



Author Database of Standardized Citation Indicators Derived from Scopus Lacks Transparency and Suggests a False Precision

HENK F. MOED 

RESEARCH



Levy
Library
Press

ABSTRACT

A critical discussion is presented for the Author Metrics Database (AMD) created by Ioannides et al. (2016, 2020) containing citation-based indicators for 165,000 authors publishing in journals indexed in Scopus. It is concluded that the AMD is a rich intermediary dataset open for further analysis to all interested users. However, its indicators suggest a false precision and lack transparency. The theoretical and statistical basis of the database's key composite impact indicator is weak, and information on whether or not underlying author publication lists were validated is lacking. The paper aims to broaden the perspective on the further development of an AMD, highlighting its bottom-up, interactive use, aptness for self-assessment and educational function for a wide user community.

POLICY HIGHLIGHTS

- Scopus diverges from Eugene Garfield's original concept of the Science Citation Index, as citation impact plays a weaker role as journal selection criterion.
- The transparency of the Article Metrics Database (AMD) is seriously hampered by the lack of information on whether the data were verified by scientists themselves.
- A complex composite indicator in the AMD decides whether or not a particular author is included. Its components are strongly statistically dependent and are largely based on the position an author has in a paper's author sequence but lack a sound theoretical foundation.
- An assessment of an individual researcher cannot be merely based on whether or not he or she is included in the AMD.
- The issue as to how to deal with multi-authored papers in research assessment of individuals can to some extent be enlightened by bibliometric indicators but cannot be solved bibliometrically. This is why the Composite Indicator suggests a false precision.
- The AMD focuses almost exclusively on senior scientists. Early career scientists and emerging research groups who will shape science and scholarship in the near future hardly appear in the AMD.

CORRESPONDING AUTHOR:

Henk F. Moed

Sapienza University of Rome,
Italy

henk.moed@uniroma1.it

KEYWORDS:

research assessment;
bibliometric indicators; author
metrics; early career scientists;
top researchers; Scopus

TO CITE THIS ARTICLE:

Moed, H. F. (2021). Author Database of Standardized Citation Indicators Derived from Scopus Lacks Transparency and Suggests a False Precision. *Scholarly Assessment Reports*, 3(1): 1, pp. 1–11. DOI: <https://doi.org/10.29024/sar.30>

- Desktop bibliometrics using the AMD as a sole source of information must be rejected. Using the AMD as a starting point in a more extensive bibliometric data collection makes it *de facto* a promotion tool for other Elsevier products.
- An alternative approach is an interactive, bottom-up bibliometric tool designed for self-assessment and educational purposes, showing how bibliometric indicators depend upon the way in which initial publication lists, author benchmark sets, subject delimitations, thresholds and evaluative assumptions are chosen.
- Research assessment is much more than just bibliometrics. It requires an overarching evaluative framework based on normative views on what constitutes research performance and which policy objectives should be achieved.

THE SUBJECT OF THIS PAPER

Recently, Ioannidis, Boyack & Baas (2020) published a “standardized citation metrics author database” derived from Scopus, Elsevier’s multi-disciplinary citation index. This database is an updated version of an earlier database published in 2019 by Ioannidis, Klavans & Boyack.¹ The Author Metrics Database, denoted throughout this article as AMD, is publicly available and contains a series of bibliometric, especially citation-based indicators for around 165,000 authors.

The AMD is beyond any doubt a valuable data source for further bibliometric research. Even though the current paper does present bibliometric analyses obtained from a statistical analysis of this database, it focuses on a different issue: What is the value of the AMD for the assessment of research performance of individual researchers? Is the information in the science-wide database actually useful? If so, in which ways?

On the one hand, the current paper profits from the transparency maintained by the creators of the AMD. On the other hand, it argues that transparency on several important issues is lacking and proposes ways to improve it. It fully acknowledges the importance of taking into account differences among subject fields and aims to fully live up to the authors’ warning that “assessing citation indicators always require caution” (Ioannidis-2020).

The current article discusses the “science-wide” AMD and the indicators it contains at two distinct analytical levels. Firstly, at the level of *bibliometrics*, technical and methodological aspects are addressed, but details are omitted; this discussion is directed toward a non-specialist audience. This part includes information on the data source underlying the indicators, the scientific literature database Scopus (2020).² At a second level, the pros and cons of the use of the database in *research assessment* are discussed from the point of view of a researcher interested in her or his own position in the database compared to that of other colleagues or from the perspective of a research manager or policy maker assessing his or her research staff.

SCOPUS CONTENT COVERAGE

MANY JOURNALS COVERED IN SCOPUS HAVE A STRONG NATIONAL ORIENTATION AND LOW CITATION IMPACT

Many scientific information scientists and research assessors may connect a citation index of scientific literature with Eugene Garfield’s vision of a multi-disciplinary core set of scientific journals selected on the basis of their citation impact, covering the best journals in science and forming the basis of his Science Citation Index (SCI), a scientific literature database launched in 1963. Soon a practice emerged that used the SCI not only for literature retrieval but also for research assessment, under the assumption that the appearance of a journal, scientific author or institution in the index can be interpreted as a sign of research quality. On many occasions, Garfield warned against over-interpretation and misuse of citation-based indicators in research assessment.

