Scientometrics
https://doi.org/10.1007/s11192-025-05338-7

How similar are field-normalized citation impact scores obtained from OpenAlex and three popular commercial databases? An empirical comparison based on large German universities

Thomas Scheidsteger[1] · Robin Haunschild[1] · Lutz Bornmann[1,2]

Received: 21 August 2024 / Accepted: 22 April 2025

Abstract
OpenAlex is a freely available bibliographic database that can be used for bibliometric studies. In this study, we compared certain field-normalized citation scores (NCS) from OpenAlex with those from three commercial databases (Web of Science, Scopus, and Dimensions). We were interested in the question whether the NCS from OpenAlex are comparable to those from the commercial databases and can be alternatively used in evaluative bibliometrics. The NCS have been calculated for nearly 335,000 papers published by 48 German universities in four main subject areas between 2013 and 2017. We found varying but overall strong agreement between the scores according to Lin's concordance correlation coefficient. Separating the publication set along the single universities and moreover along the four main subject areas involved revealed significant differences at the level of single papers but also gave indications on how to possibly mitigate outlier cases. We calculated mean normalized citation scores for the 48 universities and found that the agreements across different databases are low. On the one hand, the results suggest that comparisons of universities using NCS across different databases should be avoided. On the other hand, the difference of the concordance correlation coefficients at paper and university level is a good example for the problem of ecological fallacy in bibliometrics: The mean impact is not representative for the single papers' impact in the set.

Keywords Citation impact · MNCS · OpenAlex · Web of science · Scopus · Dimensions · German universities

✉ Thomas Scheidsteger
t.scheidsteger@fkf.mpg.de

Robin Haunschild
r.haunschild@fkf.mpg.de

Lutz Bornmann
l.bornmann@fkf.mpg.de; lutz.bornmann@gv.mpg.de

[1] IVS-CPT, Max Planck Institute for Solid State Research, Stuttgart, Germany

[2] Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Munich, Germany

Published online: 25 May 2025

Introduction

Research evaluation using bibliometric methods is frequently based on commercial bibliographic databases that have similar approaches to select journals for including papers in the database and to ensure the quality of included papers: Web of Science (WoS) and Scopus (Baas et al., 2020; Birkle et al., 2020). Dimensions is another commercial bibliographic database that provides an alternative to WoS and Scopus including many more publications (Herzog et al., 2020). With the emergence of Microsoft Academic Graph (MAG) in 2015, a free bibliographic database with an outstanding coverage (Sinha et al., 2015; Wang et al., 2020) emerged. Since Microsoft decided to discontinue MAG, the successor database OpenAlex was started by Priem, Piwowar, and Orr (2022). With the many databases that are available in principle for research evaluation purposes, the question arose whether all databases come to similar results in a certain research evaluation situation. It is especially interesting to know whether the results from the freely available database form a good alternative to the commercial databases.

In a previous case study, Scheidsteger, Haunschild, Hug, and Bornmann (2018) analyzed the publications of a computer science institute with a well-maintained publication list. They chose a bibliometric standard citation impact indicator (in the area of field-normalized indicators) and tested whether the indicator scores are similar across two different databases. Thus, they investigated the convergent validity of field-normalized indicator scores that have been generated based on MAG and WoS data. The results were encouraging (i.e., the values were in a good agreement) and motivated the present study with a significantly enlarged publication set from 48 German universities that cover a broad range of subject areas (and not only computer science, as in the first case study).

In this follow-up study of the case study by Scheidsteger et al. (2018), we leave the field of computer science and are instead interested in the convergent validity of field-normalized scores from different databases across different scientific fields. Field-normalized citation scores were calculated based on data from four different databases – three commercial databases and OpenAlex. We used a fixed common set of matching publications with a high accuracy of affiliation assignments: Do we receive the same or similar field-normalized scores when the same indicator is used or not? If we receive similar scores, one could question the praxis of using commercial databases in institutional research evaluation instead of a free database. We also investigated whether the similarity of scores differs between main scientific subject areas such as natural and social sciences.

Similarities and differences of literature databases

After the launch of Scopus in 2004 as the first serious competitor to WoS, some comparisons of WoS and Scopus have been published on a small scale with a focus on differences in the citation-based ranking of research units. Torres-Salinas et al. (2009) analyzed the 50 departments of Health Science at the University of Navarra (Spain) and found 14% more citations in Scopus than in WoS, but the rankings were similar. In a comparative study of WoS and Scopus in the field of information science, Meho and Sugimoto (2009) found significant differences in the ranking of research units up to the institutional level but not on the country level.

Stahlschmidt and Stephen (2019) took a more comprehensive view on publications from 2009 to 2015 indexed in both databases. The authors grouped the papers according to their second-level Organisation for Economic Co-operation and Development (OECD) subject categories assigned in the databases. They calculated excellence rates (shares of papers that belong to the 10% most frequently cited papers within their subject category and publication year) for the different sectors of the German science system. Their main results were three-fold: (1) Almost all entities had a higher citation-based impact in Scopus than in WoS, including the German universities. (2) There is a citation impact bias in WoS towards basic research and a bias in Scopus towards applied research. (3) The relative ranking of the sectors of the German science system changed only marginally.

In 2015, a new bibliographic database was launched in addition to WoS and Scopus: Microsoft Academic Graph (MAG) came with an outstandingly high coverage of (scientific) documents. One of its most attractive properties was that it was free to use. This triggered a series of comparative evaluations with the other databases, e.g., (Purnell, 2022; Visser et al., 2021), of which Huang et al. (2020) is the most relevant study with respect to our institutional focus. The authors compared university rankings performed in WoS, Scopus, and MAG that were built on affiliation data and found significant differences between the databases. With respect to MAG, the differences seem to stem mainly from the very much broader coverage but also a significant lack of complete affiliation data.

The emergence of Dimensions (in 2018) as a more recent serious commercial competitor to WoS and Scopus sparked another round of database assessments and comparisons. Orduña-Malea and Delgado-López-Cózar (2018) found that Dimensions shows similar metrics as Scopus but has a higher coverage. This was underlined by the comparison of all three databases by Stahlschmidt and Stephen (2022), who used matching publications between 2016 and 2018. They found similar values for citation-based indicators and a similar focus on applied research areas in Dimensions and Scopus as opposed to the more basic research-oriented content in WoS with lower citation-based impact in their respective exclusive content.

MAG's successor OpenAlex, one of the databases used in this study, inherited nearly all of MAG's publications (except patents) and their metadata (Scheidsteger & Haunschild, 2023). OpenAlex also makes the metadata and the software used to index and classify publications openly available. Significant differences between the databases concern the assignment of document types and of (subject) concepts to publications. The OpenAlex concepts are the counterparts to the Fields of Study in MAG. We found some comparative studies involving OpenAlex that focused on coverage or metadata accuracy and completeness (Céspedes et al., 2025; Ortega & Delgado-Quirós, 2024; Turgel & Chernova, 2024) but only one very recent study about citation-based comparisons: Thelwall and Jiang (2025) calculated and compared different citation-based indicators in OpenAlex and Scopus against two gold standards – one of them derived from expert scores in the UK Reference Excellence Framework of 2021. They found that a certain OpenAlex based citation indicator performs best if also information from Scopus (its finer grained document type classification) is used. They overall concluded that "OpenAlex is suitable for citation analysis in most fields".

WoS, Scopus, Dimensions, and OpenAlex use different ways of defining their indexed publications and therefore have different coverages. WoS is most restrictive in selecting "key journals" for each discipline to cover their scientific core (Birkleet al., 2020). Scopus aimed at becoming one of the largest curated bibliographic abstract and citation databases and therefore has many more journals indexed than WoS (Baas et al., 2020). However, since the expansion of WoS by the Emerging Sources Citation Index (Clarivate, 2025), the difference between WoS and Scopus in terms of number of journals has been decreasing. Dimensions takes an "inclusive" approach and indexes the widest possible range of contents and provides users with filters to narrow the publications down to suitable sets according to their own needs (Herzog et al., 2020). OpenAlex has incorporated with MAG the largest available bibliographic database at that time. OpenAlex is constantly expanding its corpus by pulling research output from registry agencies like Crossref and DataCite or from institutional and national repositories (OpenAlex, 2025; Priem et al.,2022).

In bibliometrics, the databases are mainly used for citation analyses in performance measurements. Since the coverages of publications in WoS, Scopus, Dimensions, and OpenAlex are different, differences in citation counts can be expected, in particular "greater coverage not only increases the included publications of any entities to be analyzed, but also increases the overall coverage and therefore alters the environment for any citation-based evaluation of scientific impact of the entities. Further, the changes in the overall coverage and in the coverage of an entity's publications might not be equivalent, but might result in either beneficial or adverse effects for the analysed entities" (Stahlschmidt & Stephen, 2019, p. 1699). In the interpretation of citation impact scores for scientists or institutions, one should have in mind that they depend on the database used.

Three properties of publications can lead to differences in citation impact measurements between the databases (besides their coverage): the publication year, the document type, and the subject classification. The latter both play the most important role for the definition of reference sets for the calculation of field-normalized impact indicators. The four databases use their own disciplinary classifications: WoS and Scopus assign journals to subjects; Dimensions and OpenAlex classify each item using machine learning algorithms. The algorithms solve issues with the treatment of multidisciplinary journals in journal-based assignments.

The differences between the databases (coverage, document type, subject classification) may have "important implications for the set of documents against which a publication is normalized and compared in the context of each database" (Stahlschmidt & Stephen, 2022, p. 2415). This study is intended to empirically analyze how meaningful the implications are in research evaluation practice. We used a dataset of 48 German universities and compared their field-normalized citation impact scores calculated with four different databases.

Data and methods

Selection of data sources

We used the three commercial databases Web of Science, Scopus, and Dimensions and the free database OpenAlex as sources of bibliometric data. The WoS data used were retrieved from an in-house WoS database developed and maintained by the Max Planck Digital Library and derived from the Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), and Arts and Humanities Citation Index (AHCI) provided by Clarivate (Philadelphia, Pennsylvania, USA). The database contains disambiguated and unified address information for German research institutes and universities developed by the I^{2}SoS Bibliometrics Team at the University of Bielefeld[1] and provided by the German "Kompetenznetzwerk Bibliometrie" (KB funded by the BMBF via grant 16WIK2101A, Competence Centre for Bibliometrics[2]). The Scopus data derived from Elsevier were

[1] https://www.uni-bielefeld.de/einrichtungen/i2sos/bibliometrie/index.xml

[2] http://www.bibliometrie.info

also provided by the KB. The WoS data was released in October 2021 and the Scopus data in April 2021. From Dimensions, we used a data dump from January 2022 and from OpenAlex, a snapshot from February 2022.

Field-normalized citation scores

For the comparison of the field-normalized scores across the four databases, we used the normalized citation score (NCS; Waltman et al., 2011). It is one of the most popular approaches to field-normalize citation counts (van Wijk & Costas-Comesaña, 2012). In principle, any other field-normalized indicator could have been used in this study such as percentiles (Bornmann & Williams, 2020). Although the NCS is widely used in research evaluation, it has been criticized because it is susceptible to outliers. We decided to use the indicator in this study, because the alternatives are also criticized, and we did not expect different results based on other field-normalized indicators.

The NCS is calculated as follows: The citation count of each paper is divided by the average citation count of similar papers (i.e., the reference set). Similar papers are defined as papers from the same field, publication year, and document type. The NCS is formally defined as

$$NCS_i = \frac{c_i}{e_i} \tag{1}$$

with c_i denoting the citation count of a focal paper i and e_i denoting the average citation rate of a reference set of similar papers (Lundberg, 2007; Rehn et al., 2007). In many cases, papers in the databases are assigned not only to one but to multiple fields. In this case, we calculated several NCS values for each paper. To obtain a single NCS for each paper, the multiple NCS values were averaged (Haunschild & Bornmann, 2016).

As an aggregated citation impact indicator, we also used the mean normalized citation score (MNCS) (Waltman et al., 2011), defined as the average over the NCS values of a specific research unit, here, a university.

Subject classifications

The expected citation rates for the NCS were calculated based on the different field categorization schemes used in the four databases. In WoS (Birkle et al., 2020) and Scopus (Baas et al., 2020), journals are intellectually assigned to 255 *WoS Subject Categories* (*WoSSC)* and 335 *All Science Journal Classification Codes* (*ASJC)*, respectively. In the other two databases, subjects are assigned paper-based using different taxonomies and machine learning algorithms. Dimensions (Herzog et al., 2020) has a two-level hierarchy of *Fields of Research* with 22 main categories and 154 subcategories. OpenAlex has a six-level hierarchy of *concepts* with 19 top-level categories and 284 second-level categories (Scheidsteger & Haunschild, 2023). In the case of Dimensions and OpenAlex, we used the second-level categories for the field-normalization because of their similar granularity compared to the journal-based schemes. Based on the different field categorization schemes in the databases, we received two groups of NCS values: (1) scores from a journal-based classification with NCS_WoS and NCS_Scopus, and (2) scores from a paper-based classification with NCS_Dimensions and NCS_OpenAlex.

