German Centre for Higher Education Research and Science Studies

Performance and Structures of the German Science System

Stephan Stahlschmidt, Dimity Stephen and Sybille Hinze

Studien zum deutschen Innovationssystem Nr. 5-2019

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

Februar 2019

Diese Studie wurde im Auftrag der Expertenkommission Forschung und Innovation (EFI) erstellt. Die Ergebnisse und Interpretationen liegen in der alleinigen Verantwortung der durchführenden Institute. Die EFI hat auf die Abfassung des Berichts keinen Einfluss genommen.

Studien zum deutschen Innovationssystem
Nr. 5-2019
ISSN 1613-4338
Herausgeber: Expertenkommission Forschung und Innovation (EFI)
Geschäftsstelle, c/o Stifterverband für die Deutsche Wissenschaft, Pariser Platz 6, 10117 Berlin

Kontakt und weitere Informationen:
Stephan Stahlschmidt
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)
Schützenstraße 6a
10117 Berlin
Tel. +49 30 2064177-18
Fax +49 030 2064177-99
E-Mail: stahlschmidt@dzhw.eu

Table of Contents

1 Executive Summary 1
2 Foreword 3
3 Qualitative comparison between Web of Science and Scopus 4
3.1 Introduction 4
3.2 Overview of Web of Science and Scopus 4
3.3 Business philosophy and procedural differences 5
3.4 Differences in content and the impact on analysis results 7
3.5 Conclusions 11
4 National Bibliometric Indicators comparing Web of Science and Scopus 13
4.1 Indicators on Productivity 13
4.2 Impact Indicators 23
4.2.1 Excellence Rates 24
4.2.2 Additional Segmentations 28
4.2.3 Journal-based Indicators 34
4.3 Publication profiles by universities, Fachhochschulen and research associations 39
4.3.1 Number and share of publications 39
4.3.2 Impact indicators 42
4.3.3 Discipline-specific profiles 45
4.3.4 Co-publication analysis for universities, Fachhochschulen and research associations 50
4.4 International co-publications 52
4.4.1 German national and international co-publishing 52
4.4.2 Co-publication analysis for countries and groups of countries 54
4.4.3 Key international co-authoring countries with Germany and selected countries 58
5 Micro-level comparison on diverging assessment between Web of Science and Scopus 64
5.1 Illustrating differences via duplicated and exclusive publications 64
5.2 Resulting micro variation in disciplines and on publications' valuation 66
5.3 Observing structural differences via normalised citations 70
6 The growing influence of Chinese publication on Germany's bibliometric impact 73
6.1 Chinese publications as a non-marginal effect 73
6.2 Quantifying the effect 74
6.3 Explaining the effect 75
6.4 Conclusions 77
7 Appendix A: Conference Proceedings in the Web of Science 78
8 Appendix B: Country code list 81
9 Appendix C: Methodological details 82

9.1 Whole versus fractional counting 82
9.2 Conference proceedings 82
9.3 Disciplines classification 83
9.4 Citation window 83
9.5 Self-citations 84
9.6 Scientific Regard 84
9.7 International Alignment 84
9.8 Excellence Rate 84
9.9 Differences from previous years' reports 85
10 References 86

1 Executive Summary

This study updates the annual analysis of the performance and structures of the German Science System in international comparison. Bibliometric indicators are presented and discussed for the period 2007-2017.

In addition, a comparison between the two major data sources used for bibliometric analysis, Web of Science and Scopus, is presented in order to assess the impact the choice of the data source has on the performance indicators.

Both data sources differ in their selection philosophy. While Elsevier intends for Scopus to be the largest database, Clarivate Analytics is confident that capturing high-quality journals offers sufficient data to be representative of each discipline. In general, the differing business philosophies result in differences in the thresholds at which the inclusion criteria for indexation are applied. This subsequently affects the coverage of the databases. Due to the higher number of journals and other sources covered in Scopus compared to Clarivate Analytics' Web of Science (WoS) the overall number of publications and their citations are typically higher in Scopus. Reacting to the ongoing debate of the appropriateness of the selective approach Clarivate Analytics pursues it recently introduced a new index, the Emerging Sources Index (ESCI), which, while making concessions with regard to the selection criteria, extends the coverage significantly. The ESCI sought to improve regional and field-specific coverage; however, to date neither the bias towards publications from Europe and North America nor the over-represention of the natural and medical sciences and under representation of the arts and humanities has been solved. Which also holds true for Scopus.

Both databases use a journal-based classification system, meaning that not the individual article but, the whole journal is assigned to a category. Overall, the Web of Science seems to perform significantly better than Scopus in terms of the accuracy of assigning journals to classification categories, which might affect field-normalised scores. Overall, we observe an increasing number of publication between 2007 and 2017 for most countries. This increase is due to increased publication output as such but also reflects the increased coverage of journals by WoS and Scopus. The growth is particularly countries like China and India and other countries with developing science systems such as South Africa, South Korea and Brazil. While for countries with well-established systems such as the USA, the UK, Germany and France growth rates are rather low resulting also in their decreasing publication shares. China and the USA continue to be the countries producing the highest number of publications. Germany maintained its standing as one of countries producing the highest number of publications, and publications of high quality. Compared to 2007, Germany is publishing in journals with greater international visibility, but receiving fewer citations in those journals.

With regard to the Excellence Rate Germany's position is rather stable above the threshold. China continues to improve its performance also with regard to this indicator and meanwhile almost reaching the expected value. Difference in indicator values such as Excellence Rates and proportions of uncited publications between both databases (both higher in Scopus than WoS) are due to Scopus' inclusion of journals of lesser visibility compared to those in WoS.

In general, the patterns observed in high- and low-performing countries was largely the same across the databases, despite the differences in figures.

Unbowed is the overall trend towards increasing co-authorship, reflecting increasing collaboration. Also for Germany, the proportion of publications with international collaboration has increased. Switzerland remains the country with the highest share of internationally co-authored papers. As observed in the past, international co-authorship is particular high in smaller countries like Belgium or the Nordic countries. The USA remain the most attractive partner for most countries. Obvious is – from an US-American perspective - the strong increase of the share of publications co-authored between the USA and China reflecting increasing collaborations between these two countries, while the share of co-authored publications with other countries remains mainly stable.

Looking at the different actors within Germany shows an upward trend in publications over time, which corresponds to the overall growing number of publication for Germany as whole. We observe slight changes when it comes to the performance of the sectors: the Helmholtz Association continues to produce the highest number of publications followed by the Max-Planck Society, which, however, produced fewer publications annually since 2012 and consequently its share of German publications dropped. An increasing share, though still at a low level can be observed for the Fachhochschulen. They doubled the number of their publications as well as their share of German publications over the reference period. Indicators also reflect that Fachhochschulen and each of the research associations are increasingly publishing in more highly-cited journals though their publications in these journals are receiving relatively fewer citations.

2 Foreword

This report continues an important time-series analysing the performance of Germany in the science system, internationally and internally. A variety of indicators are examined in this report, using which Germany's performance is compared against that of 22 countries, and the EU13, EU15, EU28, and OECD country groups (see Appendix B for countries and groups). The indicators include the number and share of worldwide publications by country or group annually to gauge the level of scientific contribution. We also present the Scientific Regard and International Alignment indicators which indicate, respectively, whether the country's publications in a discipline are cited more or less often than other publications in the same discipline, and whether the country publishes more or less often in highly visible journals.

We present the Excellence Rate, or the proportion of each country's publications which are in the top 10% most highly cited publications per discipline, and also the Excellence Rate when only English-language literature is considered, which provides information about the impact of publishing in languages other than English. We also provide the proportion of each country's publications which were uncited, or conversely were cited more frequently than the median or 75th quartile citations, which gives an overview of the citation distribution. The same indicators are also provided for each of the German universities and non-university research institutions, as well as the number of publications and citations per full-time equivalent research staff. Finally, we present an analysis of the rate of international collaboration, and with whom Germany and other key countries are collaborating. Several of these indicators are also differentiated by discipline which serves to normalise the indicators across publication and citation practices, and also provide greater context to the scientific performance of Germany.

Further, in particular, this report provides the same information extracted from both WoS and Scopus to enable a direct macro-level comparison of the results when holding key components, such as document types, constant. This direct comparison of the results from analyses provides insight into the sensitivity of bibliometric indicators to the database used in computing them.

The content of this report examines predominantly 'articles' and 'reviews' published in 'journals' from the entire Scopus database, and from the WoS indices Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI). Where conference papers are examined, these 'proceedings papers' are extracted from the WoS indices Conference Proceedings Citation Index and Book Citation Index for the Sciences and Social Sciences. Fractional counting is used for all data, except when examining co-publications when whole counting is used. For analyses of countries, fractional counting is conducted at the level of the author and aggregated to the country-level. As such, a proportion of each publication is attributed to each author which is then aggregated into the fractional count of publications for each author's country. Similarly, where the German universities and non-university research institutions are examined, fractional counting is conducted at the level of the author and aggregated to the level of the institution, such as the Max Planck Gesellschaft or the universities as a whole.

Publications are examined for the period 2007 to 2017. A citation window of three years is applied for citation data, as such indicators for citations include all citations of a publication which occurred within the year of its publication and the subsequent two years. Citation data are presented for the period 2007 to 2015. Self-citations have not been excluded from the data. Data for disciplines are presented using the OECD Fields of Science and Technology (FST) disciplines, which have been concorded from Scopus' All Science Journal Classification (ASJC) and WoS' 'traditional' Subject Categories classification scheme. Items in Scopus attributed to the ASJC 'multidisciplinary' category have been excluded from presentation here as this category is not mapped to any category in the FST classification. Also, a small proportion of items are unclassified in WoS and so were not able to be attributed to an FST category and were also excluded here. See Appendix C for further details about the methodology used in this report.

3 Qualitative comparison between Web of Science and Scopus

3.1 Introduction

This section provides a qualitative comparison of the Scopus and Web of Science (WoS) bibliometric databases. The comparison examines the business philosophies of the database producers, the consequent procedural differences between the databases, such as the inclusion and exclusion criteria for content sources, and the resulting content differences based on source coverage and classification structures. The information provided here is based on current published literature regarding the differences between the databases, and content guides published by the database producers, Clarivate Analytics and Elsevier. This discussion provides context to the differences in the quantitative results between the databases as seen in sections 2 and 3, and highlights the potential impact of the choice of database on analyses.

3.2 Overview of Web of Science and Scopus

Web of Science (WoS) and Scopus are both subscription-based databases of multidisciplinary bibliographic information. The foundation of WoS was the Science Citation Index developed by Eugene Garfield and launched by the Institute of Scientific Information in 1964. The Science Citation Index was later expanded and merged into what is now WoS, owned, produced and managed by Clarivate Analytics. The current WoS is comprised of a number of literature and citation search databases based on different topics or focusing on different regions. From this group, a key set of indices have been identified, titled the Core Collection (CC), which are reported by Clarivate Analytics (2018a) as the most important and useful indices according to users.

The indices in the CC are the Science Citation Index – Expanded (SCIE), the Social Science Citation Index (SSCI), the Arts & Humanities Citation Index (A&HCI), the Conference Proceedings Citation Index (CPCI), the Book Citation Index (BCI), Emerging Sources Citation Index (ESCI), and two chemistry indices for compound and reaction data, the Current Chemical Reactions (CCR), and Index Chemicus (Clarivate Analytics 2018a). The user can search the CC in its entirety or by individual index. Users can also customise the data they have access to, such as subscribing to the entire WoS platform or only selected products such as the CC or individual indices, which can either be accessed via the online platform or licensed from Clarivate Analytics as raw data.

The WoS indices cover the areas of life sciences, biomedical sciences, engineering, social sciences, and arts and humanities. As of July 2018, the collection included more than 20,300 journals, 94,000 books and an unstated number of conferences, totaling over 71 million records and more than 10 million conference proceedings (Clarivate Analytics 2018b). Content is updated daily and data are available from 1900 for the SCIE and SSCI, from 1975 for the A&HCI, from 1990 for CPCI, and from 2005 for the ESCI and BCI.

Scopus was launched by Elsevier in 2004. In contrast to WoS, Scopus amalgamates all of its abstract and citation data into a single searchable database (Elsevier 2017). Scopus covers journals, trade publications, books, conference materials, and patents across the areas of science, technology, medicine, social science, and arts and humanities. The database is updated daily and as of August 2017 was comprised of content from more than 21,950 peer-reviewed journals, 280 trade publications, 560 book series, 150,000 books, 100,000 international conferences, and patents from 5 patent offices. In total, this is more than 69 million records, 8 million conference papers, and 39 million patent records, with approximately 3 million records added each year (Elsevier 2017). The earliest record in Scopus dates back to 1788, however the majority of 62.4 million of 69 million records were published since 1969 (Elsevier 2017)

3.3 Business philosophy and procedural differences

While both WoS and Scopus seek to provide similar services to the scientific community – multidisciplinary databases suitable for retrieving relevant primarily scientific publications, allowing for the identification of key journals, papers, authors, or institutions and facilitating bibliometric analyses – their fundamental business models differ. The focus of WoS is to index only the highest quality journals. Clarivate Analytics subscribes to Garfield's law of concentration, which proposed that the majority of significant academic research is covered by a relative small number of journals. As such, Clarivate Analytics maintains that they will sufficiently capture the majority of important research simply by indexing the key journals in each discipline (Testa 2018).

To maintain its quality, Clarivate Analytics undertakes a rigorous assessment of each journal suggested for inclusion in WoS to determine if it will be indexed. Anyone may suggest a journal for inclusion through the WoS website, but priority is given to journals requested or recommended by WoS users (Testa 2018). When selecting journals, Clarivate Analytics applies different criteria based on the index in which the journals are suggested to be included. For example, the three indices which cover the most well-established and respected journals – SCIE, SSCI, and A&HCI – are held to different criteria in the WoS Core Collection Journal Selection Process than the ESCI, which covers up-and-coming journals in emerging fields of research or extends the coverage of regional journals of particular importance (Testa 2018; Huang et al. 2017). With the introduction of the ESCI in 2015, Clarivate Analytics addressed an ongoing discussion about the completeness of WoS' coverage from indexing only core sources. The ESCI is a second tier index with less strict inclusion criteria than the SCIE, SSCI and A&HCI. Journals covered in the ESCI may move up to the first tier indices over time, provided they fulfil the inclusion criteria for these products, or journals which no longer fulfil the criteria for the first tier indices may be relegated to the ESCI.

All prospective journals are considered for all applicable indices and first checked against the following minimum criteria:

1. The journal publishes peer-reviewed content
2. The journal has an International Standard Serial Number (ISSN)
3. The journal's bibliographic information and cited references are in English, or the references from non-English-languages journals are in Roman script (Clarivate Analytics 2017).

If the journal passes these criteria and is suggested for the top-tier indices, it is then assessed against the following criteria, noting that all journals are considered in terms of the norms for the field in which they publish:

1. Publishing standards: the content published is primarily scholarly, funding sources are acknowledged; the journal must subscribe to ethical publishing practices; content must be published in print or compatible electronic formats; the journal follows international editorial conventions; at least the bibliographic information and abstract is in English, if not the full text; and, of key importance, content must be published regularly and in accordance with the stated publishing schedule as lapses may indicate the journal has too little content to publish regularly.
2. Editorial content: the editors decide whether the content of the journal will enrich the database with new information, or if the content is already sufficiently covered by existing titles.
3. International or regional focus: the journal is assessed for the diversity of its editorial board and authors in the context of its intended readership, with greater diversity expected for journals aimed at an international audience than those intended for a regional or specific audience.

4. Citation analysis: because Clarivate Analytics indexes a cited reference regardless of whether the journal it comes from is indexed, the editors are able to assess suggested journals based on the citation data from records already indexed. Key data used in these assessments are total citation counts, impact factor, the citation history of authors and editors, and self-citation rates (Testa 2018; Clarivate Analytics 2017).

Each year this process is conducted for approximately 3,500 journal titles, from which around 10% of the submitted journals are accepted (Testa 2018). However, journals are often included in the ESCI first, with the possibility of being indexed in the top-tier indices later. The criteria which must be met for inclusion in the ESCI are:

1. The journal content must be peer reviewed
2. Publishing practices must be ethical
3. The journal must be published in a compatible electronic format; print-only is not accepted
4. The articles' bibliographic information must be in English
5. The journal must enhance the content of WoS
6. The timely and consistent publication of content is less important for ESCI journals than for the top-tier, however journals should be active and publish regularly (Testa 2018).

Once a journal has been accepted for any index, all of its content from that point on is indexed in WoS (Clarivate Analytics 2018b). However, Clarivate Analytics also applies ongoing evaluation to its collection to ensure its journals are maintaining high standards and remain relevant to its indices. If a journal fails to continue to meet the accepted standard, it may be downgraded to the ESCI if it was indexed in SCIE, SSCI or A&HCI, or it may be removed from WoS entirely.

Elsevier has a different philosophy for Scopus than indexing only the key journals in each discipline. In their approach, Elsevier extracted vast quantities of records from its indexing databases, such as EMBASE and GEOBASE, and then enhanced the data by indexing the records cited by those extracted records (Jacso 2005). As such, the intent for Scopus was that it should contain the largest number of records possible, which differs from WoS which chooses a reduced number of records in favour of the quality of records.

That is not to say, however, that Scopus does not control the quality of the journals it accepts. Elsevier employs 17 independent experts from a range of disciplines and geographic regions as its Content Selection and Advisory Board (CSAB), with each member responsible for a subject area. As with WoS, this board applies rigorous criteria to determine which journals are accepted into Scopus, and generally guides the development and direction of Scopus content over time (Elsevier 2017) Journals suggested for inclusion in Scopus are assessed on two levels. First, journals must meet the following minimum criteria:

1. Content must be peer-reviewed and a description of the process must be publically available
2. The journal is published regularly and has an ISSN
3. To ensure accessibility by a wide readership, the journal must publish titles and abstracts in English and references must be in Roman script
4. The publisher must have a publication ethics and malpractice statement publically available (Elsevier n.d.)

If these criteria are satisfied, the CSAB member responsible for the journal's subject area assesses the journal against the following five dimensions:

1. Journal policy: the journal has appropriate editorial policy, conducts peer reviewing, and has a diverse geographic distribution of both editors and authors
2. Content: the content makes an academic contribution to the field, has a clear abstract, is readable, of high quality, and adheres to the stated aims and scope of the journal
3. Journal standing: acceptable standing of the editor and citedness levels of articles in Scopus
4. Publishing regularity: content is published consistently and regularly
5. Online availability: full content is available online, and the journal's homepage is of high quality and available in English (Elsevier 2017).

Journals which perform well against these criteria are accepted into Scopus and their entire content is indexed from that point forward. In any given month, between 25 and approximately 250 journals are reviewed, of which 30%-70% are accepted (Elsevier 2017). Elsevier also annually re-evaluates the existing journals in Scopus to ensure they continue to be of acceptable quality. The metrics against which all accepted journals are evaluated are:

1. The self-citation rate must not exceed 200% compared to the average in the subject field
2. The total citation rate of the journal must not be less than 50% compared to the field average
3. The CiteScore for the journal must not be less than 50% compared to the field average
4. The number of articles published by the journal must not be less than 50% compared to the field average
5. The number of times a full-text link is clicked on Scopus must not be less than 50% compared to the field average
6. The number of uses of the journal's abstracts on Scopus must not be less than 50% compared to the field average (Elsevier 2017).

If any of the benchmarks for a journal are not met, the journal receives notification of this and is given one year to improve its metrics. If after this time the journal has not improved any of its metrics, the journal is re-evaluated by the relevant CSAB member against the original acceptance criteria and may be removed from Scopus (Elsevier 2017; Elsevier n.d.).

The selection processes for journals into WoS and Scopus demonstrate that both Clarivate Analytics and Elsevier are publicly committed to ensuring their databases contain high-quality journals. There are data, however, to suggest that journals are accepted into Scopus at least which do not meet the stated criteria, with journals often failing the requirement to have particular information publicly available, such as reviewer lists, ethics and malpractice information, and editorial policies (Taşkın et al. 2015). The different business philosophies between Clarivate Analytics and Elsevier have implications for the content of the databases, which in turn influences the results from analyses using the databases. These differences in content and the associated influence on analyses are discussed in the following sections.

3.4 Differences in content and the impact on analysis results

The different business philosophies of Clarivate Analytics and Elsevier influence the overall number of journals covered in each database, the number of papers covered, and the consequent number of citations recorded for each item. The database providers also use different classification structures to assign items within their databases to scientific disciplines. The effect these differences can have on bibliometric results are demonstrated in this section using previous empirical analyses, and they are also evident in the data provided in sections 2 and 3.

When Scopus was launched in 2004, it contained 14,200 journals and an additional 8,000 have since been indexed (Elsevier 2017). When a new journal is added to Scopus, content is indexed from that point on; the back catalogue of the journal is not indexed. This is the same process for WoS, however as Scopus is much newer than WoS, this resulted in a lack of historical data in Scopus. As such, Elsevier undertook two projects – one in 2004 and a second in 2014 – to increase the historical content of Scopus. Elsevier reported that by mid-2017 it had indexed an additional 195 million records published between 1970 and 1996 (Elsevier 2017). However, WoS, having been established in the 1960s and continuing to date, has a stronger collection of data prior to 1970 than does Scopus. Consequently, historical analyses are likely to be affected by this difference in coverage. See Figure 1 for the number of indexed items in WoS and Scopus in 1996 and cumulative to 2016. This period of time is presented as these are the years for which current data from both databases is available to the DZHW via the Competence Centre for Bibliometrics[1]. Please note that these figures represent only the period 1996 to 2016 and so do not match the figures noted in the section "Overview of Web of Science and Scopus" which include all items indexed in each database.

