

---

## **Quality assurance at the macro level: Comparing the current and previous Scopus snapshots**

---

Andrey Lovakov, Dimity Stephen, Stephan Stahlschmidt and Paul Donner

*October 2025*

**Editor:**

German Centre for Higher Education Research and Science Studies (DZHW) GmbH

Lange Laube 12 | 30159 Hannover | Germany | [info@dzhw.eu](mailto:info@dzhw.eu) | [www.dzhw.eu](http://www.dzhw.eu)

POB 2920 | 30029 Hannover | Germany

phone: +49 511 450670-0 | fax: +49 511 450670-960

**Chairman of the Supervisory Board:**

Ministerialdirigent Peter Greisler

**Scientific Director:**

Dr. Marcus Beiner

**Managing Director:**

Regina Oelfke

**Registration Court:**

Amtsgericht Hannover | HRB 6489

VAT No.: DE291239300

October 2025

## Contents

<b>Motivation</b>	<b>1</b>
Set of indicators . . . . .	1
Set of entities . . . . .	2
Methodological details . . . . .	2
<b>Analysis</b>	<b>3</b>
Publication counts: Total, selected countries, German sectors, and Subject Areas . . . . .	3
Journals: Total indexed and the number added or removed . . . . .	6
Excellence Rates: Selected countries and German sectors . . . . .	7
Citations: Mean 3-year citations of articles and reviews by discipline . . . . .	9
Uncited articles and reviews: Percent by selected countries and German sectors . . . . .	13
Disciplines: Changes in discipline classification . . . . .	15
Disciplines: Changes in articles and reviews by discipline . . . . .	16
Disciplines: Percentage of publications not assigned to a discipline . . . . .	17
Metadata: Changes in pubyear, doctype, pubtype and items removed . . . . .	17
Metadata: Missing metadata variables . . . . .	20
Institution and country data: Number of articles and reviews with missing data . . . . .	21
German institutions: German publications missing from KB institution coding . . . . .	22
German institutions: Changes in whole counts of articles and reviews . . . . .	22
Authors: Median number of authors by Research Area and discipline . . . . .	29
Source items: Percentage by Research Area and discipline . . . . .	30

## Motivation

The aim of the report is to identify any potential changes in data between or within database versions that may indicate quality issues. To do so it offers:

- a visual comparison
- between time-series over the last 10 years
- stemming from the current and previous KB database snapshots
- on several key indicators
- for national, sectoral and institutional entities.

The DZHW already conducts quality assurance testing at the micro-level for the KB's bibliometric databases before the tables enter the production environment. This testing is invaluable to ensuring tables and variables contain the expected content. This report supplements the current micro-level approach by examining changes in key variables between the latest two iterations of the databases at the macro-level of institutions, sectors, countries, and disciplines.

This report is not an exhaustive analysis of the databases' content, nor does it investigate any anomalies identified in the databases. However, this report probes the core variables fundamental to typical bibliometric analyses, serves as an overview of the current state of the databases, and highlights changes that may indicate issues with data quality that warrant further investigation to understand or rectify. Changes may arise through several means. For instance, the database provider may add or remove journals from indices, change the discipline classification, or change how the classification is applied. The KB may identify new or decommissioned institutions, which can affect publication output for particular disciplines, or countries may implement policies regarding publication practices that can exert a substantial influence on the content published over time. This report aims to provide users of the KB databases with an overview of any potential changes soon after the databases enter the production environment, so that these factors may be considered in analyses.

## Set of indicators

The indicators included in the report reflect the core variables in the database that are fundamental to key bibliometric analyses and indicators. We provide context to the selection of variables and what information can be determined from their examination in each of the following sections.

We make two sets of comparisons in this report. For indicators where it is important to consider trends over time, such as whole publication counts, we compare the databases for the 10 years up to the year for which both have complete data. For example, the latest common year with complete data for the scp\_b\_202404 and scp\_b\_202504 databases is 2023, as data for the absolute latest year in each database are incomplete. Similarly, where citation-based indicators are used, we present the time-series up to the latest common year with complete citation data, which is 2021 for the scp\_b\_202404 and scp\_b\_202504 databases. This comparison highlights any differences in trends between the databases for the most recent decade.

For other indicators, it is most useful to compare changes between just the most recent years of complete data in each database. For instance, we compare the number of publications per discipline in 2023 from the scp\_b\_202404 database against 2024 in the scp\_b\_202504 database. Changes between the years are expected given we are comparing two different sets of publications. However, this comparison can also provide insight into structural changes between the database iterations, such as the addition or removal of journals from indices, which may influence indicators

at the macro-level. Such comparisons are also helpful in identifying new or removed institutions or discipline categories. Further, although users will likely use the latest database to produce a complete time-series for new analyses, it is important to understand how additional years of a time-series might differ to existing time-series presented in publications and reports.

## Set of entities

We have chosen to compare the databases at the national, sectoral, and institutional levels. The countries chosen are based on those most commonly examined by the DZHW as countries against which it is useful and informative to compare Germany. We also examine the key German sectors: Universities (Uni), Fachhochschulen (FH), Max Planck Gesellschaft (MPG), Fraunhofer Gesellschaft (FHG), Helmholtz Gemeinschaft (HGF), Leibniz Gemeinschaft (WGL), the business sector (Econ), non-university hospitals (Clinic), and combined Ressortforschung-Bund and Ressortforschung-Länder (Gov). The remaining smaller sectors, such as research associations, clubs, and international and foreign organisations are grouped into an “other” category. Individual German institutions are also examined via the KB’s institutional coding for Germany. However, as there are a large number of institutions, we present data only for institutions that have shown substantial changes in the indicator of interest.

## Methodological details

We focus primarily on articles and reviews published in journals as these are the most common documents used in bibliometric analyses. As previously noted, we supply a shortened time-series for citation-based indicators to allow for a 3-year citation window. Wang (2013)<sup>1</sup> determined that at least 3 years is required for publications to reach their maximum number of citations per year, after which point the number of citations are likely representative of the publication’s long-term impact. As such, citation-based indicators include all citations received within the publication year and the subsequent two years.

Whole counting is used throughout the report. Although it is most common to use fractional counting, analysing variables using whole counts will still reveal potential changes in the variables.

Data for disciplines are presented based on either the All Science Journal Classification (ASJC) or the Research Area (RA) classification. The ASJC is a fine-grained classification that allows changes in specific disciplines to be analysed. However, given it contains over 250 categories, it is sometimes useful to use a higher level of aggregation to present an overview of the disciplines. As such, we also present some data on the RA classification. The RA consists of five broad groups: Health Sciences, Life Sciences, Physical Sciences, Social Sciences and Humanities, and Multidisciplinary. These groups are mapped from the ASJC disciplines based on a concordance supplied by Elsevier.

This report is automated. Consequently, blank tables may appear in this report, but they are nonetheless informative about the indicator under examination.

<sup>1</sup>Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851-872. DOI: 10.1007/s11192-012-0775-9

## Analysis

### Publication counts: Total, selected countries, German sectors, and Subject Areas

The count of items produced by selected entities is the most fundamental bibliometric indicator. Given publication counts form the basis of many indicators, understanding the time-series trend within and between databases can inform expectations about potential changes that may arise in other indicators. In Figure 1 we show the total number of documents of different types indexed in each database version, followed by the whole counts of articles and reviews published by selected countries and German sectors over the last 10 years in Figures 2 and 3. In Figure 4 we show the distribution of publications by RA.

Changes in publication counts over time may reflect changes made by countries, the database provider, and/or administrative decisions. For example, it is expected that the scp\_b\_202504 database contains a greater number of publications for the most recent years than the scp\_b\_202404 database due to the continued indexing of items by Elsevier past the annual point in April at which the data is cut to create the KB databases.

Increases in publications over time also result from both the continued growth of the national science systems and Scopus' ongoing indexation over time. Sharp increases for a particular country may represent an actual increase in the number of a country's articles published in Scopus-indexed journals, such as due to policy decisions, or reflect the recent indexing of region-, country-, or discipline-specific journals. Decreases may reflect the de-indexation of journals in which an entity commonly publishes or the stagnation of a sector, such as due to funding or policy decisions or the de-commissioning of an institution. Substantial deviations between databases or decreases in the current database in recent years may warrant investigation.

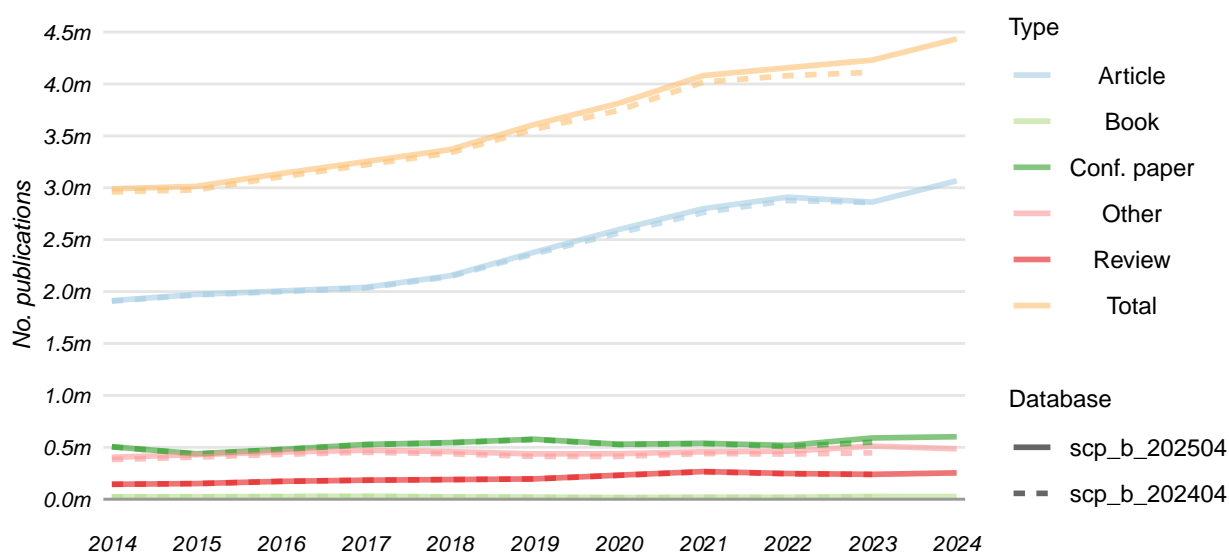


Figure 1: Number of documents in each database over time by type.

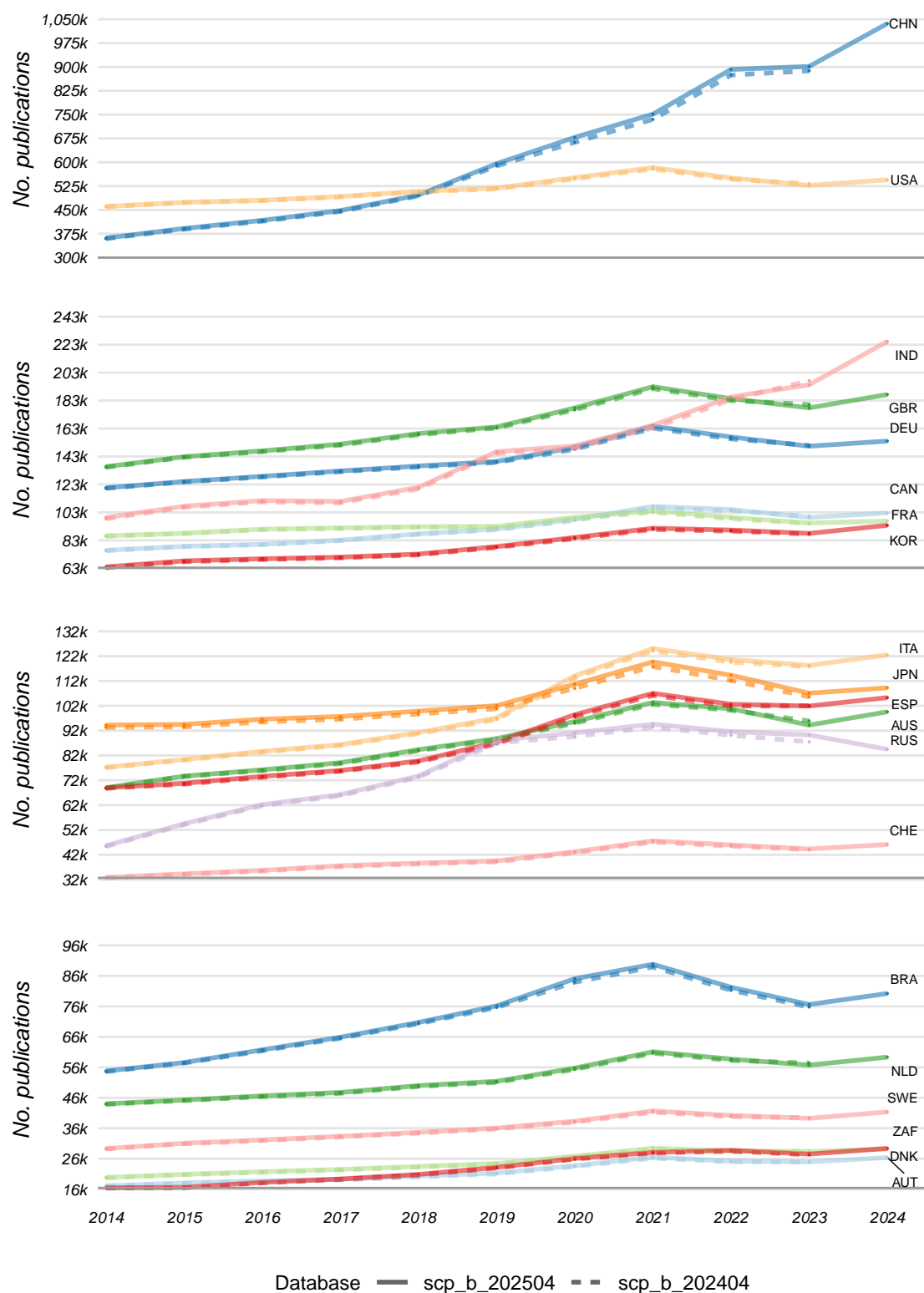


Figure 2: Whole counts of articles and reviews by country and database over time. Please note the panels' different axes.