1 The paper by Ioannidis, Boyack & Baas (2020) will be referred to as Ioannidis-2020; the publication by Ioannidis, Klavans & Boyack (2016) presenting the methodology on which the AMD is based, as Ioannidis-2016; and Ioannidis, Baas, Klavans & Boyack (2019), as Ioannidis-2019.

2 The analyses on Scopus coverage presented in this section were created by the current author using a dataset derived from Scopus kindly provided by Prof. Felix de Moya-Anegón and Prof. Vicente Guerrero-Bote from the Scimago Research Group, Spain. These are partly based on Moed et al. (2021).

Scopus diverges from Garfield’s original model, as citation impact is not the only journal selection criterion. **Table 1** provides insight into the national orientation and citation impact of journals covered. An Index of National Orientation (symbol *INO-P*) is defined as the percentage of articles published by authors from the country accounting for the largest number of articles published in that journal. The table shows that the percentage of nationally oriented journals (*INO-P* > 80) indexed in 2019 in Scopus is around 23 percent.³

NO. JOURNALS ACTIVE IN 2019	AVERAGE NO. ARTICLES PER JOURNAL	% JOURNALS WITH INO-P > 80	% JOURNALS WITH JIF3 < 0.1	% JOURNALS WITH JIF3 < 0.2
23,200	108	23%	7%	14%

The last column in **Table 1** relates to citation impact, as expressed by Journal Impact Factor (JIF3). It shows that the percentage of journals for which JIF3 is smaller than 0.1, relative to the total number of journals, amounts to 7 percent. Raising the JIF3 threshold from 0.1 to 0.2, this percentage doubles. It is assumed that a JIF3 below 0.1 is extremely low for any journal, regardless of the subject field it covers.⁴

THE EFFECT OF INDEXING POORLY CITED JOURNALS UPON AUTHOR METRICS

The effect that the inclusion of nationally oriented and/or low-impact journals may have upon citation-based author metrics depends upon the type of indicators calculated. One can distinguish two main types that are sometimes denoted as size dependent and size independent or, in terms of the key statistic calculated, as average based and sum based. A third type includes hybrid indicators, which combine elements from the size dependent and size independent approaches. **Table 2** gives typical examples from these types and presents characteristic quotes of authors defending a particular type. The next section argues that the approach adopted by Ioannides et al. is essentially size dependent, based as it is on size dependent indicators and hybrid ones positively correlating with size dependent measures.⁵

INDICATOR TYPE	EXAMPLES	RATIONALE
Size-independent/ average-based	Citations per article; Journal Impact Factor	“In view of the relation between size and citation frequency, it would seem desirable to discount the effect of size when using citation data to assess a journal’s importance” (Garfield, 1972, p. 477). The use of absolute numbers of citations favors large groups or senior authors and disadvantages small, emerging groups or junior scientists (Van Raan, 2019).
Size-dependent/ sum-based	Total citation counts; Integrated Impact Indicator	“The common assumption in citation impact analysis hitherto has been normalization to the mean. In our opinion, the results are then necessarily flawed because the citation distributions are often highly-skewed. Highly productive units can then be disadvantaged because they publish often in addition to higher-cited papers also a number of less-cited ones which depress their average performance.” (Leydesdorff & Bornmann, 2011, p. 34).
Hybrid (contains elements from both approaches)	H index	Performance must reflect both publication productivity and citation impact. Publication counts alone “do not measure importance nor impact of papers”; total citations “may be inflated by a small number of ‘big hits’, which may not be representative of the individual if he/she is coauthor”; citations per paper “rewards low publication productivity, penalizes high productivity.” (Hirsch, 2005).

Table 1 Article production, national orientation and uncitedness of Scopus journals active in 2019. Active in 2019: Scopus has indexed at least one document published by the journal in 2019. The following four document types are included in the publication counts: articles, proceedings papers, reviews and short surveys. *INO-P*: Index of National Orientation, as expressed in the geographic location of the authors publishing in a particular journal. A journal has *INO-P* > 80 if there is one country accounting for more than 80 percent of all papers published in that journal. *JIF3*: Journal Impact Factor, based on the three-year impact factor, defined as the number of citations in a particular year (e.g., 2019) to articles published in a journal during the three preceding years (e.g., 2016–2018) divided by this number of articles.

Table 2 Three main types of indicators.

³ The 8,300 journals indexed in Scopus at least in one year from 1996–2018 but not active in 2019 tend to have a stronger national orientation and a lower citation impact than periodicals active in 2019 have. This outcome suggests that in a process of re-assessment of its content coverage, the Scopus team decided to remove especially nationally oriented, low-impact journals.

⁴ The percentage of journals with *INO-P* > 80 ranges between 12% in biomedical research to 25% in clinical medicine. The percentage of journals with *JIF3* < 0.1 ranges between 3% in natural sciences to 10% in humanities and social sciences.