If a publication lacks a subject classification it is excluded from the calculation of the expected citation rates for the NCS. In WoS and Scopus, only less than 0.01% of the publications between 2013 and 2017 do not have a subject category assignment. In OpenAlex, however, 31.5% of these publications have no (second-level) concept. If we only consider the most relevant document type *journal-article* (see next section), this percentage goes down to 14.6%.

Document types

The expected citation rates for the NCS were also calculated based on different document types in the four databases: In WoS and Scopus, all items between 2013 and 2017 have a document type assigned, but in OpenAlex, 40.2% of the items do not have a document type and are therefore not included in the calculation of the expected citation rates for the NCS. Several document types that were treated separately in WoS and Scopus are grouped under one document type in Dimensions and OpenAlex. For example, the never cited reply by Rychik (2015) is assigned to the document type Letter in WoS and Scopus, but in Dimensions and OpenAlex, it has the document type Article. Another example is the never cited paper by Edwards (2015): WoS classified it as Editorial Material and Scopus as a Short Survey whereas both Dimensions and OpenAlex label it as Article. We expect that such – usually poorly cited items – decrease expected (field-specific) citation rates and thereby increase NCS values for the document type Article in Dimensions and OpenAlex as compared to WoS and Scopus.

Publication set

For the comparison of NCS values in this study, it was necessary to have the same institutional publication set from each database. To reach this goal, we started with the WoS database since we have disambiguated publication data for German universities on a high quality level. We focused on the publication years from 2013 to 2017 (to have citation windows of at least five years), and the document types Article and Review (i.e., only substantial publications). We only considered papers in the following OECD subject areas: Natural sciences, Engineering, Medicine, and Social Sciences. Subject areas defined by the OECD are broad areas that combine several WoS subject categories. In the remaining subject areas (such as Arts and Humanities), the use of bibliometrics is questionable because the coverage of the literature in the databases is mainly not given.

We restricted the publications only to those with DOIs. This focus simplified the collection of a common dataset across the four databases and missed only at most 4% of the initial dataset in each publication year. In order to have reliable data across the publication years, we chose the 48 German universities that published more than 3000 papers in total between 2013 and 2017. The final WoS dataset consisted of 363,020 publications, which were successively matched with the other databases by retaining only available and unique DOIs in the respective database. The match of the WoS data with data from the other databases using DOIs resulted in a common dataset of 334,511 papers. If a paper is assigned to at least one WoSSC associated with one of the four OECD subject areas, it is said to be assigned to that subject area. In total, there are 394,660 distinct assignments to OECD subject areas in the common dataset. In all databases, citations were counted until the end of 2020.

Table 1 Number and percentage of publications (within the common set of 334,511 DOIs) in the four databases suited for the calculation of field-normalized citation scores

Database	# Publications	% Publications
WoS	334,385	99.96
Scopus	334,227	99.92
Dimensions	316,866	94.73
OpenAlex	309,716	92.59

Of the common dataset, only publications could be considered in the comparisons of NCS values for which a second-level classification had been assigned in Dimensions and OpenAlex (see above). Furthermore, we restricted the dataset to the papers that have at least 10 documents with a mean citation count of at least 1.0 in their reference set as proposed by Haunschild et al. (2016) to have reliable reference sets for calculating the NCS. These restrictions led to the publication numbers for this study as shown in Table 1.

The gap between Scopus and Dimensions in Table 1 is mainly caused by missing subject classifications. In Dimensions, 17,084 of 334,511 DOIs (5.1%) were lacking second-level Fields of Research; in OpenAlex, 22,983 DOIs (6.9%) did not have a second-level concept.

Mutual comparisons of databases

With four databases, we could perform six comparisons of NCS values. In the empirical analysis, we either looked at all publications at once or at each university separately. As statistical key figures to assess the similarity between two databases, we used two types of correlation coefficients that were also used by Scheidsteger et al. (2018): i) Spearman's rank correlation coefficient *r_s* (applicable to monotonous relations), and ii) Lin's concordance correlation coefficient *r_ccc* (Lin, 1989, 2000; Liu, 2016).

Spearman's r_s is a non-parametric statistic that measures the strength and direction of the monotonic relationship between two variables based on the ranked ordering of the data. It assesses whether an increase in one variable tends to be associated with an increase or decrease in another variable, without making any assumptions about the linearity of the relationship or the distribution of the data. In contrast, Lin's r_ccc evaluates the agreement between two continuous measurements by assessing both precision (the closeness of the data points to the fitted line) and accuracy (the closeness of the fitted line to the line of perfect concordance at 45 degrees).

Lin's coefficient quantifies how closely the observed values conform to the line of perfect agreement, thereby combining measures of correlation and bias correction. While Spearman's coefficient is sensitive to consistent but non-linear relationships and is suitable for ordinal data or data with outliers, Lin's r_ccc is specifically designed to assess the degree to which pairs of observations fall on the line of perfect agreement, making it more appropriate for evaluating the interchangeability of measurement methods such as the field-normalized measure of citation impact.

Results

Results at the level of individual papers

Table 2 shows Spearman's r_s for the six comparisons between the databases (and the number of publications considered). The consistently high r_s of at least 0.88 demonstrate *high correlations* (Cohen, 1988) between NCS values from the databases.

Table 3 displays Lin's r_ccc for the comparisons together with the associated confidence intervals (confidence level 95%). According to Koch and Spörl (2007), values of r_ccc between 0.8 and 1.0 mean an *almost complete agreement,* which is only reached by the comparisons of Dimensions with OpenAlex and Scopus, respectively. The other comparisons reach values between 0.6 and 0.8, pointing to a *strong agreement*.

Scatterplots allow graphical assessments of comparisons between NCS values from different databases in more detail than a correlation coefficient. As an example, Fig. 1 shows a scatterplot for the comparison of Scopus with WoS including the outcomes of a linear regression and the correlation coefficients. We present the scatterplot for only one comparison in the main text, since the scatterplots for the other comparisons are very similar (the other comparisons can be found in the Appendix as Figs. 13, 14, 15, 16, and 17). The overall tendency to higher NCS values in Scopus compared to WoS obviously–according to Eq. 1–stems from a broader inclusion of less highly-cited papers in the Scopus reference sets: For the document type Article (that is dominant in the common dataset: 94% in WoS; 92% in Scopus), we found a reduction of the overall average citation rate from 14.8 in WoS to 12.9 in Scopus for the period 2013 to 2017. For the document type Review (with a share in the common dataset of over 6% in both databases), we found a reduction by one third from 35.5 for WoS to 23.7 for Scopus. Figure 1 also reveals that in the realm of higher NCS values, there are many papers with a strong discrepancy in terms of their NCS values in the respective databases displayed by a strong deviation from the regression line.

In order to assess the robustness of the results in Table 2 and Table 3, we compared the correlation coefficients for the datasets including *all* publications against the whole set excluding those papers with NCS values among the top 1% in either database. We

Table 2 Spearman's r_s (below the diagonal) with respect to NCS values from four databases and numbers of publications considered (above the diagonal)

Database	WoS	Scopus	Dimensions	OpenAlex
WoS	*1*	334,135	316,809	309,647
Scopus	0.931	*1*	316,672	309,504
Dimensions	0.884	0.886	*1*	294,811
OpenAlex	0.882	0.884	0.912	*1*

Table 3 Lin's r_ccc (below the diagonal) together with the respective confidence intervals (above the diagonal) with respect to NCS values from four databases (r_ccc values higher than 0.8 are printed in bold)

Database	WoS	Scopus	Dimensions	OpenAlex
WoS	*1*	[0.7567;0.7591]	[0.7703; 0.7723]	[0.6769; 0.6793]
Scopus	0.758	*1*	**[0.8537; 0.8555]**	[0.7842; 0.7866]
Dimensions	0.771	**0.855**	*1*	**[0.8756; 0.8772]**
OpenAlex	0.678	0.785	**0.876**	*1*

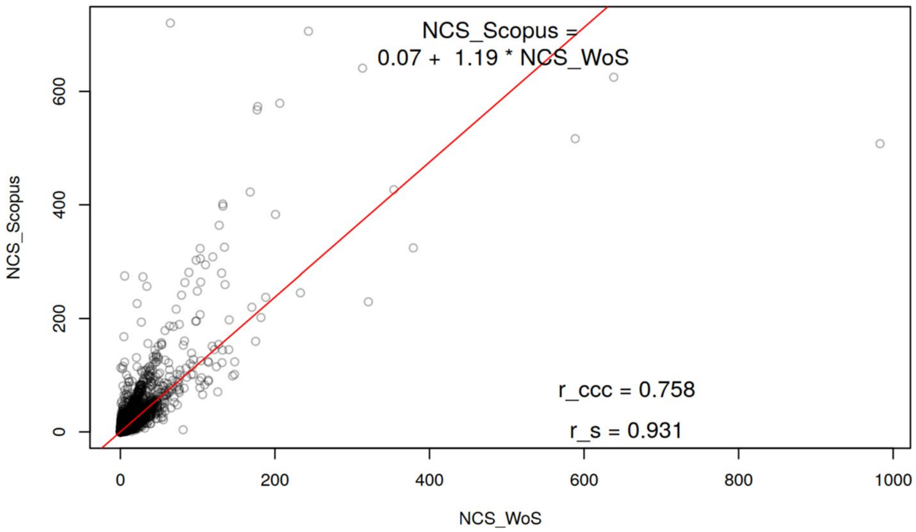

Fig. 1 Scatterplot of the NCS values of Scopus vs. WoS with parameters of a linear regression as well as the values of Spearman's r_s and Lin's r_ccc

analyzed the robustness of the results, since we observed stronger deviations between the databases at higher NCS values (see Fig. 1). After the removal of the top 1% papers, Spearman's r_s values remain very similar: Since the decrease is less than 0.005, the results are not shown. The changes in Lin's r_ccc values are shown in Table 4.

All changes in the coefficients are statistically significant (the confidence intervals do not overlap for the coefficients including and excluding the top 1% papers), but the effect sizes of the changes are small. Lin's r_ccc increases by 0.1 for the comparison of WoS and Scopus—the two databases with journal-based subject classification and the smallest coverage. The agreement improves from *strong* to *almost complete*. For the comparison of Dimensions and OpenAlex—the two databases with paper-based subject classification and the highest coverage—the coefficient decreases slightly by about 0.02. For the comparison of Dimensions and Scopus, the decrease by 0.08 leads to a change from *almost complete* to only *strong* agreement. For the other three comparisons, the decrease of Lin's concordance is smaller and does not change the assessment of agreement.

Since the rank correlation and concordance differences between the datasets including all papers and the restricted set of papers can be denoted as small, the results can be considered as robust.