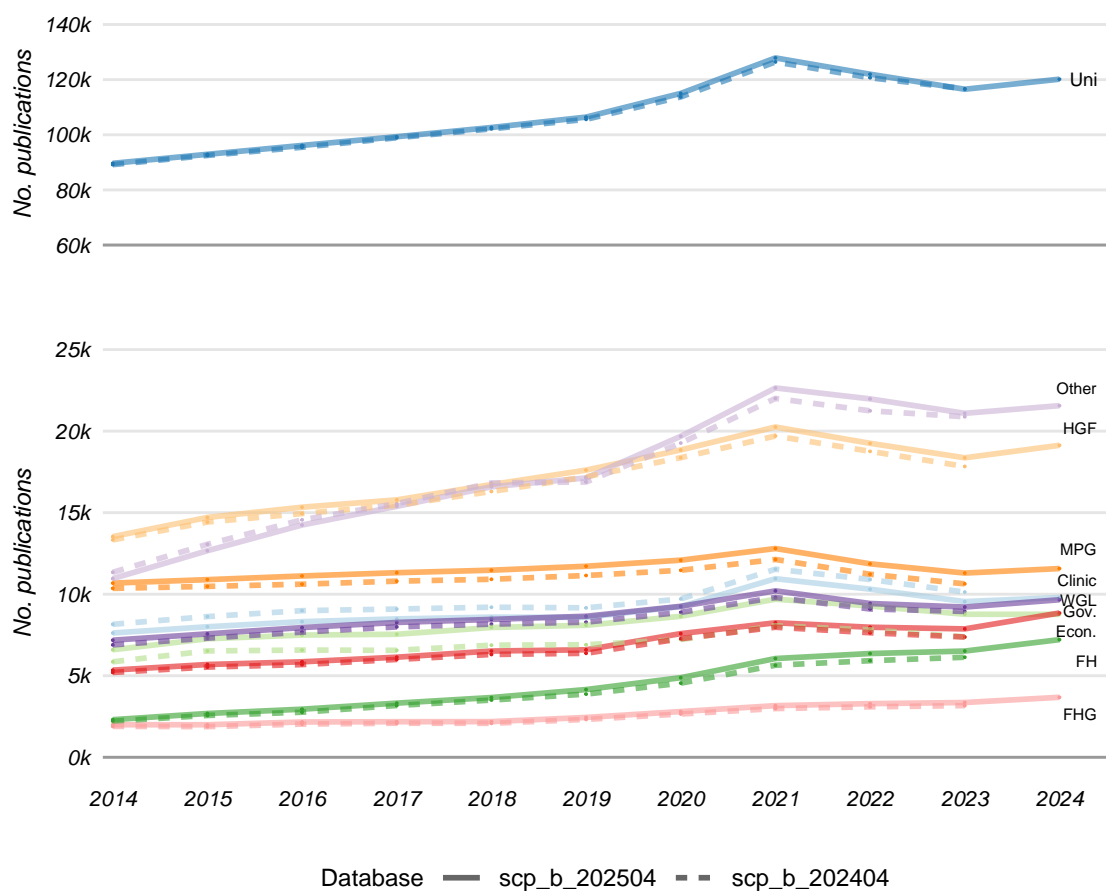


Figure 3: Whole counts of articles and reviews by German sector and database over time. Please note the panels' different axes.

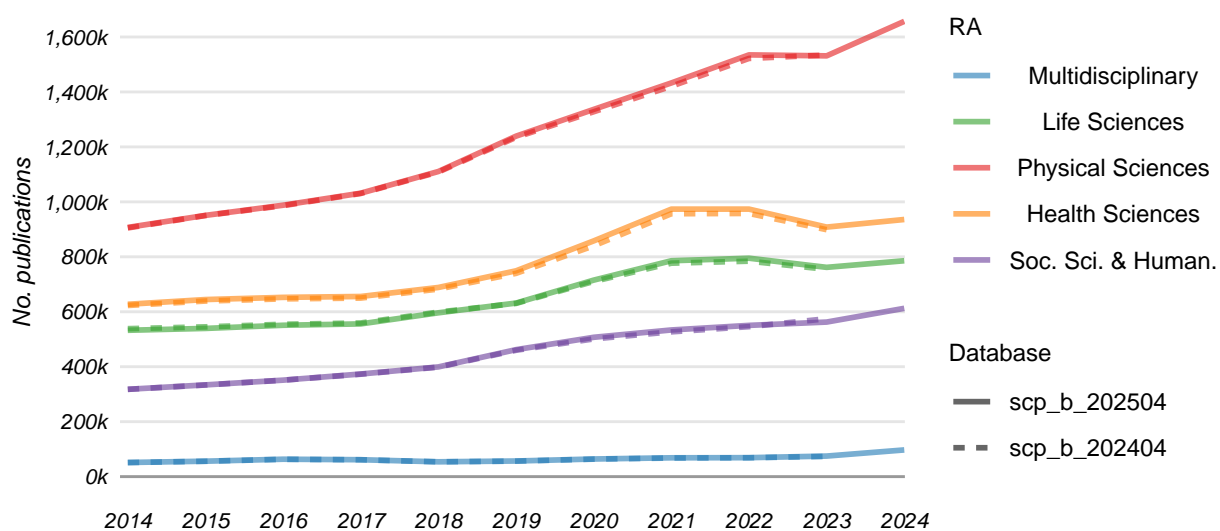


Figure 4: Whole counts of articles and reviews by RA and database over time.

## Journals: Total indexed and the number added or removed

The journals indexed constitute the foundation of the database. Year to year changes in the journals indexed reflect the database provider's curation procedures to introduce new content and remove content no longer meeting indexation criteria. The amount of and changes in content indexed can influence bibliometric indicators, particularly if changes are concentrated in specific disciplines. Figure 5 shows the total number of journals in each database over time, while Figure 6 shows the number of journals added and removed in each RA.

Changes in the journals indexed were identified by matching the titles of all journals indexed in 2023 in the scp\_b\_202404 database to those with 2024 content in the scp\_b\_202504 database. Titles were used as all journals have titles recorded, while some journals are missing ISSNs. Titles in scp\_b\_202404 but not in scp\_b\_202504 were considered removed, while titles in scp\_b\_202504 but not in scp\_b\_202404 were considered added. In total, 1923 journals were added and 1138 were removed. These data may include a small number of journals that changed titles. Some double-counting of journals between RAs may also occur when a journal maps to two or more RAs.

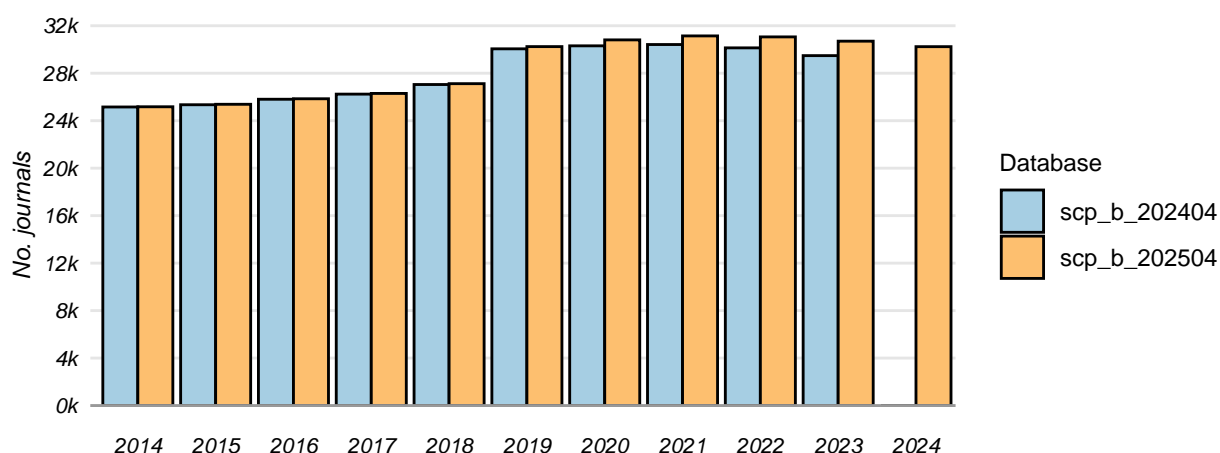


Figure 5: The number of journals indexed in each database over time.

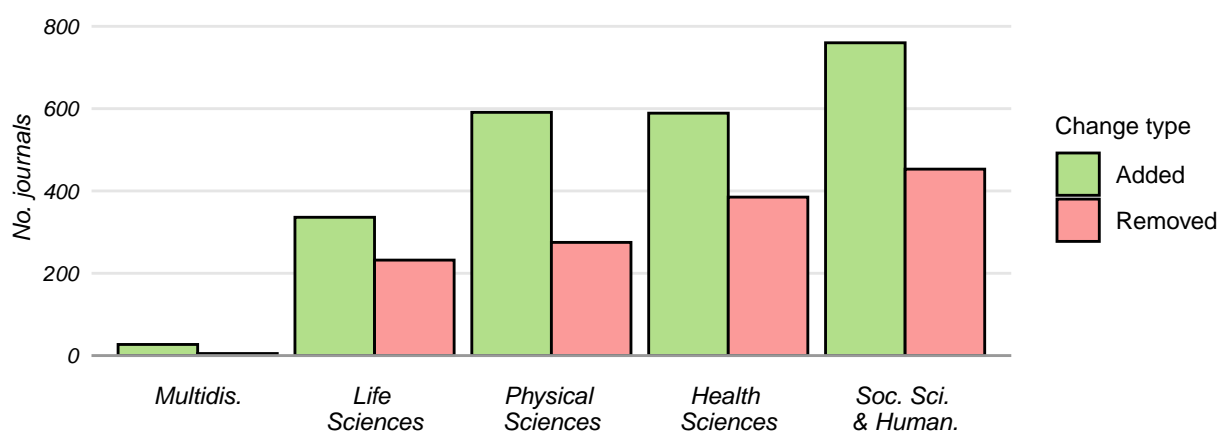


Figure 6: The number of journals added or removed between 2023 in scp\_b\_202404 and 2024 in scp\_b\_202504 by RA.

## Excellence Rates: Selected countries and German sectors

Excellence Rates (ER) identify the percentage of an entity's publications that are in the 10% most highly cited publications from each discipline and could be considered of excellent quality on this basis. ERs are a common indicator used to assess an entity's performance, with an ER exceeding the expected 10% threshold interpreted as better than expected performance. ERs for the most recent years from the two databases are presented for German sectors in Figure 7 and for countries in Figure 8. As with whole counts of publications, we would expect general agreement between the databases, particularly in the earlier years of the time-series, so substantial deviations may warrant further analysis.