With regard to business philosophies, given that Elsevier has a more inclusive approach to indexing journals in Scopus than does Clarivate Analytics for WoS, the number of journals covered and subsequent bibliometric indicators such as the number of publications and their citations, are typically higher in Scopus than in the core WoS indices most often used (SCIE, SSCI and A&HCI). The influence the introduction of the ESCI has had on WoS' level of coverage is discussed later in this report.

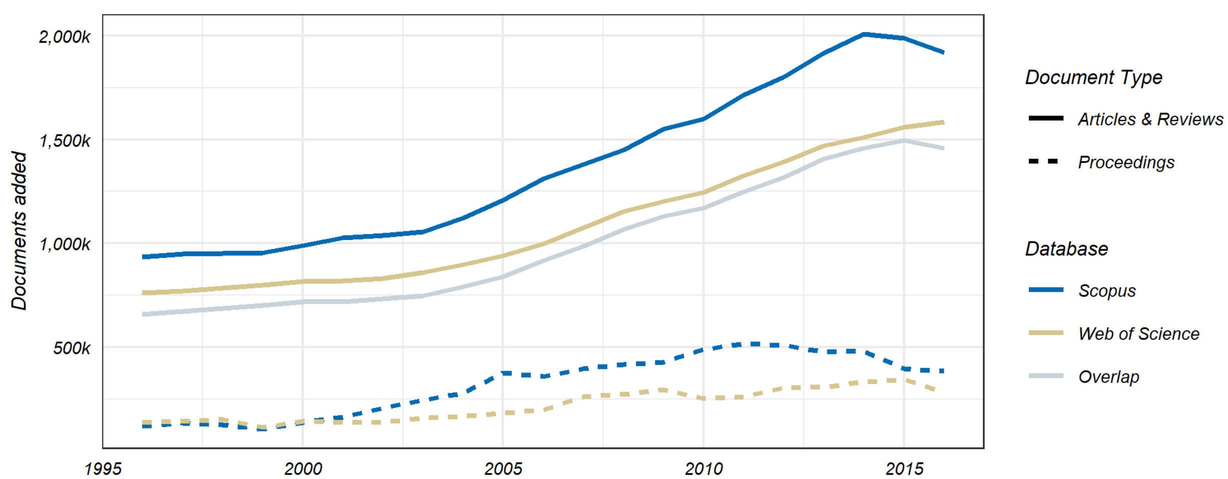


Figure 1 The number of conference proceedings (broken line), articles & reviews (whole line) added yearly to the Web of Science (SCIE, SSCI, AHCI and CPCI) or Scopus. Overlap of documents added jointly to both databases.

Mongeon and Paul-Hus (2016) undertook an extensive comparison of the journals covered by the three core WoS indices and Scopus. They matched the lists of indexed journals from WoS (13,607 journals in the SCIE, SSCI and A&HCI) and Scopus (20,464 journals) with Ulrichsweb – a database of more than 300,000 journals and other periodicals – to determine how well each field, publishing country and language was represented in each of the databases compared to its coverage in Ulrichsweb. As expected,

[1] http://bibliometrie.info/

Scopus exceeded the coverage of WoS, covering nearly a third of all Ulrichsweb journals (32.7%) compared to the 22% covered by WoS.

A study conducted by Bar-Ilan (2017) shows how this difference in coverage affects results for specific analyses. Bar-Ilan (2017) examined how many articles published between 2013 and 2016 were retrieved when searching for the term "information retrieval" in Scopus and WoS (SCIE, SSCI, A&HCI, ESCI, and CPCI). After cleaning, 5,458 articles were retrieved from Scopus and 4,264 were retrieved from WoS. When counting citations, both WoS and Scopus include only the citations which come from items indexed within their own databases, which tends toward Scopus producing higher citation counts due to its greater number of indexed items. In the Bar-Ilan (2017) study, the articles from Scopus had a total of 12,087 citations, an average of 2.2 citations per paper, while the WoS articles had 6,674 citations, an average of 1.6 citations per paper (Bar-Ilan 2017). Evidently, the higher coverage of journals in Scopus than WoS can impact the number of records and citations reported in analyses.

Bar-Ilan (2017) also looked at the overlap in coverage between the databases. Of the records retrieved by WoS, 58% were also indexed in Scopus. Conversely, 45% of the records retrieved by Scopus were indexed in WoS. When the comparison between coverage was opened to any content from journals published between 1996 and 2013, Donner (2016) found that 91% of the items from WoS (SCIE, SSCI and A&HCI) were also present in Scopus, while 73% of the Scopus content was present in WoS. When analysed by citation index, 93% of SCIE items, 92% of SSCI items, and 61% of A&HCI items were indexed in Scopus (Donner 2016).

Also, in the Bar-Ilan (2017) study, the top 3 most highly cited papers were different between the databases. One of the top 3 from WoS was indexed in Scopus, but none of the top 3 from Scopus were indexed in WoS (Bar-Ilan 2017). Similar results – higher records retrieved and citation counts for Scopus, and little agreement on most highly cited publications – were found for another study in the area of inclusive education research (Shah, Mahmood, and Hameed 2017), demonstrating that differences in results arise across subject areas.

Indeed, Mongeon and Paul-Hus (2016) noted similar effects when they examined the coverage of WoS (SCIE, SSCI and A&HCI) and Scopus by subject area. They divided the journals in each of the three databases into four areas: Natural Sciences and Engineering (NSE), Biomedical Research (BR), Social Sciences (SS), and Arts and Humanities (AH). They found the greatest differences in coverage occurred for BR, where Scopus covered 47% of Ulrichsweb content while WoS covered only 28%, and SS where Scopus covered 25% of Ulrichsweb journals and WoS covered 13%. Donner (2016) also found that Scopus' coverage exceeded that of WoS (SCIE, SSCI, AHCI, CPCI) by an average of 25% across the 7 disciplines he examined (law, psychology, environmental science, engineering, nursing, education, and computer science) when compared to a third source, in this case the Australian Research Council's list of validated journals. These findings indicate that the same analysis in a discipline would yield substantially different counts of items depending on which database was used.

Further, Mongeon and Paul-Hus (2016) found that the relative distribution of journals in WoS and Scopus to subject areas was not representative of the general distribution of subject areas according to Ulrichsweb. When the distributions were compared, they found that both Scopus and WoS overrepresented NSE and BR journals and underrepresented SS and AH journals. Forty-three percent of WoS journals and 33% of Scopus journals were from the NSE field, compared to 28% in Ulrichsweb, while 45% of Scopus' journals and 30% of WoS' journals were from BR, compared to 21% in Ulrichsweb. AH was similarly underrepresented at around 9% of all journals in WoS and Scopus, compared to 15% in Ulrichsweb, and SS accounted for 28% of Scopus journals and 21% of WoS journals, compared to 36% in Ulrichsweb (Mongeon and Paul-Hus 2016).

Moed (2005) also sought to gauge the adequacy of the coverage of disciplines in WoS by examining the cited references of indexed items. The argument being that, if a large proportion of the cited items in the papers of a given discipline are also indexed in WoS, then WoS sufficiently covers the literature of that discipline. Moed (2005) found that disciplines such as the biological sciences, clinical medicine, chemistry and physics had excellent coverage, with more than 75% of cited references also indexed, while arts and humanities and non-health-related social sciences had the poorest coverage, with between 7% and 33% of cited references also indexed.

With regard to representation of countries, Mongeon and Paul-Hus (2016) noted greater representation in both WoS (SCIE, SSCI and A&HCI) and Scopus of journals published in the United States, the United Kingdom, the Netherlands, France, Germany and Switzerland compared to Ulrichsweb's distribution of countries. However this should be interpreted cautiously as countries were assigned based on the country in which the journal was published which does not necessarily represent the geographic location of the journal or its authors. For example, the online and print versions of the journal may be published in different countries which reflects business processes rather than the national identity of the journal (Mongeon and Paul-Hus 2016). Aman (2016), however, when examining journal content between 1996 and 2013 based on the author's location, found that both Scopus and WoS had a much stronger coverage of items published by authors in Europe, North America and Asia, and these three regions comprised 90% of the content on each database. Also, not unexpectedly, Mongeon and Paul-Hus (2016) noted journals published in English were overrepresented in both WoS and Scopus and in all fields compared to Ulrichsweb. All other languages were underrepresented, except for Dutch, French and German in NSE, and French in AH (Mongeon and Paul-Hus 2016).

These studies demonstrate that both Scopus and the key indices of WoS have tendencies to overrepresent content from Europe, North America and Asia, and content published in English. Further, WoS and Scopus both emphasise publications from the natural and medical sciences and underrepresent publications for the social sciences and humanities. To combat this underrepresentation, both databases have introduced indices for books as a significant proportion of work in these fields is published in books (Mongeon and Paul-Hus 2016). However, 96% of the books covered by the Book Citation Index in WoS between 2005 and 2012 were in English (Torres-Salinas et al. 2014), suggesting underrepresentation of the social sciences and humanities would not be alleviated for non-English-speaking countries. These findings suggest that results from analyses using these databases may still be less accurate for the social sciences and humanities than other fields, and for certain regions such as Africa, Oceania and Central and South America, and for content published in languages other than English.

Clarivate Analytics sought to address some of this underrepresentation by introducing the ESCI into the WoS CC in November 2015 as a means of increasing the coverage of WoS in emerging fields and important regional topics. Due to its only recent introduction, there are still relatively few studies which have examined the impact of the ESCI of the coverage of WoS, however one study found that the ESCI added more than 6,000 journals to the CC, 49% of which are also indexed in Scopus; that the ESCI had more open access journals (36%) than the SCIE, SSCI and A&HCI (5-12%) or Scopus (14%); and the ESCI typically had slightly higher representation of non-English-speaking countries, in particular Spain, than the CC indices or Scopus, although a substantial proportion of 35% were still from the United States and England (Somoza-Fernández, Rodríguez-Gairín, and Urbano 2018). Another analysis by Huang et al. (2017) of 2015 articles and reviews in ESCI similarly found that the ESCI had higher rates of open access articles compared to the SCIE, SSCI and A&HCI, and also had greater coverage of disciplines in the arts and humanities and the social sciences than these indices. However, the majority of journals still originated from Europe and North America, which Huang et al. (2017) suggested inadequately accounted for the contributions of countries such as China, Japan and Korea which lead the world in some disciplines. As such, the ESCI has increased the coverage of the CC by around 6,000 journals, although Scopus also covers half of these journals, and has gone some way in improving the visibility of research from non-English-speaking countries, however the ESCI still predominantly focuses on Europe and North American research.

The second key area which can spur differences in analysis outcomes is the classification of disciplines. WoS and Scopus implement both different methods of assigning their content to disciplines and different classifications of disciplines. Both, however, assign classifications based on the content of the journal – not the individual article – and both allow each journal to have more than one classification (Wang and Waltman 2016).

WoS contains two classifications: a higher-level classification of approximately 150 research areas and a more detailed classification of around 250 categories across the sciences, social sciences, and arts and humanities. The classification used in Scopus is the All Science Journal Classification (ASJC) system. At its most detailed level, it contains 304 categories which can be aggregated into 27 broader categories (Wang and Waltman 2016).

Researchers have long held concerns about the accuracy of the classification systems in both databases, in part because of the lack of information about how categories are assigned to journals. Clarivate Analytics is believed to assign categories in WoS via an algorithm based on both cited and citing data, while Scopus has not explained how its classifications are assigned (Wang and Waltman 2016). It is important that journals are accurately assigned within classifications because several indicators, such as citation rates, are often normalised against the field in which the journal is published to avoid presenting misleading comparisons between fields with different citation and published practices. For example, Leydesdorff and Bornmann (2016) examined field-normalisation for the fields of library and information sciences, and science and technology studies in WoS, and found field-normalisation would be potentially "harmful" to the analysis of these fields due to the way in which journals in these areas are allocated to categories.

Wang and Waltman (2016) found that WoS slightly out-performed Scopus in accurately assigning journals to classification categories. Their results showed that, while both databases accurately classified journals when a journal had a strong affiliation with a particular subject, both databases – and particularly Scopus – too leniently assigned journals to multiple categories with which they were only loosely affiliated. Donner (2016) similarly found Scopus leniently classified journals. For instance, Scopus categorised 131,000 items as from the law discipline and 65% of those items were also covered by WoS, however, of the items that both databases covered and Scopus defined as law, only 38% were also classified by WoS as from the law discipline. Across disciplines, the percentage of content covered and similarly classified by both databases ranged between 16% in nursing to 65% in psychology (Donner 2016). The results from these studies suggest that while field-normalisation continues to be used – and rightfully so as it is an important means of reducing the disparity between fields to enable comparison – differences in results will arise between analyses from WoS and Scopus due to both the different classifications used and how journals are ascribed to classifications.

3.5 Conclusions

Both Clarivate Analytics and Elsevier take measures to ensure the quality of the content indexed in WoS and Scopus respectively. However, the companies have taken different approaches to developing their databases: While the applied inclusion criteria do not differ to a large extent from the outset, their actual application and interpretation in the unobservable inclusion process result in a substantial difference between WoS and Scopus. In particular, the WoS SCIE, SSCI and A&HCI jointly index fewer journals than Scopus on the basis that Clarivate Analytics believes the majority of important research will be captured by covering the key journals in each field. ESCI was introduced in a bid to increase the coverage of emerging fields and areas of regional importance, which it has achieved to a certain extent however the index remains centred on Europe and North America. In contrast, Elsevier intends for Scopus to be the largest collection of high-quality content possible and accordingly applies a more liberal inclusion policy. These different business philosophies have resulted in differences in the coverage of journals between the databases, both in terms of the time-frames and number of journals covered, and differences also arise through the classification of journals to disciplines.

WoS has stronger historical content than does Scopus, while Scopus' higher coverage since 1970 tends to produce a higher number of items and citations in analyses than does WoS, and journals are also more leniently assigned to classification categories in Scopus. This means that, for example, the nursing category will include journals with looser affiliations to nursing and which may not follow the same citing and publishing practices as core nursing journals, which could have implications for field-normalised indicators. WoS tends to classify journals more accurately, which aligns with its intent to only cover core journals to each discipline. However, both databases overrepresent the natural and medical sciences, content in English, and content from Europe, North America, and Asia. The differences both between WoS and Scopus, and between these databases and third-party databases, such as Ulrichsweb or the Australian Research Council's list, should be taken into consideration when interpreting results from bibliometric analyses. Furthermore new data sources like *Dimensions* or *1Science* will bring up new and more comprehensive perspectives in the near future. Consequently the described difference in positioning Germany in the global science system by using either WoS or Scopus might be confirmed or challenged by these new sources.

4 National Bibliometric Indicators comparing Web of Science and Scopus

4.1 Indicators on Productivity

The number of publications produced continued to grow between 2007 and 2017 for most countries due to increased publication output and increased coverage of journals by WoS and Scopus. As shown in Figures 2 and 3, Scopus generally recorded a higher number of publications for each country or group of countries, however the ranking of countries based on number of publications is mostly consistent. China and the USA were the highest producing countries of publications worldwide during the reference period. In both databases, the number of Chinese publications increased dramatically since 2007 with an average growth rate of 13.9% per year between 2007 and 2017 in WoS and 9.4% between 2007 and 2017 in Scopus when the world average growth rate was 4-5% (see Table 1). Notably however, in Scopus, China overtook the USA to become the highest producing country in 2016.

The other notable difference in rankings between WoS and Scopus based on number of publications is India. As shown in Figure 3, Scopus recorded an additional 7,000-36,000 publications from India per year than WoS which placed India as the third-highest producing country behind China and the USA since 2014. In WoS India was ranked 6th since 2014. Aside from these differences, WoS and Scopus consistently recorded the United Kingdom, Germany and Japan as the highest producing countries of publications after the USA and China, and India in Scopus.

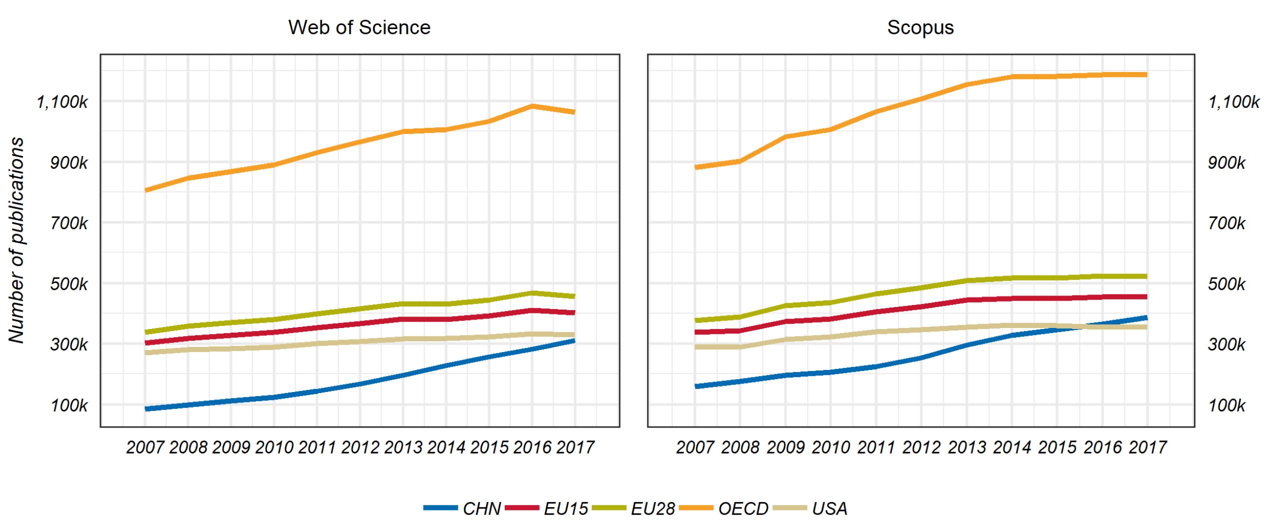


Figure 2 The fractional counts of publications from China, USA, and the EU15, EU28 and OECD countries.

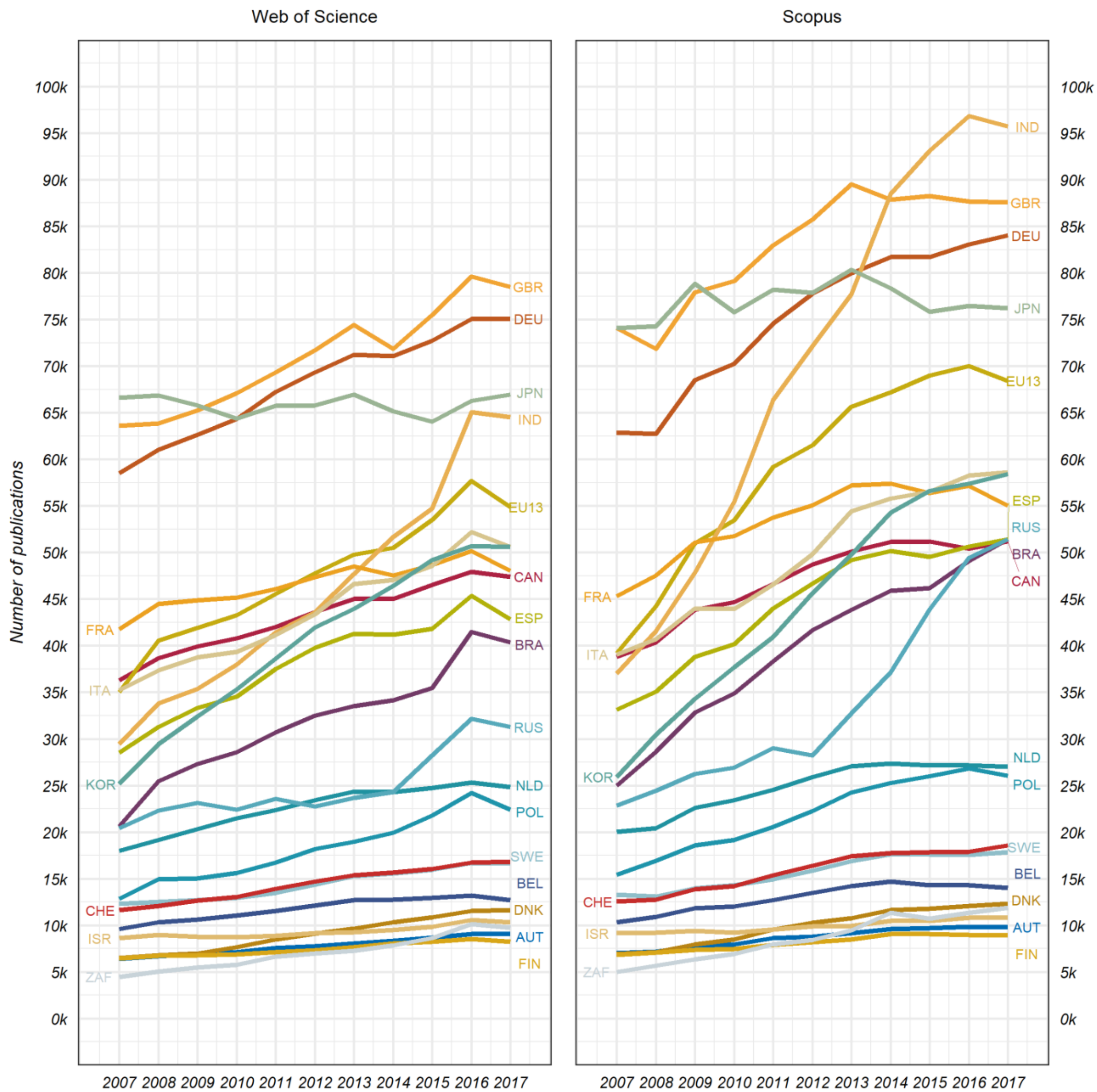


Figure 3 The fractional counts of publications from selected countries and groups between 2007 and 2017.