Results based on papers grouped by universities

Since many research evaluations are not at the document level but at the institutional level, we undertook analogous comparisons at the university level. Because of many papers based on collaborations between German universities, the 334,511 papers occur 424,267 times in total in the separate evaluations of the universities. At first, we looked at the range of the correlation coefficients across all universities for each combination of databases as

Table 4 Effect of removing the top 1% NCS values of each database in each comparison on Lin's r_ccc for the six database comparisons indicated by the r_ccc point estimates, the confidence intervals, and the overall percentages of removed papers. Scores above 0.8 are printed in bold face

	Scopus vs. WoS	Dimensions vs. WoS	OpenAlex vs. WoS	Dimensions vs. Scopus	OpenAlex vs. Scopus	OpenAlex vs. Dimensions
Lin's r_ccc for all papers	0.758 [0.7567;0.7591]	0.771 [0.7703;0.7723]	0.678 [0.6769;0.6793]	**0.855** **[0.8537;0.8555]**	0.785 [0.7842;0.7866]	**0.876** **[0.8756;0.8772]**
Lin's r_ccc without top 1% papers	**0.867** **[0.8661;0.8677]** 1.28%	0.704 [0.7032;0.7058] 1.31%	0.622 [0.6202;0.6230] 1.35%	0.778 [0.7771;0.7795] 1.35%	0.716 [0.7149;0.7176] 1.37%	**0.857** **[0.8561;0.8579]** 1.28%
Change in Lin's r_ccc	0.109	− 0.067	− 0.056	− 0.077	− 0.069	− 0.019

Table 5 Min–max intervals of the correlation coefficients of the 48 universities separately. Spearman's r_s is given below, Lin's r_ccc above the diagonal

Database	WoS	Scopus	Dimensions	OpenAlex
WoS	*1*	[0.64; 0.97]	[0.54; 0.97]	[0.18; 0.93]
Scopus	[0.91; 0.95]	*1*	[0.66; 0.96]	[0.29; 0.94]
Dimensions	[0.86; 0.91]	[0.87; 0.90]	*1*	[0.41; 0.97]
OpenAlex	[0.86; 0.90]	[0.86; 0.90]	[0.89; 0.92]	*1*

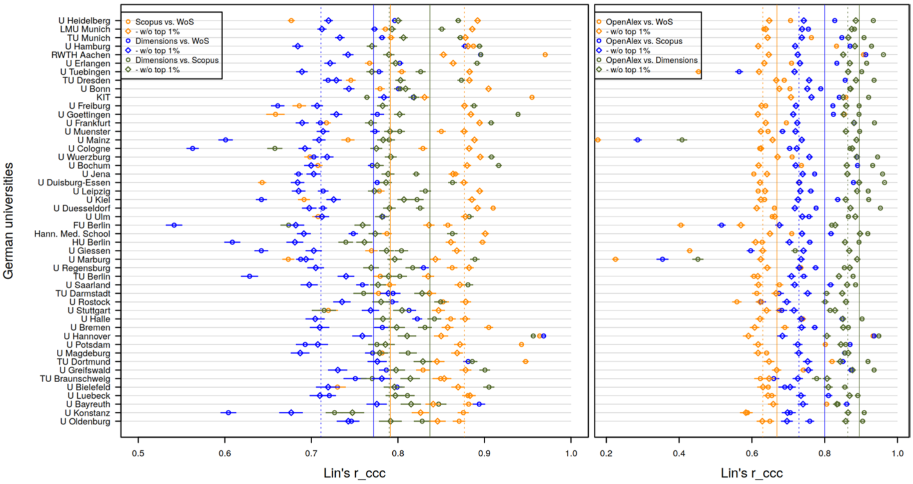

Fig. 2 Lin's r_ccc with confidence intervals for 48 German universities (ordered by publication output) in mutual comparisons of the three commercial databases on the left and of OpenAlex with the commercial databases on the right–considering either all documents or all documents without the top 1% papers in each database. Vertical lines indicate the median over all universities with all documents (solid) and all documents without the top 1% papers (dashed line), respectively. The Karlsruhe Institute of Technology is abbreviated as KIT

collected in Table 5. For example, in the upper right table cell for the comparison of WoS with OpenAlex we have obtained as smallest r_ccc value 0.18 for U Mainz and as largest r_ccc value 0.93 for U Hannover (see also the orange circles in the right panel of Fig. 2).

That the values of Spearman's *r_s* consistently show a *high* to *nearly perfect correlation* in small intervals centered around the values from Table 2 comes with no surprise. Lin's *r_ccc* displays a more diverse pattern. To assess the distribution of *r_ccc* values across the universities and possibly detect outliers, stripcharts may be helpful. Stripcharts display the values of r_ccc for every single university in each database comparison. This has been done in Fig. 2 for the full publication data (like in Table 5) denoted by circles as well as after removal of the top 1% papers (see the previous section) denoted by diamonds. The left stripchart in Fig. 2 compares the three commercial databases with one another: In each case, the distributions of *r_ccc* values (denoted by circles) are relatively homogeneous. We consistently have a *strong* to *almost complete agreement*, with the exception of Dimensions vs. WoS, where two universities have r_ccc values below 0.6 (FU Berlin 0.54, U Cologne 0.56), and six other universities range below 0.65 (U Mainz 0.60, U Konstanz 0.60, HU Berlin 0.61, TU Berlin 0.63, U Giessen 0.64, U Kiel 0.64).

The stripchart also displays a robustness check at the university level: the distribution of r_ccc with each university's top 1% papers removed–denoted by diamonds. This removal reduced the spread of r_ccc values across the universities drastically. The comparison between Dimensions and WoS has no longer r_ccc values below 0.67 but also none over 0.8. The most prominent example for a significant *reduction* is U Hannover with a decrease in r_ccc of 0.21 from about 0.95. Significant *increases* can be exemplified by, e.g., FU Berlin: + 0.14, U Cologne: + 0.13, TU Berlin, and U Mainz: + 0.11. For Dimensions vs. Scopus, the values vary in a small range between 0.75 and 0.83. Only in the case of the smaller Scopus and WoS databases (with a journal-based subject classification), the median over the universities increases (by nearly 0.1) and *all* r_ccc values can be seen as pointing to an *almost complete agreement*. For three universities, the increase in r_ccc is even greater than 0.2 (U Duisburg-Essen 0.23; U Goettingen 0.23; U Heidelberg 0.22); four (more technical) universities experience a decrease in r_ccc of at least 0.1 (KIT 0.12; RWTH Aachen 0.12; U Hannover 0.11; TU Dortmund 0.10).

The largest spreads of r_ccc values in Table 5 occur at comparisons including OpenAlex. Comparing OpenAlex with the three commercial databases in the right stripchart of Fig. 2 (marked by circles) reveals in each case several outliers separated from a more or less homogeneous majority field. Two outliers are among the most extreme ones in each comparison: U Mainz and U Marburg have very low *r_ccc* values of about 0.2 (OpenAlex vs. WoS), of about 0.3 (OpenAlex vs. Scopus), and of about 0.4 (OpenAlex vs. Dimensions). In three other cases, at least the values for comparisons of OpenAlex vs. WoS or Scopus are below 0.6: FU Berlin (OpenAlex vs. WoS: 0.40; OpenAlex vs. Scopus: 0.52), U Giessen (OpenAlex vs. WoS: 0.43; OpenAlex vs. Scopus: 0.60), and U Tuebingen (OpenAlex vs. WoS: 0.45; OpenAlex vs. Scopus: 0.56).

The removal of top 1% papers in the robustness check changes the picture significantly. The median over the universities (marked by diamonds) decreases by between 0.03 and 0.07. There are only three universities with values slightly below 0.6 remaining for OpenAlex vs. WoS: FU Berlin with r_ccc = 0.57, U Konstanz with 0.58, and U Hannover with 0.59. For OpenAlex vs. Dimensions, *all* r_ccc values are greater than 0.8 and therefore even point to an *almost complete agreement.*

Results at the university level based on institutional MNCS values

In many institutional research evaluations, universities are compared using MNCS values. The results of Waltman et al. (2012) point out that single papers with very high NCS values can have a stronger influence on the MNCS and thus on correlations of the scores between databases. Figures 3 and 4 show the MNCS values for the 48 universities based on data from the four databases–*with* all publications and *without* the top 1% publications. For each university in both comparisons, the MNCS values follow a strict order: MNCS_WoS < MNCS_Scopus < MNCS_Dimensions < MNCS_OpenAlex. This is plausible especially because of different influences on both the numerator and the denominator of the NCS formula, see Eq. (1).

The numerator is increased overall by an increasing number of potentially citing publications, and the denominator is decreased by an increasing number of poorly cited publications. The visual impression of a high correlation in the order of MNCS values is corroborated by Spearman's r_s listed in Table 6. The highest r_s values occur between WoS and Scopus as well as OpenAlex and Dimensions. For both comparisons, the correlation

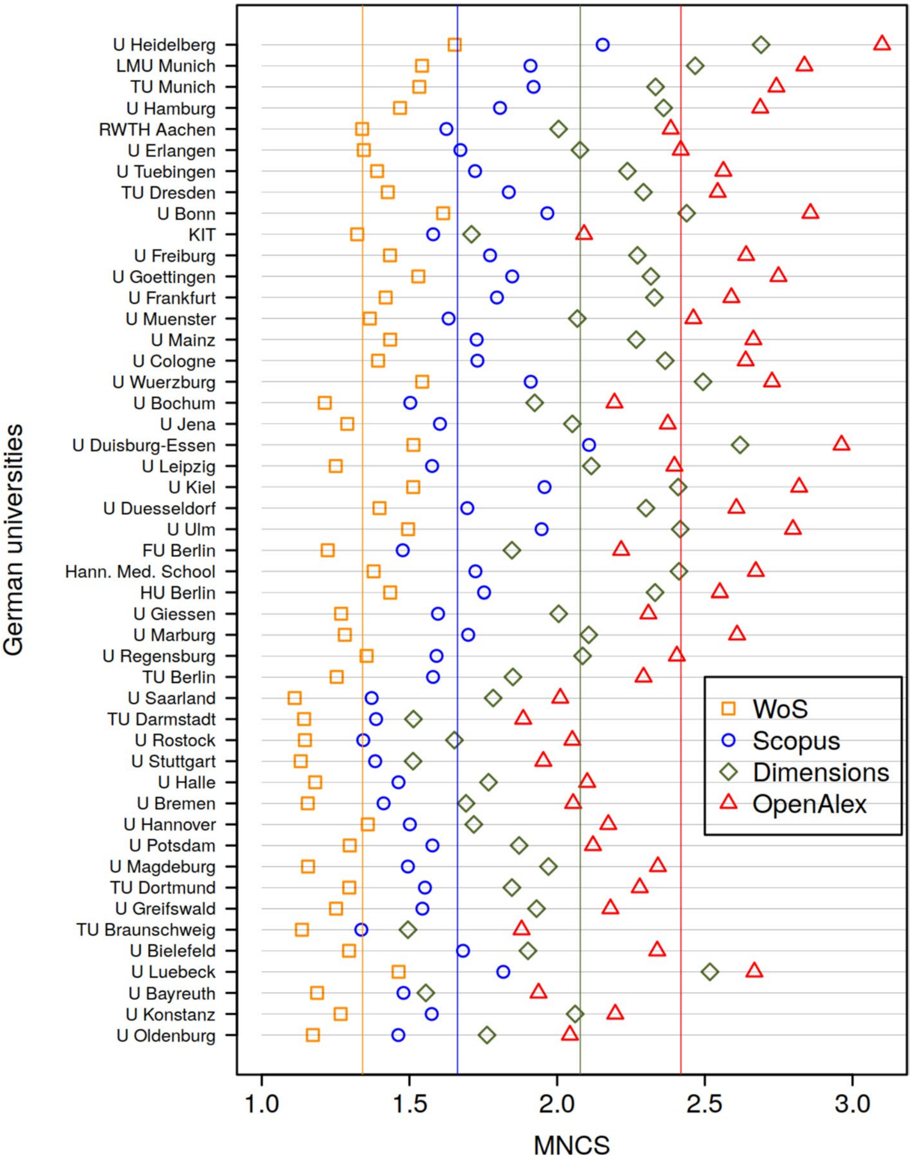

Fig. 3 MNCS values across the 48 German universities (ordered by publication output) calculated within each of the four databases. The vertical solid lines represent the mean values across the universities for each database

coefficients are preserved after removing the top 1% papers whereas for the other comparisons they have decreased–but not by more than 0.1. Viewed through the MNCS indicator lens, the differences between the universities are similarly represented in the four databases, but the impact level is different. The concordance coefficients that have been calculated based on the MNCS values of the German universities are rather low