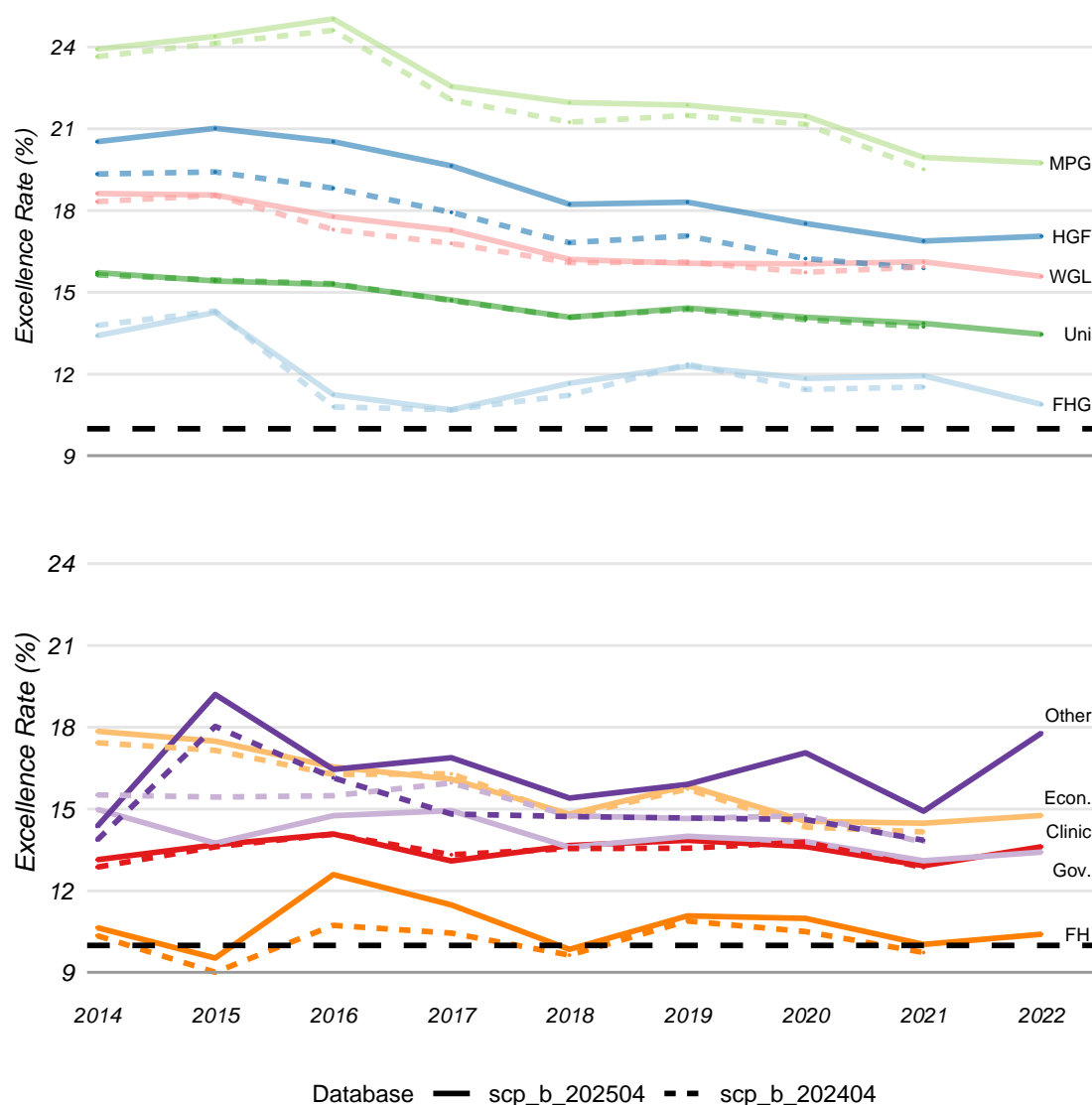


Figure 7: ERs, based on whole counts, by German sector and database over time. The black line is the expected 10% threshold.

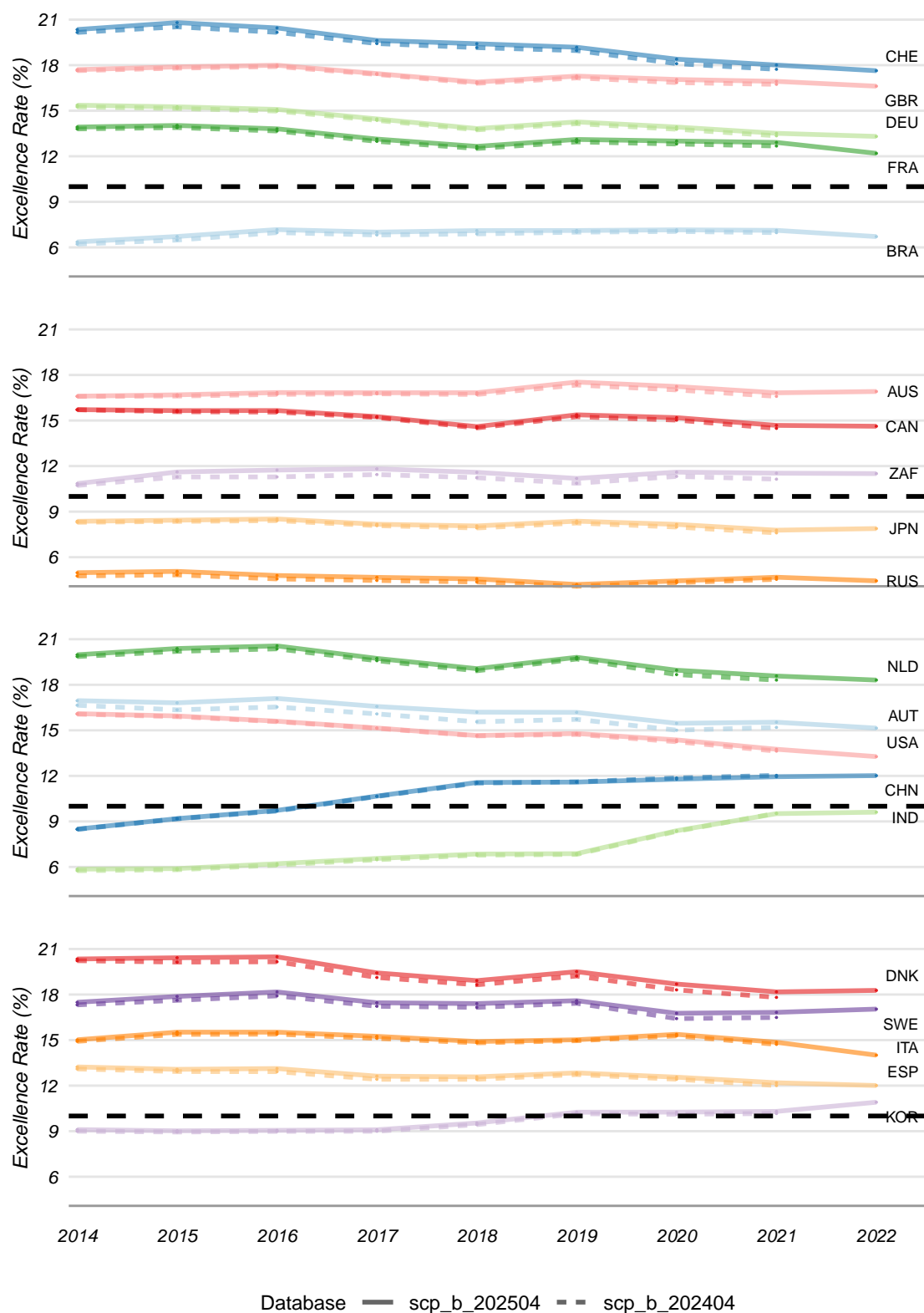


Figure 8: ERs, based on whole counts, by selected country and database over time. The black line is the expected 10% threshold.

## Citations: Mean 3-year citations of articles and reviews by discipline

The number of citations a publication could be expected to receive is dependent to an extent on its discipline. As such, we examine here the mean 3-year citations of articles and reviews by discipline. Mean 3-year citations (MC3) are the mean citations publications in each discipline accrued in the first 3 years after publication. We examine here in Figure 9 the last common year in both databases (top panels) to assess the retroactive effects stemming from changes made in the latest database, and the latest complete year in both databases (bottom panels) to assess potential structural changes and updates to the time-series. A greater deviation of disciplines from the central line indicates a greater degree of change in the mean citations of a discipline's items between years. The outlying disciplines from the bottom panels of Figure 9 are shown in Tables 1 and 2, along with disciplines where the previous threshold was zero. We use a threshold of a current MC3 of at least 1 for articles and 3 for reviews to remove disciplines with spurious changes due to low levels of citations.

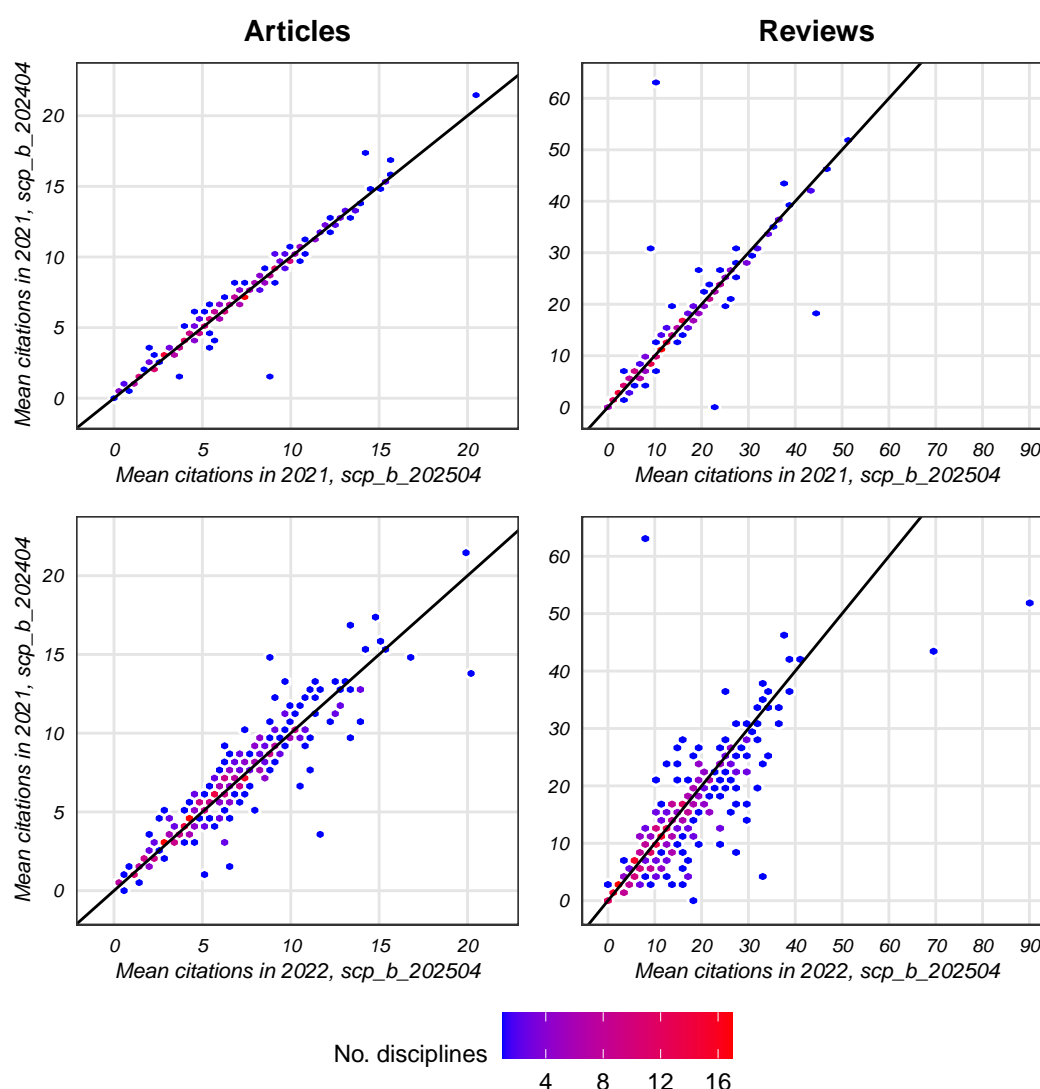


Figure 9: The MC3 for articles and reviews in each discipline between databases, where colour denotes the number of disciplines with this combination of citations.

Table 1: Articles: Disciplines with a current MC3 of at least 1, where the MC3 decreased by over 20% or increased by over 50% between 2021 in scp\_b\_202404 and 2022 in scp\_b\_202504, or the previous MC3 was 0.

Discipline	Previous MC3	Current MC3	No. currnt pubs.	Perc. diff.
Decision Sciences (misc.)	1.3	6.3	43	376.2
Colloid & Surface Chemistry	1.3	5.1	11	284.0
Materials Chemistry	3.8	11.6	106	205.7
Biochemistry, Genetics & Molecular Biology (misc.)	2.9	6.0	819	108.2
Health, Toxicology & Mutagenesis	3.3	6.2	322	86.0
Statistics, Probability & Uncertainty	0.8	1.4	54	75.9
Energy (misc.)	6.5	10.4	781	61.1
Mgmnt of Technology & Innovation	5.1	8.1	462	59.2
Media Technology	13.6	20.2	1013	48.4
Architecture	3.2	4.6	6324	44.9
Drug Guides	0.8	1.1	30	43.6
Filtration & Separation	7.7	11.0	4	42.7
Surfaces, Coatings & Films	4.2	5.9	1776	42.4
Fundamentals & Skills	1.5	2.2	288	41.2
Computer Graphics & Computer-Aided Design	9.8	13.5	793	37.5
Process Chemistry & Technology	4.7	6.4	1521	35.6
Automotive Engineering	10.5	14.1	7225	34.4
Optometry	4.7	6.2	42	32.1
Mgmnt, Monitoring, Policy & Law	6.1	4.3	414	-30.5
Dental Hygiene	2.6	1.8	90	-30.9
Respiratory Care	4.9	2.9	18	-39.6
Hematology	14.8	8.6	9423	-42.0
Nursing (misc.)	4.9	2.5	321	-48.4
Dentistry (misc.)	3.7	1.8	862	-50.3

Table 2: Reviews: Disciplines with a current MC3 of at least 3, where the MC3 decreased by over 20% or increased by over 60% between 2021 in scp\_b\_202404 and 2022 in scp\_b\_202504, or the previous MC3 was 0.