⁵ According to **Table 2** below, the Hirsch index (H index) correlates strongly with indicators based on total citations (Pearson’s *R* = 0.92) and number of publications (*R* = 0.49), consistent with earlier bibliometric indicator studies. This outcome provides evidence that this indicator is more a size dependent than a size independent measure.

AUTHOR METADATA IN THE AMD

AUTHOR DATA ARE ONLY PARTIALLY VALIDATED BY SCIENTISTS THEMSELVES

Numerous experiences collected in the past decennia with the calculation of bibliometric indicators at the level of individuals have shown that the identification of all publications of a given individual researcher in a scientific literature database is highly sensitive to errors. The most important source of error is the occurrence of homonyms—different people with the same name, e.g., Smith, Jones, Lee, Liu, Andersen—and synonyms—different names for the same person, for instance, due to differences between full first name and nicknames, mixing up first name and family name, different transliterations of Cyrillic and other non-Latin names and name changing if a person assumes the name of a partner.

Although Ioannidis-2016 states that “Scopus author IDs were used for all author-based analyses,” they do not provide any information on how these Scopus IDs are created. Ioannidis-2020 refers to an article by Baas et al. (2020) describing how author profiles are created in Scopus. This information is the same as that given in Ioannidis-2019. Although Baas et al. (2020) do not give details on the author-clustering routine that underlies the author profiling and its ownership, they indicate three sources through which curation of these can be achieved: via ORCID, via the Scopus Author Feedback Wizard and via a special commercial Elsevier service.⁶

However, the AMD does *not* contain an indication as to whether the publications assigned to a particular author were actually verified by the person represented by this author. Hence, it is unknown how many author clusters included in the AMD are actually verified. Both this result and the lack of information about clustering software substantially *reduce the transparency* of the data included in the AMD.

AUTHOR INSTITUTIONAL AFFILIATION IS BASED ON AN AUTHOR’S MOST RECENT PUBLICATION

The AMD indicates for each author an institutional affiliation, derived from an author’s *most recent* publication. For instance, an author with a 30-year career at University A who moved in the last year to University B (and who indicated his new affiliation in an article published in this year) is assigned in the AMD to B, *not* to A. As a result, an analysis of an *institution* based on the authors linked to it in the AMD may at best provide an indication of the past performance of the academic staff *currently employed* at that institution but does *not necessarily* give an impression of how the research staff *appointed at*—and in most cases funded by—an institution has collectively performed over the years.

INDICATORS CALCULATED IN THE AMD

THE COMPOSITE INDICATOR AND ITS COMPONENTS

The current paper focuses on a composite indicator that plays a key role in the inclusion of authors in the AMD and their ranking. It is presented in **Figure 1**. Its symbol is **c** and is calculated for each author in the AMD. It is defined as the sum of six components, each of which is basically calculated as the ratio of a specific citation indicator for a particular author to the maximum value of this indicator across all authors in the AMD. Rather than using straight counts, logarithmic values are calculated for the indicator value (plus one) both in a component’s numerator and in its denominator. Logarithmic values were used as the underlying citation distributions across authors are very skewed, a phenomenon that is clearly illustrated in **Table 3** below.

Figure 1 Composite indicator in the Author Metrics Database (AMD). Source: Ioannidis, Boyack & Baas (2020). NC: Total number of citations. H: H Index (Hirsch, 2005). Hm: Hm Index, similar to H index but accounting for multi-authored papers (Schreiber, 2008). NCS: Number of citations to single-authored papers. NCSF: Number of citations to single- and first-authored papers. NCSFL: Number of citations to single-, first- and last-authored papers. Index *i* indicates a particular author. Log: Natural logarithm. Maxlog: The natural logarithm of the maximum score on a particular indicator in the entire AMD. The six components have, statistically speaking, equal weights.

$$c_i = \frac{\log(NC_i + 1)}{\max \log(NC + 1)} + \frac{\log(H_i + 1)}{\max \log(H + 1)} + \frac{\log(Hm_i + 1)}{\max \log(Hm + 1)} + \frac{\log(NCS_i + 1)}{\max \log(NCS + 1)} + \frac{\log(NCSF_i + 1)}{\max \log(NCSF + 1)} + \frac{\log(NCSFL_i + 1)}{\max \log(NCSFL + 1)}$$

The citation indicator in the first component, **NC**, counts the total number of citations in a given year to all publications by a particular author. Components 4, 5 and 6 take into account the

⁶ Baas et al. (2020) claim that publications in author profiles currently have 98.1% average precision and 94.4% average recall and that “All above efforts combined have led to approximately 1.8 million Scopus author profiles that have been manually enhanced.”

number of co-authors in a given author’s papers or his or her *position in the author list*. **NCS** is based on citations to papers on which a given author is the sole author (single-authored papers) and **NCSF** on papers on which he or she is either single or first author, while **NCSFL** counts citations to single-, first- or last-authored articles. The second component is the *H* Index (Hirsch, 2005), and the third its variant, the **Hm** index (Schreiber, 2008).