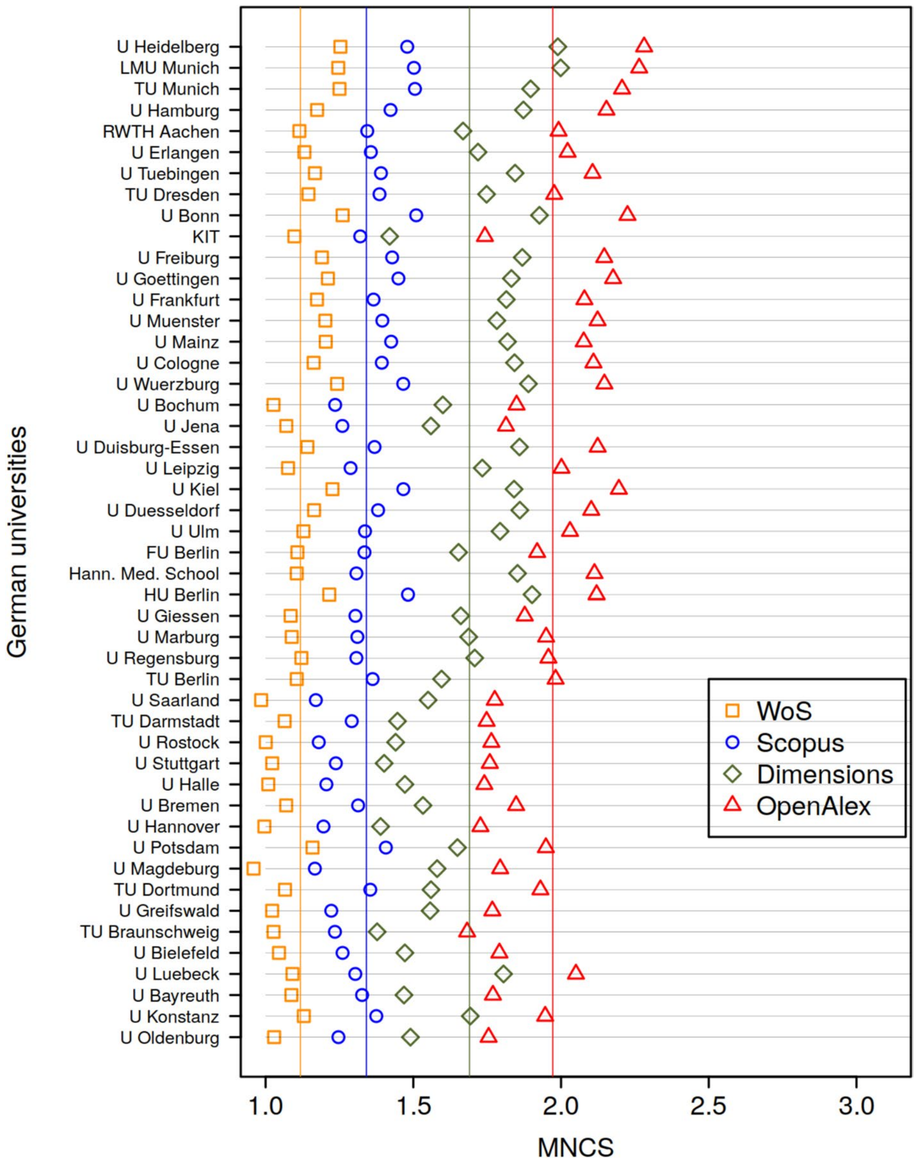

Fig. 4 MNCS values across the 48 German universities (ordered by publication output) calculated within each of the four databases after removing the top 1% publications. The vertical solid lines represent the mean values across the universities for each database

correspondingly: Although the concordance coefficients that have been calculated at the level of single papers are high (they indicate at least a strong agreement), many coefficients that have been calculated at the aggregated level are low.

The differences in the university rankings between the databases according to the MNCS values in Fig. 3 are collected in Table 11 in the Appendix and visualized in

Table 6 Spearman's r_s including and excluding the top 1% publications and Lin's r_ccc for the database comparisons of the MNCS values for 48 German universities

	Scopus vs. WoS	Dimensions vs. WoS	OpenAlex vs. WoS	Dimensions vs. Scopus	OpenAlex vs. Scopus	OpenAlex vs. Dimensions
Spearman's r_s	0.96	0.89	0.92	0.93	0.95	0.96
Spearman's r_s without top 1%	0.97	0.87	0.90	0.82	0.86	0.96
Lin's r_ccc	0.32	0.12	0.06	0.37	0.16	0.60
Lin's r_ccc without top 1%	0.22	0.07	0.03	0.16	0.06	0.41

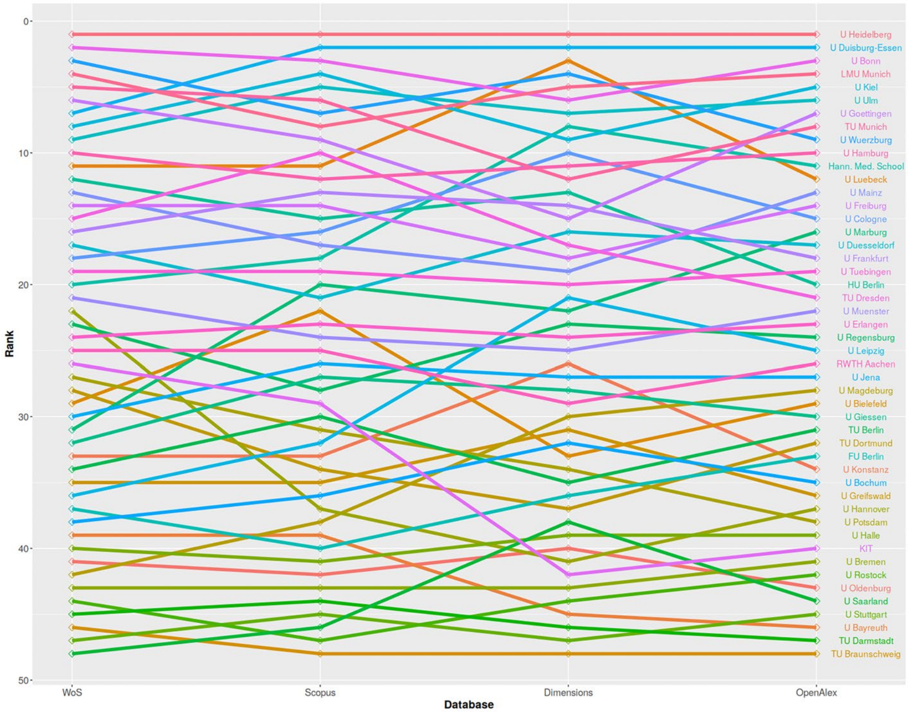

Fig. 5 MNCS ranks for the 48 German universities in the four databases

Fig. 5. For most of the universities, the changes from one database to another amount only up to a few rank positions. Only a few experience considerable jumps of even more than 15 places like the KIT or U Hannover.

Results based on papers grouped by four subject areas

By performing analyses grouped by four OECD subject areas (Natural Sciences, Medicine, Engineering, and Social Sciences), we investigated systematic differences in the

Table 7 Numbers and percentages of distinct assignments of papers in WoS to the four OECD subject areas (in bold) and their mutual overlap in the respective OECD subject areas. The numbers are given in and below the diagonal, and the percentages in each of the respective compared OECD subject areas are given above the diagonal in the order of (row; column). For example, the overlap between Natural Sciences and Medicine is 7.5% with respect to all papers in Natural Sciences but 12.1% with respect to all papers in Medicine

OECD subject area	Natural Sciences	Medicine	Engineering	Social Sciences
Natural Sciences	**188,094 (47.7%)**	(7.5; 12.1)%	(18.0; 55.2)%	(1.8; 11.6)%
Medicine	14,103	**116,750 (29.6%)**	(3.2; 6.0)%	(4.8; 19.7)%
Engineering	33,792	3,695	**61,222 (15.5%)**	(2.1; 4.6)%
Social Sciences	3,322	5,636	1,304	**28,593 (7.2%)**

distribution and correlation of NCS values. Table 7 lists the numbers of distinct assignments in the four OECD subject areas as well as their mutual overlap, i.e., the number of distinct DOIs assigned to at least two OECD subject areas.

In contrast to Stahlschmidt and Stephen (2019), we do not determine the OECD subject area assignments for each paper based on the native classification inside the respective database. We hold on to the WoS assignments, which have been decisive in the compilation of the common publication set. To get a first impression of the influence of the subject area (from WoS) on the similarity of NCS values, we produced scatterplots for the comparisons of NCS values in Scopus and WoS for the four OECD subject areas (see Fig. 6).

The results in the figure reveal very similar NCS values (on average) with the slope of nearly 1.0 for Natural Sciences and Engineering. On the contrary, due to the slope of 1.83 for Medicine we have for high-impact papers nearly twice as high NCS values in Scopus than in WoS. The respective slopes are similar for the comparisons of OpenAlex with WoS (Natural Sciences: 1.34; Medicine: 2.11) and Dimensions with WoS (Natural Sciences: 1.18; Medicine: 1.92) and point to a systematic effect. It reminds of the findings in the case study of a university's health department by Torres-Salinas et al. (2009), who found 14% more citations in Scopus than in WoS for this research area. According to Stahlschmidt and Stephen (2019), this bias can be traced back to the stronger focus on applied research including Medicine in Scopus. This is also true for Social Sciences but not that pronounced.

Figure 7 shows Spearman's r_s for the six comparisons and the four OECD subject areas. All coefficients are on a high level. The highest values occur for the comparison of Scopus vs. WoS (both with journal-based subject classification and lower coverage) with r_s greater than 0.9 for all OECD subject areas as could be expected from Table 2 displaying r_s based on *all* papers without disciplinary distinction. The next highest coefficients are associated with the comparison of Dimensions vs. OpenAlex (both with article-based subject classification and higher coverage) with an r_s of 0.91 for all OECD subject areas except Social Sciences. Social Sciences has an r_s less than 0.88– the minimal value in Table 2–like for all other comparisons besides Scopus vs. WoS. The smallest coefficient might be caused by the small overlap of the covered papers in this OECD subject area in the different databases. With slightly more than 7% of all distinct assignments, this OECD subject area has only a small influence on the overall assessment.

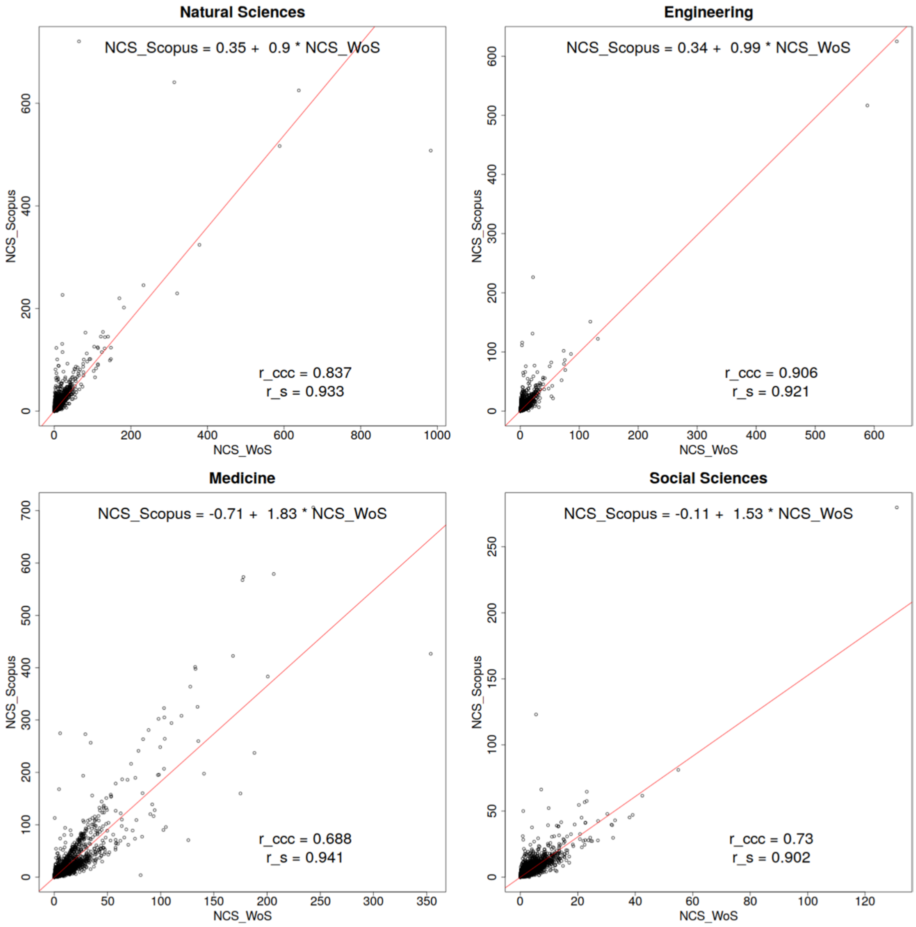

Fig. 6 Scatterplots of the NCS values of all publications separated by their OECD subject area in WoS

Figure 8 displays Lin's *r_ccc values* for the six comparisons and the four OECD subject areas–together with the associated confidence intervals. Like in the case of Spearman's r_s, Lin's r_ccc values are smallest for the Social Sciences–except for the comparison of Scopus vs. WoS. For the comparison of Scopus vs. WoS, Medicine shows an even lower r_ccc value, which has a significant influence on the overall r_ccc. Medicine has a share of 30% of all distinct assignments of publications to OECD subject areas.

In the comparison with the WoS, there is a descending order of the r_ccc values from Scopus via Dimensions to OpenAlex. In these three comparisons, we observe the highest spread of the r_ccc values across the OECD subject areas, and the r_ccc values of the two pairs Social Sciences and Medicine as well as Natural Sciences and Engineering are clearly grouped. In the comparisons with OpenAlex, there is an ascending order of the r_ccc values from WoS via Scopus to Dimensions. In the latter two cases, we observe that the r_ccc values of the OECD subject areas other than the Social Sciences are very close to each other.