The surges in publication numbers from China and India resulted in very high average growth rates and also drew the world average growth rate up to 4.9% in WoS and 4.6% in Scopus, as presented in Table 1. Other countries with developing science systems such as South Africa, South Korea and Brazil also demonstrated strong rates of growth compared to the world average, while Denmark and Poland were the only European countries to exceed the world growth rate, although Spain was close. Countries with well-established systems such as the USA, the UK, Germany and France had amongst the lowest growth rates.

Table 1 The CAGR for selected countries and groups between 2007 and 2017 for Web of Science and Scopus, based on fractional counting.

COUNTRY	WoS	Scopus
AUT	3.61	3.38
BEL	2.85	3.12
BRA	6.96	7.48
CAN	2.71	2.80
CHE	3.76	3.99
CHN	13.92	9.36
DEU	2.53	2.94
DNK	6.00	5.91
ESP	4.15	4.49
FIN	2.52	2.66
FRA	1.41	1.96
GBR	2.13	1.69
IND	8.16	9.97
ISR	1.82	1.70
ITA	3.69	4.14
JPN	0.05	0.28
KOR	7.24	8.46
NLD	3.26	3.03
POL	5.71	5.37
RUS	4.34	8.44
SWE	3.07	2.98
USA	1.99	2.15
ZAF	8.15	9.01
EU13	4.59	5.73
EU15	2.86	3.02
EU28	3.05	3.33
OECD	2.83	3.03
WORLD	4.94	4.55

The strong growth in the number of Chinese publications has also caused a substantial increase in China's share of worldwide publications over time and corresponding decreases for most other countries in both WoS and Scopus (see Tables 2 and 3). However, Scopus' initial stronger coverage than WoS of Chinese publications has magnified this effect in WoS compared to Scopus. The proportion of worldwide publications from Chinese authors has increased by nearly 7 percentage points in Scopus from 12.0% to 18.8% between 2007 and 2017 compared to an increase of 10 percentage points in WoS from 8.1% in 2007 to 18.4% in 2017.

Accordingly, both databases recorded decreases in the shares from other countries that typically held high proportions of worldwide publications, although these decreases were emphasised in WoS. For example, the USA's share fell from 21.9% to 17.4% in Scopus but from 25.7% to 19.4% in WoS; the UK's share fell from 6.1% to 4.6% in WoS and from 5.6% to 4.3% in Scopus; and Germany's share fell from 5.6% to 4.4% in WoS and from 4.8% to 4.1% in Scopus. Japan in particular had a particularly large decrease in shares, from 6.4% to 4.0% in WoS and from 5.6% to 3.7% in Scopus, because in addition to the influence of increased Chinese publications, Japan's had no growth in publication counts over the reference period. Conversely, as with China, India's share in worldwide publications rose throughout the reference period from 2.8% to 3.8% in WoS and from 2.8% to 4.7% in Scopus, which suggests India substantially increased its production of publications or that both databases have increased their coverage of journals in which Indian authors publish.

Table 2 The shares of selected countries and groups of world publications between 2007 and 2017, based on fractional counting, from Web of Science.

COUNTRY	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUT	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.5	0.5
BEL	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.7
BRA	2.0	2.3	2.3	2.3	2.4	2.4	2.3	2.3	2.3	2.4	2.4
CAN	3.5	3.4	3.4	3.3	3.2	3.2	3.1	3.0	3.0	2.8	2.8
CHE	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.0	1.0	1.0	1.0
CHN	8.1	8.6	9.5	10.0	11.0	12.1	13.5	15.2	16.3	16.7	18.4
DEU	5.6	5.4	5.3	5.3	5.2	5.0	4.9	4.8	4.6	4.4	4.4
DNK	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7
ESP	2.7	2.8	2.8	2.8	2.9	2.9	2.8	2.8	2.7	2.7	2.5
FIN	0.6	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5
FRA	4.0	4.0	3.8	3.7	3.5	3.4	3.3	3.2	3.1	3.0	2.8
GBR	6.1	5.7	5.5	5.5	5.3	5.2	5.1	4.8	4.8	4.7	4.6
IND	2.8	3.0	3.0	3.1	3.2	3.2	3.3	3.5	3.5	3.8	3.8
ISR	0.8	0.8	0.7	0.7	0.7	0.7	0.6	0.6	0.6	0.6	0.6
ITA	3.4	3.3	3.3	3.2	3.2	3.2	3.2	3.1	3.1	3.1	3.0
JPN	6.4	5.9	5.6	5.3	5.0	4.8	4.6	4.4	4.1	3.9	4.0
KOR	2.4	2.6	2.7	2.9	3.0	3.1	3.0	3.1	3.1	3.0	3.0
NLD	1.7	1.7	1.7	1.8	1.7	1.7	1.7	1.6	1.6	1.5	1.5
POL	1.2	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.4	1.4	1.3
RUS	2.0	2.0	2.0	1.8	1.8	1.7	1.6	1.6	1.8	1.9	1.8
SWE	1.2	1.1	1.1	1.1	1.0	1.0	1.1	1.0	1.0	1.0	1.0
USA	25.7	24.9	24.0	23.6	22.9	22.3	21.7	21.2	20.5	19.6	19.4
ZAF	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6
EU13	3.3	3.6	3.6	3.5	3.5	3.5	3.4	3.4	3.4	3.4	3.2
EU15	28.9	28.2	27.7	27.5	27.0	26.7	26.3	25.4	24.9	24.2	23.6
EU28	32.2	31.8	31.3	31.1	30.5	30.2	29.8	28.8	28.3	27.7	26.9
OECD	76.8	75.1	73.5	72.7	71.2	70.2	68.9	67.2	65.9	64.0	62.7
WORLD	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Despite the difference in figures between Scopus and WoS consequent to variations in journal coverage, the results from the analysis are consistent: China and the USA continued to produce the highest number of publications and accounted for the greatest shares of world publications, followed by the UK, Germany, India and Japan. China and India's strong growth rates have increased their shares of world publications while nearly all other countries experienced corresponding declines in their shares, and emerging science systems continued to grow strongly while more established systems displayed less growth. Germany remained a strong contributor to the worldwide science system with high numbers and shares of publications, despite relatively low growth rates and the influence of China on world shares.

Table 3 The shares of selected countries and groups of world publications between 2007 and 2017, based on fractional counting, from Scopus.

COUNTRY	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUT	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
BEL	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.7	0.7
BRA	1.9	2.1	2.2	2.3	2.3	2.4	2.4	2.4	2.3	2.4	2.5
CAN	3.0	2.9	3.0	2.9	2.8	2.8	2.7	2.6	2.6	2.5	2.5

CHE	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
CHN	12.0	12.6	13.1	13.4	13.6	14.5	15.8	16.9	17.5	17.9	18.8
DEU	4.8	4.5	4.6	4.6	4.5	4.5	4.3	4.2	4.1	4.1	4.1
DNK	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
ESP	2.5	2.5	2.6	2.6	2.7	2.7	2.6	2.6	2.5	2.5	2.5
FIN	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.4
FRA	3.5	3.4	3.4	3.4	3.2	3.2	3.1	3.0	2.9	2.8	2.7
GBR	5.6	5.2	5.2	5.1	5.0	4.9	4.8	4.5	4.5	4.3	4.3
IND	2.8	3.0	3.2	3.6	4.0	4.1	4.2	4.6	4.7	4.8	4.7
ISR	0.7	0.7	0.6	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5
ITA	3.0	2.9	3.0	2.9	2.8	2.9	2.9	2.9	2.9	2.9	2.9
JPN	5.6	5.4	5.3	4.9	4.7	4.5	4.3	4.0	3.8	3.8	3.7
KOR	2.0	2.2	2.3	2.5	2.5	2.6	2.7	2.8	2.9	2.8	2.9
NLD	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.4	1.4	1.3	1.3
POL	1.2	1.2	1.2	1.2	1.2	1.3	1.3	1.3	1.3	1.3	1.3
RUS	1.7	1.8	1.8	1.8	1.7	1.6	1.8	1.9	2.2	2.4	2.5
SWE	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
USA	21.9	20.7	21.1	20.9	20.4	19.9	19.1	18.6	18.2	17.4	17.4
ZAF	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.6	0.5	0.6	0.6
EU13	3.0	3.2	3.4	3.5	3.6	3.5	3.5	3.5	3.5	3.4	3.3
EU15	25.6	24.7	25.1	24.8	24.4	24.2	23.8	23.2	22.7	22.3	22.1
EU28	28.6	27.9	28.6	28.3	28.0	27.7	27.4	26.7	26.2	25.7	25.5
OECD	67.1	65.0	66.1	65.4	64.3	63.4	62.1	60.9	59.8	58.3	58.0
WORLD	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Examining Germany's publications disaggregated by scientific discipline highlights those disciplines in which Germany is strongest in terms of number of publications, shares of worldwide publications and growth over time. As such, the annual fractional counts of German publications in each discipline from WoS are presented in Table 4 and from Scopus in Table 5. Germany's shares of worldwide publications per discipline are shown in Figure 4, and growth rates per discipline in Figure 5. To provide international context, the shares from China and the USA of worldwide publications per discipline are presented in Figures 6 and 7.

In both databases, clinical medicine had by far the highest number of publications, accounting for around 20% of all German publications. Germany also demonstrated a strong publishing trend in the natural sciences of biological sciences, physical sciences and astronomy, and chemical sciences, with these disciplines having the next highest number of publications. These well-established disciplines are stable over time with growth rates of between -0.6% and 1.6% per year between 2007 and 2017. Germany held between 4-6% of worldwide publications in these disciplines in 2017, however the shares for these disciplines have all decreased over time by 1-2 percentage points.

Table 4 The fractional count of German publications in each OECD discipline between 2007 and 2017, from Web of Science.

Discipline	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Agriculture, forestry, fisheries	547	700	663	642	650	645	707	669	697	711	709
Animal and dairy science	176	185	196	187	171	207	184	171	159	160	156
Art	90	115	185	186	189	199	174	175	169	188	206
Basic medical research	4402	4722	4815	4834	4691	4952	4913	4895	4932	4958	4892
Biological sciences	7303	7419	7301	7684	7695	7847	7654	7961	7800	7740	7611
Chemical engineering	584	539	570	651	649	635	668	633	710	752	794
Chemical sciences	5900	5906	6081	6238	6482	6528	6664	6639	6897	6717	6890
Civil engineering	383	368	372	382	418	439	426	478	514	551	535

Clinical medicine	13725	13988	14291	14570	14578	14539	14784	13986	14401	14959	14737
Computer and information sci.	784	846	863	872	1050	1020	1039	1095	1112	1147	1181
Earth and related environ. sci.	2311	2622	2669	2784	2890	2998	3274	3293	3517	3681	3494
Economics and business	722	907	1059	1147	1256	1436	1583	1658	1763	1943	1869
Educational sciences	170	207	192	249	262	249	309	338	354	393	409
Electrical eng., electronic eng.	838	879	911	886	954	997	998	1034	1096	1151	1181
Environmental biotechnology	534	489	470	583	571	569	642	696	699	634	683
Environmental engineering	368	372	443	433	597	573	735	799	892	1026	1039
Health sciences	1464	1600	1643	1765	1841	1966	2022	2124	2136	2275	2370
History and archaeology	236	294	343	371	406	417	431	428	441	488	489
Industrial biotechnology	84	92	91	126	104	123	189	181	181	193	191
Languages and literature	305	423	435	458	580	521	595	568	576	708	682
Law	76	110	126	142	172	166	109	95	120	170	120
Materials engineering	2044	1999	2082	2123	2297	2266	2203	2336	2428	2503	2563
Mathematics	2015	2171	2228	2072	2313	2345	2442	2415	2599	2583	2572
Mechanical engineering	983	959	1076	960	1123	1073	1182	1164	1260	1331	1399
Media and communications	81	88	100	90	103	108	124	141	159	212	204
Medical engineering	380	379	403	441	456	456	473	487	472	512	486
Nano-technology	234	285	283	304	344	350	363	428	476	486	465
Other agricultural sciences	295	298	311	332	364	366	349	370	369	393	354
Other eng. and technologies	1084	1098	1161	1153	1219	1273	1204	1269	1267	1355	1397
Other humanities	54	67	68	61	99	89	66	97	65	153	91
Other natural sciences	460	575	541	833	1388	2124	2674	2700	2858	3214	3570
Other social sciences	60	77	103	115	127	126	169	186	170	267	219
Philosophy, ethics and religion	168	214	245	263	269	324	309	336	409	415	420
Physical sciences and astronomy	7253	7472	7541	7568	7938	8091	7942	7568	7246	7109	6951
Political science	290	285	343	343	310	371	362	417	430	525	512
Psychology	933	1056	1118	1173	1314	1411	1595	1718	1739	1788	1863
Social and economic geography	171	190	253	298	318	342	435	458	554	613	644
Sociology	246	320	314	321	364	401	444	438	506	513	515
Veterinary science	593	567	596	551	567	625	558	496	463	466	499
Overall	58346	60883	62485	64191	67119	69167	70994	70940	72636	74983	74962

Table 5 The fractional count of German publications in each OECD discipline between 2007 and 2017, from Scopus.

Discipline	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Agriculture, forestry, fisheries	914	912	934	1019	1210	1519	1668	1623	1592	1530	1601
Animal and dairy science	568	529	521	505	553	558	537	543	530	525	552
Art	28	50	73	76	90	150	109	106	123	130	123
Basic medical research	4378	4385	4230	4341	4423	4616	4628	4476	4516	4698	4524
Biological sciences	7846	7770	7914	7989	8400	8835	8845	8768	8688	8528	8569
Chemical engineering	864	830	843	884	938	936	1022	1105	1166	1425	1331
Chemical sciences	4955	5040	5422	5381	5591	5627	5802	5776	5970	5899	6199
Civil engineering	146	168	172	177	196	200	207	250	303	306	413
Clinical medicine	15592	14726	16446	17121	17805	17915	18305	17899	17868	17963	17932
Computer and information sci.	1133	1292	1467	1419	1589	1567	1622	1781	1676	1704	1893
Earth and related environ. sci.	3780	4162	4378	4767	4879	5169	5402	5586	5666	5756	5759
Economics and business	972	1095	1395	1590	1832	1915	2093	2260	2164	2306	2214
Educational sciences	184	229	280	333	337	387	487	504	496	526	589
Electrical eng., electronic eng.	641	647	644	620	674	677	723	746	811	772	785
Environmental engineering	66	84	90	95	132	144	177	209	235	261	309
Health biotechnology	87	87	115	129	143	383	194	164	168	181	191

Health sciences	1631	1582	1839	1950	2151	2258	2293	2369	2369	2484	2570
History and archaeology	207	212	300	334	365	416	421	446	524	568	525
Languages and literature	246	299	448	486	598	641	781	704	642	772	785
Law	139	123	203	238	273	261	283	280	287	297	318
Materials engineering	2035	2106	2394	2403	2541	2569	2580	2736	2859	2777	2906
Mathematics	2132	2290	2504	2335	2625	2694	2828	2923	2912	2964	2949
Mechanical engineering	634	631	717	726	791	875	910	942	992	1114	1105
Media and communications	22	29	31	39	46	55	53	74	89	103	111
Medical engineering	325	269	231	240	257	518	310	297	305	316	324
Other agricultural sciences	86	61	32	29	47	49	57	48	40	44	47
Other eng. and technologies	720	731	880	873	1002	1052	1032	1122	1257	1508	1526
Other humanities	125	149	191	194	235	275	258	304	294	315	342
Other medical sciences	1112	1007	1365	1604	1360	1462	1709	2886	2174	1345	1102
Other natural sciences	754	770	918	919	987	1042	1113	1155	1219	1304	1421
Other social sciences	727	784	864	851	994	1058	1089	1137	1155	1301	1318
Philosophy, ethics and religion	137	164	230	278	364	374	428	575	556	530	590
Physical sciences and astronomy	7019	7086	7367	7143	7663	7892	7894	7662	7469	7537	7256
Political science	127	107	185	216	193	214	254	284	310	309	317
Psychology	1068	1028	1149	1198	1424	1496	1685	1719	1695	1850	1957
Social and economic geography	261	200	213	230	266	273	315	318	383	373	389
Sociology	473	447	618	664	723	780	851	968	957	970	1029
Veterinary science	492	409	563	574	550	551	512	500	450	427	486
Overall	62626	62490	68166	69970	74247	77403	79477	81245	80910	81718	82357

This decrease may be due to the influence of China which has expanded its share of publications in all disciplines in WoS and all but seven disciplines in Scopus during the reference period. For instance, China increased its share of worldwide publications in clinical medicine by 75.7% in WoS and 44.7% in Scopus over the reference period. Other large increases were observed for basic medical research (78.8% in WoS, 60.6% in Scopus), biological sciences (66.1% in WoS, 59.9% in Scopus), and chemical sciences (37.7% in WoS, 35.6% in Scopus). The USA's shares of worldwide publications have also been affected by China's increased output, with decreases in shares in every discipline across both databases between 2007 and 2017, except for the disciplines of art, languages and literature, and other medical sciences in Scopus.

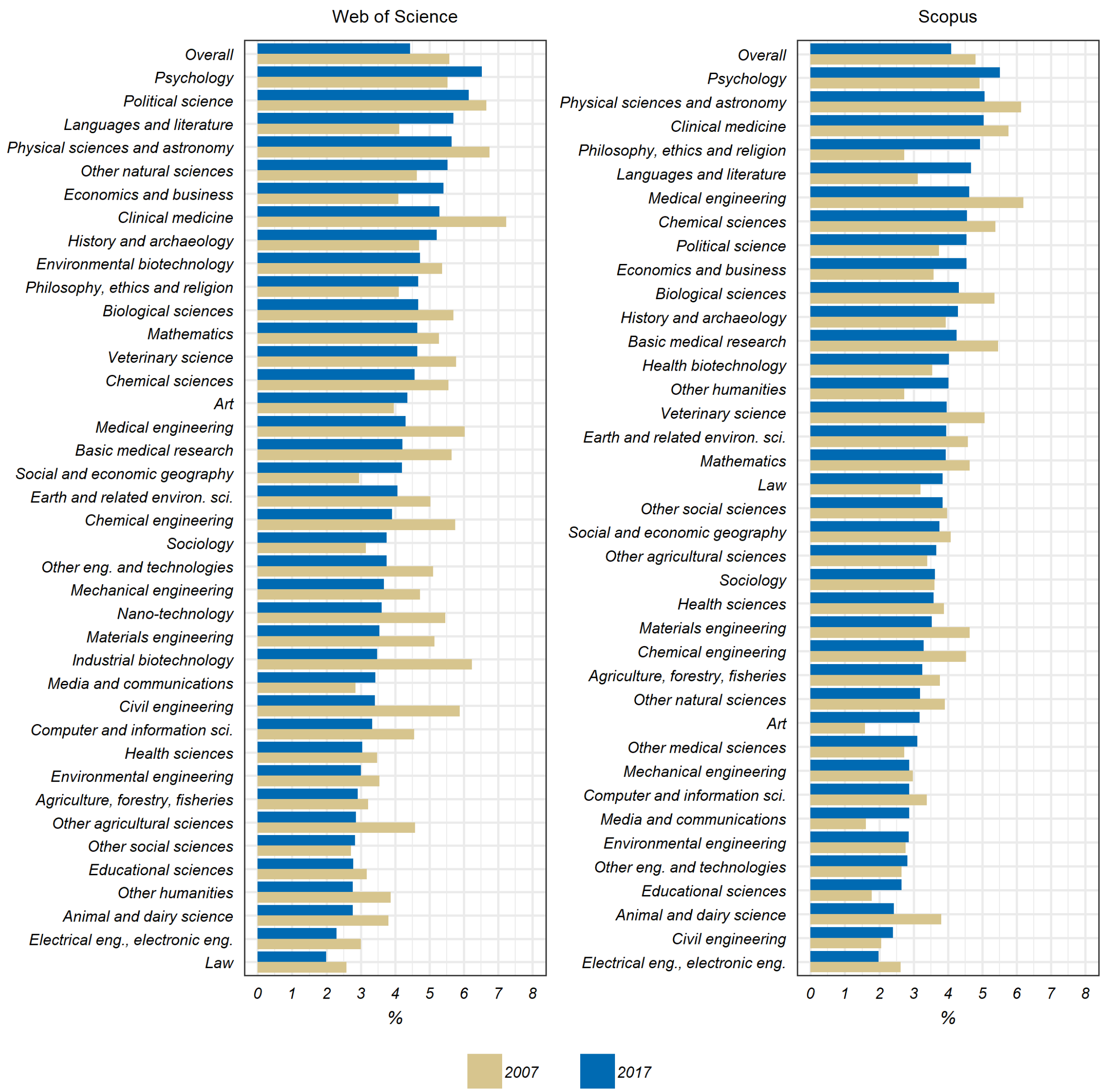


Figure 4 The German share of world publications in 2007 and 2017 in each OECD discipline based on fractional counting.