KEY STATISTICS					PEARSON CORRELATION COEFFICIENTS							
VAR	MEAN	MEDIAN	MIN	MAX	VAR	LOG NC	LOG H	LOG HM	LOG NCS	LOG NCSF	LOG NCSFL	LOG NP
NC	1,148	719	13	67,118	log NC	1.00	0.92	0.57	-0.24	0.23	0.71	0.57
H	14	13	1	99	log H	0.92	1.00	0.65	-0.26	0.21	0.66	0.49
Hm	6.8	6.4	0.34	45.8	log Hm	0.57	0.65	1.00	0.17	0.28	0.73	0.46
NCS	37.5	10	0	13,437	log NCS	-0.24	-0.26	0.17	1.00	0.24	0.02	-0.09
NCSF	208.1	137	0	28,269	log NCSF	0.23	0.21	0.28	0.24	1.00	0.44	-0.12
NCSFL	504.1	336	10	46,567	log NCSFL	0.71	0.66	0.73	0.02	0.44	1.00	0.42
NP	180.0	134	2	4460	log NP	0.57	0.49	0.46	-0.09	-0.12	0.42	1.00

THE COMPOSITE INDICATOR IS BASED ON STATISTICALLY DEPENDENT ELEMENTS

One may argue in favor of the Composite Indicator that an indicator based on *all* of an author’s publications does not reveal well how his or her single-author papers are performing. However, the number of single-authored articles and their citations (**NCS**) is included in the count of single- and first-authored papers and citations (**NCSF**), which in turn are included in the count based on single-, first- and last-authored publications (**NCSFL**) and on the total number of citations (**NC**). As a result, the numbers of citations to an author’s various groups of papers are statistically dependent. This is clearly illustrated in the correlation matrix in **Table 3**. This table also shows that not only these elements of the Composite Indicator are statistically dependent. Pearson R values for correlations among indicators **NC**, **H**, **Hm** and **NCSFL** are all above 0.5 and range between 0.57 (**NC** versus **Hm**) and 0.92 (**NC** versus **H**). These four indicators all show R values above 0.4, with the total number of publications (**NP**) revealing that they are all size dependent.

As in **Table 1** in Ioannidis-2016, all values in the correlation analysis in **Table 3** are log transformed. Moreover, **Table 1** is based on citations in the year 2019, obtained from Table-S7-singleyr-2019 in Ioannides-2020, while **Table 1** in Ioannides-2016 is based on citations in a single year as well, namely the year 2013. In this way, the tables are both based on single-year studies and can therefore be compared. Generally speaking, Pearson R values in **Table 3** are much higher than they are in the corresponding table in Ioannides-2016. This underlines the statistical dependence between the elements in the Composite Indicator.⁷

Allowing users to assess distinct categories of papers makes sense, even though it is argued below that indicators based on author sequences have a limited validity. In addition, Ioannidis-2016 states that equal weights were given to all six log-transformed indicators included in the composite for parsimony reasons and that “if, for whatever reason, one or more of these indicators are considered more essential in a particular field, one can weigh them more compared with the others.” However, it is questionable whether this consideration provides sufficiently valid grounds for including statistics for series of partially overlapping sets in a composite indicator that plays such an important role in the AMD. After all, it is the measure on which authors are ranked and is used to expand the AMD beyond the set of the top 100,000 authors.

Table 3 Data were obtained from the dataset Table-S7-singleyr-2019. The key statistics are based on the absolute values of the indicators, and the correlation coefficients on their logarithmic values. NC: Total number of citations from 2019. H: H Index for the year 2019. Hm: Hm Index for 2019. NCS: Number of citations to single-authored papers. NCSF: Number of citations to single- and first-authored papers; NCSFL: Number of citations to single-, first- and last-authored papers. NP: Total number of publications between 1960 and 2019. Calculations are based on all 161,441 authors in the single-year 2019 dataset. Pearson correlation coefficients between six indicators included in the Composite Indicator (single-year dataset for 2019).

⁷ Striking differences can be observed between the correlations obtained in **Table 3** in the current paper and those presented in **Table 1** in Ioannidis-2016. The largest differences are found for the correlation between the following pairs of indicators: NC and NCSFL (0.71 in **Table 3**, -0.04 in **Table 1** in Ioannidis-2016), H and Hm (0.65 vs 0.19), H and NCSF (0.21 vs -0.12), H and NCSFL (0.66 vs 0.04), Hm and NCSF (0.28 vs 0.72) and NCSFL and NP (0.72 vs 0.06). It seems improbable that these differences can be explained by the use of different though partially overlapping author populations and time periods (161,000 authors in citing year 2019 vs. 84,000 authors in citing year 2013).

STANDARDIZATION FACTOR IN THE COMPOSITE INDICATOR DOES NOT ACCOUNT FOR DIFFERENCES AMONG SUBJECT FIELDS

All six indicators included in the composite measure are log transformed and standardized. Ioannidis-2016 argue that “log-transformations ensure that there are no major outlier values.” Their standardization method gives a value of 1 to the author with the highest raw value for a particular indicator. Ioannidis-2020 rightly underlines that “comparisons of citation metrics are more meaningful when done within the same subdiscipline.” However, their standardization method uses the highest raw value across all subject fields, while there are good reasons to use subject field-dependent highest raw values.