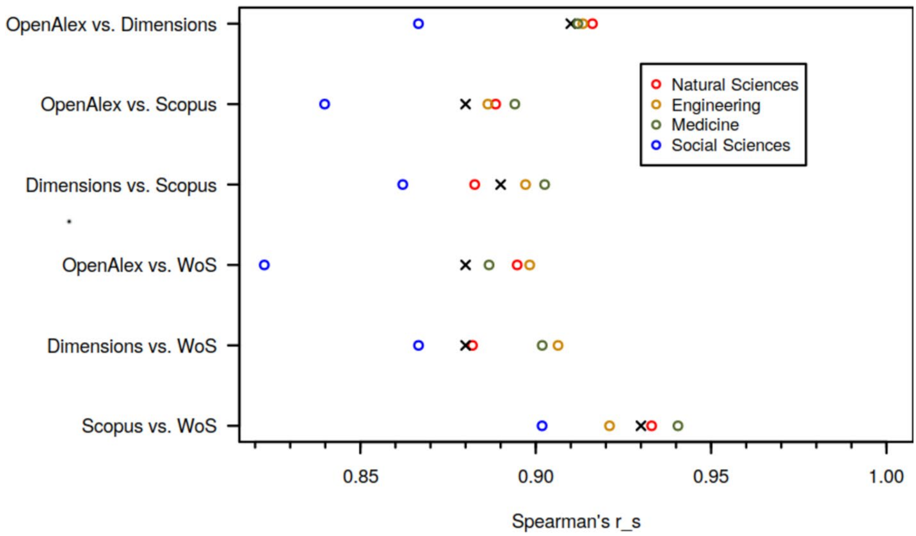

Fig. 7 Spearman's r_s for the six comparisons and the four OECD subject areas, separately. The crosses indicate the r_s values for the whole publication set that are indicated as reference values

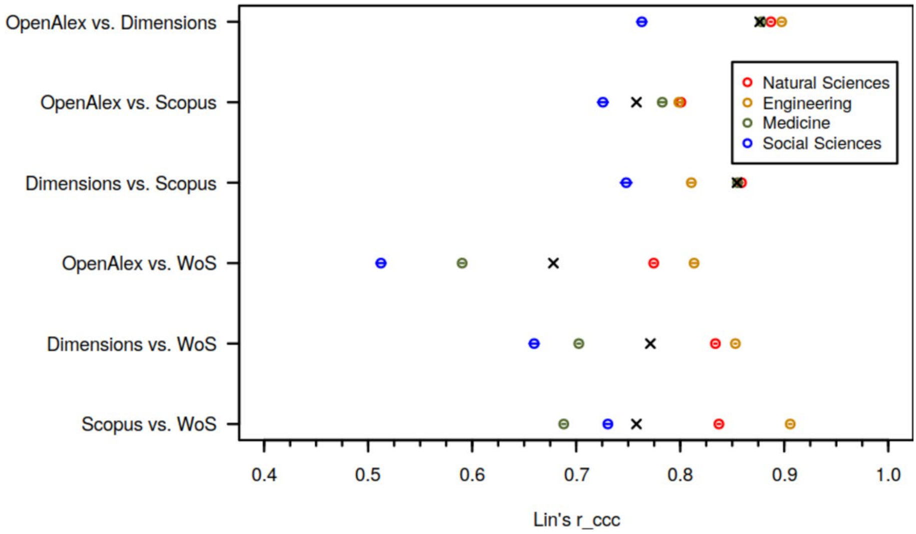

Fig. 8 Lin's r_ccc with confidence intervals for the six comparisons and the four OECD subject areas. The crosses indicate the respective r_ccc values for all papers that are indicated as reference values

Figure 9 displays Lin's *r_ccc* values for the six comparisons and the four OECD subject areas–together with the associated confidence intervals–after excluding the top 1% papers from each OECD publication subset as a robustness check. The comparison of the results with Fig. 8 shows that Medicine has a significant increase of the r_ccc value by 0.2 in Scopus vs. WoS. Medicine seems to dominate the overall increase of r_ccc in that comparison.

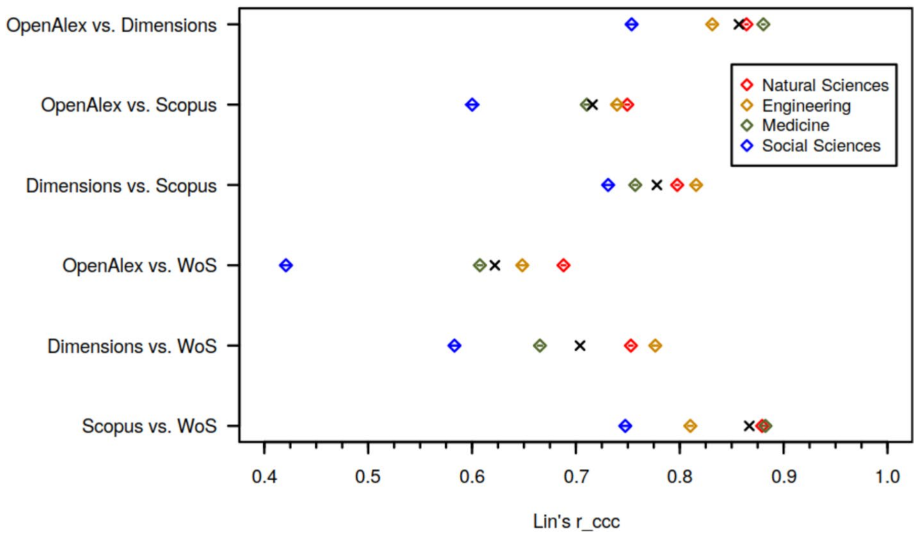

Fig. 9 Lin's r_ccc with confidence intervals for the six comparisons and the four OECD subject areas excluding the top 1% papers. The crosses indicate the r_ccc values for all papers that are indicated as reference values

In the other comparisons, the r_ccc values of Medicine remain at the same level or have decreased, e.g., by 0.1 in Dimensions vs. Scopus (where all OECD subject areas are now closer together).

Social Sciences has (again) the smallest r_ccc values–now even in three comparisons *below the strong agreement* threshold: Dimensions vs. WoS, OpenAlex vs. Scopus, and OpenAlex vs. WoS. In the latter case, r_ccc dropped by 0.1 to 0.4. The r_ccc values of Natural Sciences and Engineering change their order in some cases but remain the largest overall. Consistent with Table 4, the r_ccc value of OpenAlex vs. Dimensions (both with article-based subject classifications) has undergone the smallest change due to the removal of the top 1% papers–sustaining the degree of agreement in all OECD subject areas.

Results based on papers grouped by four subject areas and by universities

Figures 10 and 11 show stripcharts analogous to Fig. 2 but separated along the OECD subject areas Natural Sciences and Medicine in the order of their share of papers. The two subject areas Engineering and Social Sciences with the smallest shares of papers are not discussed here but shown in the Appendix (Figs. 18 and 19). The shares of the OECD subject areas for each university are displayed in Fig.20 in the Appendix. In the following stripcharts, the universities are ordered by the overall number of papers in the whole dataset. To present this information, we write for each university in brackets the average number of publications involved in the six database comparisons.

In the comparisons including OpenAlex in Fig. 10 and Fig. 11, we examine the five outlier cases from Fig. 2 considered earlier that displayed the overall Lin's r_ccc values for the 48 German universities without any grouping by subjects. We ask which OECD subject areas contribute the most to the low r_ccc values for those five universities.

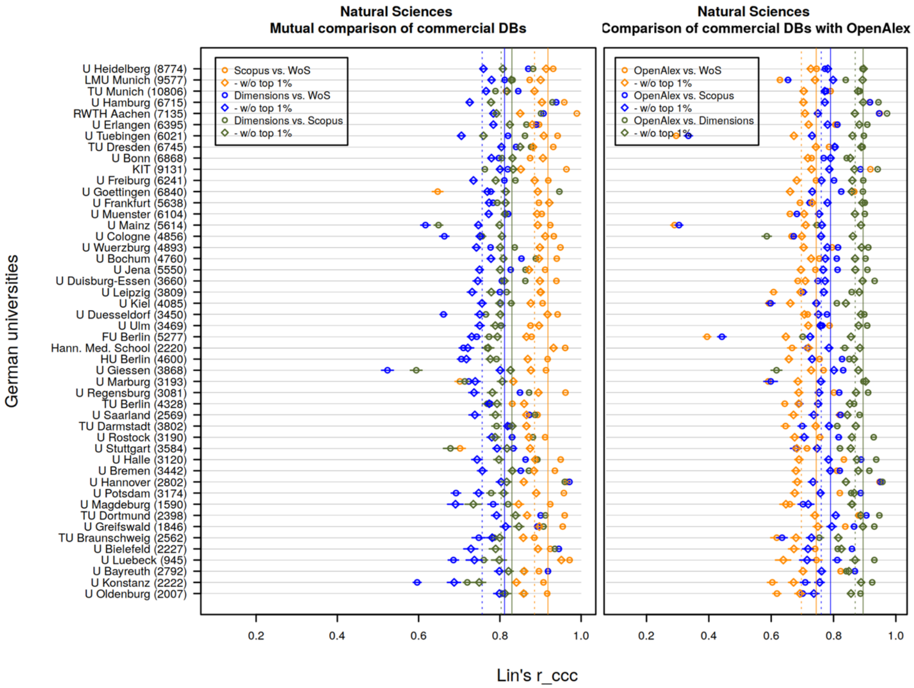

Fig. 10 Lin's r_ccc with confidence intervals for 48 universities (ordered by publication output) and the OECD subject area Natural Sciences in mutual comparisons of the three commercial databases on the left and of OpenAlex with the commercial databases on the right–considering either all documents or the documents without the top 1% papers. Vertical lines indicate the median over all universities with all documents (solid) and the documents without the top 1% papers (dashed line), respectively. Numbers in brackets indicate the average numbers of publications involved in all six database comparisons

Table 8 collects the relevant r_ccc values shown in the respective figures. As Table 8 shows, Medicine contributes the most to the low r_ccc values of U Mainz and U Marburg. In the case of U Mainz, the subject area Natural Sciences seems to play a role, too, at least concerning WoS and Scopus. Natural Sciences may be also responsible for the overall poor r_ccc values in the comparisons OpenAlex vs. WoS and OpenAlex vs. Scopus for U Tuebingen and FU Berlin. It seems that Medicine is decisive in this respect for the results of U Giessen.

Looking at the other three comparisons (only of the commercial databases) in Figs. 2, 10, and 11, the first of the two most notable results in Fig. 2 have been the small r_ccc values for Dimensions vs. WoS. Table 9 collects the respective r_ccc values.

In the case of the three Berlin universities, Medicine (see Fig. 11) is the strongest contributor to their result (FU 0.39, HU 0.39, TU 0.43)–even with shares of below 20%. This applies to a lesser extent also to U Cologne (0.53)–with a share of 40% of Medicine papers–and to U Kiel (0.61). For U Kiel, the value for Social Sciences is even lower (0.57) but this category has only a share of about 10% as opposed to more than 30% for Medicine. For U Giessen (0.52) and U Konstanz (0.60), the Natural Sciences with shares of about 50% account the most for the overall small r_ccc.

For U Mainz, the influences of Natural Sciences (50%) and Medicine (40%) seem to be of equal importance (0.62 resp. 0.58). Looking at U Mainz in more detail via scatterplots

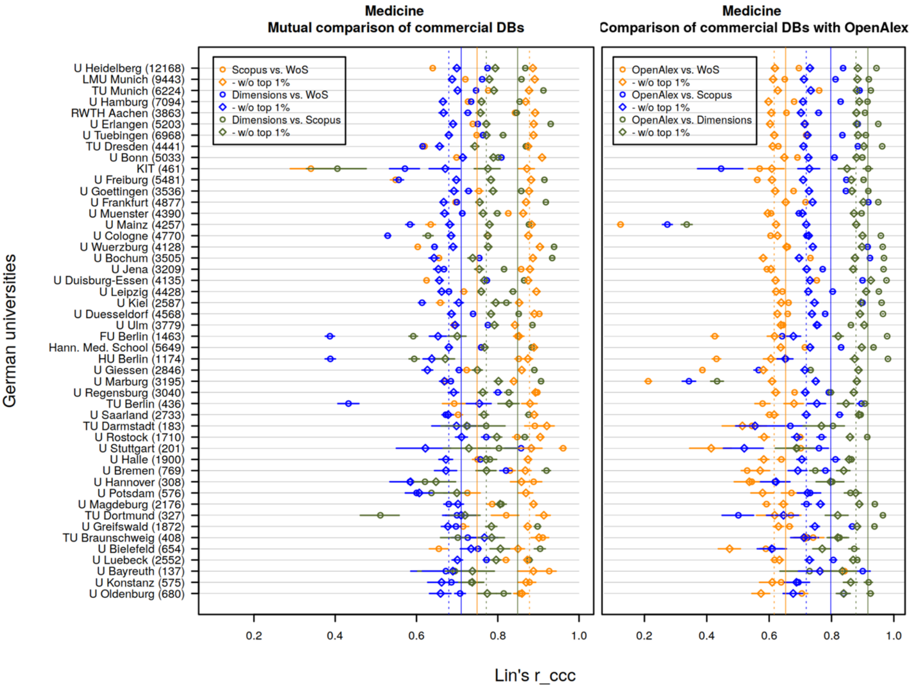

Fig. 11 Lin's r_ccc with confidence intervals for 48 universities (ordered by publication output) and the OECD subject area Medicine in mutual comparisons of the three commercial databases on the left and of OpenAlex with the other ones on the right–considering either all documents or the documents without the top 1% papers. Vertical lines indicate the median over all universities with all documents (solid) and without the top 1% papers (dashed line), respectively. Numbers in brackets indicate the average numbers of publications involved in all six database comparisons

for OpenAlex vs. WoS in Fig. 12, the results corroborate this interpretation as of the three top outlier papers–with respect to OpenAlex alone–two belong to the OECD subject area Medicine (Agha et al., 2016; Kleinstäuber et al., 2014) and one belongs to Natural Sciences (Haak et al., 2015). The removal of the three papers would propel the agreement up to 0.7.