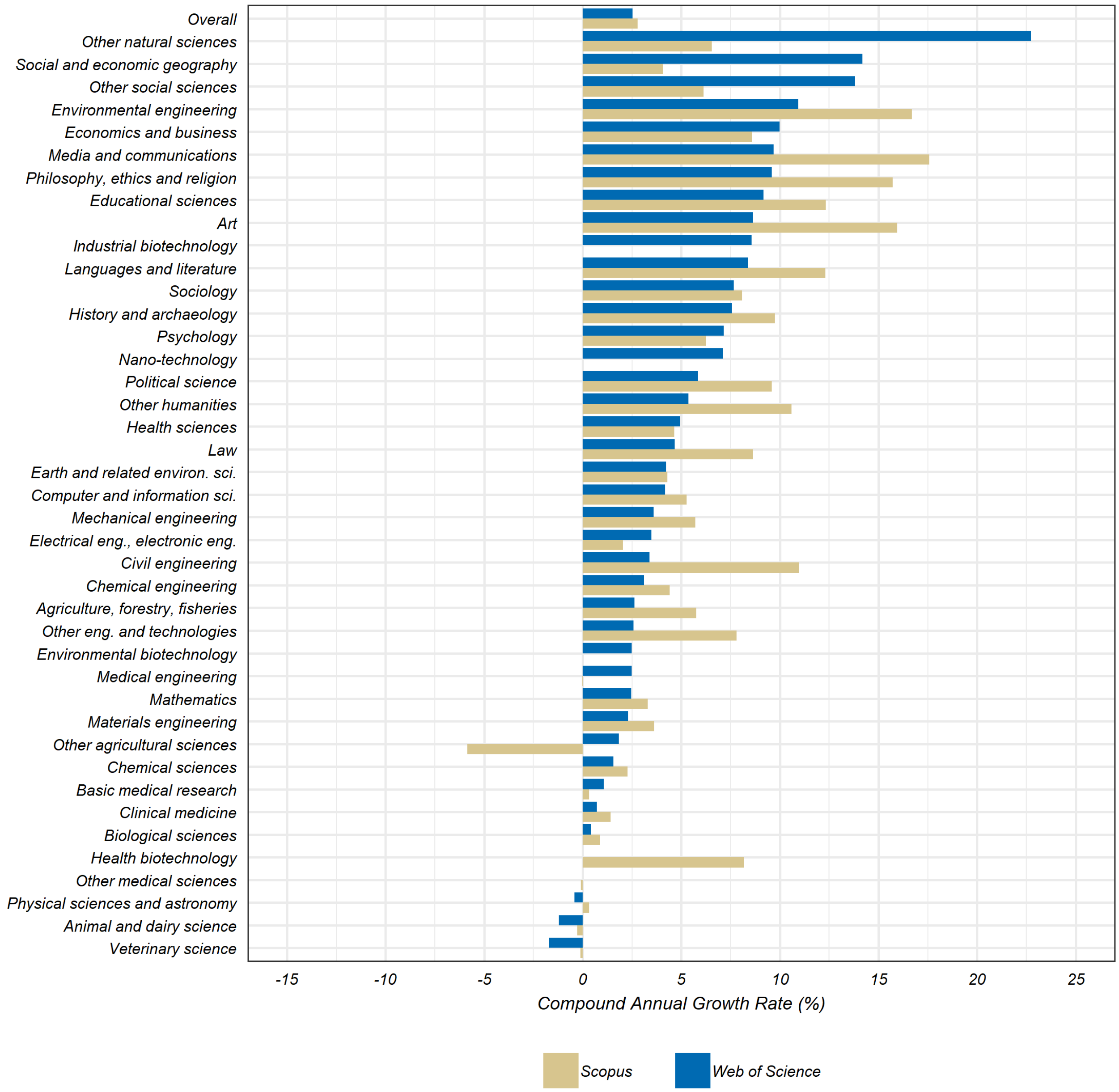


Figure 5 The CAGR of German publications by OECD discipline between 2007 and 2017 for Web of Science and Scopus, based on fractional counting.

The USA's largest contributions to the science system came from the 'soft' sciences, where it accounted for up to 40% of worldwide shares of publications in disciplines such as psychology, sociology, law, and media and communications. Germany held its highest shares in the soft sciences of psychology (6.5% of worldwide publications), political sciences (6.1%), and languages and literature (5.7%) in WoS, and in psychology (5.6%), physical sciences and astronomy (5.3%), and clinical medicine (5.0%) in Scopus. Conversely, China's largest contributions were in engineering and technology disciplines. Scopus' stronger and earlier coverage of Chinese publications is evident in Figure 6 where much larger increases in shares per discpline are seen between 2007 and 2017 in WoS than in Scopus, and Chinese publications in Scopus account for higher proportions of worldwide publications in both years.

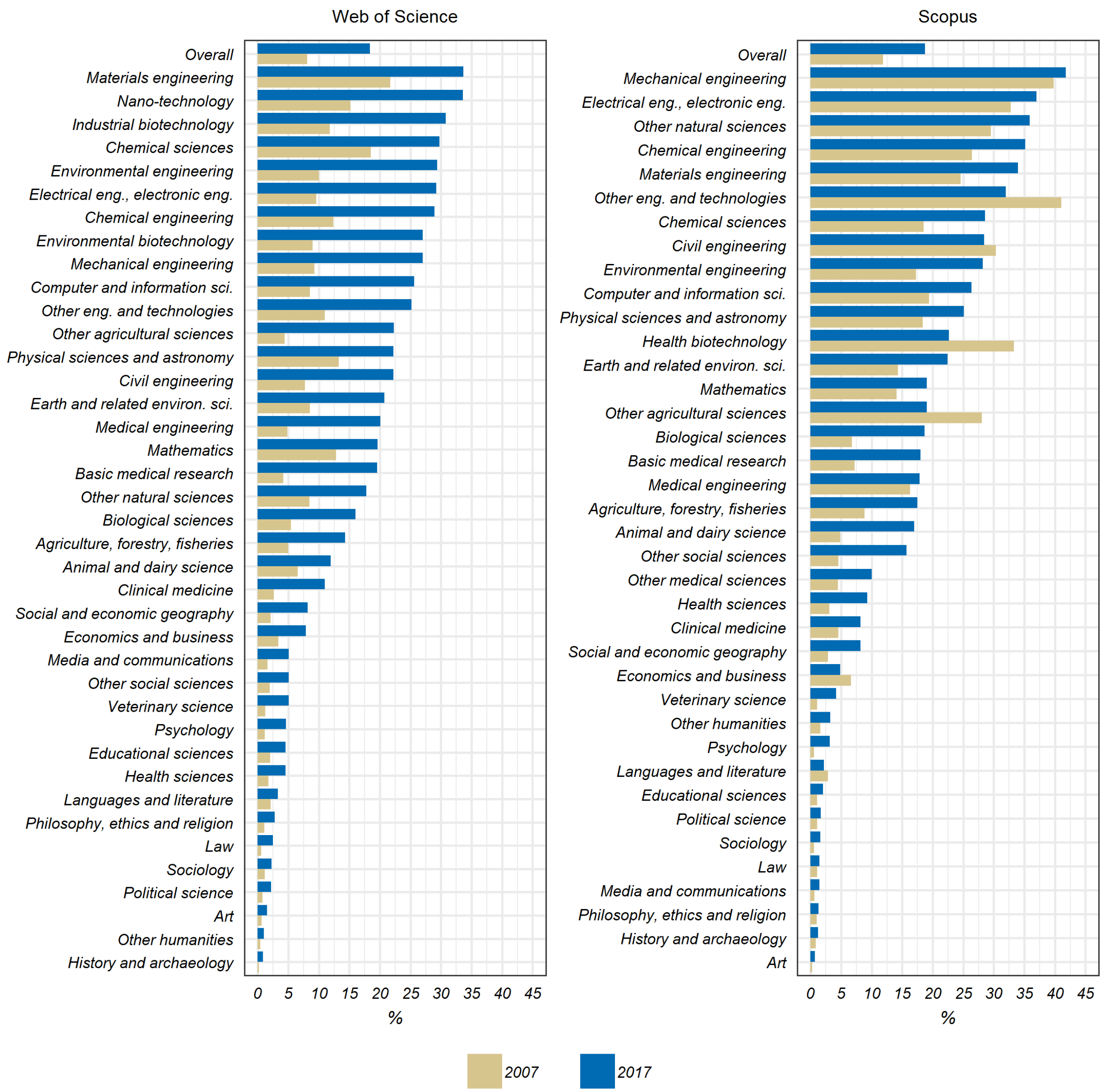


Figure 6 The Chinese share of world publications in 2007 and 2017 in each OECD discipline based on fractional counting.

The growth rates of disiplines should be interpreted cautiously, particularly between databases. Disciplines with the strongest growth rates in both databases often had a low base of publications in 2007 on which to build. Also the fastest-growing disciplines in Scopus were primarily from the arts and humanities so their high growth is likely partially due to greater coverage of arts and humanities journals in the database over time. Further, there are differences in the mapping of the internal classification structures of the databases to the FOS structure presented here. For example, the discipline 'other natural sciences' had the highest growth rate in WoS where publications in 'multidisciplinary sciences' are assigned to this discipline, while 'other natural sciences' had a mid-range growth rate in Scopus where it is based on publications in the ASJC categories energy, fuel and materials sciences. These differences in disciplines and databases should be considered when analysing these results.

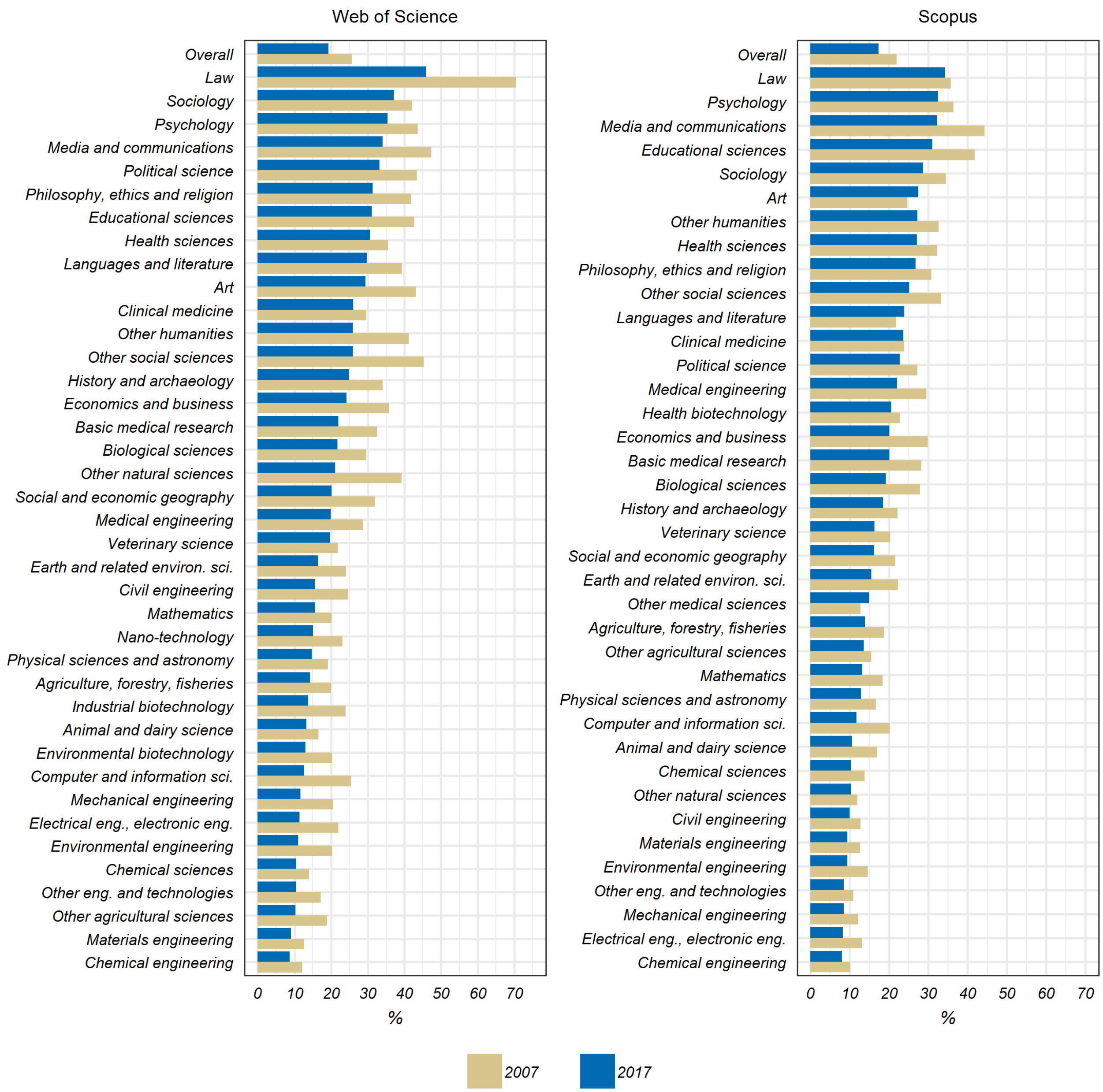


Figure 7 The US share of world publications in 2007 and 2017 in each OECD discipline based on fractional counting.

4.2 Impact Indicators

The following section presents and discusses the Scientific Regard (SR), International Alignment (IA), and Excellence Rate (ER) indicators by country and group of countries, and the proportions of uncited publications, publications cited more frequently than the median or 75th quartile by country and group. We also present the ER for countries when only literature published in English is considered. Together these indicators provide information about Germany's performance within the science system in relation to other countries and groups of countries.

4.2.1 Excellence Rates

ER is the proportion of a country's publications that are in the 10% most highly cited publications of each scientific discipline, or the proportion of publications that could be considered 'excellent'. An ER of 10% is the expected rate and higher values indicate a country had a higher proportion of publications in the subset of 'excellent' publications and performed better than expected. The ER for each country and group is presented in Figures 8 and 9. ERs were consistently higher by up to 3 percentage points for nearly all countries in Scopus than in WoS. For example, Germany's ER in WoS between 2007 and 2014 was consistently between 11% and 12% while in Scopus it was consistently between 12% and 13%. The exceptions were China, which had lower ERs by up to 3 percentage points in Scopus, and India, Poland and Russia which had lower ERs by up to 1 percentage point in Scopus.

Although there were differences in the ER figures, the patterns of countries' rankings was similar. In both databases, Switzerland had the highest ERs with more than 14% of its publications in the 'excellent' band in 2014/2015, followed by the USA, the Netherlands, the United Kingdom, and Denmark, all with 13-14%. Several countries had less ERs of less than 10%, including Poland, Israel, Korea, Japan, and the BRICS countries, with Brazil and Russia lowest in WoS at 3.8% and 2.7% respectively, and lowest in Scopus were Brazil and Poland at 4.6% and Russia at 3.0% in 2015.

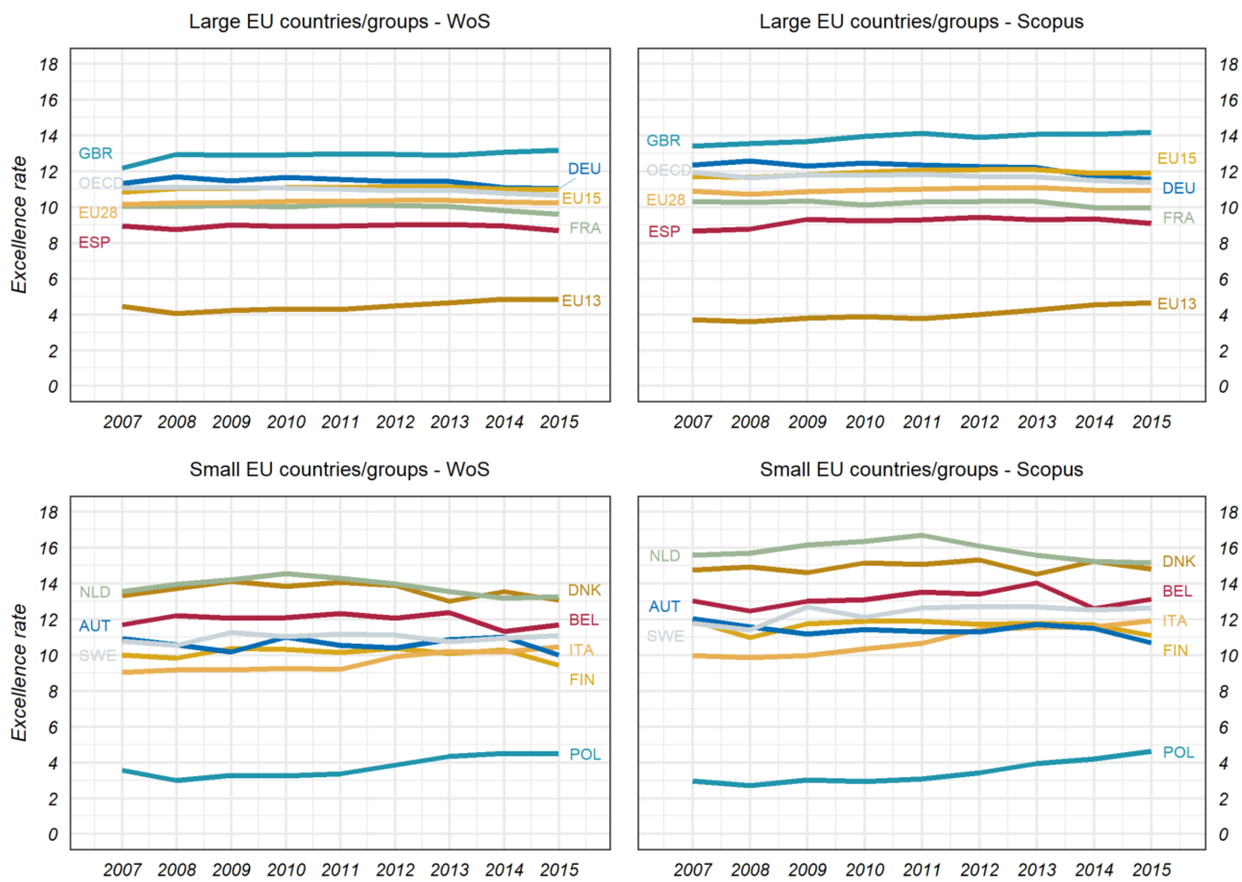


Figure 8 The Excellence Rates for EU and OECD countries and groups, based on fractional counting, from Web of Science and Scopus, 2007-2015.

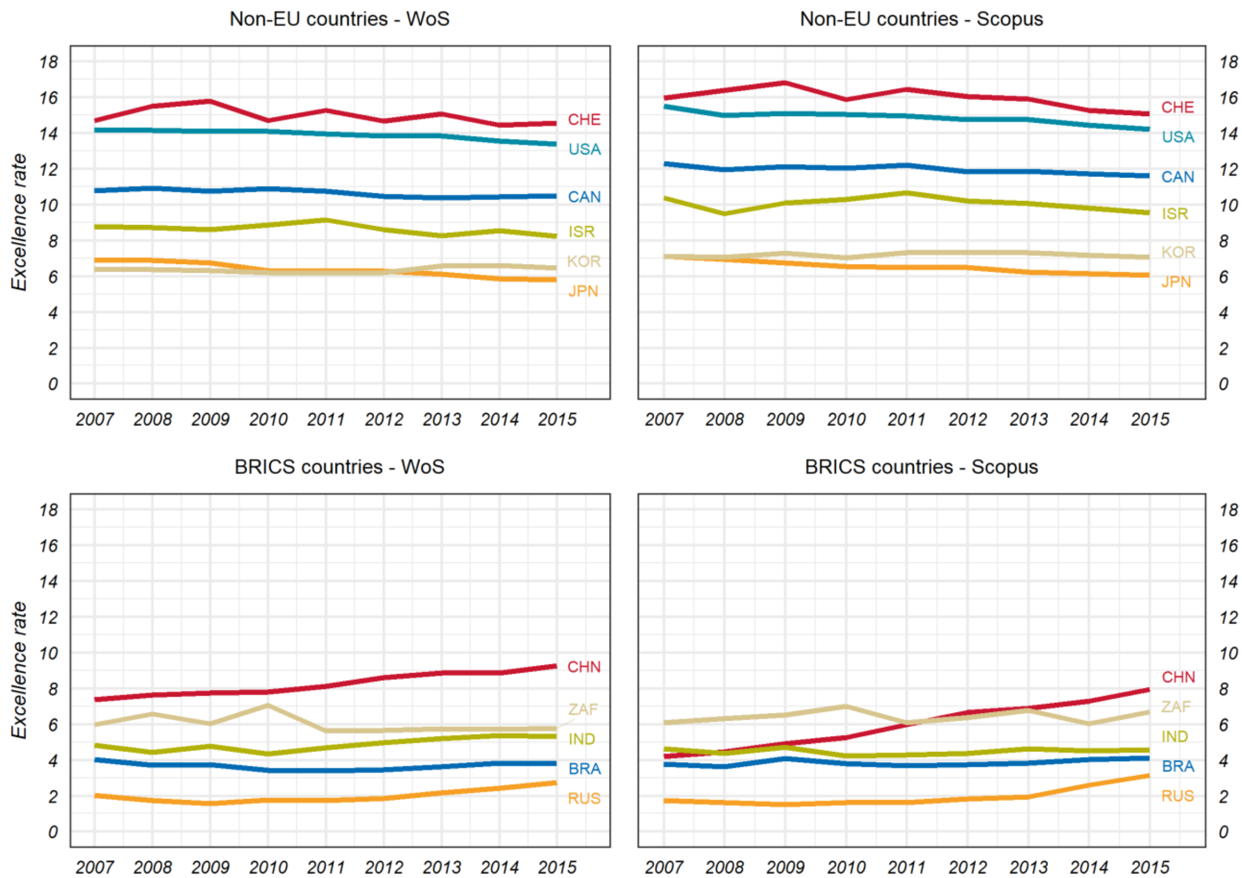


Figure 9 The Excellence Rates for non-EU and BRICS countries, based on fractional counting, from Web of Science and Scopus, 2007-2015.