As expected, each indicator reveals substantial differences in these maximum values across subject fields. Calculating for each author a new composite measure based on maximum values *per subject field* using the Science Metrix classification into 174 subfields and correlating it with the original measure included in the AMD, the two composite indicators show a Pearson correlation of 0.77. Using the Science-Metrix classification into 20 main fields Pearson’s R amounts to 0.86. These outcomes show that applying a field-normalized standardization factor rather than one single factor across all subject fields does make a difference.⁸

THERE IS HARDLY A THEORETICAL BASIS FOR WEIGHING A SCIENTIST’S CONTRIBUTION TO A PAPER BASED ON AUTHOR SEQUENCE

The underlying basic assumption of the AMD is that one can derive an indication of the contribution an author has made to a multi-authored paper from the paper’s author *sequence*. The indicators in the AMD seem to be based on the assumption that in a multi-authored paper, the first and the last authors make the largest contribution to the paper. Indeed, there is evidence that research groups in experimental fields such as Physics and Chemistry often adopt an authoring practice according to which the first author is the PhD student conducting the experiment and the last author the supervisor responsible for—and often the intellectual owner of—the research program in which the PhD project is included. However, three essential limitations should be underlined:

- The situation becomes more complex when two collaborating research groups make equal contributions. If the two supervisors obtain the semi-last and last positions and the two PhD students the first and second, there is no reason to give a higher weight to the first and last authors only. The only currently available model for author weighting in multi-team collaboration gives a special status to the research group delivering the *reprint* author, who is assumed to function as the team’s *research guarantor*⁹ (Moya-Anegón et al., 2013).
- One may claim that even if a uniform author weight parameter may be inadequate in individual cases, deviations from an assumed “true” author weight tend to cancel out if an assessed author has published a sufficiently large number of papers. However, this argument is invalid especially in the case of citation analysis, in which citation distributions are known to be skewed and only few papers are responsible for the biggest part of an author’s or a group’s citation impact. The key question then is: what is the contribution of the various author to *these* papers?
- There is evidence that especially in Mathematics and Social Sciences & Humanities, distinct authoring conventions exist, based on lexicographical ordering of authors or on rotating first authorship. In this case, there is no justification for giving a special status to the first and last authors. This limitation is also mentioned in Ioannidis-2016. Using

⁸ It follows that the subfield-normalized composite measure explains only 60 percent of the variance (R-square) in the Ioannidis-2020 composite indicator and the main field-based measure 74%. Ioannidis-2016 and Ioannidis-2020 are fully aware that their composite indicator does not account for differences among subject fields and that one should interpret rankings based on this measure only on a field-by-field basis, comparing an author with authors from the same subject field. The observed field dependence of their standardization has *no* implications for rankings *within* subject fields. The current author does not claim that the use of field-dependent normalization factors would correct for all disturbing differences among subject fields. However, it would be worthwhile to consider it as an alternative to the current solution, in which a standardisation factor is fully determined by the extreme score—possibly a statistical outlier—of one single author across all subject fields.

⁹ The hypothesis of Moya et al. (2013) explicitly relates the notion of guarantorship to a *research group*, not to an *individual* author (the so-called reprint author).

data from the AMD, the current author observed an over-representation of first authors in the upper part of the lexicographically ordered full author list in Visual & Performing Arts (18%), Philosophy & Theology (7%), Communication & Textual Studies (3%) and Mathematics & Statistics (2%).¹⁰ It must be noted that articles resulting from multi-team collaborations in “hot” fields in natural and biomedical sciences may use alphabetical author ordering as well.

COMPOSITE INDICATOR VALUE DECIDES WHETHER OR NOT A PARTICULAR AUTHOR IS INCLUDED IN THE AMD

Ioannidis-2020 states that in a first step, the top 100,000 authors are selected across all subject fields based on the Composite Indicator. In a second step, this set is complemented with authors not among the top 100,000 but still among the top 2 percent of their main subject field and publishing at least five papers. Although Ioannides et al. (2020) put the Composite Indicator into perspective by underlining that different components may be included or that different weights may be assigned to an indicator, it is clear that the Composite Indicator as defined in *Figure 1* above plays the key role in deciding whether or not a particular author is included in the AMD. Hence, one should realize that analyzing the AMD and experimenting with the selection of indicators and weights can only be applied to those authors who are already in the AMD and can therefore not be used, for instance, to examine the effect of changes in the formula of the Composite Indicator upon the inclusion of authors in the AMD. It follows that in the assessment of an individual author, an evaluator cannot simply assume that one is making a correct judgment if it is based on whether or not an author is included in the AMD.

USEFULNESS OF THE AMD IN RESEARCH ASSESSMENT

DESKTOP BIBLIOMETRICS USING THE AMD AS THE SOLE DATA SOURCE MUST BE REJECTED

One type of use of the AMD in the assessment of an individual researcher, for instance, for hiring or promotion purposes, is to look up the author entry in the AMD with the same name as the assessed researcher. Next, an assessment criterion is defined, for instance, being included in the AMD or having a Composite Indicator score in the top quartile of this indicator’s distribution. Finally, a decision is made purely on the basis of the thus-obtained outcome, without taking into account any other sources of information. *This* type of use can be denoted as *desktop bibliometrics*. The creators of the AMD make clear that they are strongly opposed to this type of use. So is the current author (Moed, 2017, 2020). Judgment of an individual’s performance by applying assessment criteria based on thresholds for a particular bibliometric indicator is indefensible not only if the validity of the indicator is questionable but also if threshold values themselves are not well founded.