The second most conspicuous result in Fig. 2 is the strong increase of r_ccc values for the Scopus vs. WoS comparison after removal of the top 1% papers. From the perspective of the OECD subject areas, Medicine has the strongest effect here on U Duisburg-Essen (by 0.25 from 0.62 to 0.87) and U Heidelberg (by 0.23 from 0.64 to 0.87)–with shares of about 40% resp. 50% of Medicine papers. U Goettingen seems to profit in its r_ccc increase primarily from Natural Sciences (0.24 from 0.0.65 to 0.89) with a share of about 55%.

Another striking result in all six comparisons are the very high r_ccc values of between 0.95 and 0.97 for U Hannover–but only in the Natural Sciences. This result seems to be due to a handful of publications by the Nobel-prize-awarded LIGO cooperation. The Laser Interferometer Gravitational-Wave Observatory (LIGO) is a groundbreaking scientific facility and observatory that aims to identify gravitational waves originating from cosmic events and to establish their detection as a fundamental method for astronomical exploration. This cooperation has very high and relatively similar NCS values in all four databases. Removing them–together with the other top 1% papers–results in various reductions of r_ccc but without leaving the realm of good agreement.

Table 8 Collection of Lin's r_ccc values (overall and for the two main OECD subject areas) for the five outlier cases discussed in connection with Fig. 2 for the comparisons of OpenAlex with the three commercial databases. The small overall r_ccc values up until 0.6 as well as those values from the OECD subject areas that seem to have the largest contribution to the small overall r_ccc values are printed in bold type

	U Mainz	U Marburg	U Tuebingen	FU Berlin	U Giessen
Overall (from Fig. 2*)*					
WoS	**0.18**	**0.23**	**0.45**	**0.40**	**0.43**
Scopus	**0.29**	**0.35**	**0.56**	**0.52**	**0.60**
Dimensions	**0.41**	**0.45**	0.90	0.82	0.72
Natural Sciences (from Fig. 10*)*					
WoS	**0.29**	0.59	**0.30**	**0.39**	0.77
Scopus	**0.30**	0.60	**0.33**	**0.44**	0.83
Dimensions	0.75	0.90	0.90	0.70	0.62
Medicine (from Fig. 11*)*					
WoS	**0.12**	**0.21**	0.72	0.43	**0.39**
Scopus	**0.27**	**0.34**	0.83	0.64	**0.57**
Dimensions	**0.34**	**0.43**	0.91	0.98	0.73

Table 9 Collection of Lin's r_ccc values (overall and for the two main OECD subject areas) for the eight universities with the smallest r_ccc values in the comparison of Dimensions with WoS from Fig. 2. The small overall r_ccc values as well as those values from the OECD subject areas that seem to have the largest contribution to the former are printed in bold type

FU Berlin	HU Berlin	TU Berlin	U Cologne	U Kiel	U Giessen	U Konstanz	U Mainz
Overall (from Fig. 2*)*							
0.54	**0.61**	**0.63**	**0.56**	**0.64**	**0.64**	**0.60**	**0.60**
Natural Sciences (from Fig. 10*)*							
0.74	0.71	0.78	0.66	0.80	**0.52**	**0.60**	**0.62**
Medicine (from Fig. 11*)*							
0.39	**0.39**	**0.43**	**0.53**	**0.61**	0.70	0.69	**0.58**

Discussion

The use of field-normalization scores is common in evaluative bibliometrics, but different sources of data are used for calculating the scores. In this study, we addressed the question whether the NCS values from different databases are similar (and can be compared with one another) or are different. The study followed several previous studies that investigated differences and similarities between various databases. Our results are not directly comparable to previous studies, since they either focus on other databases or other indicators and evaluated units (see subsection Similarities and differences of literature databases). Our results are consistent with the results of previous studies in that they also are mixed: The databases lead to similar but also to different results with respect to (field-normalized) citation impact measurements.

In this study, we were especially interested in the results for OpenAlex: Does this free database produce similar results as the commercial databases? We calculated correlation coefficients for six mutual comparisons of NCS values for a common set of nearly

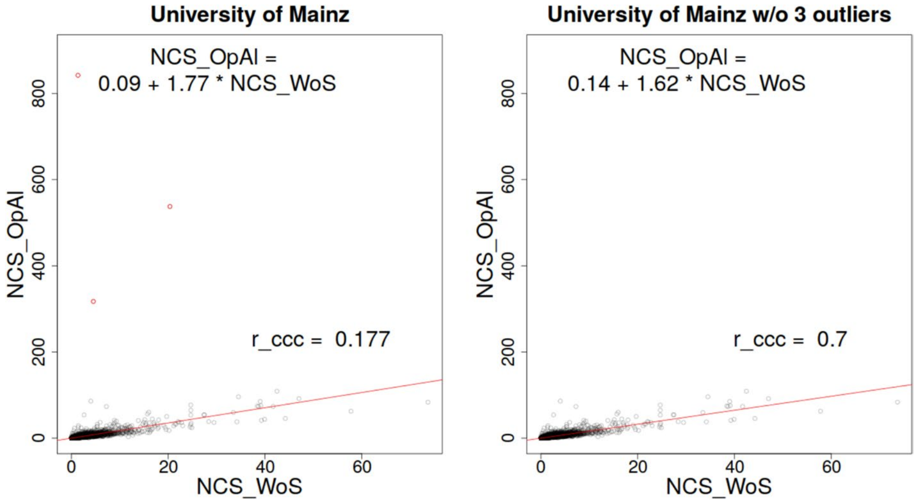

Fig. 12 Scatterplots of NCS values with linear regression and Lin's r_ccc of U Mainz for the OpenAlex vs. WoS comparison with and without the three top outliers

335,000 publications from 48 German universities in four databases. The results for Lin's concordance correlation coefficient r_ccc show (at the level of single papers) that *all* comparisons reveal *almost complete* but at least *strong agreement*–following the guidelines of (Koch & Spörl, 2007). We assessed the robustness of the results by removing the papers within the top 1% of NCS values in the databases. These additional analyses led to a small decrease in most cases with coefficients indicating at least a *strong agreement*.

Looking at the 48 German universities separately (but still at the paper level), we found in nearly all cases a *strong* to *almost complete agreement* between the three commercial databases, but several cases of very low r_ccc values in their comparisons with the free database OpenAlex. The removal of the top 1% papers, including those with strongly deviating NCS values across the respective databases, in most cases led to *strong* or *almost complete agreement*. We also compared aggregated MNCS values (at the level of single universities) and found a highly correlated representation of inter-university differences between all databases. However, the concordance between the MNCS values of the German universities is rather low. Both results indicate that relative comparisons between different universities are valid only within either of the databases. On the one hand, the results reveal that MNCS values which have been calculated using data from different databases should not be compared. On the other hand, the difference of the concordance coefficients at the paper and university level is a good example for the problem of the ecological fallacy in bibliometrics: The mean impact is not representative for the single papers' impact in the set.

The separation of the publication set along four OECD subject areas yielded some further insights. Concerning the overall agreement values, Natural Sciences and Engineering are in most comparisons close together with high agreement values, whereas Medicine follows with lower values, and the values for Social Sciences are far smaller. Only in the comparison of Scopus vs. WoS, the removal of the top 1% papers leads to the highest agreement for Medicine and a nearly as high r_ccc value for Natural Sciences. Because of the large share of both OECD subject areas in the whole set of subject assignments, this results is an increase of agreement for the whole dataset. A strong influence of those two OECD subject areas can also be seen in the distribution of agreement values across the 48 German universities. Very low agreement values can be traced back to the r_ccc values in those two OECD subject areas.

In this study, we investigated the measurement of citation performance based on NCS values from different databases but for the same set of publications from single universities. The reason for the investigation was a practical consideration: Does it make a difference which database is used in research evaluation? We know that citation counts for a single paper are different in the databases; but does this hold for NCS values? Since we designed the study from the perspective of research evaluation practice, we have not applied methods that can be used to standardize NCS values from different databases. For example, the values can be standardized if the definition of reference sets (and document type classification) is kept constant in the calculation. We know from Haunschild, Daniels and Bornmann (2022) and Haunschild and Bornmann (2022) that switching granularity in a classification scheme (concerning the field or document type), or switching from a (journal) classification system to an item-by-item-based approach (where fields may be defined based on citation networks) can have a significant effect on NCS values (within the same database).

There are some limitations concerning the generalizability of our results: (1) We were able to use disambiguated affiliation data for German universities. Since institutional data with disambiguation on a very high quality level may be scarcely available for other countries, it will be difficult to undertake similar studies for other countries. These studies are important yet to see whether our results are country-specific or can be generalized. (2) NCS and MNCS are susceptible to outliers. Future studies may focus thus on other field-normalized indicators that are not susceptible to outliers such as percentiles to possibly confirm our results.

Conclusions

Our results suggest (1) that comparisons of universities using NCS across different databases should be avoided, and that (2) institutional performance differences are similarly reflected within each database. The first conclusion from our study is thus that the suitability of OpenAlex for bibliometric evaluations seems to be similar to that of the established commercial databases (keeping in mind the possibility of its stronger distorting effects in the high NCS realm than in analyses based on data from commercial databases). This is confirmed by the conclusions of Thelwall and Jiang (2025). It remains to be seen if the changes in OpenAlex since the snapshot from January 2022, e.g., of the subject classification scheme, turn out to change our conclusions (Priem & Piwowar, 2022).

Our study also suggests (3) that the user of institutional citation impact values should be aware of the differences between the databases in the general citation impact levels–also in cases where field-normalized values are applied. It is common in many evaluation processes (e.g., of funding organizations) that the evaluated unit (a single scientist or an institution) is asked to deliver certain citation impact values. Since these values depend on the underlying database (as we know from this study), it is either important to request values from the same database or to request not only the values but also the name of the used database. If the name is known, the user is able to frame the NCS values and to properly compare them with other values.

Appendix

See Figs. 13, 14, 15, 16, 17, 18, 19, and 20 as well as Tables 10 and 11.