Discipline	Previous MC3	Current MC3	No. currnt pubs.	Perc. diff.
Decision Sciences (misc.)	0.0	18.4	18	Inf
Health, Toxicology & Mutagenesis	4.1	32.2	32	685.2
Mathematics (misc.)	2.2	16.3	31	648.1
Economics, Econometrics & Finance (misc.)	3.2	14.3	39	341.7
Mgmnt of Tech. & Innovation	3.9	17.2	19	339.4
Information Systems & Mgmnt	4.0	16.1	16	301.6
Logic	2.5	8.3	3	233.3
Mgmnt Science & Operations Research	3.8	12.3	13	222.8

Psychology (misc.)	8.7	27.7	83	218.8
Podiatry	2.2	6.7	22	206.7
Architecture	5.9	15.3	222	157.9
Industrial Relations	3.8	9.6	22	152.7
Communication	4.2	10.2	220	142.0
Finance	7.5	17.7	129	136.5
Public Administration	5.1	12.1	31	136.5
Space & Planetary Science	10.3	23.1	7	125.3
Conservation	3.0	6.2	88	103.4
Business, Mgmt & Accounting (misc.)	5.3	10.6	111	101.0
Computers in Earth Sciences	14.4	28.9	8	101.0
Media Tech.	5.7	11.4	9	99.8
Business, Mgmt & Accounting (all)	8.1	15.9	211	97.3
Anthropology	1.8	3.5	75	92.1
Development	6.8	13.1	107	91.3
Geophysics	10.3	19.7	171	90.8
Environmental Science (misc.)	12.3	23.3	301	89.0
Mgmt Information Systems	12.6	23.2	145	83.8
Accounting	7.8	14.3	115	82.4
Applied Mathematics	5.1	9.3	23	81.3
Life-span & Life-course Studies	10.5	18.7	7	78.2
Agricultural & Biological Sciences (misc.)	7.7	13.7	255	78.2
Energy (misc.)	17.0	29.8	120	75.5
Computer Vision & Pattern Recognition	51.6	89.2	19	72.8
Nuclear & High Energy Physics	15.1	26.0	132	71.9
Materials Chemistry	19.3	33.0	5	70.7
Economics & Econometrics	7.8	13.1	145	69.0
Earth-Surface Processes	16.1	27.0	56	67.4
Veterinary (misc.)	7.4	12.3	30	67.1
Gender Studies	2.6	4.2	201	65.8
Community & Home Care	3.8	6.4	109	65.7
Nuclear Energy & Engineering	6.2	10.2	95	64.7
Physiology (medical)	2.2	3.7	66	64.0
Mathematical Physics	5.7	9.2	38	61.6
Business & International Mgmt	9.5	15.4	349	61.1
Artificial Intelligence	24.4	19.4	151	-20.3
Health Professions (all)	5.2	4.1	7	-21.1
Occupational Therapy	4.0	3.2	77	-21.4
Critical Care Nursing	6.7	5.3	108	-21.4
Electrical & Electronic Engineering	23.5	18.4	321	-21.8
Microbiology (medical)	24.5	19.1	517	-21.8
Neuroscience (misc.)	16.0	12.4	359	-22.4
Medicine (all)	12.7	9.7	14166	-23.1

---

Epidemiology	16.8	12.7	805	-24.4
Dentistry (misc.)	6.6	4.9	99	-25.5
Physics & Astronomy (misc.)	25.9	19.2	78	-25.9
Building & Construction	25.3	18.7	185	-26.2
Acoustics & Ultrasonics	17.0	12.4	35	-26.9
Computational Mathematics	8.7	6.3	15	-27.3
Histology	21.1	15.0	48	-28.6
Process Chemistry & Tech.	8.1	5.7	112	-29.1
Biological Psychiatry	36.6	25.5	58	-30.2
Applied Microbiology & Biotechnology	14.8	10.3	158	-30.8
Radiation	11.5	7.4	214	-35.1
Geochemistry & Petrology	15.7	10.1	49	-35.3
Endocrinology	12.1	7.8	162	-35.6
Marketing	23.8	14.9	48	-37.6
Optometry	11.0	6.8	4	-38.6
Gerontology	12.5	7.2	104	-42.1
Sensory Systems	24.0	13.5	107	-43.7
Urban Studies	28.3	15.6	23	-44.8
Physical & Theoretical Chemistry	26.4	14.4	767	-45.4
Ocean Engineering	11.7	6.0	42	-48.9
Nursing (misc.)	7.6	3.6	52	-52.1
Numerical Analysis	20.7	9.3	27	-55.1
Computational Theory & Mathematics	62.7	7.6	9	-88.0

---

### Uncited articles and reviews: Percent by selected countries and German sectors

While ERs represent the most highly cited publications and mean citations tell us about what's average, the percentage of uncited publications can tell us about the entities at the tail end of the citation distribution. When examining uncited publications, we expect to see a decreasing trend in uncited publications over time. This occurs because citation counts are based on the items indexed in each database and so, as the database provider continues to index journals, the likelihood increases that any publication will have been cited by the indexed items. In particular, we would expect that the percentage of uncited publications in the last common year would be lower in the current database than the previous database, as data added in the latest iteration "complete" the incomplete last year of the previous database. An increase in uncited publications in the latest year may reflect processing issues that require investigation. We present in Figures 10 and 11 the percentage of articles and reviews per German sector and selected country that remained uncited 3 years after they were published.

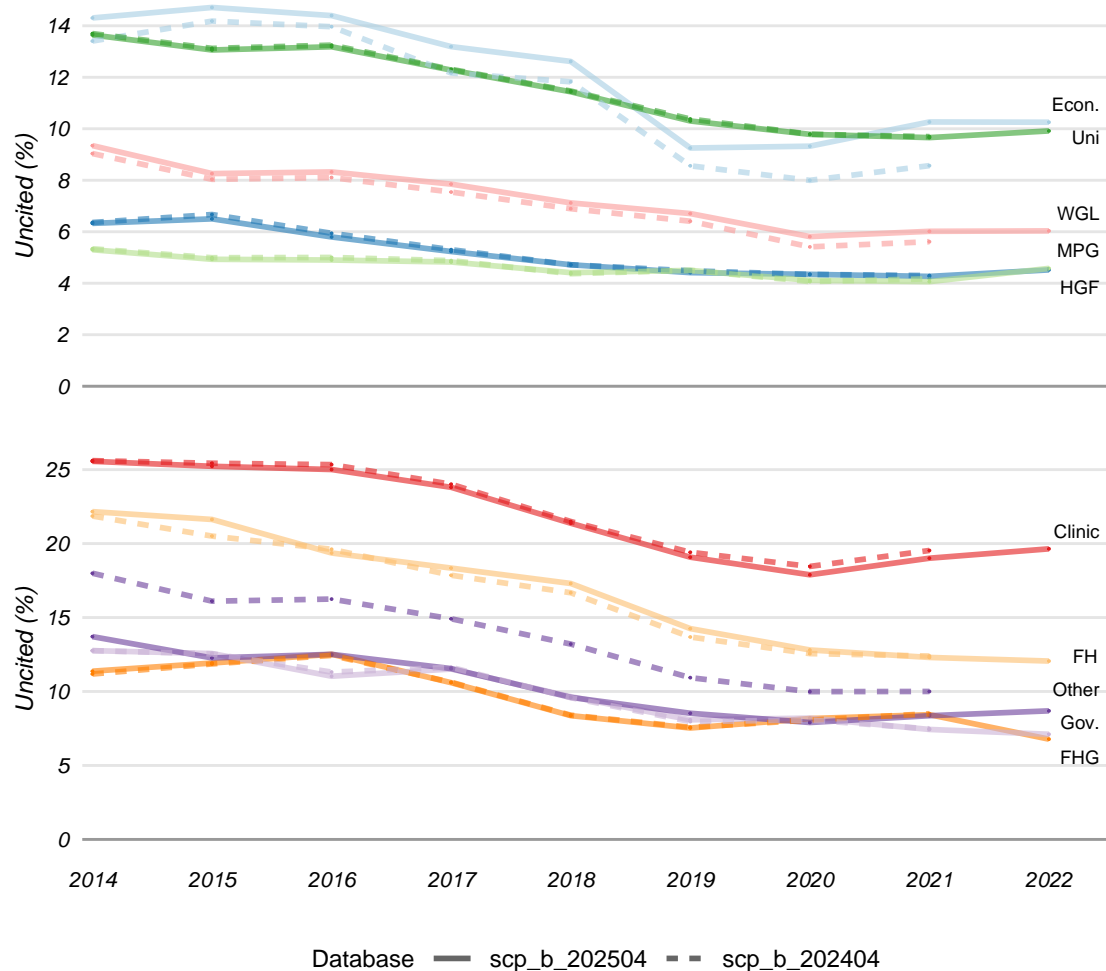


Figure 10: The percentage of uncited publications in each database over time by German sector.

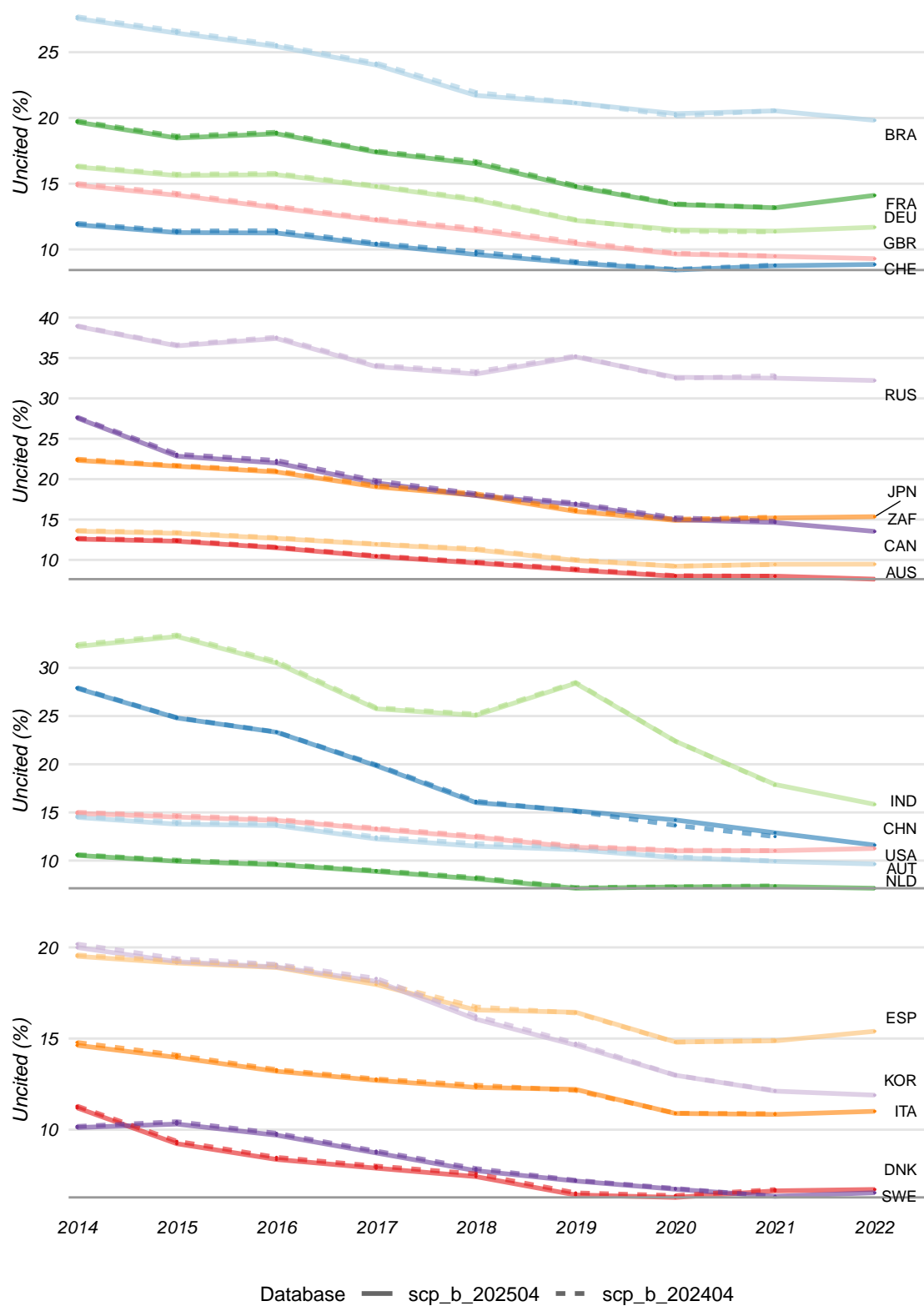


Figure 11: The percentage of uncited publications in each database over time by selected countries.