PERFORMANCE OF AN INDIVIDUAL AND THE CITATION IMPACT OF HIS OR HER PAPERS RELATE TO TWO DISTINCT ANALYTICAL LEVELS

The AMD creators rightly point out that multiple co-authorship is a rule rather than an exception, especially in the natural and life sciences. As a consequence, publications (co-)authored by an individual researcher are often, if not always, the result of research to which other scientists have contributed as well, sometimes even dozens of them. The crucial issue is how one should relate the citation impact of a *team’s* papers to the performance of an *individual* working in that team. It is fully appropriate that the creators of the AMD dedicate so much attention to this issue. However, one must realize that performance of an individual and the citation impact of his papers relate to two distinct analytical levels.

¹⁰ The current author adopted the following approach. In a first step, all authors were divided on a field-by-field basis into two groups of approximately equal size based on the first character of their last name. Overall, 48.5 percent of author names started with characters A-K, and 51.5 percent with characters L to Z, but there are differences among subject fields. If lexicographical ordering of authors plays a role in a field, one would expect to find among first authors a higher fraction of authors whose names start with A-K than there are in the total population of authors publishing in that field. Using the Science-Metrix main field classification, an over-representation of A-K first authors was found for the fields mentioned in the main text. For all other fields, it was zero. The outcomes do not allow one to estimate the actual number of papers using alphabetical authorship. It must be noted that an observed alphabetical order in a paper does not necessarily imply that the authors decided to order their names alphabetically.

THE USE OF BIBLIOMETRIC INDICATORS FOR INDIVIDUAL SCIENTISTS SUGGEST A FALSE PRECISION

The current author defends the position that a valid assessment of the research performance of individuals can be properly made only on the basis of sufficient knowledge of the particular role they played in the research presented in their publications, for instance, whether this role has been leading, instrumental or technical. In addition, other manifestations of research performance should be taken into account as well. Calculating indicators at the level of an individual and claiming they measure by themselves an individual's performance, statistically sophisticated as they may be, suggests an accuracy of measurement that cannot be justified. This is especially also true for the AMD Composite Indicator. Ultimately, its validity does *not* depend upon the number of components in the indicator or on the level of sophistication of their weights.

LACK OF INFORMATION ON DATA VERIFICATION BY SCIENTISTS THEMSELVES SERIOUSLY HAMPERS THE TRANSPARENCY OF THE AMD

The very existence of a database with “top” researchers invites evaluators and other interested users to use the information for their own evaluative purposes. The AMD creators explicitly refer to the entities analyzed in the AMD as *scientists*, not as authors, thus emphasizing the *personal* rather than the *statistical* nature of the data. As outlined above, for part of the 165,000 authors, the publication lists have to some extent been verified by the corresponding scientist, but there is *no* information available on how large this fraction actually is. In addition, the AMD does *not* include for each author a flag indicating whether or not the underlying data were verified. The current author believes that the lack of this information seriously hampers the transparency of the AMD that such a flag must be included.

What is more, it would have been much more appropriate to include only scientists whose algorithmically generated publication lists were verified and who explicitly have given their consent. The fact that the statistical de-duplication of author names and assignment of documents has already taken place in Scopus and has not been contested by subjected scientists does *not* justify the creation of the AMD in its current form, as Scopus is primarily a scientific literature search tool in which author names are *content descriptors*, not scientists subjected to a performance evaluation, many of whom may not even be aware that they are included in the AMD.

EARLY CAREER SCIENTISTS AND MEMBERS OF EMERGING RESEARCH GROUPS HARDLY APPEAR IN THE AMD

The analysis of Scopus content coverage revealed that this database indexes a substantial number of nationally oriented journals with a low citation impact. Although there is evidence that once they are indexed in Scopus, many of these journals internationalize and increase their citation impact (Moed et al., 2021), their inclusion may distort size independent, average-based citation indicators. In terms of the distinction between indicator types made above, the decision made by Ioannidis et al. (2016; 2020) to apply size dependent or hybrid indicators is well defensible. However, this choice has its limits.

The AMD focuses almost exclusively on *senior* scientists. Early career scientists and members of emerging research groups who will shape science and scholarship in the near future hardly appear in the AMD. A rough indication of the extent to which early career researchers (ECRs) are covered in the AMD can perhaps be based on the assumption that an ECR would publish only papers as a first or single author, and not yet publications as a last author. There appear to be only about 1,000 authors meeting this criterion, accounting for 0.6 percent of the total number of AMD authors. The AMD aims to cover “top” researchers. The Composite Indicator and its components are size dependent and strongly biased in favor of senior authors with long scientific careers.¹¹ The current author wishes to underline that this size dependence is a choice made by the creators themselves. In the field of bibliometrics, there is also a line of development of size independent or “relative” indicators.¹²

11 Using a qualification often used in other domains of society, one could facetiously characterize the AMD as a dataset primarily about “(baby) boomers.”