Although most publications are published in English to ensure the broadest audience possible, many countries also publish a substantial share of their publications in their local languages. As English is generally considered to be the lingua franca in science allowing the respective content of such publications to be carried and understood across borders, non-English publications might face a language barrier restricting their dissemination and citability irrespective of the content. To gauge whether publishing in English influences the ER of the country, the ER for each country for only English publications in WoS was calculated and is presented in Figure 10 and can be compared to Figures 8 and 9 which include publications published in any language in WoS.

A negative change in ER when considering English-only publications indicates that publications in the local language are more often considered ‘excellent’ in terms of citations than are publications in English or, in the case of English-speaking countries, that the environment has become more competitive. Conversely, a positive change in ER for only English publications indicates that the country’s English publications attract more citations than those in the local language. Especially for non-English speaking countries with a relatively large national science system this observation indicates that the general ER statistics are held back by the aforementioned language barrier. For most countries, the influence of publishing in local languages affected their ERs very little, as shown in Figure 11. Only Germany, France, Spain and China saw a positive increase of 0.5-1 percentage points in their ERs when publishing in English compared to other languages.

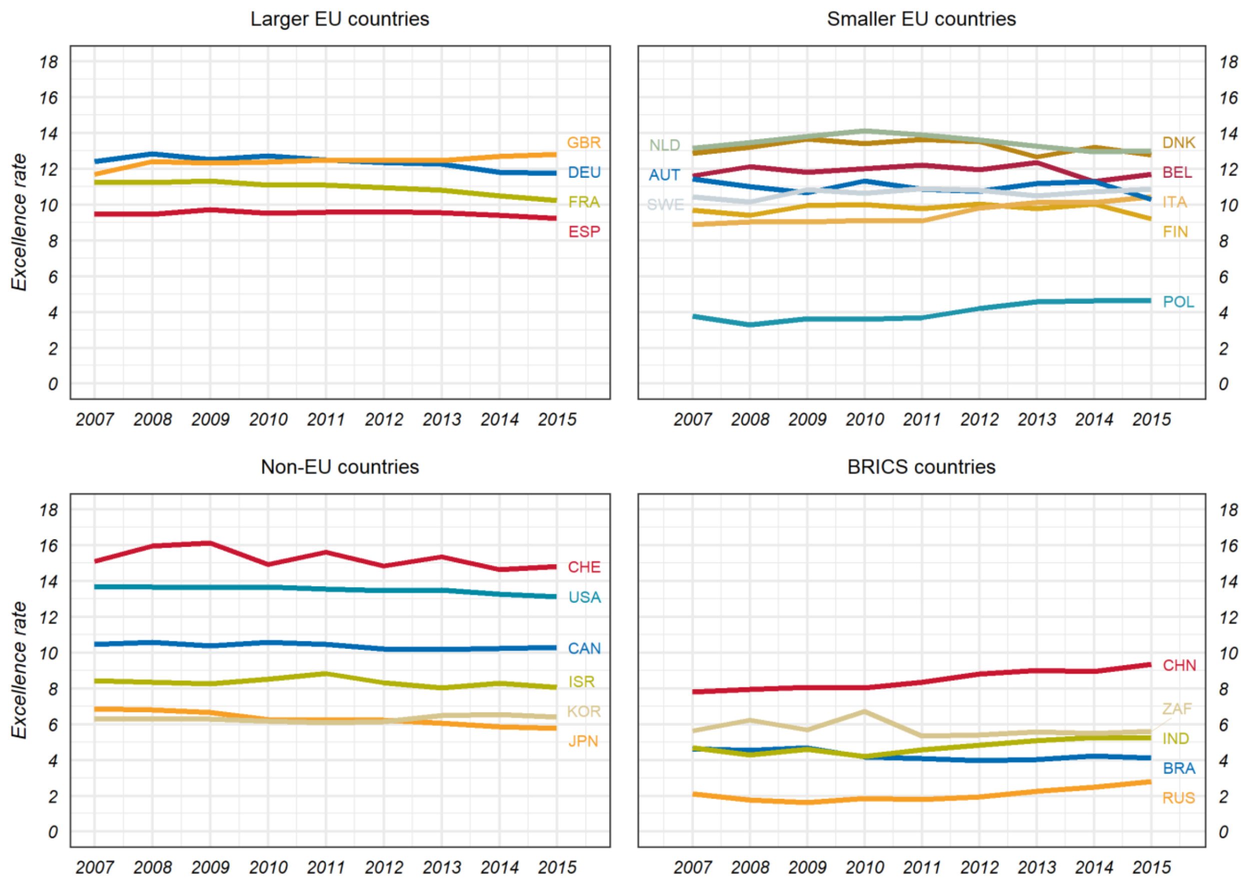


Figure 10 The Excellence Rates for selected countries and groups based on fractional counting of papers written in English from Web of Science (2007-2015).

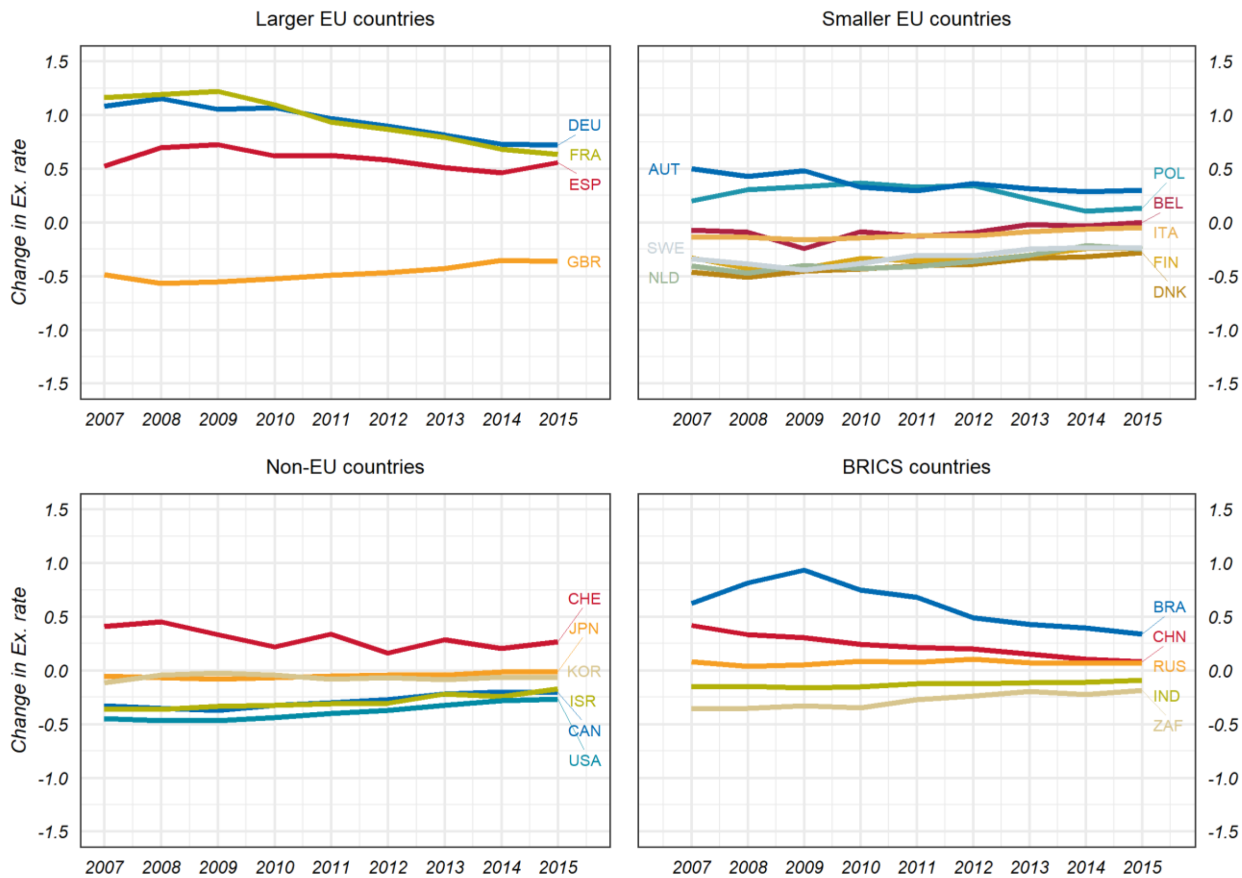


Figure 11 The change in Excellence Rates when only English publications are considered for selected countries based on fractional counting from Web of Science (2007-2015).

The ER for German publications disaggregated by discipline in 2007 and 2015 for WoS and Scopus are presented in Figure 12. The ERs are the proportion of publications in each discipline that are considered 'excellent' as measured by the number of citations and identify disiplines in which German publications are well- or poorly-cited.

In both databases the majority of disciplines attained or exceeded the expected threshold of 10% of publications in the excellent band, however there was little agreement between WoS and Scopus with regard to either the ER of a discipline or its ranking. In Scopus, the highest performing disciplines were other agricultural sceinces, educational sciences, biological sceinces, and political sciences with ERs of more than 12%. In WoS, other humanities (which consisted of multidisciplinary humanities publications), agriculture, forestry and fisheries, and biological sciences had the highest ERs at more than 13%.

Conversely, in Scopus several disciplines did not reach the expected 10% threshold: mechanical engineering, vertinary sciences, other medical sciences, philosophy, ethics and religion, health sciences and law had the lowest share of excellent publications. Law, mechanical egineering and philosophy, religion and ethics were also amongst the lowest ER rates in WoS, along with chemical engineering, civil engineering and sociology.

These inconsistencies in ERs reflect a number of differences between WoS and Scopus, including variations in coverage of disciplines between the databases, the more selective acceptance of journals into WoS which tends to raise the number of citations required to exceed the 90th percentile citations threshold, and the differences in the composition of the disciplines due to variations in mapping to the FOS disciplines from WoS and Scopus' individual subject classifications.

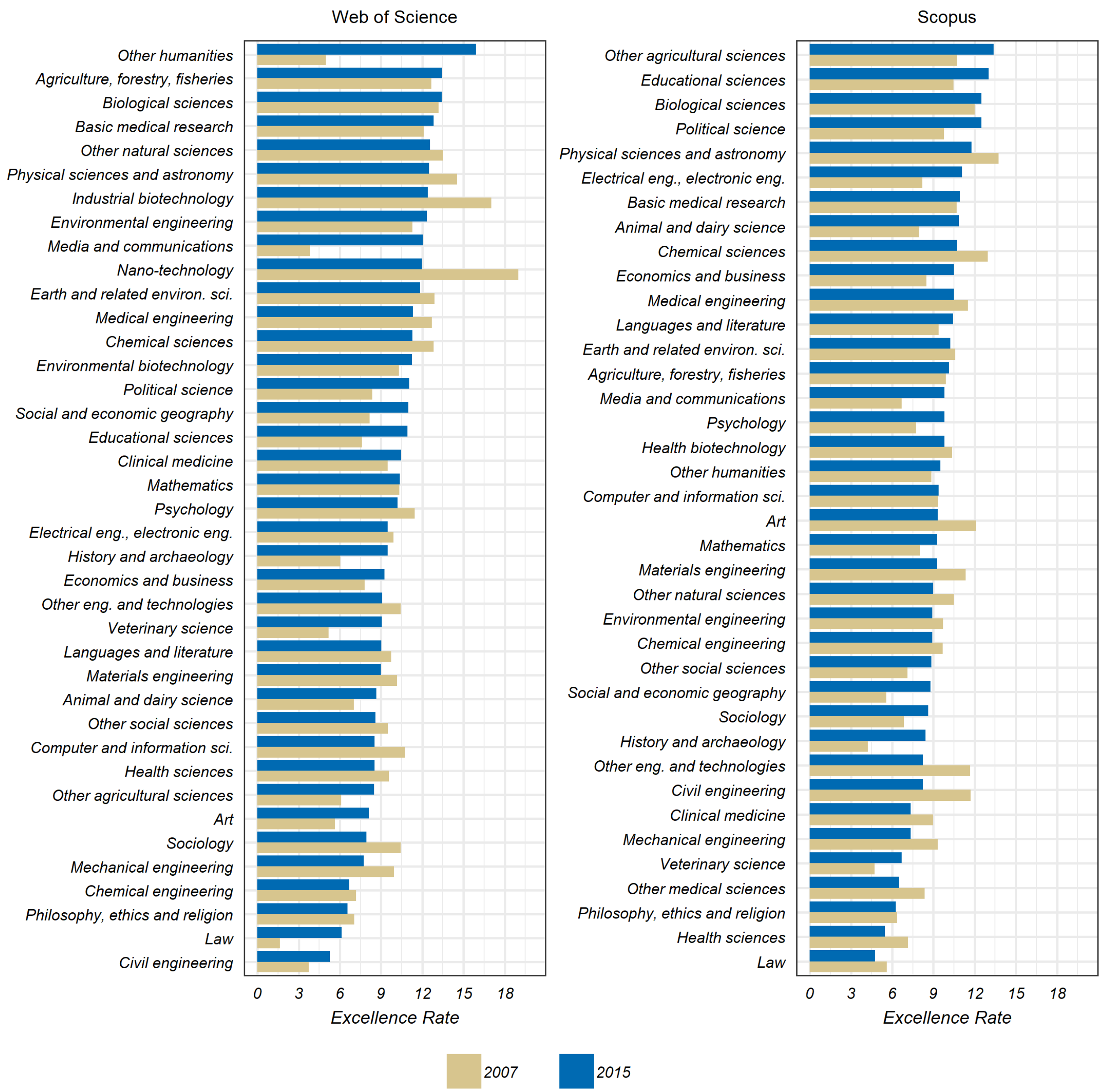


Figure 12 The Excellence Rate of German publications by OECD discipline, 2007 and 2015, based on fractional counting.

4.2.2 Additional Segmentations

While ERs assist in identifying high-performing countries or disciplines, they provide no information regarding performance across the remaining 90% of the citation distribution. To provide a more complete overview of Germany's performance across the citation distribution, the proportion of publications which received no citations during the 3-year citation window for each country are presented in Tables 6 and 7, and the proportions of publications which were cited more often than the median number of citations or the 75th percentile citations are shown in Figures 13 to 16.

Table 6 The percentage of uncited publications from selected countries and groups between 2007 and 2015, based on fractional counts, from Web of Science.

Country	2007	2008	2009	2010	2011	2012	2013	2014	2015
AUT	20.2	21.2	20.4	18.2	18.8	18.0	17.2	15.6	15.6
BEL	18.9	18.7	17.1	18.0	17.4	17.0	16.2	15.5	15.0
BRA	31.8	33.8	33.5	32.0	32.0	31.7	29.3	27.1	25.8
CAN	19.8	19.2	19.0	18.1	17.6	16.9	16.4	15.6	16.0
CHE	15.9	16.1	14.9	14.7	13.8	13.3	12.7	12.3	13.0
CHN	28.2	25.7	24.2	23.0	22.1	20.4	19.4	18.1	16.8
DEU	20.9	20.1	19.5	18.3	18.1	17.0	16.6	15.8	15.2
DNK	13.9	13.5	13.1	12.6	11.9	11.8	11.8	10.8	11.0
ESP	21.0	22.1	21.5	20.2	20.1	19.6	19.4	18.2	18.9
FIN	16.8	17.2	15.4	15.2	14.7	14.2	14.4	13.6	14.0
FRA	22.9	23.2	22.2	21.4	20.0	19.5	18.4	17.3	17.4
GBR	20.0	19.2	18.5	18.2	18.3	17.7	17.6	16.0	16.0
IND	33.6	34.2	30.7	29.5	28.0	25.8	24.9	21.8	21.3
ISR	22.0	21.6	21.9	20.8	20.1	19.9	20.6	18.8	19.5
ITA	19.7	19.4	18.6	17.9	17.0	15.5	14.9	14.1	13.7
JPN	22.4	21.3	20.9	20.8	20.1	19.5	19.3	18.3	18.0
KOR	24.6	25.0	25.5	24.1	23.8	22.1	20.8	20.1	19.4
NLD	14.2	14.2	13.4	12.4	12.4	11.8	11.7	11.3	11.3
POL	35.7	39.5	36.0	34.5	31.6	30.4	26.1	24.0	23.4
RUS	52.3	53.0	52.0	49.7	48.3	45.2	40.5	37.3	34.8
SWE	15.6	15.1	14.4	13.7	14.1	14.2	13.4	12.0	12.5
USA	17.9	17.7	17.2	16.7	16.3	16.2	15.7	15.2	15.5
ZAF	33.2	32.4	31.6	29.0	32.4	30.5	28.2	27.5	28.3
EU13	34.5	37.6	35.6	33.8	32.2	30.9	28.4	26.1	25.6
EU15	19.5	19.5	18.7	17.9	17.5	16.8	16.3	15.2	15.2
EU28	20.9	21.5	20.6	19.7	19.1	18.4	17.7	16.5	16.4
OECD	19.8	20.0	19.4	18.8	18.4	17.9	17.2	16.3	16.4

The proportion of uncited publications fell over time for all countries in WoS, which is likely in part due to increased indexing of journals. This occurs as citations are based on the items indexed in each database and so the larger number of journals indexed over time increases the likelihood of any publication having been cited by the indexed items. In 2015, 15.2% of Germany's publications were uncited, ranking it eighth of the individual countries. Denmark had the lowest proportion of uncited publications at 11.0%, followed by the Netherlands at 11.3%, and Sweden at 12.5%. Conversely, 34.8% of Russia's publications were uncited (down from 52.3% in 2007), along with 25.8% of publications from Brazil and 23.4% from Poland.

In Scopus data also the proportion of uncited publications decreases for most countries between 2007 and 2015. Most countries also had a higher proportion of uncited publications in Scopus than in WoS, likely due to the less selective inclusion of journals in Scopus. The ranking of countries was similar between the databases, with the lowest proportion of uncited publications coming from the Netherlands, Finland, Sweden and Denmark, and the highest from Russia, India, Poland and Brazil in 2015 in Scopus.

Table 7 The percentage of uncited publications from selected countries and groups between 2007 and 2015, based on fractional counts, from Scopus.

Country	2007	2008	2009	2010	2011	2012	2013	2014	2015
AUT	22.0	22.3	21.8	20.9	21.2	20.5	20.1	19.6	19.5
BEL	20.1	19.0	19.3	19.3	19.3	18.6	18.6	19.0	17.7
BRA	33.2	32.8	32.8	32.0	32.4	32.9	32.4	31.5	31.0
CAN	18.8	17.8	18.4	17.9	17.0	16.7	16.3	16.6	16.5
CHE	18.9	17.4	16.8	16.8	16.5	16.1	16.2	16.5	16.3
CHN	48.3	43.9	41.3	38.6	34.8	32.3	32.2	29.7	27.1
DEU	22.5	20.4	21.4	20.7	20.5	20.2	19.8	20.4	19.9
DNK	13.9	13.3	15.5	14.4	14.3	13.8	13.6	12.6	11.8
ESP	25.4	24.5	24.6	24.1	24.0	23.8	23.9	24.1	24.2
FIN	15.0	15.5	14.9	14.5	15.0	13.9	14.2	14.5	14.2
FRA	26.5	25.0	26.2	26.5	26.2	25.2	24.8	24.9	24.2
GBR	21.7	20.2	20.9	20.0	19.9	19.1	19.1	18.9	18.4
IND	37.7	36.5	35.4	35.1	35.8	35.1	34.0	34.7	36.4
ISR	20.9	20.0	20.0	19.0	18.5	19.1	18.9	18.7	19.1
ITA	22.2	20.6	21.5	19.8	19.5	18.2	17.6	17.5	17.2
JPN	26.4	25.3	28.0	28.0	28.0	26.2	26.1	25.3	24.6
KOR	25.1	24.7	24.6	23.5	23.0	21.7	21.3	22.5	22.0
NLD	15.5	14.6	15.0	14.2	13.8	13.0	13.6	13.7	13.6
POL	41.1	41.4	39.7	38.9	36.9	34.4	33.2	30.7	28.3
RUS	59.1	58.3	56.8	54.0	52.1	48.1	45.2	44.1	41.2
SWE	14.9	13.8	14.6	15.0	14.7	13.9	13.4	13.4	13.8
USA	18.4	17.2	17.9	17.6	17.3	17.0	16.7	16.9	16.8
ZAF	31.7	31.5	32.0	30.6	32.2	30.8	31.5	35.4	30.1
EU13	39.8	39.8	39.4	38.4	39.0	36.7	35.6	33.6	33.0
EU15	21.8	20.5	21.2	20.6	20.4	19.6	19.4	19.5	19.0
EU28	23.6	22.7	23.3	22.7	22.7	21.7	21.4	21.3	20.9
OECD	21.7	20.7	21.5	21.0	20.8	20.1	19.8	19.8	19.5

When looking at performance at the mid-range, 54% of Germany's publications in WoS in 2015 were cited more frequently than the median number of citations, which was slightly more than expected and similar to the proportions recorded for Austria, Finland and Italy. Sixty percent of Dutch publications were cited above the median, as were 59% of Swiss and Danish publications. The BRICS countries, the collective EU13 countries, Israel, Poland, Japan and Korea all had fewer than the expected level of half of their publications cited above the median, with Russia trailing at 26.0%.