## Disciplines: Changes in discipline classification

This section shows in Table 3 any changes that have been made to Scopus' discipline classification, the ASJC. This could include splits, aggregations or removals of a discipline, or the inclusion of a new discipline to reflect new and emerging topics. We identify changes in the classification structure by comparing the number of articles and reviews attributed to each discipline in the latest years of each database and selecting those disciplines where the number was zero in one year but not in the other. Disciplines with no prior publications but some in the current year suggest the discipline may have been recently added, while the opposite suggests the discipline may have been removed or merged. Changes may also reflect changes in spelling or punctuation of the discipline name. Any changes should be checked with Elsevier's published classification structure.

Table 3: Changes in the ASJC discipline classification structure between the previous and current databases

Classification	Previous pubs	Current pubs
----------------	---------------	--------------

Figure 12 shows the number of publications assigned to specific disciplines identified to have changed in recent versions of the database.

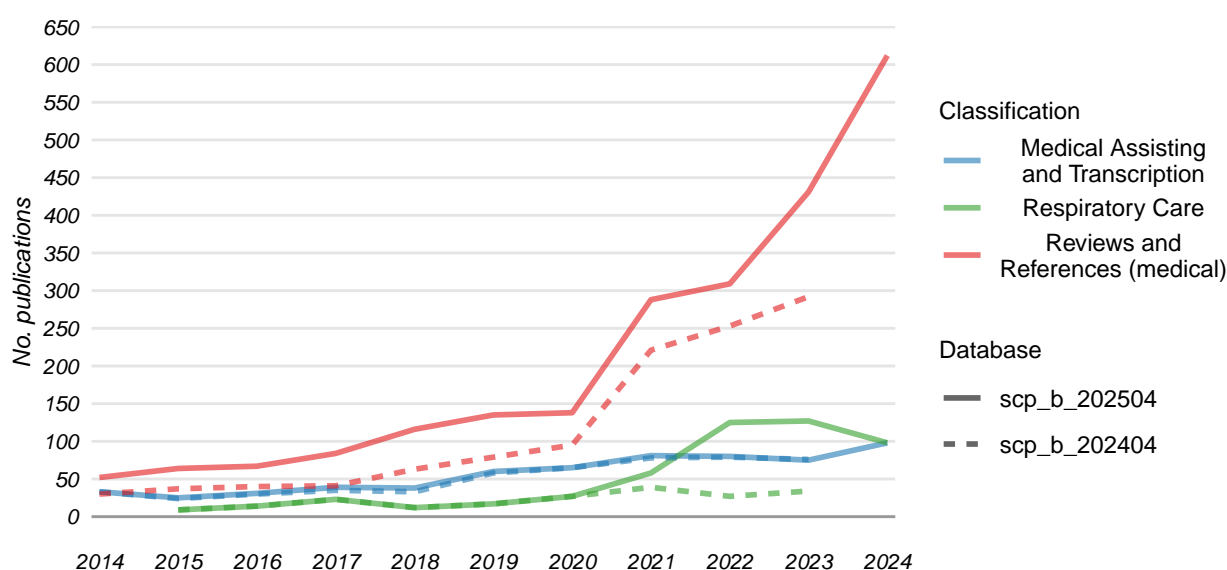


Figure 12: Time-series of ASJC disciplines previously observed to have changed.

## Disciplines: Changes in articles and reviews by discipline

This section identifies the disciplines that had a substantial change in the number of publications assigned to them between the latest years in each database. Changes in counts of publications per discipline may reflect changes in the journals indexed, the classification structure, and any potential processing issues. As such, any large changes shown here may be worth examining.

We show in Figure 13 the 40 disciplines with the highest percentage increases and decreases in publication counts between 2023 in scp\_b\_202404 and 2024 in scp\_b\_202504. The number shown next to each bar is the numerical change in publication counts. We have used whole counting. Disciplines previously identified as being new or removed have not been included here.

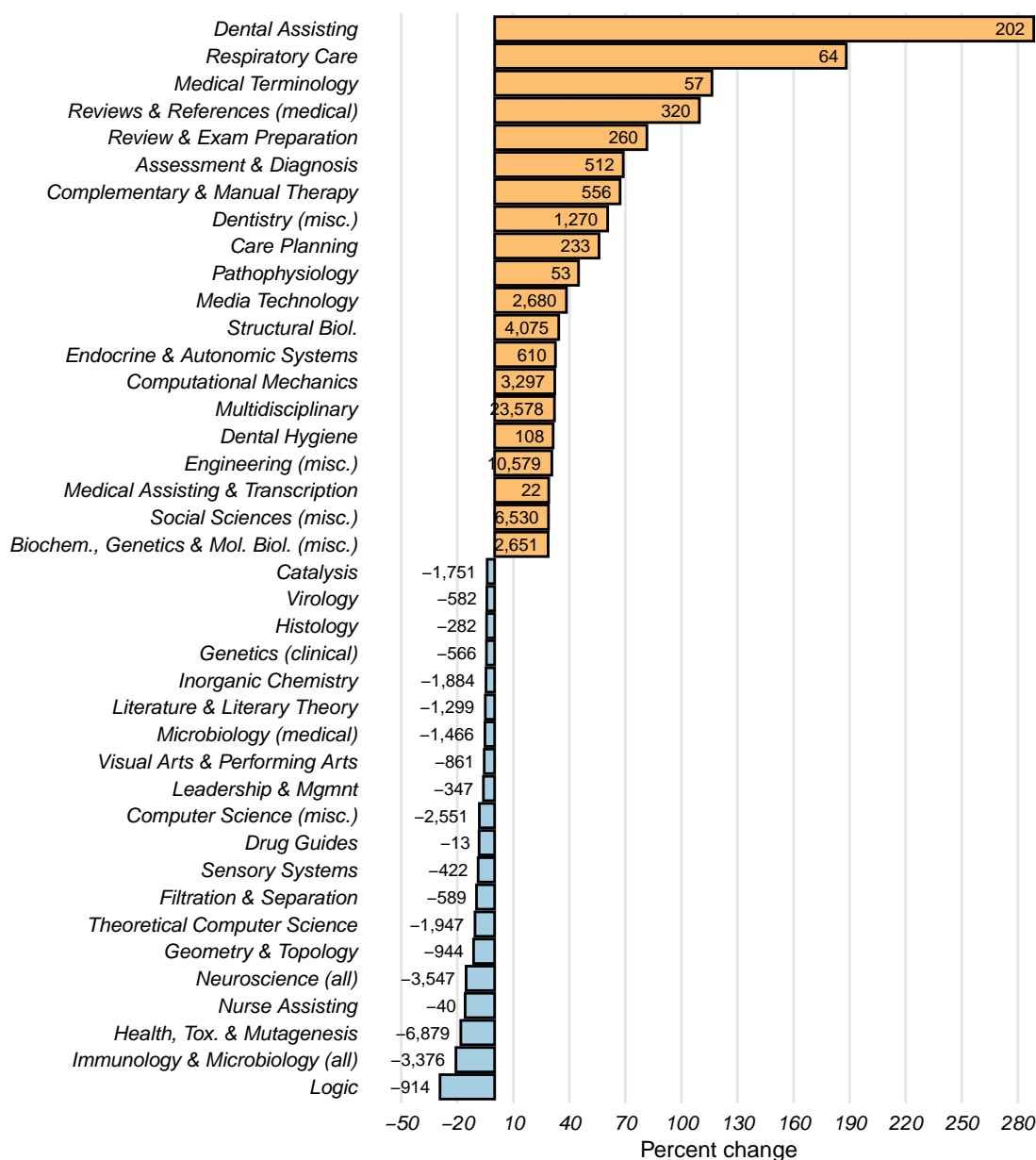


Figure 13: The 40 disciplines with the highest percentage change in publication counts between 2023 in scp\_b\_202404 and 2024 in scp\_b\_202504, with numerical difference in counts.

## Disciplines: Percentage of publications not assigned to a discipline

Figure 14 shows the percentage of publications in each database that were not assigned to a discipline over the previous 11 years. Complete assignment of publications to disciplines is important as citation-based indicators typically use field-normalisation to account for differences in citation practices between disciplines. As such, items missing discipline information are excluded from such analyses and so large percentages of, or large changes in, unclassified items should be investigated.

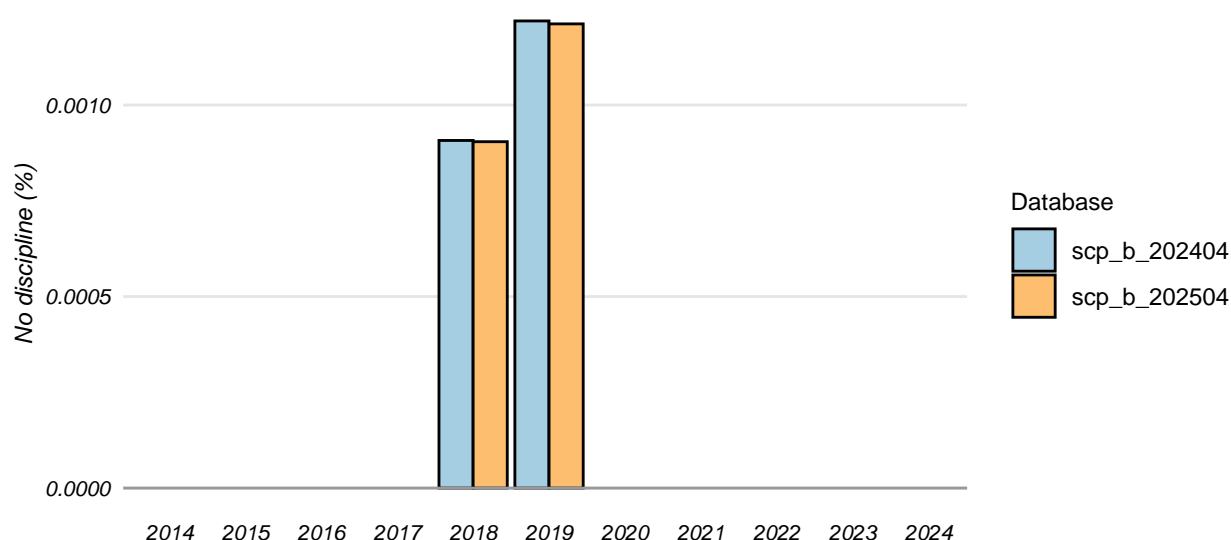


Figure 14: The percentage of publications in each database that do not have a discipline classification.

## Metadata: Changes in pubyear, doctype, pubtype and items removed

This section details the number of items for which changes were made to key metadata in the latest iteration of the database or the items were removed. We look at changes in the recorded publication year, document type and publication type as these three variables are typically the key inclusion criteria for bibliometric analyses. We also examine the number of items that were present in the scp\_b\_202404 database but not in the scp\_b\_202504 database (removed items), and items present in the scp\_b\_202504 database but not in the scp\_b\_202404 database (added items). A change in metadata for a large number of items may be problematic, particularly if the changes are not randomly distributed, such as adjustments having been made to items from a particular journal or set of publications, which may affect counts and indicators for specific entities. Some changes can be expected as the database provider updates or corrects items. However, changes to or removal of a large number of items may require investigation. Notably, the documents examined are not restricted to articles and reviews, but any document type.

We identified changes in the metadata of in-scope items by first matching items between the scp\_b\_202404 and scp\_b\_202504 databases using the item\_id identifier and then calculating the number of items that were added, removed, or had different metadata. The results are shown in Table 4.

Table 4: The number of items with changes in metadata between  
scp\_b\_202404 and scp\_b\_202504.