12 See, for instance, the notes on size dependent versus size independent indicators related to the Leiden Ranking, available at <https://www.leidenranking.com/information/indicators>, or related to SCImago Institutions Rankings, available at <https://www.scimagoir.com/methodology.php>.

CONSIDERING THE AMD AS A STARTING POINT IN BIBLIOMETRIC DATA COLLECTION MAKES IT DE FACTO A PROMOTION TOOL FOR OTHER ELSEVIER PRODUCTS

It was argued above that the issue as to how one should deal in research assessment with multi-authored papers can to some extent be *enlightened* by bibliometric indicators but *cannot* be solved *bibliometrically*. It was concluded that it is problematic to justify an evaluative judgment merely based on quantitative indicators and threshold, as they suggest a false precision. The following question then arises: how can the AMD be used in a proper manner? One could argue that the data on authors presented in the AMD represent only a first step in an assessment and that additional bibliometric data can be retrieved from other data sources, especially the online version of Scopus and the special online bibliometric tool SciVal created by Elsevier. Although the *intentions* of the creators of the AMD are beyond any doubt to calculate bibliometrically founded indicators at the author level available to a wide audience, *this* argument would underline that the AMD is *de facto* a *promotion tool* for these two Elsevier products.

AN ALTERNATIVE APPROACH: AN INTERACTIVE, EDUCATIONAL, BIBLIOMETRIC SELF-ASSESSMENT TOOL

The creators of the AMD have indeed made an important step toward a bibliometric assessment tool by creating a rich intermediary dataset with bibliometric indicators, open for further analysis to all interested people. Although the Composite Indicator is the key measure, its components can be used as separate indicators as well. In addition, the AMD contains other interesting features not discussed in the current paper, such as the possibility to analyze citation counts including or excluding author self-citations.

Technically, it seems feasible to add in a follow-up version of the AMD the verification status of the publication data relating to a particular author, as this information is available in the Scopus system, or to include only authors who have validated their data. However, making the database interesting for ECRs by increasing the number of included authors and adding size independent indicators seems hardly doable within the framework of the current AMD model. The current author would like to broaden the perspective and bring in *three* new elements that could play a role in the further development of AMDs.

Firstly, the AMD is perhaps still too much based on more classical data-handling approaches developed during the past decades and does not yet fully profit from tools to create *interactive and flexible bottom-up* applications enabling interested users to go back to the raw data, decompose existing indicators and generate new, more fit-to-purpose measures if needed. *Secondly*, the key function of a new version could be to deliver bibliometric data informing an author *self-assessment*. It would enable a scientist to select and verify his or her own publication data; next, it creates a set of ‘candidate’ benchmark authors or groups using algorithms similar to those proposed by Eugene Garfield for evaluating faculty (Garfield, 1983a, 1983b). It may also offer a flexible benchmarking feature for users as the practical realization of Robert K. Merton’s notion of a reference group, i.e., the group they “do not necessarily belong but aspire to” (Merton, 1996).

Thirdly, it could also function as an *educational tool* to become more acquainted with the ins and outs of bibliometric indicators by making users aware of the technical and evaluative *choices* that have to be made in a bibliometric analysis. It could stimulate the user to specify at least some of the elements from an evaluative framework overarching the self-assessment, thus stimulating the user to reflect upon this framework. It would reveal to a user how outcomes of bibliometric assessment depend upon the way initial publication lists, author benchmark sets, author position weights and subject delimitations are being defined and upon the role of particular evaluative assumptions and setting of thresholds. It could contribute to the transparency of a research assessment process by enabling those subjected to an assessment in their external professional environment to critically follow this process and could defend them against inaccurate calculation, misinterpretation or inappropriate use of indicators.¹³

13 The application should be created by experienced and independent researchers assisted by professional developers and fully free to share whatever information they find relevant during the development process with the wider research community, including theoretical issues related to its functionality, preliminary functional designs, mock-ups, intermediate versions and the formation of test and focus groups. It would be essential to conduct well-designed tests of the tool and to monitor its use after its introduction. If it would appear that the tool leads to an even more inconsiderate use of bibliometrics in research assessment, its introduction would be counter-productive, and its design may have to be reconsidered.

Obviously, research assessment is much more than just bibliometrics. Research assessment requires an overarching evaluative framework based on normative views on what constitutes research performance and which policy objectives should be achieved. Informetricians should comply in their scientific work with the methodological principle to maintain a neutral position toward an assessment's constituent policy issues, the criteria specified in the evaluative framework and the goals and objectives of the assessed subject. As professional experts, their competence lies primarily in the development and application of *analytical* models given the established evaluative framework. They may contribute to a productive combination of qualitative and bibliometric tools. In addition, as more and more bibliometricians have been involved as actor, advisor or observer in actual assessment processes using bibliometric indicators, they can report on their experiences in these processes to a wide scholarly and policy audience.