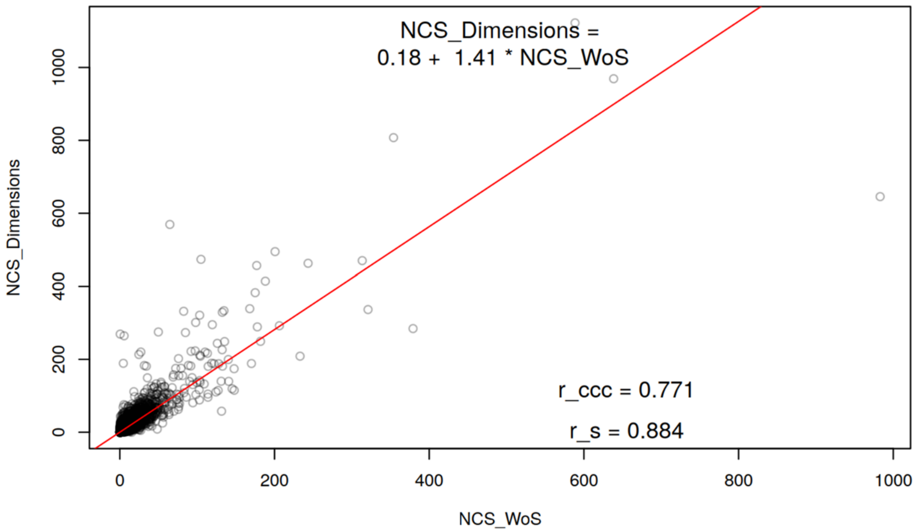

Fig. 13 Scatterplot of the NCS values of Dimensions vs. WoS with parameters of a linear regression as well as the values of Spearman's r_s and Lin's r_ccc

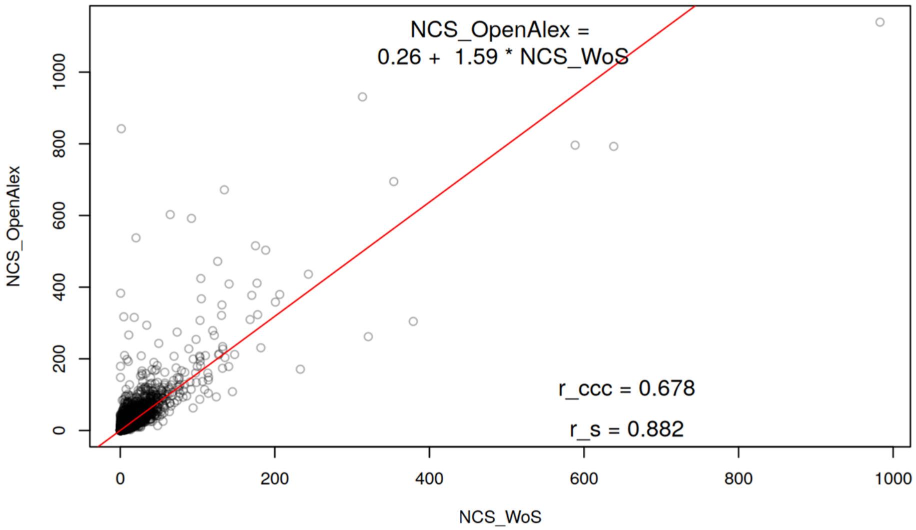

Fig. 14 Scatterplot of the NCS values of OpenAlex vs. WoS with parameters of a linear regression as well as the values of Spearman's r_s and Lin's r_ccc

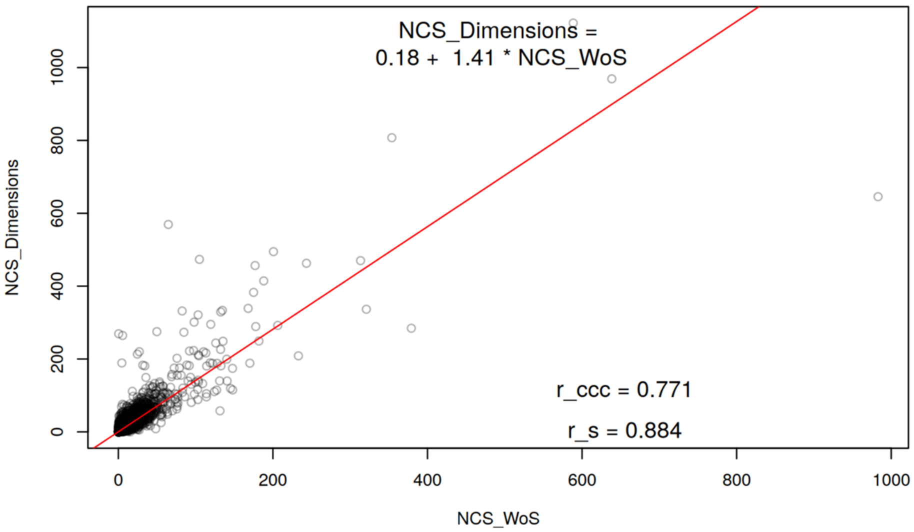

Fig. 15 Scatterplot of the NCS values of Dimensions vs. Scopus with parameters of a linear regression as well as the values of Spearman's r_s and Lin's r_ccc

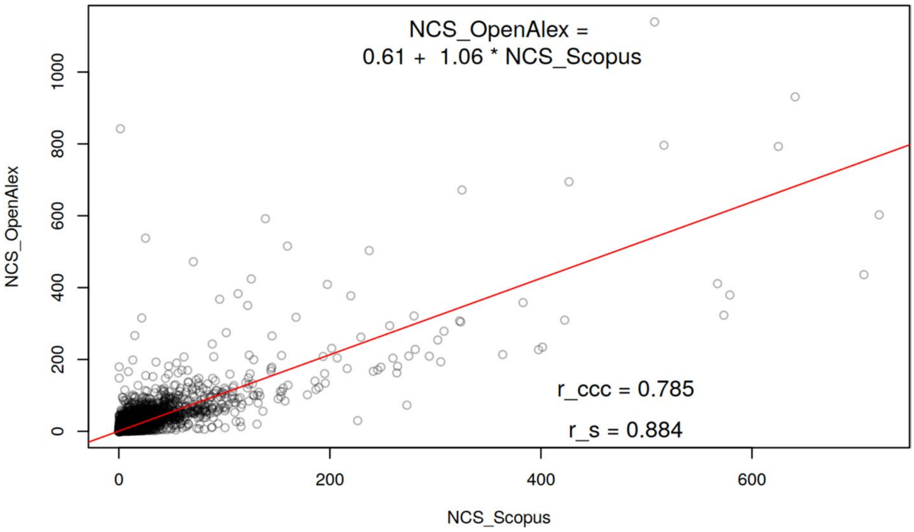

Fig. 16 Scatterplot of the NCS values of OpenAlex vs. Scopus with parameters of a linear regression as well as the values of Spearman's r_s and Lin's r_ccc

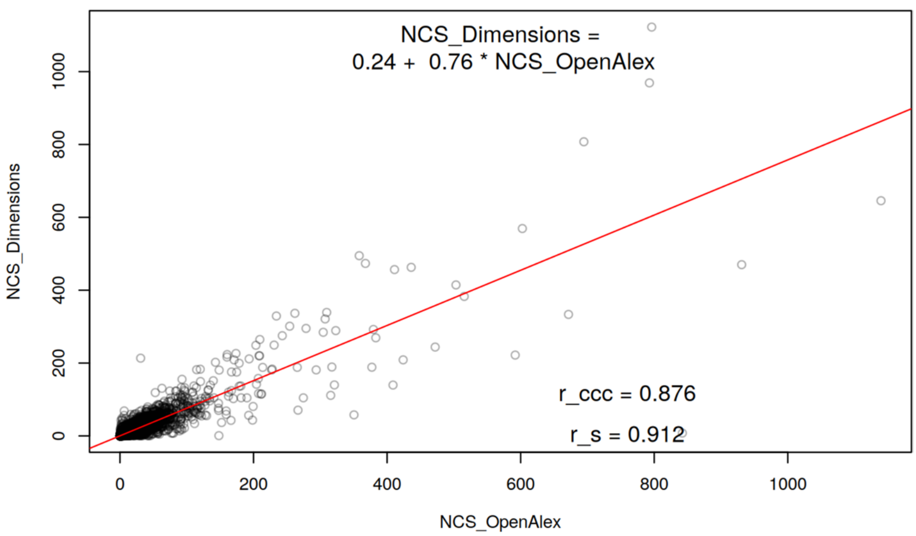

Fig. 17 Scatterplot of the NCS values of Dimensions vs. OpenAlex with parameters of a linear regression as well as the values of Spearman's r_s and Lin's r_ccc

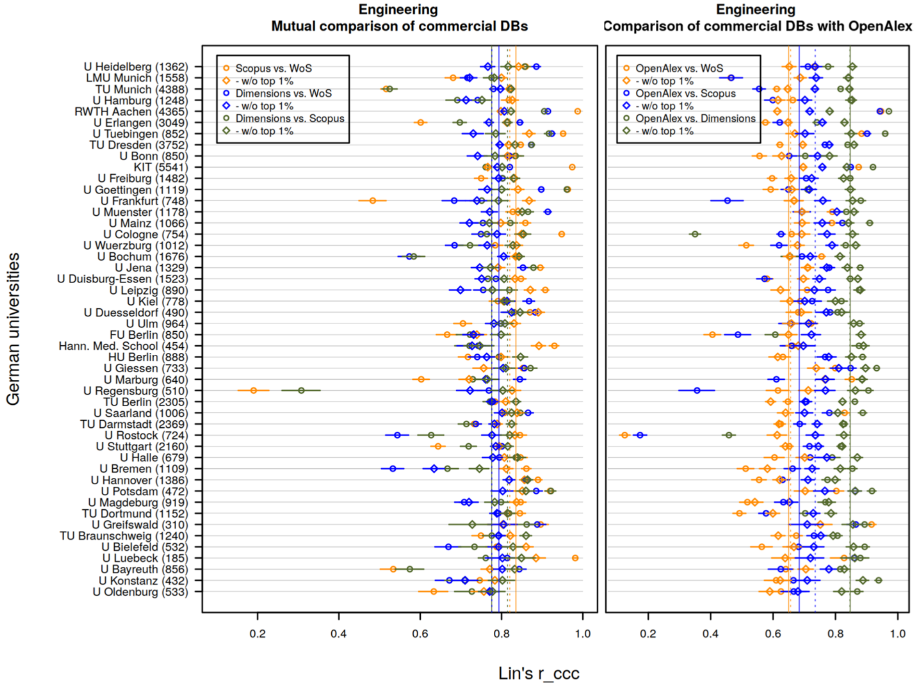

Fig. 18 Lin's r_ccc with confidence intervals for 48 universities (ordered by publication output) and the OECD subject area Engineering in mutual comparisons of the three commercial databases on the left and of OpenAlex with the other ones on the right–considering either all documents or without the top 1% papers in each database. Vertical lines indicate the median over all universities with all documents (solid) and without the top 1% papers (dashed line), respectively. Numbers in brackets indicate the average numbers of publications involved in all six database comparisons

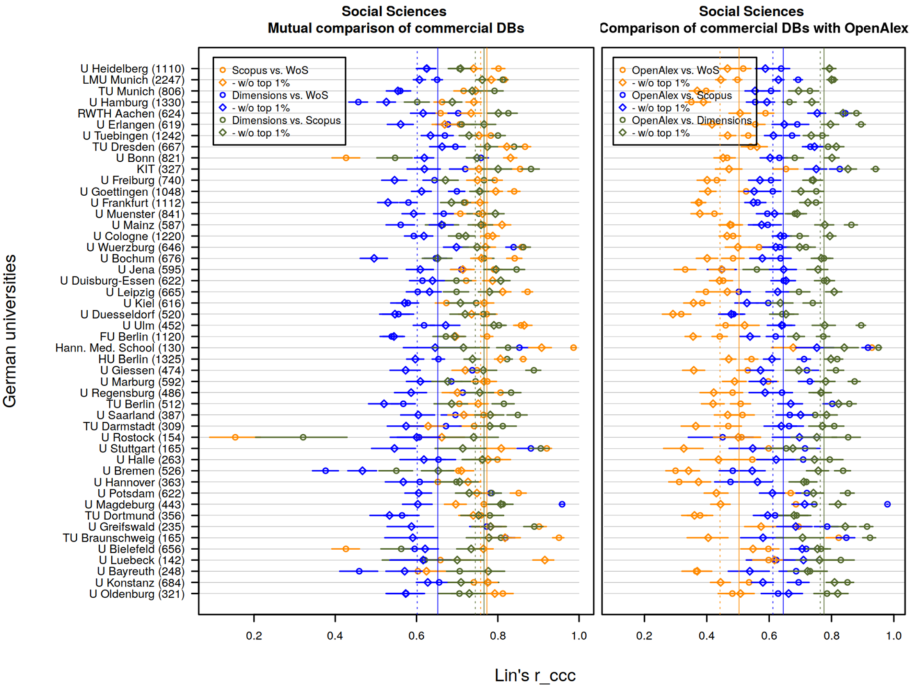

Fig. 19 Lin's r_ccc with confidence intervals for 48 universities (ordered by publication output) and the OECD subject area Social Sciences in mutual comparisons of the three commercial databases on the left and of OpenAlex with the other ones on the right–considering either all documents or without the top 1% papers in each database. Vertical lines indicate the median over all universities with all documents (solid) and without the top 1% papers (dashed line), respectively. Numbers in brackets indicate the average numbers of publications involved in all six database comparisons