Crrnt year	Prvs year	Diff. year	Diff. src type	Diff. doc type	Added	Removed
2016	2016	0	59	387	0	0
2016	2017	2	0	0	0	0
2016	2018	1	0	0	0	0
2016	2019	1	0	0	0	0
2016	2020	1	0	0	0	0
2016	2021	2	0	0	0	0
2016		0	0	0	33520	0
2017	2016	72	0	0	0	0
2017	2017	0	781	1149	0	0
2017	2018	20	0	0	0	0
2017	2019	12	0	0	0	0
2017	2020	1	0	0	0	0
2017		0	0	0	31472	0
2018	2016	1	0	0	0	0
2018	2017	121	0	4	0	0
2018	2018	0	31	578	0	0
2018	2019	258	0	218	0	0
2018	2020	14	0	0	0	0
2018	2021	13	0	2	0	0
2018		0	0	0	33064	0
2019	2018	108	1	1	0	0
2019	2019	0	61	1657	0	0
2019	2020	39	0	6	0	0
2019	2021	10	0	0	0	0
2019	2022	8	0	0	0	0
2019		0	0	0	50149	0
2020	2018	1	0	0	0	0
2020	2019	1371	0	84	0	0
2020	2020	0	54	1600	0	0
2020	2021	99	7	13	0	0
2020	2022	107	0	0	0	0
2020	2023	6	0	0	0	0
2020		0	0	0	70749	0
2021	2017	1	0	1	0	0
2021	2018	1	0	0	0	0
2021	2019	102	0	8	0	0
2021	2020	580	0	30	0	0
2021	2021	0	50	2150	0	0
2021	2022	671	0	11	0	0

2021	2023	11	0	0	0	0
2021		0	0	0	71724	0
2022	2018	1	0	0	0	0
2022	2019	23	0	0	0	0
2022	2020	317	0	7	0	0
2022	2021	3110	0	48	0	0
2022	2022	0	28	6590	0	0
2022	2023	301	57	75	0	0
2022		0	0	0	99397	0
2023	2018	30	0	0	0	0
2023	2019	12	0	1	0	0
2023	2020	108	1	1	0	0
2023	2021	513	0	29	0	0
2023	2022	3971	2	74	0	0
2023	2023	0	203	5309	0	0
2023		0	0	0	223803	0
2024	2017	2	0	0	0	0
2024	2018	1	0	0	0	0
2024	2019	135	0	5	0	0
2024	2020	731	0	24	0	0
2024	2021	2292	0	78	0	0
2024	2022	13057	1	202	0	0
2024	2023	85131	0	652	0	0
2024		0	0	0	4333094	0
	2016	0	0	0	0	1185
	2017	0	0	0	0	595
	2018	0	0	0	0	1033
	2019	0	0	0	0	2381
	2020	0	0	0	0	2837
	2021	0	0	0	0	4481
	2022	0	0	0	0	8936
	2023	0	0	0	0	25032

## Metadata: Missing metadata variables

Figure 15 shows the annual percentage of publications in each database that are missing particular metadata, including page numbers, journal issue and volume information, DOIs, titles, references, abstracts, and keywords. We could reasonably expect improvements over time in missing metadata, such as for DOIs through increasing uptake of this identifier, however increasing missing metadata should be investigated. Empty graphs indicate there were no items missing this metadata.

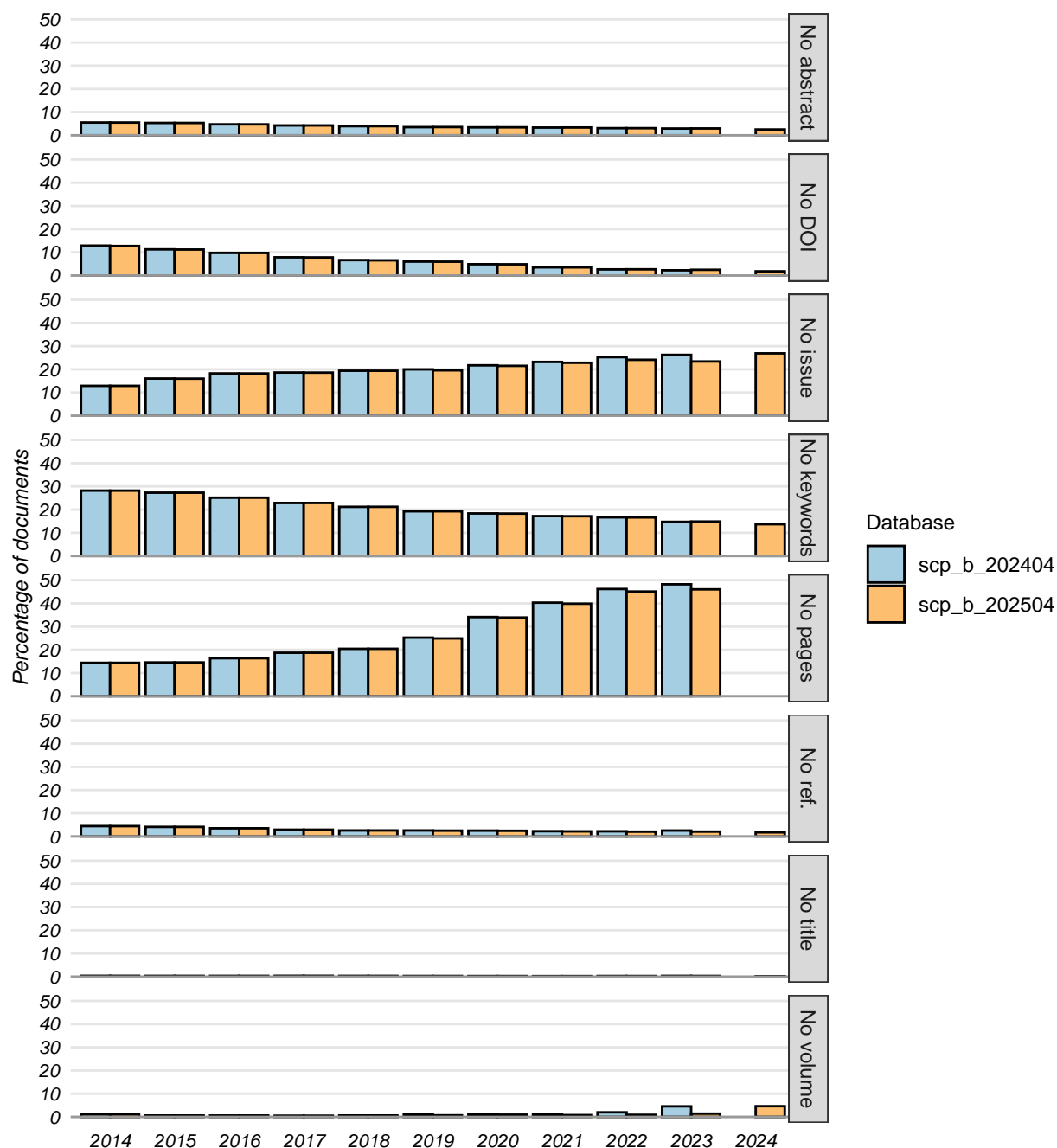


Figure 15: The percentage of items with missing metadata over time by database.

### Institution and country data: Number of articles and reviews with missing data

Bibliometric analyses often examine indicators at the level of institutions or countries. Further, fractional counting can be applied based on institutions, with articles apportioned according to authors' affiliations. It is imperative for accurate indicators that most, if not all, items have institution and country data, as missing information removes otherwise valid items from analyses.

The items table of the KB databases holds a record of all available items, while the associated data about authors' affiliations are held, in part, in the items\_affiliations tables. We have operationalised missing institution information here as publications that appear in the items table but have no corresponding information in the items\_affiliation tables. We present in the top panel of Figure 16 the number of items in each database between 2014 and 2024 with no institution information. Additionally, items can have institution information but no country code – from which country counts are derived – and these are shown in the bottom panel of Figure 16. Large disparities between the databases or substantial increases in missing information should be investigated.

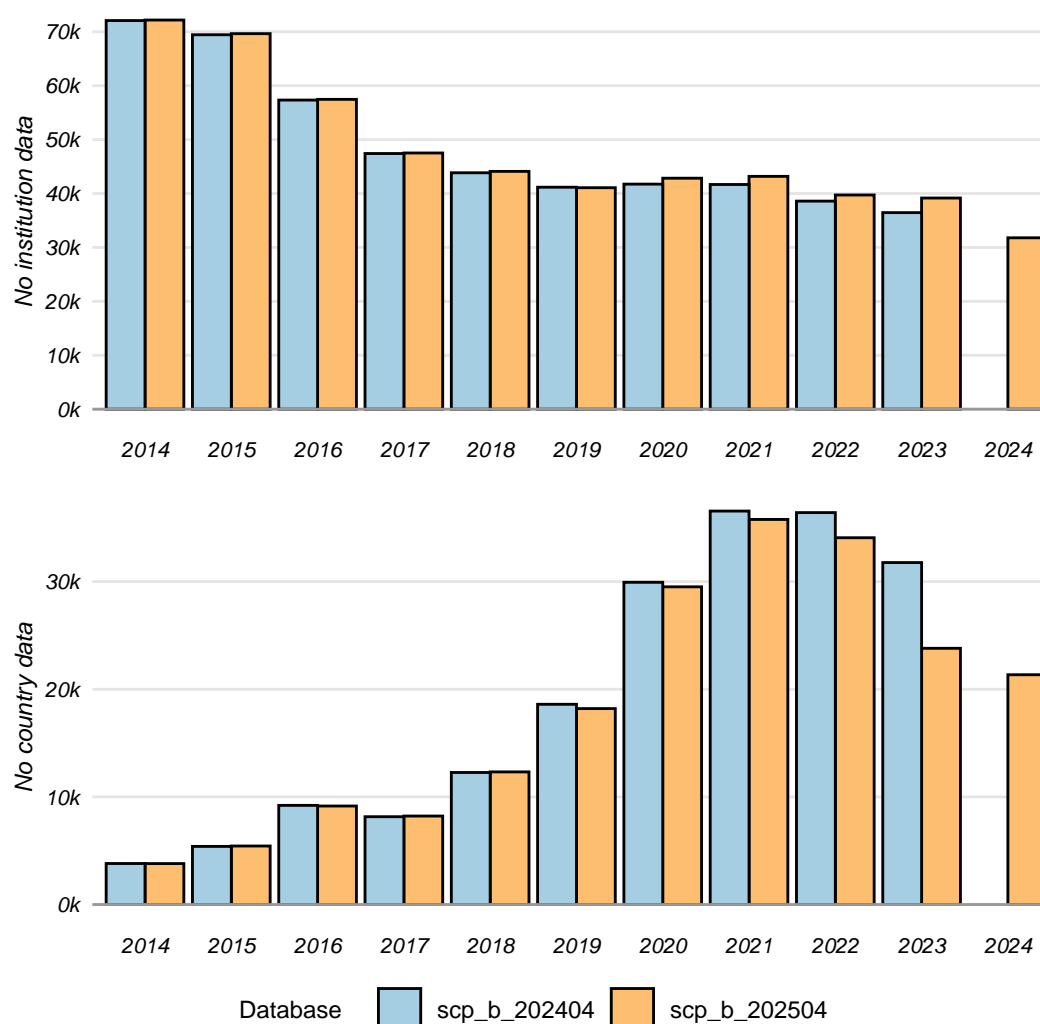


Figure 16: The number of items with missing institution information (top) and the additional items that have institution information but no country code (bottom) over time by database.

## German institutions: German publications missing from KB institution coding

In Figure 17 we show the annual percentage of German publications, i.e. those where the German indicator is TRUE, that were not assigned a KB institution code through the institution coding process. Increases over time may be due to the foundation of new institutions that have not yet been integrated into the coding process. However, publications without KB institutions are typically excluded from sector-level analyses, so it is important to understand the extent of missing institution information.

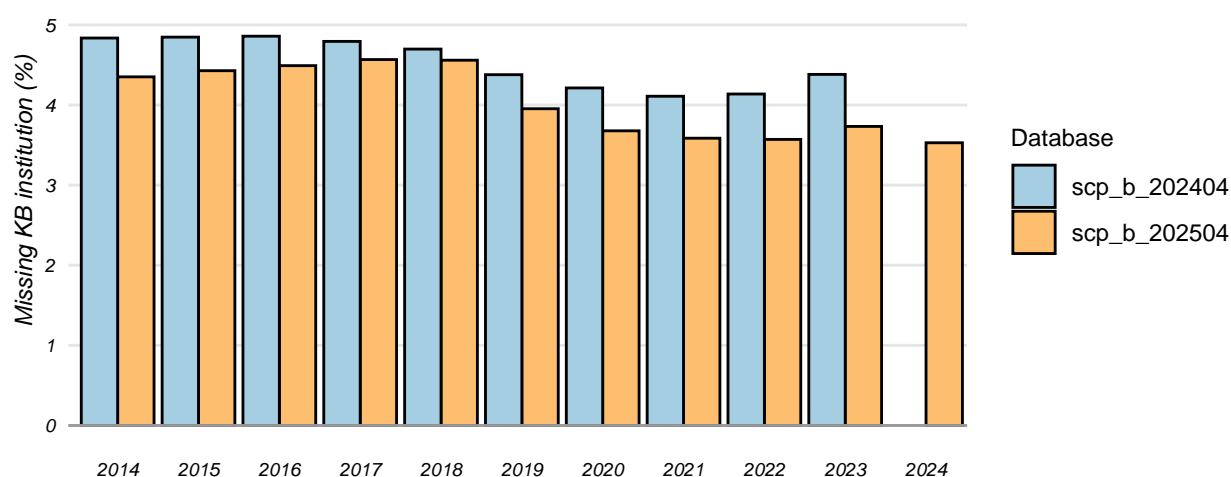


Figure 17: The percentage of German publications in each database that are missing a KB institution.

## German institutions: Changes in whole counts of articles and reviews

This section compares changes in the number of articles and reviews published by German institutions between the latest years available in each database. These tables can assist in identifying institutions for which substantial numbers of publications have been added, removed or otherwise changed in the latest database. They can also aid in assessing the degree of change in publication numbers for larger institutions, which may require further examination if considered unusual or excessive.