ACKNOWLEDGEMENTS

The author wishes to thank two anonymous reviewers for their valuable comments on an earlier version of this paper.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR INFORMATION

Former senior staff member (1981–2009) at the Centre for Science and Technology Studies, Leiden University, and former senior scientific advisor (2010–2014) with Elsevier, Amsterdam, the Netherlands. The author is currently an independent researcher and a scientific advisor to the SCImago Research Group, Spain.

AUTHOR AFFILIATION

Henk F. Moed  orcid.org/0000-0003-2446-905X
Sapienza University of Rome, Italy

REFERENCES

- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R.** (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1, 377–386. DOI: https://doi.org/10.1162/qss_a_00019
- Garfield, E.** (1972). Citation Analysis as a tool in journal evaluation. *Science*, 178, 471–479. DOI: <https://doi.org/10.1126/science.178.4060.471>
- Garfield, E.** (1983a). How to use citation analysis for faculty evaluation, and when is it relevant. Part 1. *Current Contents*, 44, 5–13, October 31, 1983. In: *Essays of an Information Scientist*, 6, 354–362. Philadelphia: ISI Press.
- Garfield, E.** (1983b). How to use citation analysis for faculty evaluation, and when is it relevant. Part 2. *Current Contents*, 45, 5–13, November 7, 1983. In: *Essays of an Information Scientist*, 6, 363–372. Philadelphia: ISI Press.
- Hirsch, J. E.** (2005). An index to quantify an individual's scientific research output. *PNAS*, 102, 16569–16572. DOI: <https://doi.org/10.1073/pnas.0507655102>
- Ioannidis, J. P., Klavans, R., & Boyack, K. W.** (2016). Multiple Citation Indicators and Their Composite across Scientific Disciplines. *PLoS Biol*, 14(7), e1002501. DOI: <https://doi.org/10.1371/journal.pbio.1002501>
- Ioannidis, J. P. A., Baas, J., Klavans, R., & Boyack, K. W.** (2019). A standardized citation metrics author database annotated for scientific field. *PLoS Biol*, 17(8), e3000384. DOI: <https://doi.org/10.1371/journal.pbio.3000384>
- Ioannidis, J. P. A., Boyack, K. W., & Baas, J.** (2020). Updated science-wide author databases of standardized citation indicators. *PLoS Biol*, 18(10), e3000918. DOI: <https://doi.org/10.1371/journal.pbio.3000918>

- Leydesdorff, L., & Bornmann, L.** (2011). Integrated Impact Indicators (I3) compared with Impact Factors (IFs): An alternative research design with policy implications. *Journal of the American Society for Information Science and Technology*, 62, 2133–2146. DOI: <https://doi.org/10.1002/asi.21609>
- Merton, R. K.** (1996). The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of intellectual property. In: Merton, R. K., *On Social Structure and Science*. Chicago: The University of Chicago Press, 318–336. Also in *ISIS*, 79, 607–623, 1988. DOI: <https://doi.org/10.1086/354848>
- Moed, H. F.** (2017). *Applied Evaluative Informetrics*. Cham, Switzerland: Springer, 312 pp. DOI: <https://doi.org/10.1007/978-3-319-60522-7>
- Moed, H. F.** (2020). Appropriate Use of Metrics in Research Assessment of Autonomous Academic Institutions. *Scholarly Assessment Reports*, 2(1). DOI: <https://doi.org/10.29024/sar.8>
- Moed, H. F., de Moya-Anegón, F., Guerrero-Bote, V., Lopez-Illescas, C., & Hladchenko, M.** (2021). Bibliometric Assessment of National Journals. *Scientometrics*. DOI: <https://doi.org/10.1007/s11192-021-03883-5>
- Moya-Anegón, F., Guerrero-Bote, V. P., Bornmann, L., & Moed, H. F.** (2013). The research guarantors of scientific papers and the output counting: A promising new approach. *Scientometrics*, 97, 421–434. DOI: <https://doi.org/10.1007/s11192-013-1046-0>
- Schreiber, M.** (2008). A modification of the h-index: The hm-index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2(3), 211–216. DOI: <https://doi.org/10.1016/j.joi.2008.05.001>
- Scopus.** (2020). Scopus Source Title list. Version April 2020. Available at <https://www.elsevier.com/solutions/scopus/how-scopus-works/content>.
- Van Raan, A. F. J.** (2019). Measuring Science: Basic Principles and Applications of Advanced Bibliometrics. In: W. Glanzel, H. F. Moed, M. Thelwall & U. Schmoch (Eds.), *Springer Handbook of Science and Technology Indicators*. Cham, Switzerland: Springer, pp. 237–273. DOI: https://doi.org/10.1007/978-3-030-02511-3_10

TO CITE THIS ARTICLE:

Moed, H. F. (2021). Author Database of Standardized Citation Indicators Derived from Scopus Lacks Transparency and Suggests a False Precision. *Scholarly Assessment Reports*, 3(1): 1, pp. 1–11. DOI: <https://doi.org/10.29024/sar.30>

Submitted: 19 January 2021

Accepted: 17 March 2021

Published: 21 April 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Scholarly Assessment Reports is a peer-reviewed open access journal published by Levy Library Press.