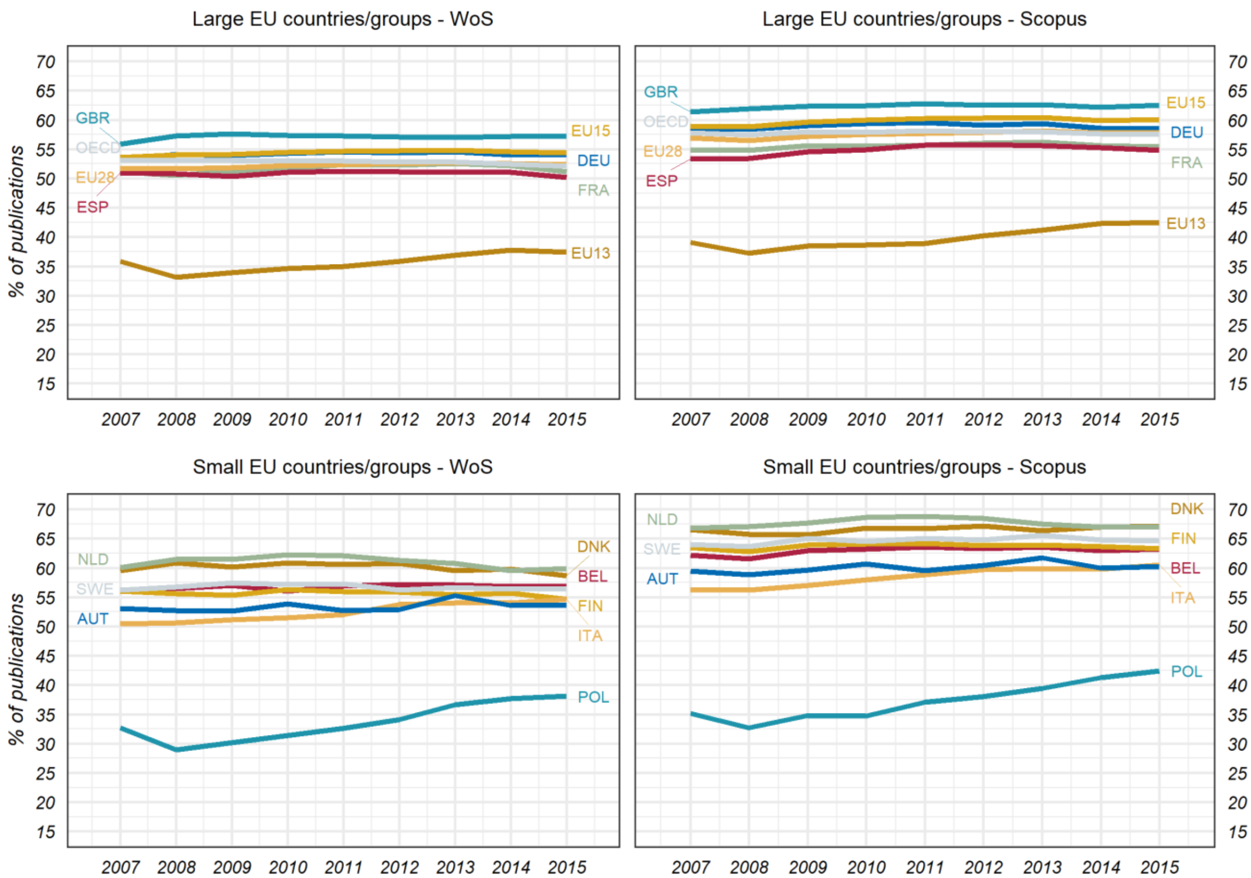


Figure 13 The percentage of publications from EU and OECD countries and groups which were cited more frequently than the median, based on fractional counting, from Web of Science and Scopus, 2007-2015.

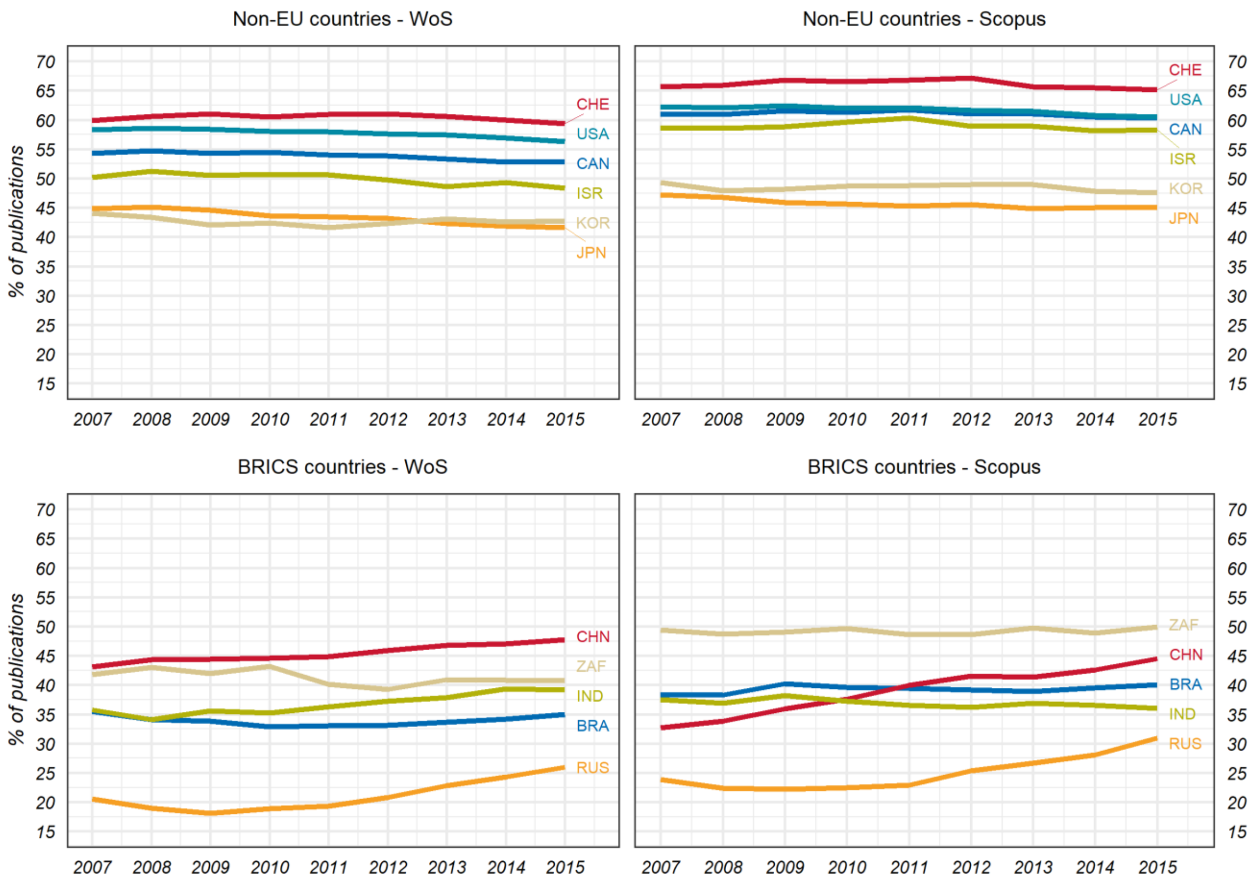


Figure 14 The percentage of publications from non-EU and BRICS countries which were cited more frequently than the median, based on fractional counting, from Web of Science and Scopus, 2007-2015.

The rankings were similar for publications cited more frequently than the 75th quartile. Russia had only 9% of publications cited above the 75th quartile compared to expected rate of 25%, while the other BRICS countries, Spain, the EU13 countries, Israel, Japan, Korea, and Poland also had less than 25% cited over the threshold. Switzerland, Denmark, the United Kingdom, the Netherlands, and the USA all had more than 30% of publications cited more than the 75th quartile, while Germany had 28%.

In both cases of the proportions above the median and 75th quartile, once again Scopus recorded higher proportions for most countries than did WoS as Scopus' less selective policy lowers the number of citations which constitute the median and 75th quartile thresholds compared to WoS. The countries that performed well and poorly were the same in both databases. However, notable differences occurred for China and India where these countries had lower proportions of publications performing above the median and 75th quartile in Scopus than in WoS despite the lower thresholds in Scopus. This suggests that Scopus coverage includes journals with particularly poor-performing publications from China and India.

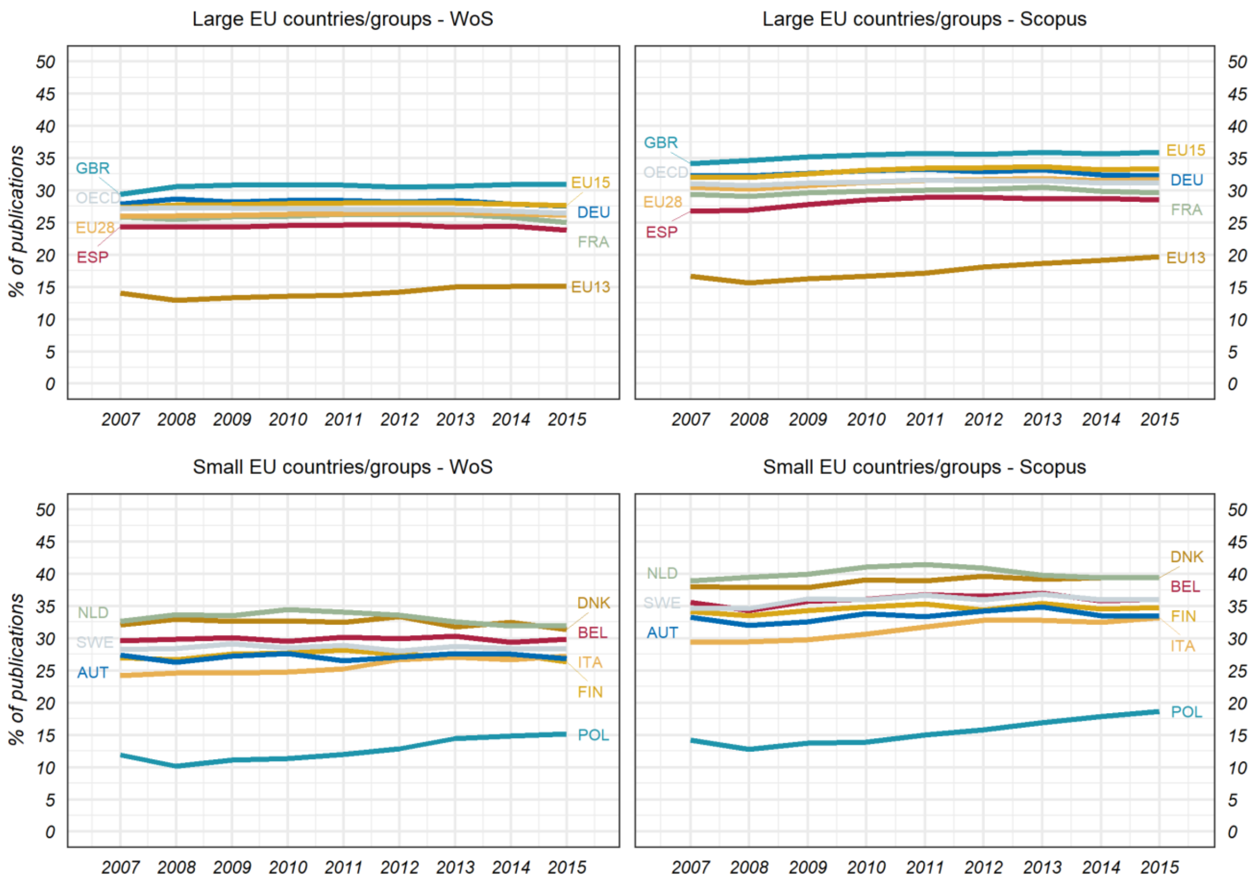


Figure 15 The percentage of publications from EU and OECD countries and groups which were cited more frequently than the 75th quartile, based on fractional counting, from Web of Science and Scopus, 2007-2015.

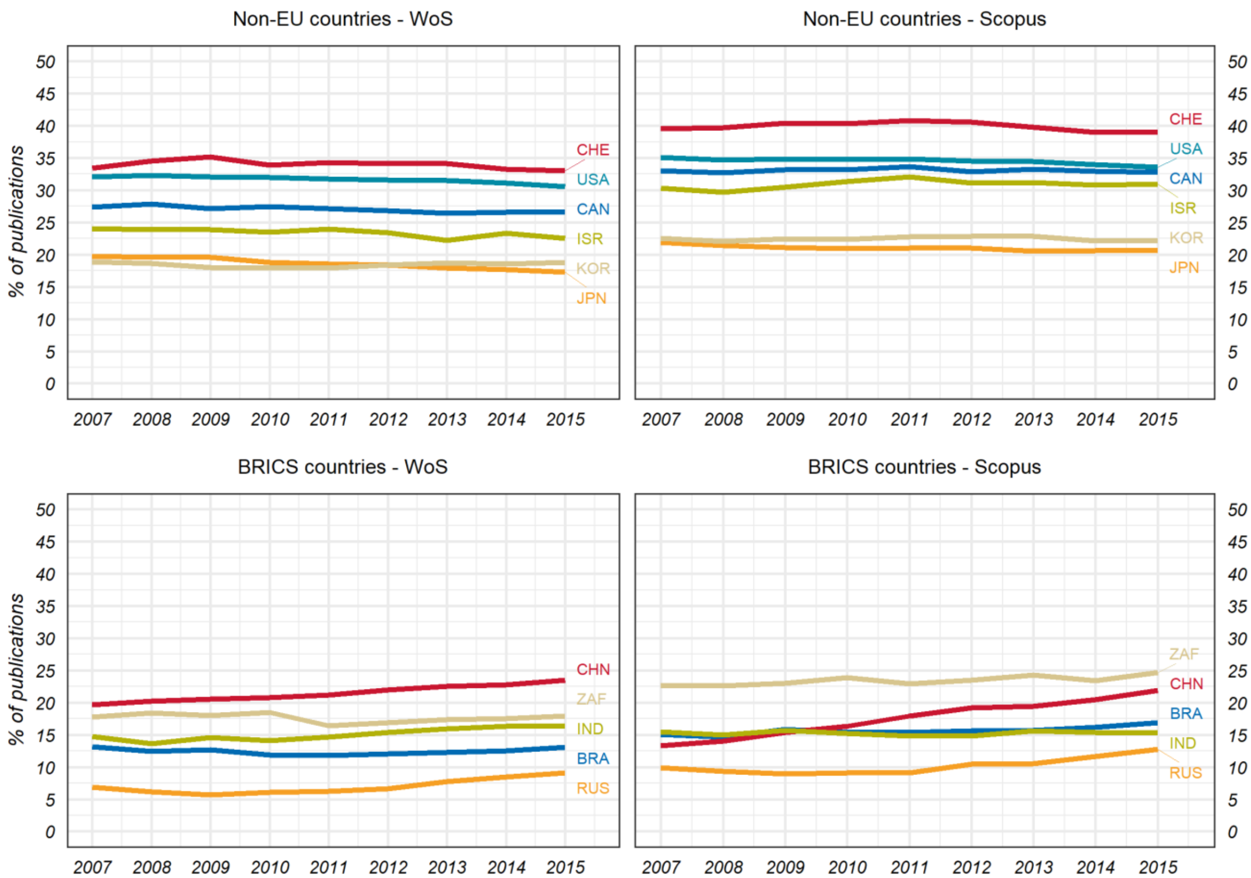


Figure 16 The percentage of publications from non-EU and BRICS countries which were cited more frequently than the 75th quartile, based on fractional counting, from Web of Science and Scopus, 2007-2015.

4.2.3 Journal-based Indicators

SR is a measure of how often a country's publications are cited compared to other publications in the same journals. SR values above 1 indicate the country's publications are cited more frequently than the average of publications in the same journals, while SR values below 1 indicate the publications are cited less frequently than average. The complementary IA indicator describes whether the journals the country's authors publish in are cited more or less often than the world average. An IA above zero indicates the country publishes in journals of higher visibility and thus higher impact than the world average, while values below zero indicate the journals the country publishes in are of less visibility than the average. Together SR and IA provide insight into whether a country is publishing in well-respected journals, and whether its' publications are well-received within those journals.

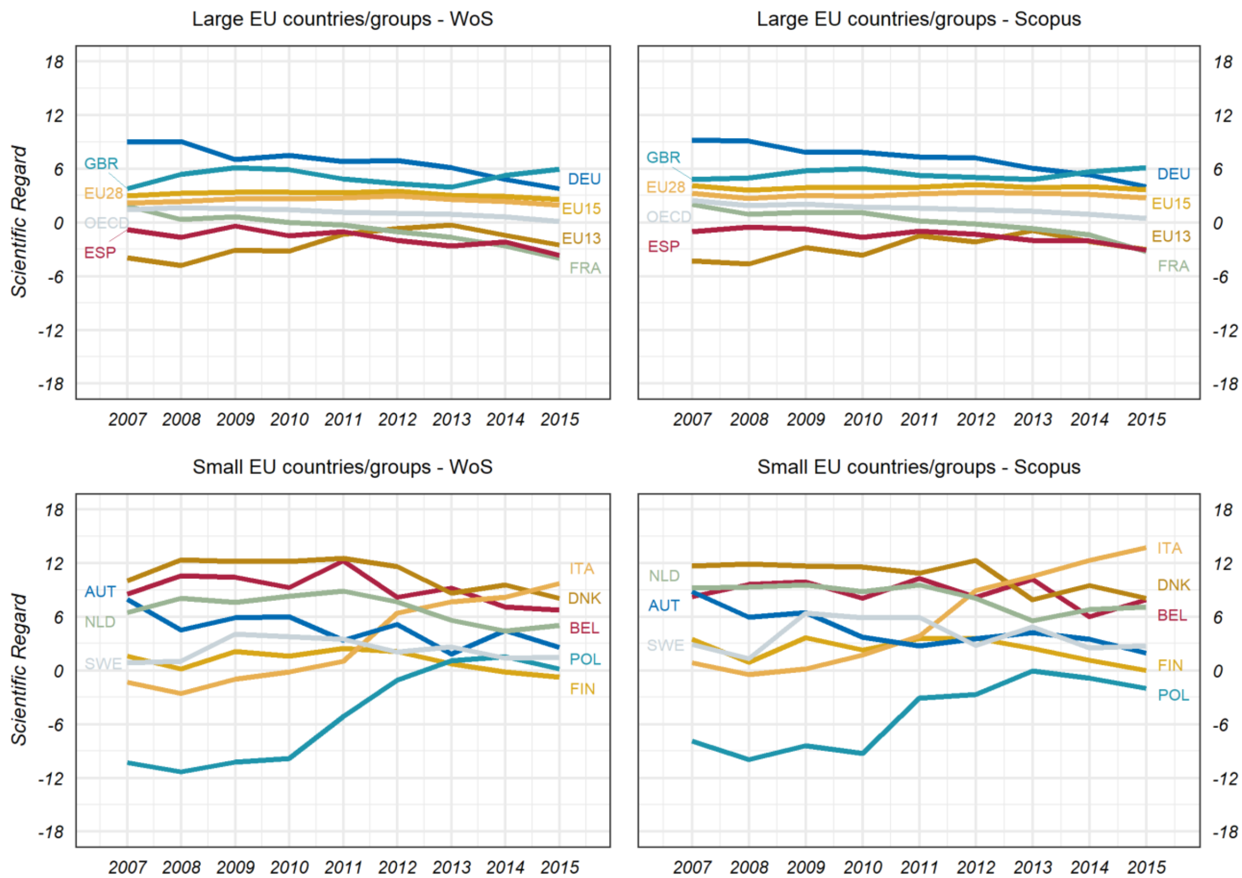


Figure 17 The Scientific Regard values for EU and OECD countries and groups, based on fractional counting, from Web of Science and Scopus, 2007-2015.

The SR values for each country and group for WoS and Scopus are presented in Figures 17 and 18. Most countries' SR values decreased over the reference period. For example, Germany's SR value fell from 9.0 in both databases in 2007 to 3.8 in Scopus in 2015 and 3.7 in WoS in 2015. This put Germany among the countries with the largest decrease in SR, alongside Austria and France in WoS. Increases in SR between 2007 and 2015 in WoS were observed for only eight of the 23 countries examined: China, the United Kingdom, India, Italy, Poland, Russia, Sweden and South Africa. The largest increases were seen for Italy with 11.0 and Poland with 10.5.

The data for Scopus were somewhat different. For example, the countries with improved SRs over time included Finland, the United Kingdom, Poland, and in particular Russia and Italy had much higher SRs between 2007 and 2015. Germany's decline in SR in Scopus was among the mid-range values rather than the upper end as in WoS, as Austria, Israel and Japan all had sharper declines. However, the key message from both databases is that most countries are receiving fewer citations over time for the journals in which they are publishing, although some countries which typically received fewer citations such as Russia and Poland are improving over time.