Table 5 presents potentially new institutions – these had no publications in 2023 in the scp\_b\_202404 database but more than five publications in 2024 in the scp\_b\_202504 database. Conversely, Table 6 shows the institutions that had at least five publications in 2023 in the scp\_b\_202404 database but no publications recorded in 2024 in the scp\_b\_202504 database. We also highlight in Tables 7 and 8 the larger institutions (with at least 20 publications) that had a change in publication counts of more than 40% between 2023 and 2024 in the scp\_b\_202404 and scp\_b\_202504 databases.

Table 5: Institutions with more than 5 publications in 2024 in scp\_b\_202504 that had no publications in 2023 in scp\_b\_202404.

Inst ID	Name	Previous pubs	Current pubs
3987	Restkategorie Deutschland (keine Zuordnu	0	960
6329	Deutsches Zentrum für Psychische Gesundh	0	387

5542	Munich Center for Quantum Science and Te	0	247
6415	Heinz Maier-Leibnitz Zentrum	0	162
6347	Center for Mind, Brain and Behavior	0	161
6583	Berufsgenossenschaftliches Universitätsk	0	154
6385	Center for Integrated Quantum Science an	0	122
6358	Center for Scalable Data Analytics and A	0	114
6356	Cluster of Excellence on Plant Sciences	0	113
5682	Frankfurt Cancer Institute	0	112
1502	E.ON AG	0	74
6297	Berlin School of Economics	0	71
6589	LABOKLIN GMBH & CO.KG	0	63
6444	Leibniz Gesundheitstechnologien: Ein For	0	57
6296	Lamarr - Institut für Maschinelles Lerne	0	56
6587	Sana Klinikum Offenbach GmbH	0	50
504	Berufsgenossenschaftliches Unfallkranken	0	46
6596	BG Klinikum Unfallkrankenhaus Berlin gGm	0	44
6324	Forschungscampus STIMULATE	0	43
6332	Niedersächsisches Zentrum für Biomedizin	0	40
6445	OCM Klinik GmbH	0	40
6369	Helmholtz AI	0	39
5765	Technische Universität Nürnberg	0	37
6591	Schwarzwald-Baar Klinikum Villingen-Schw	0	37
6363	Deutsche Forschungsgemeinschaft e.V.	0	30
293	Schön Klinik Roseneck	0	28
6349	Medical Park SE	0	28
6375	Tübingen AI Center	0	28
6592	St. Vincenz-Krankenhaus GmbH	0	28
6532	Naturschutzbund Deutschland (NABU) e.V.	0	26
6586	Schön Klinik Bad Aibling SE & Co. KG	0	26
6585	Schön Klinik Vogtareuth SE & Co. KG	0	25
6351	General Electric Deutschland Holding Gmb	0	24
6345	Bremen International Graduate School of	0	23
6582	BSBI-Berlin School of Business and Innov	0	23
4705	Schön Klinik Hamburg Eilbek	0	21
6588	Parkinson-Klinik Ortenau GmbH & Co. KG	0	21
6360	Google Deutschland	0	20
4465	Werkzeugmaschinenlabor-Forum gGmbH an de	0	19
6401	Hessische Zentrum für Künstliche Intelli	0	19
6449	CeGaT GmbH	0	19
6563	Deutsche Gesellschaft für Chirurgie e.V.	0	19
6342	Rigaku Europe SE	0	18
6584	Schön Klinik Berchtesgadener Land SE & C	0	18
6594	Klinikum Passau	0	17

6598	Schön Klinik München Harlaching SE & Co.	0	17
6590	Fachklinik Hornheide e.V.	0	16
6331	MVZ Medizinische Genetik Mainz	0	15
392	HSK, Dr. Horst Schmidt Kliniken GmbH	0	14
4569	Bundesverband der Deutschen Industrie e.	0	14
6409	EFCNI - European Foundation for the Care	0	14
6595	Klinikum Wolfsburg	0	14
6306	Johnson & Johnson GmbH	0	13
6572	RIFCON GmbH	0	13
6311	Cytel Inc.	0	12
6336	BasCat - UniCat BASF JointLab	0	12
6340	BioMed X GmbH	0	12
6374	Microsoft Germany	0	12
6435	Helmholtz Information & Data Science Aca	0	12
6552	Materialprüfungsamt Nordrhein-Westfale	0	12
5232	Rotkreuzklinikum München	0	11
6359	PharMetrX	0	11
6530	Ramboll Germany	0	11
6501	Voith GmbH & Co. KGaA	0	10
6506	AIR LIQUIDE Deutschland GmbH	0	10
5985	Deutsche Nationalbibliothek	0	9
6300	Klinikum Mutterhaus der Borromäerinnen g	0	9
6420	Nord-Ostdeutsche Gesellschaft für Gynäko	0	9
6429	Vandage GmbH	0	9
6431	Kassenärztliche Vereinigung Bayerns (KVB	0	9
6432	Lumileds Germany GmbH	0	9
6535	International Solar Energy Research Cent	0	9
6537	Kipu Quantum GmbH	0	9
6538	Sysmex Deutschland GmbH	0	9
6540	Hans-Böckler-Stiftung	0	9
6553	Institut für Grenzgebiete der Psychologi	0	9
3830	Thomson Adressfehler: Länderzuordnung	0	8
6348	NVIDIA Deutschland	0	8
6377	Brockmann Consult GmbH	0	8
6390	Deutsche Gesellschaft für Ernährung e.V.	0	8
6402	Steinbeis-Stiftung für Wirtschaftsförder	0	8
6511	Corteva Agriscience Germany GmbH	0	8
6528	HELLA GmbH & Co. KGaA	0	8
5858	BSP Business School Berlin	0	7
6298	Andrologicum München	0	7
6299	Zoologische Gesellschaft Frankfurt von 1	0	7
6317	Muehlhan Holding GmbH	0	7
6327	TÜV RHEINLAND AG	0	7

6407	NVision Imaging Technologies GmbH	0	7
6513	Daimler Truck AG	0	7
6541	Institut für Luft- und Kältetechnik geme	0	7
6570	PETA Science Consortium International e.	0	7
286	Schön Klinik Neustadt	0	6
2988	Zentrum für ambulante Neuropsychologie u	0	6
6085	ISDC - International Security and Develo	0	6
6305	kbo-Kinderzentrum München gemeinnützige	0	6
6428	Servier Deutschland GmbH	0	6
6441	Deutsche Multiple Sklerose Gesellschaft,	0	6
6523	KfW - Kreditanstalt für Wiederaufbau	0	6

Table 6: Institutions with no publications in 2024 in scp\_b\_202504 that had more than 5 publications in 2023 in scp\_b\_202404.

Inst ID	Name	Previous pubs	Current pubs
813	Europäisches Laboratorium für Molekularb	385	0
5160	Universitäts Herzzentrum Freiburg Bad Kr	198	0
5477	Leibniz-Institut für Photonische Technol	159	0
3540	DWI an der RWTH Aachen e.V.	126	0
2827	Klinikum Augsburg	93	0
548	Hochschule für angewandte Wissenschaften	90	0
1126	Fraunhofer-Institutszentrum Schloss Birl	32	0
372	Klinikum Augsburg	16	0
2978	Gemeinschaftspraxis für Hämatologie und	8	0
6124	AVL Deutschland Netzwerk	8	0
193	Bayerische Zentrum für Angewandte Energi	7	0
1535	Deutsche Edelstahlwerke GmbH	7	0
3836	SPECS Surface Nano Analysis GmbH	6	0
6103	Vector Consulting Services (VCS) GmbH	6	0

Table 7: Institutions with more than 20 publications in 2023 in scp\_b\_202404 that increased in publication counts by over 40% in 2024 in scp\_b\_202504.

Inst ID	Name	Previous pubs	Current pubs	Perc. diff.
704	Institut für Photonische Technologien e.	31	242	680.6
1082	Max-Planck-Institut für Chemie (Otto-Hah	210	807	284.3
6250	Excellence Cluster Cardio-Pulmonary Inst	27	98	263.0
355	Klinikum der Stadt Ludwigshafen gGmbH	33	119	260.6
1085	Max-Planck-Institut für Biophysik	27	90	233.3
6210	Sammelkategorie Vereine, Stiftungen	30	95	216.7

205	Naturkundemuseum Karlsruhe	24	69	187.5
1094	Max-Planck-Institut für Sonnensystemfors	87	249	186.2
6203	Munich Center for Machine Learning	47	133	183.0
417	Berufsgenossenschaftliche Kliniken Bergm	28	79	182.1
1175	Fraunhofer-Institut für Angewandte Infor	41	104	153.7
6005	Facility for Antiproton and Ion Research	25	62	148.0
1042	Max-Planck-Institut für medizinische For	35	85	142.9
1049	Max-Planck-Institut für biologische Kybe	44	106	140.9
5879	HMU Health and Medical University Potsda	66	157	137.9
541	Westsächsische Hochschule Zwickau	23	53	130.4
6155	Bayerische Zentrum für Krebsforschung (B	84	192	128.6
5352	Psychologische Hochschule Berlin gGmbH (	25	54	116.0
5325	Auditory Valley	27	58	114.8
474	Städtisches Klinikum Dessau	26	53	103.8
189	Diabetes Zentrum Mergentheim	23	45	95.7
5721	Kerckhoff-Klinik GmbH	66	127	92.4
1144	Fraunhofer-Institut für Physikalische Me	37	71	91.9
895	Nordwestdeutsche Forstliche Versuchsanst	23	44	91.3
670	Fachhochschule Aachen	52	94	80.8
1214	UCB Pharma GmbH	26	47	80.8
841	Bundesamt für Naturschutz (BfN)	25	45	80.0
4421	ExtreMe Matter Institute EMMI	56	99	76.8
1029	Max-Planck-Institut zur Erforschung von	35	61	74.3
5894	Hochschule Weihenstephan-Triesdorf	49	85	73.5
443	Berufsgenossenschaftliche Unfallklinik	37	64	73.0
5674	BMW GROUP	76	131	72.4
4766	Helmholtz-Institut Mainz (HIM)	105	180	71.4
1053	Max-Planck-Institut für Herz- und Lungen	92	157	70.7
30	IPN - Leibniz-Institut für die Pädagogik	74	126	70.3
1096	European Space Agency	39	65	66.7
5913	Max Planck School of Cognition	27	45	66.7
6018	Deutscher Verein des Gas- und Wasserfach	24	40	66.7
618	Hochschule für angewandte Wissenschaft u	32	53	65.6
610	Technische Hochschule Ostwestfalen-Lippe	168	278	65.5
1129	Fraunhofer-Institut für Werkstoffmechani	27	44	63.0
601	Hochschule Kaiserslautern	24	39	62.5
674	Wuppertal Institut für Klima, Umwelt, En	24	39	62.5
5715	Hochschule für Gesundheit	32	52	62.5
581	Hochschule für angewandte Wissenschaften	109	177	62.4
5975	European Virus Bioinformatics Center	37	60	62.2
1240	Springer Nature AG & Co. KGaA	89	144	61.8
1019	Max-Planck-Institut für Physik (Werner-H	203	327	61.1
1152	Fraunhofer-Institut für Lasertechnik (IL	46	74	60.9

1020	Max-Planck-Institut für Ornithologie	35	56	60.0
5771	Max-Planck-Institut für biologische Inte	44	70	59.1
731	FIDAM GmbH - Forschungsinstitut Diabetes	26	41	57.7
1169	Fraunhofer-Institut für Angewandte Polym	33	52	57.6
582	Hochschule Mittweida, University of Appl	30	47	56.7
5552	Fraunhofer-Einrichtung für Individualisi	30	47	56.7
645	Hochschule Darmstadt	73	114	56.2
5487	Fraunhofer-Institut für Gießerei-, Compo	22	34	54.5
36	Leibniz-Institut für ökologische Raument	48	74	54.2
1125	Fraunhofer-Institut für Zelltherapie und	81	124	53.1
1005	Restkategorie MPG, ohne eindeutige Insti	48	73	52.1
1052	Max-Planck-Institut für Kernphysik	204	310	52.0
5660	Heidelberger Institut für Radioonkologie	62	94	51.6
5652	MSB Medical School Berlin - Hochschule f	98	148	51.0
1157	Fraunhofer-Institut für Graphische Daten	26	39	50.0
6208	Institute for Lung Health (ILH)	46	69	50.0
49	Leibniz-Institut für Agrarentwicklung in	59	88	49.2
40	Leibniz-Institut für innovative Mikroele	76	113	48.7
327	Berufsgenossenschaftliche Unfallklinik M	50	74	48.0
971	Forstliche Versuchs- und Forschungsansta	23	34	47.8
5906	Technische Hochschule Aschaffenburg	23	34	47.8
5958	Münchener Rückversicherungs-Gesellschaft	40	59	47.5
514	Deutsches Herzzentrum Berlin	263	387	47.1
23	Institut für Angewandte Geophysik	43	63	46.5
643	Fachhochschule Dortmund	43	63	46.5
884	Staatliche Naturwissenschaftliche Sammlu	118	172	45.8
3750	Technische Hochschule Würzburg-Schweinfu	67	97	44.8
566	Hochschule Pforzheim - Gestaltung, Techn	34	49	44.1
5707	Evangelische Hochschule Rheinland-Westfa	41	59	43.9
148	Europa-Universität Viadrina Frankfurt (O	71	102	43.7
1382	Miltenyi Biotec B.V. & Co. KG	30	43	43.3
230	Sammelkategorie Praxen	672	961	43.0
563	Fachhochschule Potsdam	21	30	42.9
441	Frankfurter Diakonie-Kliniken gemeinnütz	47	67	42.6
608	Hochschule für Technik und Wirtschaft Dr	47	67	42.6
5837	Zentrum für Individualisierte Infektiosm	26	37	42.3
5610	Helmholtz-Institut Erlangen-Nürnberg für	142	201	41.5
791	Sanitätsakademie der Bundeswehr, Ernst-v	63	89	41.3
1058	Max-Planck-Institut für Infektionsbiolog	32	45	40.6
12	RWI - Leibniz-Institut für Wirtschaftsfo	37	52	40.5
1011	Max-Planck-Institut für Softwaresysteme	37	52	40.5
1609	BASF	223	313	40.4

Table 8: Institutions with more than 20 publications in 2023 in scp\_b\_202404 that decreased in publication counts by over 40% in 2024 in scp\_b\_202504.