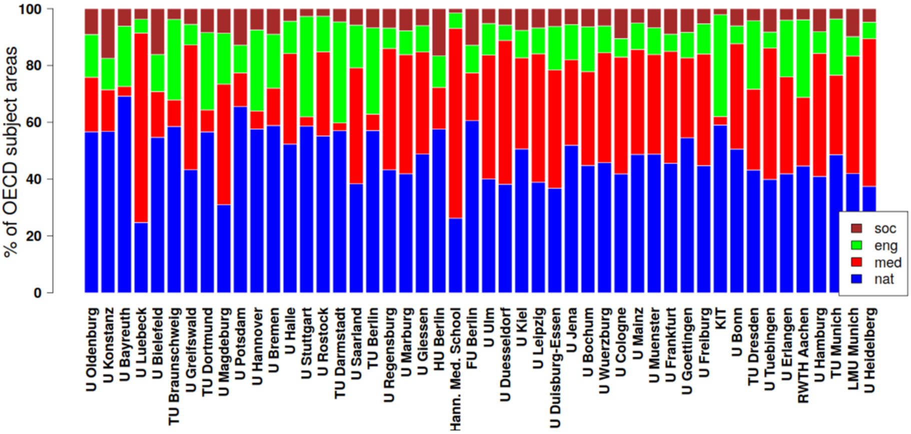

Fig. 20 Shares of the four OECD subject areas (soc = Social Sciences, eng = Engineering, med = Medicine, and nat = Natural Sciences) in the publications of the 48 German universities ordered by publication output

Table 10 Spearman's r_s (below the diagonal) with respect to NCS values from four databases and numbers of publications considered (above the diagonal) after removing the top 1% NCS values of each database

Database	WoS	Scopus	Dimensions	OpenAlex
WoS	*1*	329,872	312,645	305,466
Scopus	0.929	*1*	312,406	305,263
Dimensions	0.880	0.882	*1*	291,037
OpenAlex	0.877	0.880	0.909	*1*

Table 11 Differences in MNCS rank for the 48 German universities in the four databases ordered by their rank in OpenAlex

University	WoS	Scopus	Dimensions	OpenAlex
U Heidelberg	1	1	1	1
U Duisburg-Essen	7	2	2	2
U Bonn	2	3	6	3
LMU Munich	4	8	5	4
U Kiel	8	4	9	5
U Ulm	9	5	7	6
U Goettingen	6	9	15	7
TU Munich	5	6	12	8
U Wuerzburg	3	7	4	9
U Hamburg	10	12	11	10
Hann. Med School	20	18	8	11
U Luebeck	11	11	3	12
U Mainz	13	17	19	13
U Freiburg	14	14	18	14
U Cologne	18	16	10	15
U Marburg	31	20	22	16
U Duesseldorf	17	21	16	17
U Frankfurt	16	13	14	18
U Tuebingen	19	19	20	19
HU Berlin	12	15	13	20
TU Dresden	15	10	17	21
U Muenster	21	24	25	22
U Erlangen	24	23	24	23
U Regensburg	23	28	23	24
U Leipzig	36	32	21	25
RWTH Aachen	25	25	29	26
U Jena	30	26	27	27
U Magdeburg	42	38	30	28
U Bielefeld	29	22	33	29
U Giessen	32	27	28	30
TU Berlin	34	30	35	31
TU Dortmund	28	34	37	32
FU Berlin	37	40	36	33

Table 11 (continued)

University	WoS	Scopus	Dimensions	OpenAlex
U Konstanz	33	33	26	34
U Bochum	38	36	32	35
U Greifswald	35	35	31	36
U Hannover	22	37	41	37
U Potsdam	27	31	34	38
U Halle	40	41	39	39
KIT	26	29	42	40
U Bremen	43	43	43	41
U Rostock	44	47	44	42
U Oldenburg	41	42	40	43
U Saarland	48	46	38	44
U Stuttgart	47	45	47	45
U Bayreuth	39	39	45	46
TU Darmstadt	45	44	46	47
TU Braunschweig	46	48	48	48

Acknowledgements The WoS and Scopus data were provided by the German “Kompetenznetzwerk Bibliometrie” (KB), which was funded by the BMBF via grant 16WIK2101A). We thank Digital Science for providing us with Dimensions data for temporary storage in an in-house PostgreSQL database for use in this research. We thank OurResearch for preserving and continually building upon the MAG corpus and providing the OpenAlex data freely for all purposes.

The present paper is an extended version of an STI 2023 conference paper and available as a preprint (Scheidsteger, Haunschild, & Bornmann, 2023). It has been augmented by the analyses concerning the OECD subject areas as well as a more detailed analysis of the universities’ MNCS values.

Author contributions T.S.: Conceptualization, Data curation, Formal Analysis, Methodology, Visualization, Writing–original draft, Writing–review & editing. R.H.: Conceptualization, Methodology, Visualization, Writing–review & editing. L.B.: Conceptualization, Methodology, Writing–review & editing.

Funding Open Access funding enabled and organized by Projekt DEAL. The KB database with affiliation information for German research institutions has been sponsored by the German Federal Ministry of Education and Research (BMBF) under the funding code “16WIK2101A”.

Declarations

Competing interests RH serves as an Associate Editor for Scientometrics. LB and RH are members of the Distinguished Reviewers Board of Scientometrics.

References

Agha, R. A., Fowler, A. J., Saeta, A., Barai, I., Rajmohan, S., Orgill, D. P., & Rosin, D. (2016). The SCARE statement: Consensus-based surgical case report guidelines. *International Journal of Surgery, 34*, 180–186. https://doi.org/10.1016/j.ijsu.2016.08.014

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*(1), 377–386. https://doi.org/10.1162/qss_a_00019

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies, 1*(1), 363–376. https://doi.org/10.1162/qss_a_00018

Bornmann, L., & Williams, R. (2020). An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics, 124*(2), 1457–1478. https://doi.org/10.1007/s11192-020-03512-7

Céspedes, L., Kozlowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., & Larivière, V. (2025). Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24979

Clarivate. (2025). Emerging sources citation index (ESCI), 2025, from http://info.clarivate.com/ESCI

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Edwards, N. (2015). What to do about failing NHS organisations. *BMJ, 351*, h6972. https://doi.org/10.1136/bmj.h6972

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., & Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature, 522*(7555), 207–211. https://doi.org/10.1038/nature14317

Haunschild, R., & Bornmann, L. (2016). The proposal of using scaling for calculating field-normalized citation scores. *El Profesional De La Información, 25*(1), 1699–2407. https://doi.org/10.3145/epi.2016.ene.02

Haunschild, R., Schier, H., & Bornmann, L. (2016). Proposal of a minimum constraint for indicators based on means or averages. *Journal of Informetrics, 10*(2), 485–486. https://doi.org/10.1016/j.joi.2016.03.003

Haunschild, R., Daniels, A.D., & Bornmann, L. (2022). Scores of a specific field-normalized indicator calculated with different approaches of field-categorization: Are the scores different or similar? *Journal of Informetrics*, *16*(1), 101241. https://doi.org/10.1016/j.joi.2021.101241

Haunschild, R., & Bornmann, L. (2022). Relevance of document types in the scores' calculation of a specific field-normalized indicator: Are the scores strongly dependent on or nearly independent of the document type handling? *Scientometrics*, *127*(8), 4419–4438. https://doi.org/10.1007/s11192-022-04446-y

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies, 1*(1), 387–395. https://doi.org/10.1162/qss_a_00020

Huang, C.-K.K., Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies, 1*(2), 445–478. https://doi.org/10.1162/qss_a_00031

Kleinstaeuber, M., Witthoeft, M., Steffanowski, A., van Marwijk, H., Hiller, W., & Lambert, M. J. (2014). Pharmacological interventions for somatoform disorders in adults. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD010628.pub2

Koch, R., & Sporl, E. (2007). Statistical methods for comparison of two measuring procedures and for calibration: Analysis of concordance, correlation and regression in the case of measuring intraocular pressure. *Klinische Monatsblatter fur Augenheilkunde, 224*(1), 52–57. https://doi.org/10.1055/s-2006-927278

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*(1), 255–268. https://doi.org/10.2307/2532051

Lin, L. I. (2000). A note on the concordance correlation coefficient. *Biometrics, 56*(1), 324–325. https://doi.org/10.1111/j.0006-341X.2000.00324.x

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: Overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry, 28*(2), 6. https://doi.org/10.11919/j.issn.1002-0829.216045

Lundberg, J. (2007). Lifting the crown - citation *z*-score. *Journal of Informetrics, 1*(2), 145–154. https://doi.org/10.1016/j.joi.2006.09.007

Meho, L. I., & Sugimoto, C. R. (2009). Assessing the scholarly impact of information studies: A tale of two citation databases—scopus and web of science. *Journal of the American Society for Information Science and Technology, 60*(12), 2499–2508. https://doi.org/10.1002/asi.21165

OpenAlex. (2025). How does OpenAlex work? Retrieved 24 Feb 2025, from https://help.openalex.org/hc/en-us/articles/28932712154391-How-does-OpenAlex-work

Orduña-Malea, E., & Delgado-López-Cózar, E. (2018). Dimensions: Re-discovering the ecosystem of scientific information. *Profesional De La Información, 27*(2), 420–431. https://doi.org/10.3145/epi.2018.mar.21

Ortega, J. L., & Delgado-Quirós, L. (2024). The indexation of retracted literature in seven principal scholarly databases: A coverage comparison of dimensions, openalex, pubmed, scilit, scopus, the lens and web of science. *Scientometrics, 129*(7), 3769–3785. https://doi.org/10.1007/s11192-024-05034-y

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. Paper presented at the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain. doi: https://doi.org/10.5281/zenodo.6936227.

Priem, J., & Piwowar, H. (2022). Openalex-concept-tagging. Retrieved June 27, 2022, from https://github.com/ourresearch/openalex-concept-tagging

Purnell, P. J. (2022). The prevalence and impact of university affiliation discrepancies between four bibliographic databases—Scopus, web of science, dimensions, and microsoft academic. *Quantitative Science Studies*, *3*(1), 99–121. https://doi.org/10.1162/qss_a_00175

Rehn, C., Kronman, U., & Wadskog, D. (2007). *Bibliometric indicators – definitions and usage at Karolinska Institutet*. Karolinska Institutet University Library.

Rychik, J. (2015). Reply. *Ultrasound in Obstetrics & Gynecology, 46*(6), 746–747. https://doi.org/10.1002/uog.15748

Scheidsteger, T., Haunschild, R., Hug, S., & Bornmann, L. (2018). *The concordance of field-normalized scores based on web of science and microsoft academic data: A case study in computer sciences.* Paper presented at the 23th International Conference on Science, Technology and Innovation Indicators (STI 2018), Leiden. The Netherlands. https://hdl.handle.net/1887/65358.

Scheidsteger, T., Haunschild, R., & Bornmann, L. (2023). *How similar are field-normalized scores from different free or commercial databases calculated for large German universities?* Paper presented at the 27th International Conference on science, technology and innovation indicators (STI 2023), Leiden, The Netherlands. https://doi.org/10.55835/6441118c643beb0d90fc543f (URL of last version: https://dapp.orvium.io/deposits/64f83f76269bc6dc8000269c/view).

Scheidsteger, T., & Haunschild, R. (2023). Which of the metadata with relevance for bibliometrics are the same and which are different when switching from microsoft academic graph to openAlex? *Profesional De La Información*. https://doi.org/10.3145/epi.2023.mar.09

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J. P., & Wang, K. (2015). *An Overview of Microsoft Academic Service (MAS) and Applications.* Paper presented at the 24th International conference on world wide web (WWW '15 Companion), Florence, Italy. https://doi.org/10.1145/2740908.2742839.

Stahlschmidt, S., & Stephen, D. (2019). *Varying resonance chambers: A comparison of citation-based valuations of duplicated publications in web of science and scopus.* Paper presented at the 17th International Conference of the International Society for Scientometrics and informetrics (ISSI 2019).

Stahlschmidt, S., & Stephen, D. (2022). From indexation policies through citation networks to normalized citation impacts: web of science, scopus, and dimensions as varying resonance chambers. *Scientometrics, 127*(5), 2413–2431. https://doi.org/10.1007/s11192-022-04309-6

Thelwall, M., & Jiang, X. (2025). Is OpenAlex suitable for research quality evaluation and which citation indicator is best? https://doi.org/10.48550/arXiv.2502.18427.

Torres-Salinas, D., Lopez-Cózar, E. D., & Jiménez-Contreras, E. (2009). Ranking of departments and researchers within a university using two different databases: Web of science versus scopus. *Scientometrics, 80*(3), 761–774. https://doi.org/10.1007/s11192-008-2113-9

Turgel, I. D., & Chernova, O. A. (2024). Open science alternatives to scopus and the web of science: A case study in regional resilience. *Publications, 12*(4), 43. https://doi.org/10.3390/publications12040043

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic. *Quantitative Science Studies, 2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., & Wouters, P. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419–2432. https://doi.org/10.1002/asi.22708

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37–47. https://doi.org/10.1016/j.joi.2010.08.001

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies, 1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

van Wijk, E., & Costas-Comesaña, R. (2012). *Bibliometric study of FWF Austrian science fund 2001-2010/11*. Leiden, the Netherlands: Center for science and technology studies (CWTS). URL: https://zenodo.org/record/17851.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.