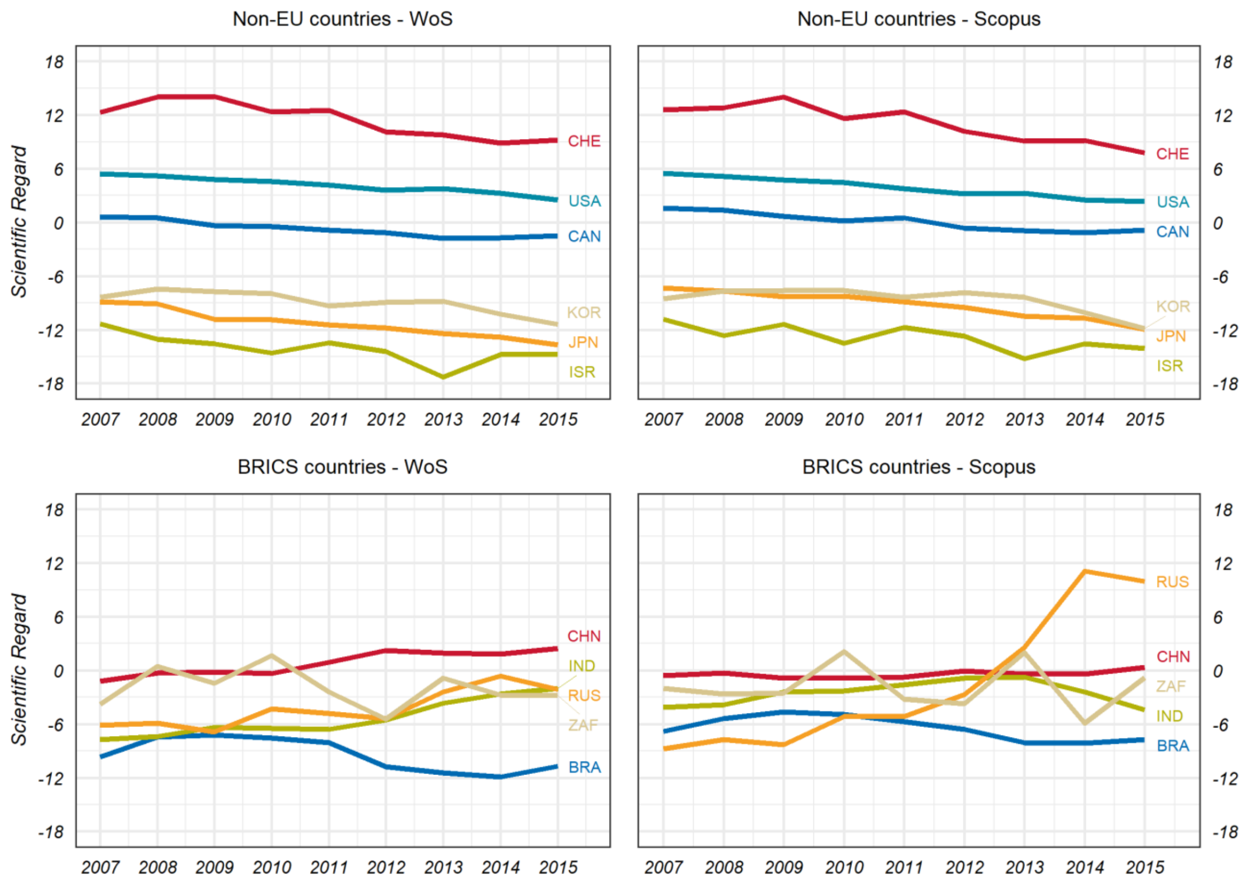


Figure 18 The Scientific Regard values for non-EU and BRICS countries, based on fractional counting, from Web of Science and Scopus, 2007-2015.

The IA values for each country and group for WoS and Scopus are presented in Figures 19 and 20. The IA values calculated using Scopus data were substantially higher than those from WoS for most countries, although the patterns in IA over time are similar for most countries. Of note is the sharp increase in IA observed for China in both databases from -30.7 to -2.6 between 2007 and 2015 in WoS and from -60.7 to -10 between 2007 and 2014 in Scopus. This trajectory suggests the improvement in IA will continue over time, however China's currently negative IA indicates it publishes in journals of lesser visibility than average. The influence of China's strong increase in number of publications on bibliometric indicators for itself and also Germany are explored more thoroughly in the following Section of this report.

Similarly, the BRICS countries, Korea and Poland had consistently negative IA values throughout the reference period, while Spain and Japan fluctuated around zero. Switzerland, the USA, and the Netherlands published in the most high-visibility journals, with IAs of 28.9, 22.0, and 20.7 respectively, followed by Denmark and the United Kingdom (each with an IA of 15.3) and Germany (15.0) when using WoS data. The rankings in Scopus are somewhat different with the following order: Switzerland (39.5), the USA (35.0), the Netherlands (34.8), Denmark (31.6), Sweden (27.0), Israel (26.9), and Germany (25.7) in 7th place. This difference may result from greater coverage in Scopus of particular regional journals that are well-cited.

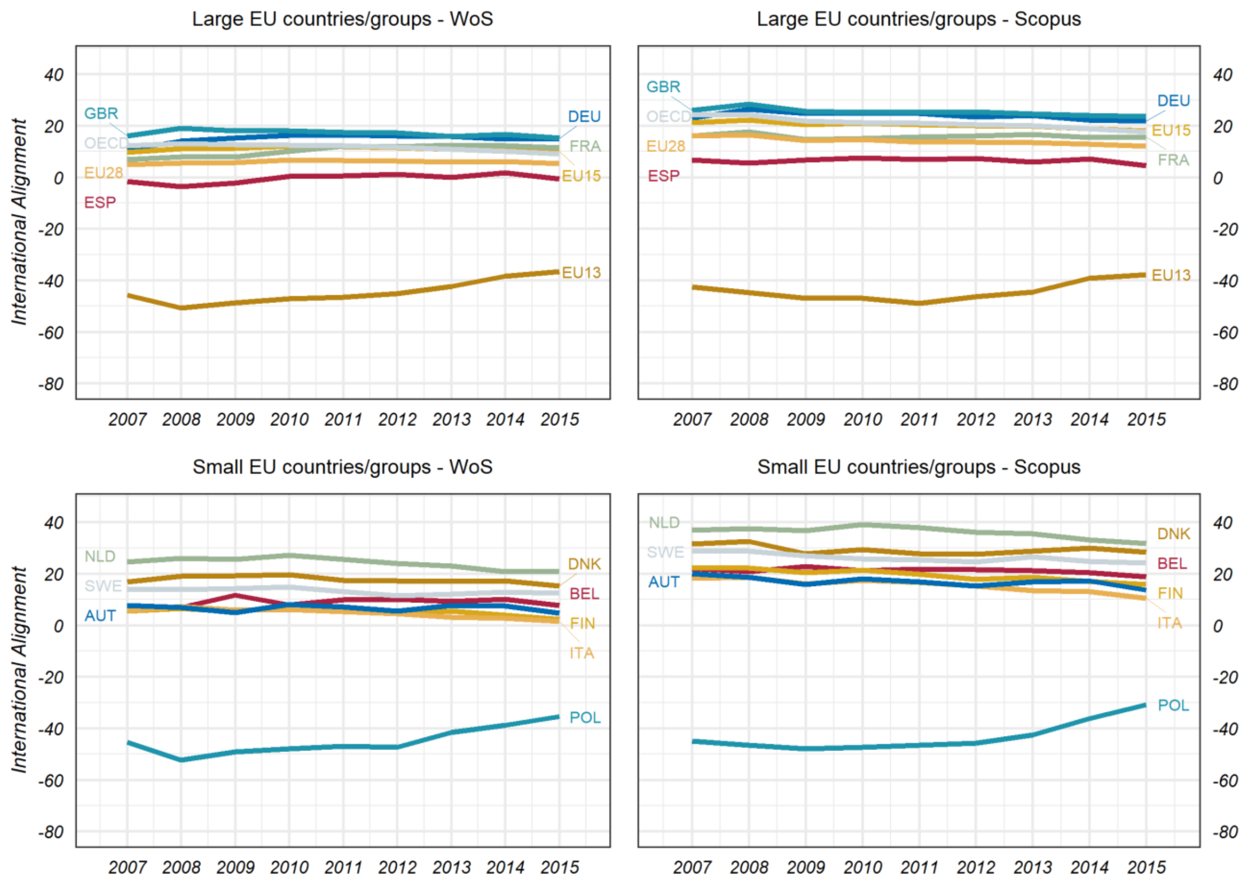


Figure 19 The International Alignment values for EU and OECD countries and groups, based on fractional counting, from Web of Science and Scopus, 2007-2015.

The SR and IA values are plotted together for Germany, the United Kingdom, the Netherlands, the USA and China in Figure 21 and include data for 2007 and 2014 for Scopus and WoS. The direction of the arrow in the graphs shows the movement in values between 2007 and 2015 and assists in understanding the combined change in these related indicators over the reference period. Comparisons are made with the USA and China as the countries with the largest scientific contribution, and with the United Kingdom and the Netherlands as EU countries with a similar scientific standing to Germany.

Germany published in more visible journals in 2015 compared to 2007, but received fewer citations for its publications than it had in 2007 in both Scopus and WoS. In 2015 the United Kingdom received a greater number of citations in journals of similar visibility compared to 2007. Conversely, the USA and the Netherlands both showed a similar pattern of decreased values in SR and IA, indicating these countries published in less visible journals and received fewer citations in 2015 than in 2007, although there were smaller decreases for the Netherlands than the USA. The IA values for both countries remained high however, indicating that they are still publishing in well-cited journals. In 2015 in Scopus, China published in substantially more visible journals than in 2007. In WoS China clearly improved both its visibility and rate of citations. Viewing these indicators together provides a deeper insight into how countries perform in terms of the kinds of journals they publish in and the reception of their publications within these journals.

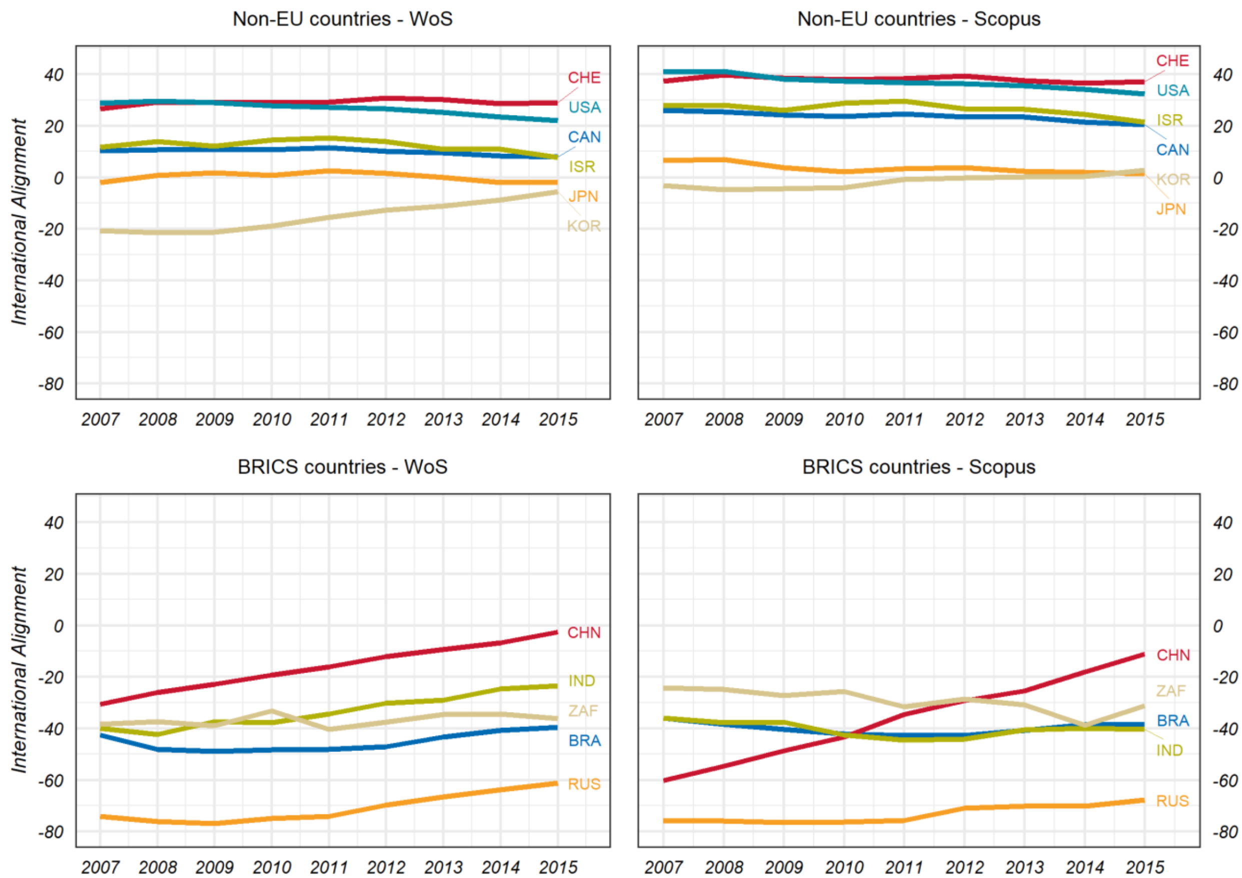


Figure 20 The International Alignment values for non-EU and BRICS countries, based on fractional counting, from Web of Science and Scopus, 2007-2015.

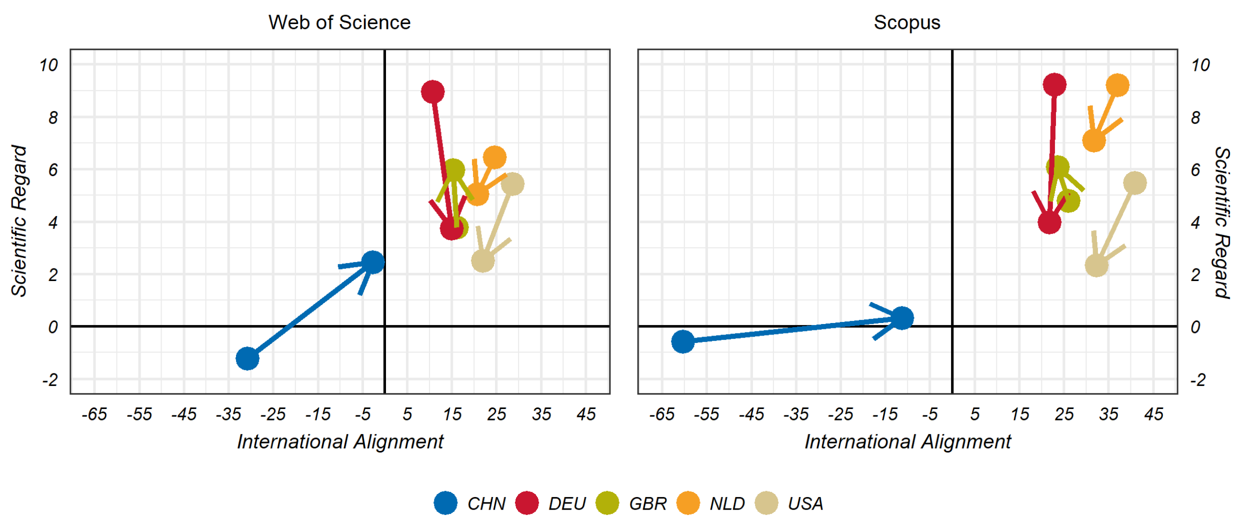


Figure 21 The International Alignment and Scientific Regard values for Germany, USA, China, the United Kingdom, and the Netherlands, based on fractional counting, in 2007 and 2015 from Web of Science and Scopus. The direction of the arrow denotes the change in values for each country between 2007 and 2015.

4.3 Publication profiles by universities, Fachhochschulen and research associations

The science system in Germany can be considered in terms of six different groups, each with a different focus. The universities undertake both teaching and research in the full array of disciplines. The Fachhochschulen focus on technical application and less on conducting original research. There are also four non-university research associations: the Helmholtz Association (HGF) which focuses on health, energy, earth and physical sciences research, often on a large scale; the Leibniz Association (WGL) conducts both basic and applied research in social sciences, health, natural sciences, mathematics and engineering; the Max Planck Society (MPG) focuses on basic research, and the Fraunhofer Society (FhG) focuses on applied research.

This section provides information on the publication and citation indicators for these groups, however these data should be considered in the context that the associations have different publishing objectives due to their varied missions, research interests and profiles. Data are presented for all of the universities as a group and all of the Fachhochschulen as a group.

4.3.1 Number and share of publications

The number of publications from the universities, Fachhochschulen and each of the research associations are presented in Figure 22 and their share of the overall set of German publications is shown in Table 8. WoS and Scopus record similar numbers of publications for the research associations and the Fachhochschulen, although the number of publications for the universities is somewhat higher in Scopus, particularly during 2011-2015.

All groups show an upward trend in publications over time, which corresponds with a relatively stable share in German publications over time. The exception is the MPG which produced fewer publications annually after 2012 and consequently had an approximately 20% decrease in its share of German publications over the reference period. The MPG remains the second-highest producing group however, behind only the HGF. The Fachhochschulen doubled both its number of publications and its share of German publications over the reference period, resulting in a average growth rate of more than 9% (see Figure 23), however it and the FhG continue to account for only approximately 1% of German publications each.

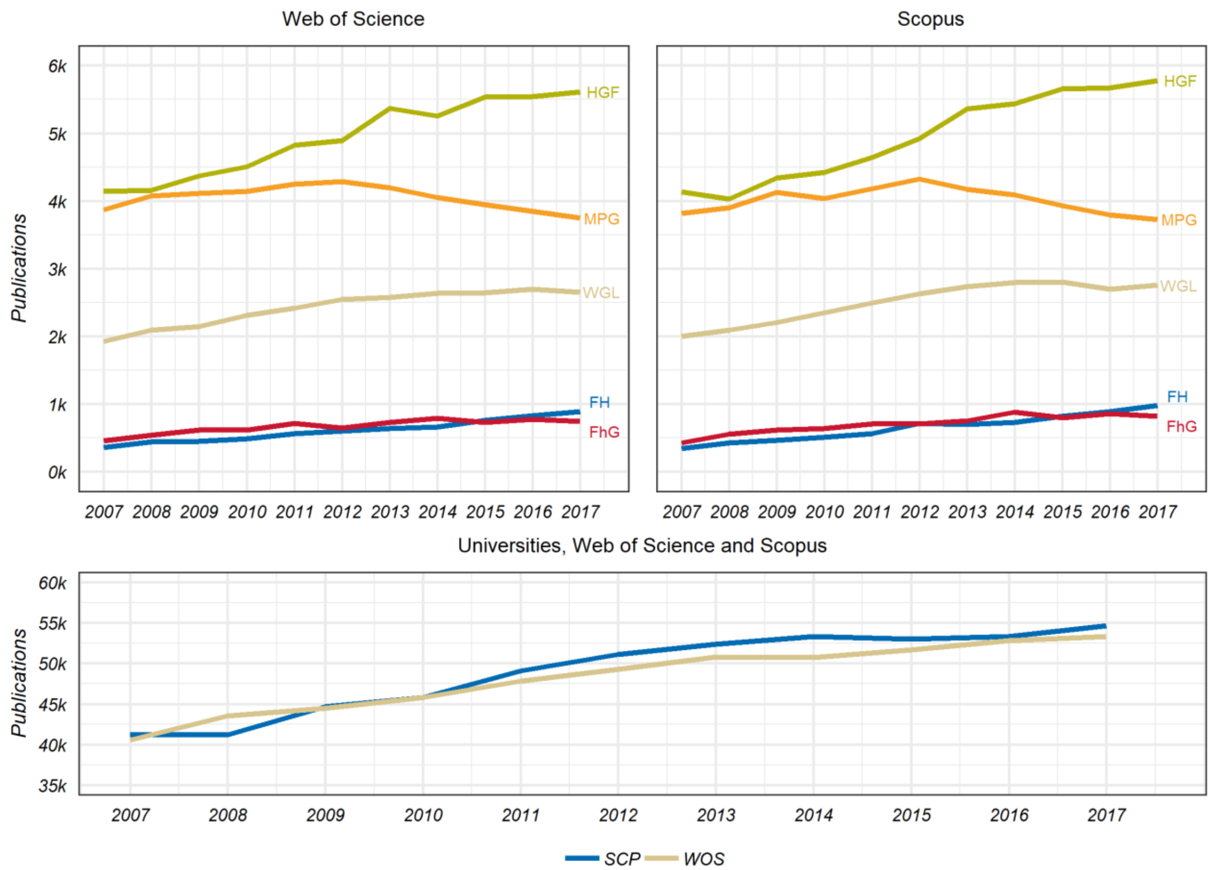


Figure 22 The fractional count of publications from German universities, Fachhochschulen and research associations, from Web of Science and Scopus, 2007-2017.

Table 8 The universities', Fachhochschulen's and research associations' shares of German publications, based on fractional counting, from Web of Science and Scopus, 2007-2017.

