasets may be used or misused. Again, we advise cautious use of citation metrics and careful perusal of the methods that have been used to generate the metrics in our datasets.