Inst ID	Name	Previous pubs	Current pubs	Perc. diff.
4192	Institut für Transfusionsmedizin und Imm	39	22	-43.6
1064	Max-Planck-Institut für Hirnforschung	65	35	-46.2
452	St.-Antonius-Hospital Eschweiler	41	22	-46.3

## Authors: Median number of authors by Research Area and discipline

The median number of authors on a paper can be informative about patterns of collaboration and their potential implications for fractional counting. For instance, increasing levels of inter-sector or international collaboration could result in decreased publication counts for individual sectors or countries when using fractional counting. As such, understanding changes in authorship patterns can provide some insight into potential macro-level changes for entities.

We show in the left panel of Figure 18 the median number of authors per discipline in 2023 in both databases, and in the right panel the median number of authors per discipline in 2023 in the scp\_b\_202404 database compared to 2024 in the scp\_b\_202504 database.

While little change is expected to be seen in the left-hand panel of Figure 18 as the number of authors on a paper is unlikely to change between databases, differences in the right-hand panel indicate potential changes in disciplines' collaboration patterns. Disciplines for which the median number of authors changed by more than 1, based on the right-hand panel of Figure 18, are shown in Table 9.

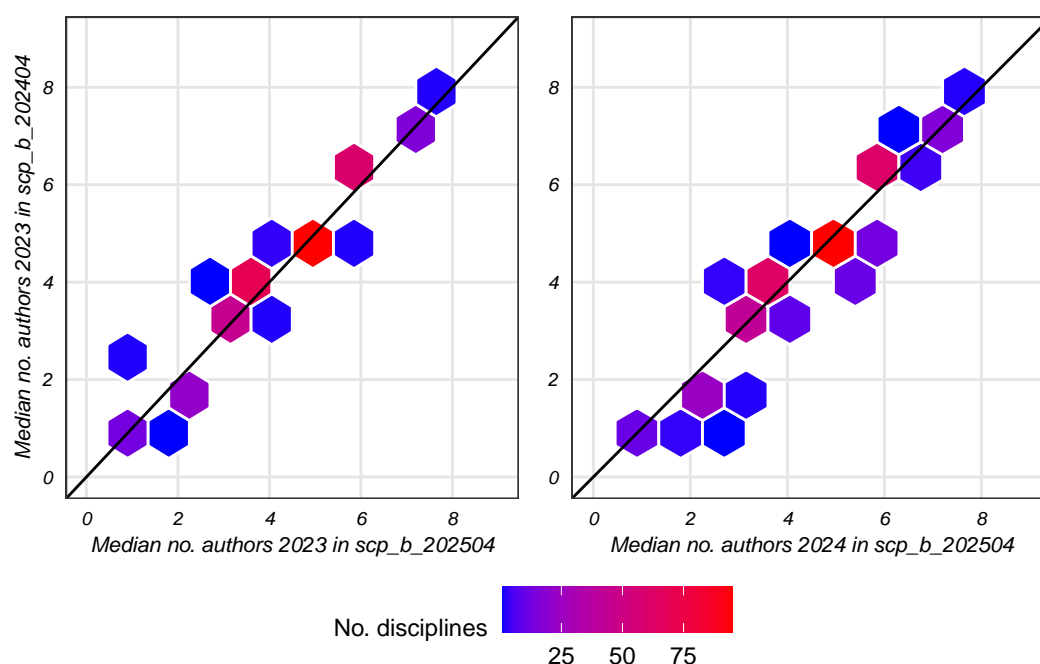


Figure 18: Median number of authors per discipline between databases, where colour denotes the number of disciplines with this combination of median authors.

Table 9: Disciplines where the median number of authors changed by more than 1 between 2023 in scp\_b\_202404 and 2024 in scp\_b\_202504.

Discipline	Previous median authors	Current median authors	Diff.
Medical Terminology	1	3	2

## Source items: Percentage by Research Area and discipline

Source items refer to whether the publications on the reference list of an indexed publication are also indexed in the database, as opposed to non-source items that are not indexed. Only source items are included in citation counts and so understanding the percentage of items cited that are also source can give an indication of the depth of Scopus' coverage of a discipline. That is, if a large number of indexed items' sources are not indexed, the reverse is also likely true and a large number of citations of indexed items are also missing, which has the effect of reducing citation counts for disciplines with lower coverage, such as the arts and humanities.

The percentage of references that are source items is expected to increase over time as the database provider continues to index journals and makes efforts to improve coverage of journals from disciplines with known low coverage. The percentage is not likely to ever reach 100% however, as authors will continue to cite items outside of the scope or coverage of Scopus.

We show in the left-hand panel of Figure 19 the percentage of references that are source items per discipline in 2023 in both databases, and in the right-hand panel the percentage of references that are source items per discipline in 2023 in the scp\_b\_202404 database compared to 2024 in the scp\_b\_202504 database.

It is in the right-hand panel that the effect of recently indexed journals may become apparent, where an increase in the percentage of source items may be seen if the journal is often cited within a discipline. The disciplines with a change in the percentage of indexed references of more than five percentage points between databases, based on the right-hand panel of Figure 19, are shown in Table 10. Longer term trends can be seen in Figure 20 where we present the percentage of reference that are source items per Research Area over the last ten common years of both databases.

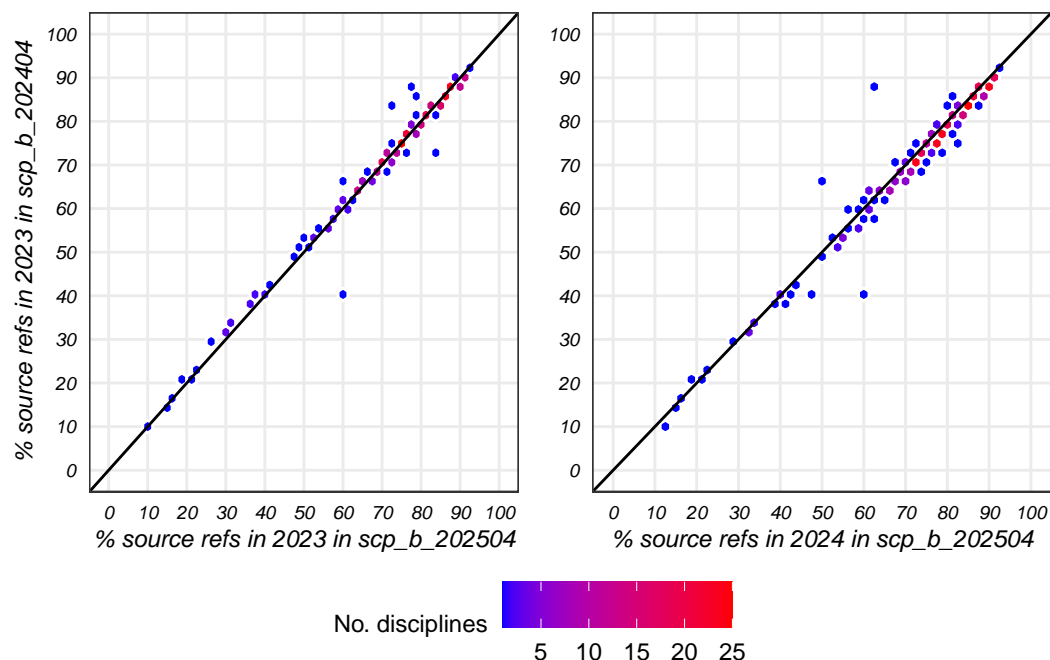


Figure 19: The percentage of cited items that are source items per discipline by database, where colour denotes the number of disciplines with this combination of source references.

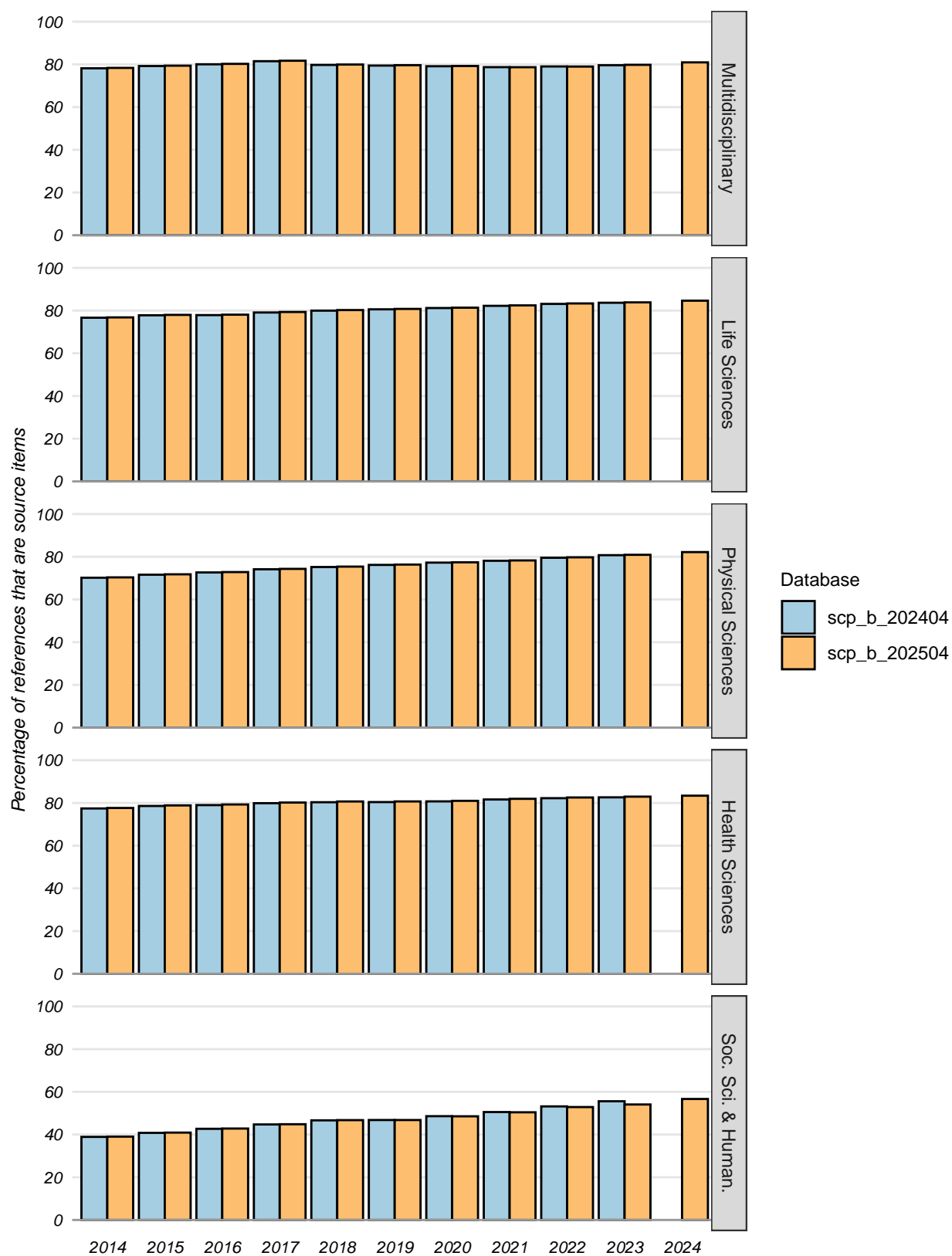


Figure 20: The percentage of references that are source items by Research Area and database over time.

Table 10: Disciplines where the percentage of indexed references changed by 9 or more percentage points between 2023 in scp\_b\_202404 and 2024 in scp\_b\_202504.

Discipline	Prvs % source	Crrnt % source	Change
Medical Terminology	41.0	59.8	18.8
Pathophysiology	73.8	82.9	9.0
Review and Exam Preparation	67.0	48.8	-18.2
Reviews and References (medical)	88.3	63.3	-25.0