DATABASE	SUBSEC	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
WoS	FH	0.61	0.72	0.71	0.75	0.84	0.86	0.90	0.93	1.05	1.11	1.18
WoS	FhG	0.77	0.88	0.98	0.96	1.06	0.93	1.03	1.11	1.00	1.03	0.99
WoS	HGF	7.08	6.81	6.98	7.00	7.17	7.05	7.54	7.39	7.61	7.38	7.47
WoS	MPG	6.62	6.67	6.56	6.44	6.32	6.18	5.89	5.70	5.42	5.13	4.99
WoS	UNI	69.36	71.31	71.02	71.20	71.10	71.08	71.35	71.42	71.06	70.37	70.96
WoS	WGL	3.29	3.43	3.42	3.59	3.59	3.67	3.62	3.70	3.63	3.59	3.53
WoS	DEU	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Scopus	FH	0.55	0.68	0.68	0.73	0.75	0.91	0.87	0.89	1.01	1.07	1.17
Scopus	FhG	0.68	0.88	0.89	0.90	0.95	0.91	0.94	1.08	0.97	1.03	0.97
Scopus	HGF	6.58	6.42	6.33	6.29	6.22	6.33	6.71	6.66	6.93	6.83	6.88
Scopus	MPG	6.07	6.21	6.02	5.75	5.61	5.56	5.22	5.01	4.81	4.57	4.43
Scopus	UNI	65.67	65.71	65.20	65.23	65.78	65.74	65.47	65.29	64.90	64.23	65.04
Scopus	WGL	3.18	3.33	3.22	3.34	3.35	3.38	3.42	3.42	3.43	3.25	3.28
Scopus	DEU	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

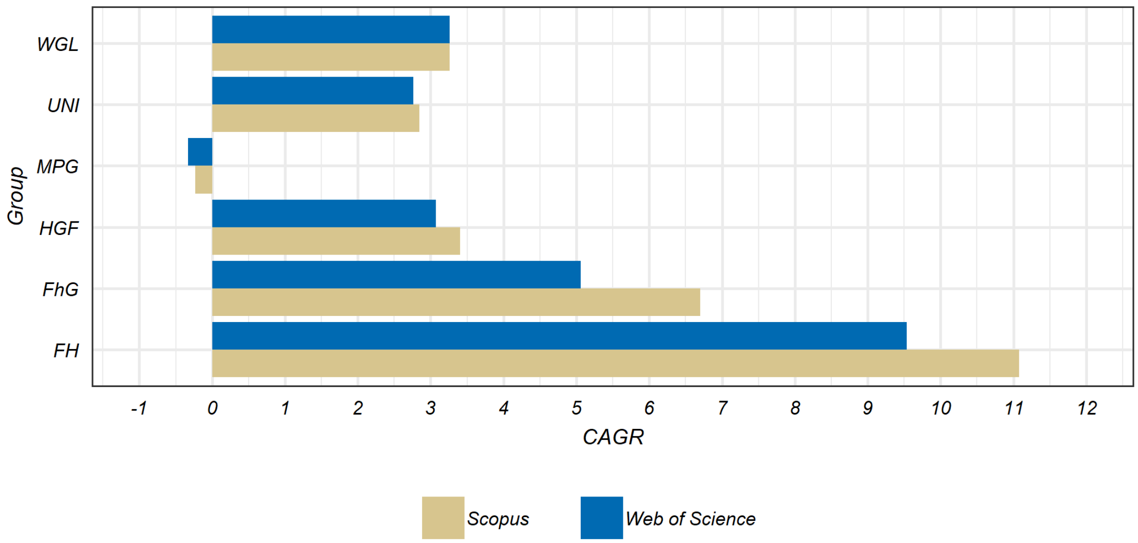


Figure 23 The CAGR of publications from German universities, Fachhochschulen and research associations, based on fractional counting, from Web of Science and Scopus, 2007-2017.

To consider the groups on a more even footing, the number of publications and citations for the universities, Fachhochschulen and each research association per full-time equivalent (FTE) research staff are presented in Figure 24 for the years 2007 to 2016. The number of FTE staff per group were sourced from the Federal Ministry of Education and Research Data Portal and were available up to 2016 (see Table 1.7.8)[2]. It should be noted that the number of publications and citations per FTE staff do not account for discipline-specific differences in staffing requirements which may arise between groups due to their varied missions and research specialties.

Most of the groups produced a stable number of publications over time and received a stable number of citations. Notably however, the WGL increased its publications per FTE staff over time and also its citations in WoS. The MPG produced fewer publications and received fewer citations over the reference period, except for a peak in 2014 which occurred due to a reduction in the number of FTE research staff while the number of publications and citations remained in line with previous years. Despite the MPG's reduced output, it still produced the most publications and in particular received a substantially higher number of citations per FTE than the universities, Fachhochschulen or other associations. The same is true for Scopus.

---

[2] http://www.datenportal.bmbf.de/portal/en/Table-1.7.8.html

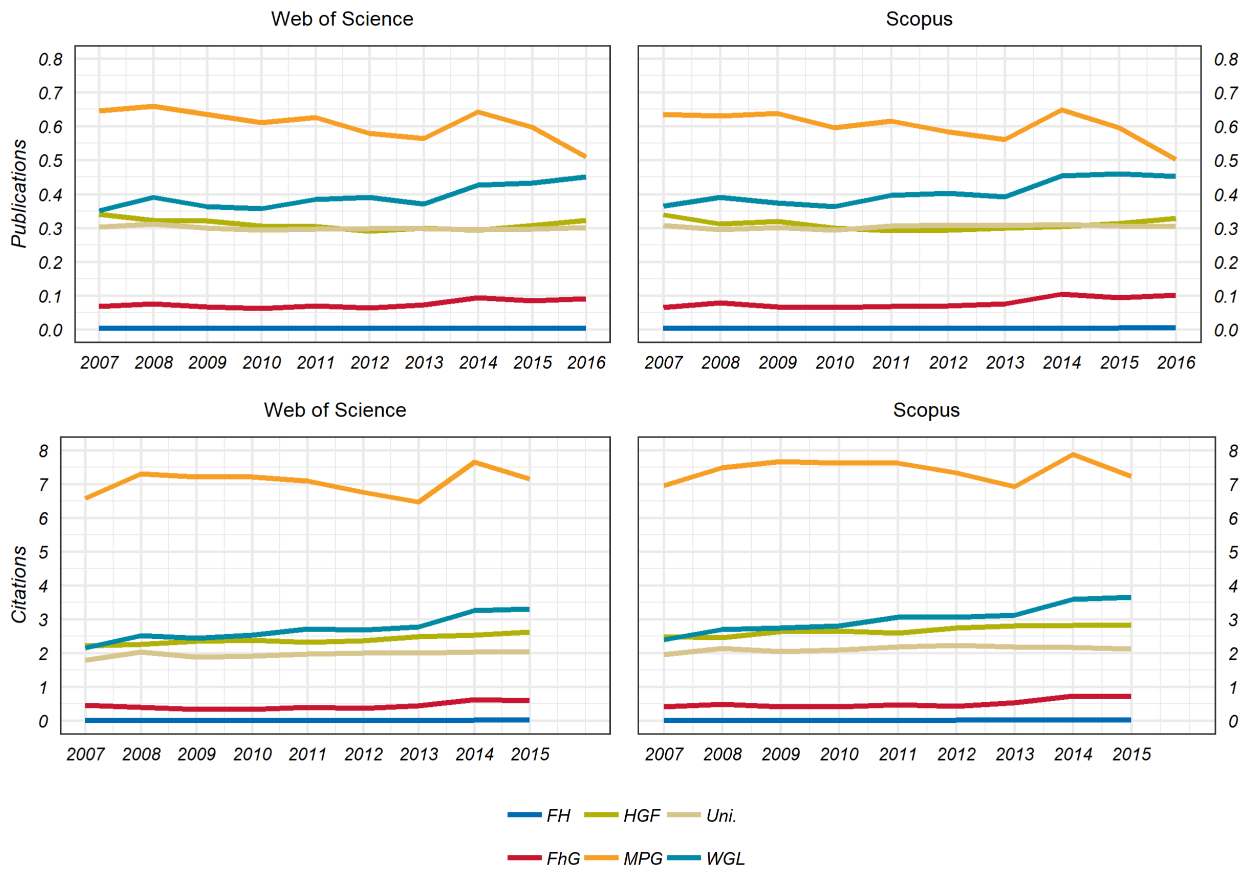


Figure 24 The number of publications (top) and the number of citations (bottom) per FTE research staff from German universities, Fachhochschulen and research associations from Web of Science and Scopus.

4.3.2 Impact indicators

The SR, IA and ER values for the universities, Fachhochschulen and research associations are shown in Figure 25. The IA for the universities was stable over the reference period, while the Fachhochschulen and each of the research associations increased their IA values indicating that they are increasingly publishing in more highly-cited journals. The IA values were slightly higher in Scopus than WoS but the same patterns were observed.

Similarly for SR, the patterns for the groups between WoS and Scopus were similar, however the trends were less overt. The SR values for the universities, the HGF, and the MPG all decreased over time in both WoS and Scopus indicating their publications are receiving fewer citations, although the citations are still above average, particularly for the MPG and the HGF. The SR for the WGL tended to increase over time but not in a stable fashion, and the FhG and Fachhochschulen had particularly unstable SR values.

This combination of SR and IA values suggests that while the research associations and universities are all publishing in more internationally-visible journals, their publications in these journals are receiving fewer citations. This result may be particularly exacerbated for the Fachhochschulen and the FhG because these groups also publish a relatively small number of publications each year meaning there is greater variability from differences in number of publications and the citation practices of the journals they publish in which causes their SRs to fluctuate more than the groups with higher number of publications.

Publications from German universities and research associations are often amongst the 10% most highly cited in their fields, as demonstrated by the ER values shown in Figure 25. From the Scopus data, all of the groups except the Fachhochschulen met or exceeded the expected 10% threshold of publications in the top 10% of publications, in particular the MPG which had up to 23% in the top 10% although this decreased to 10% by 2015. The trends for WoS data were similar, although the proportions were lower, the MPG appears to increase its ER again in 2015, and the FhG dipped below 10% in the top 10% since 2012.

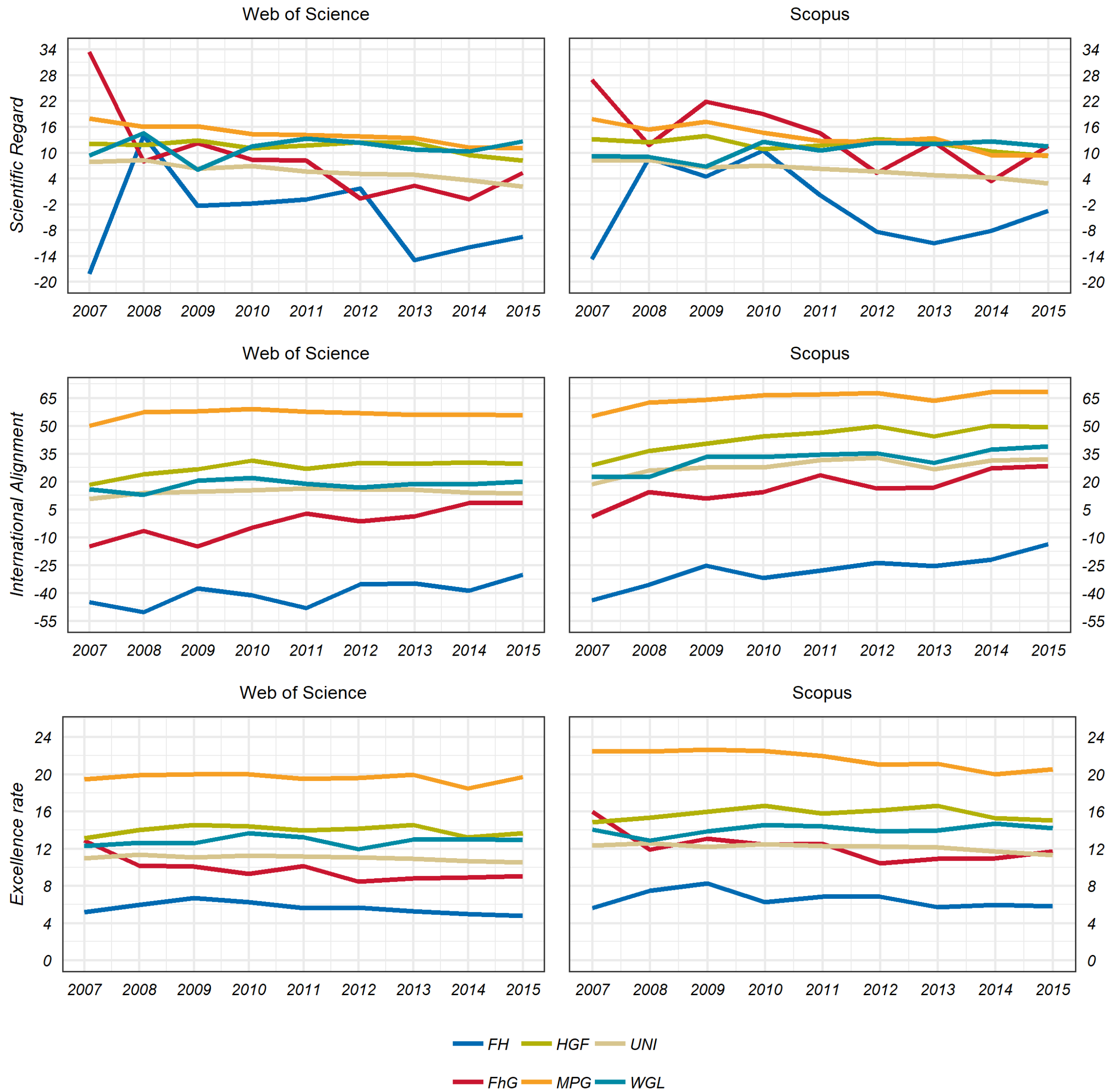


Figure 25 The SR values (top), IA values (middle) and ERs (bottom) for German universities, Fachhochschulen and research associations, based on fractional counting, from Web of Science and Scopus, 2007-2015.

The proportion of publications that received no citations during the 3-year citation window, and the proportion of publications that were cited more frequently than the median number or 75th quartile citations are presented in Figure 26. The proportion of uncited publications steadily decreased for all groups in the data from WoS with fewer than 15% of publications uncited after the 3 years by 2015, except for those from the Fachhochschulen which remained at approximately 25% uncited.

The trends for the proportion of publications cited more frequently than the median or the 75th quartile were also stable and slightly higher in Scopus than WoS. The MPG had the most highly-cited publications, with consistently more than 65% cited more frequently than the median citations and more than 40% cited more frequently than the 75th quartile citations.

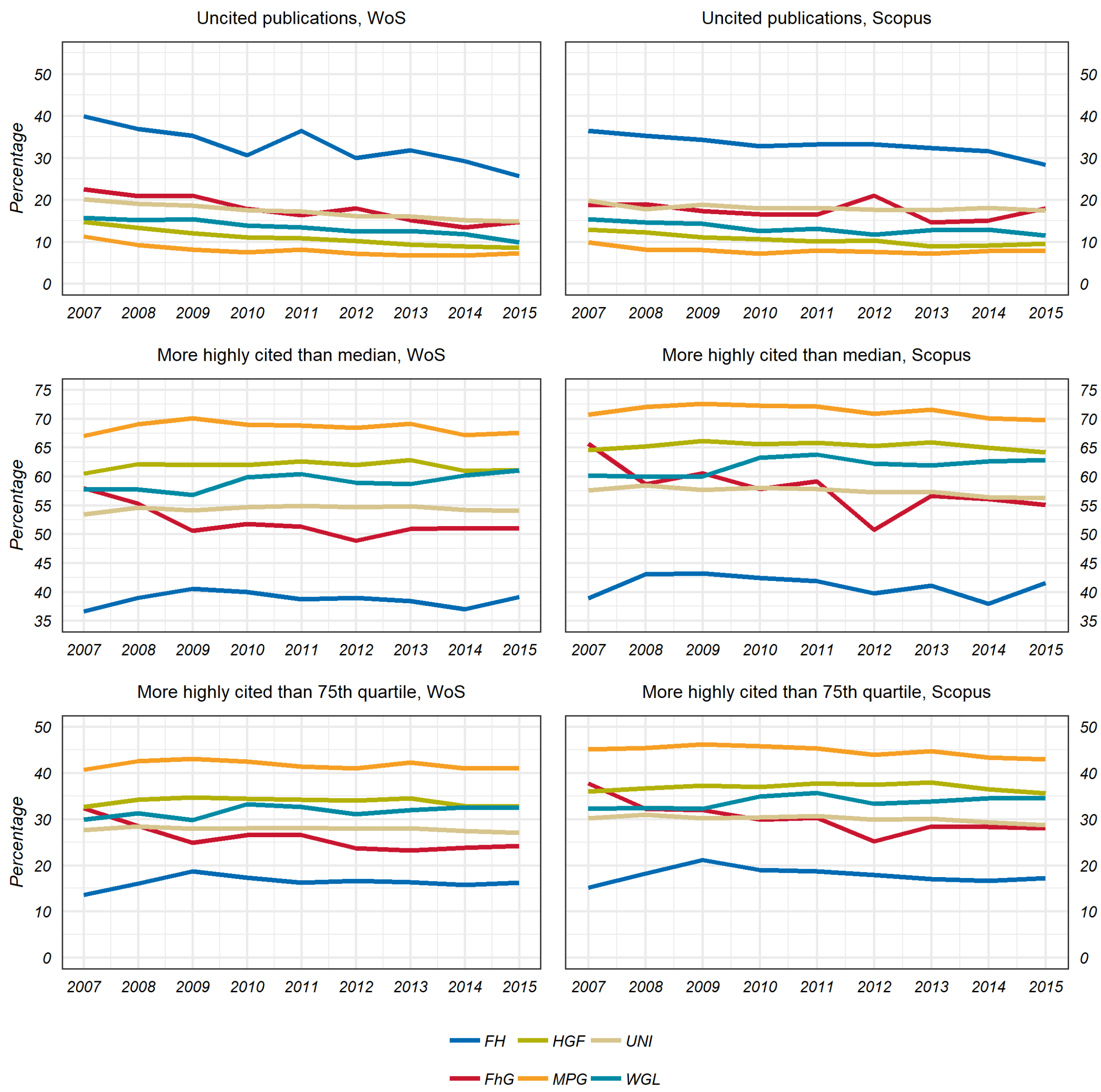


Figure 26 The percentage of uncited publications, and publications cited more frequently than the median or 75th quartile from German universities, Fachhochschulen and research associations, based on fractional counting, from Web of Science and Scopus, 2007-2015.

The proportion of publications in each discipline in the periods 2007-2010 and 2014-2017 are presented for the universities, Fachhochschulen and each research association in Figures 27 to 32. The differences in the key disciplines in which the groups publish highlight the different focuses of the groups, however biological, chemical and physical sciences are consistently amongst the disciplines with the highest proportion of publications for each group. Further, the disciplines in which each group publishes the most of their publications is mostly consistent between WoS and Scopus.

4.3.3 Discipline-specific profiles

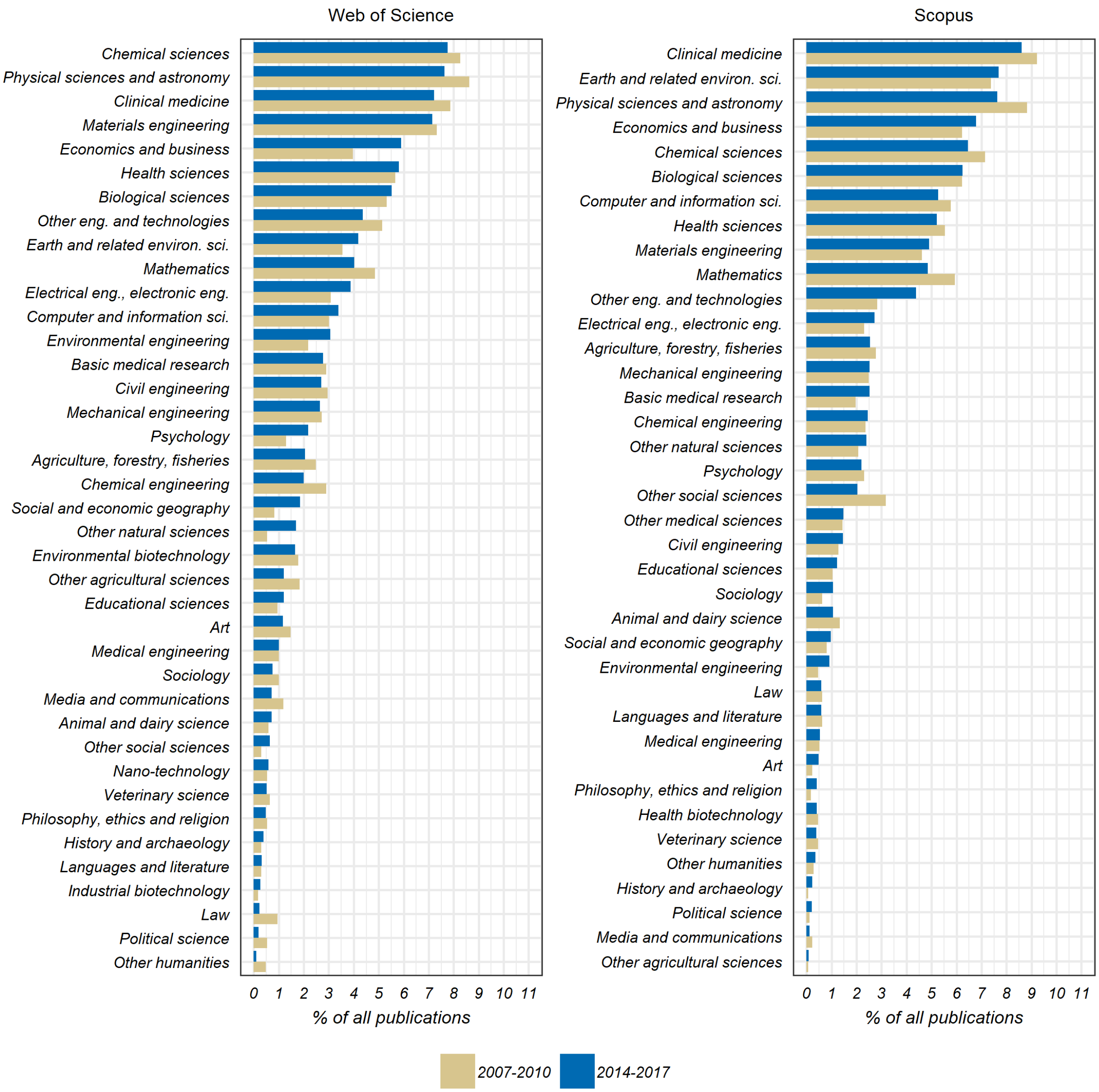


Figure 27 The percentage of total Fachhochschulen publications accounted for by each OECD discipline, based on fractional counting, grouped years 2007-2010 and 2014-2017.

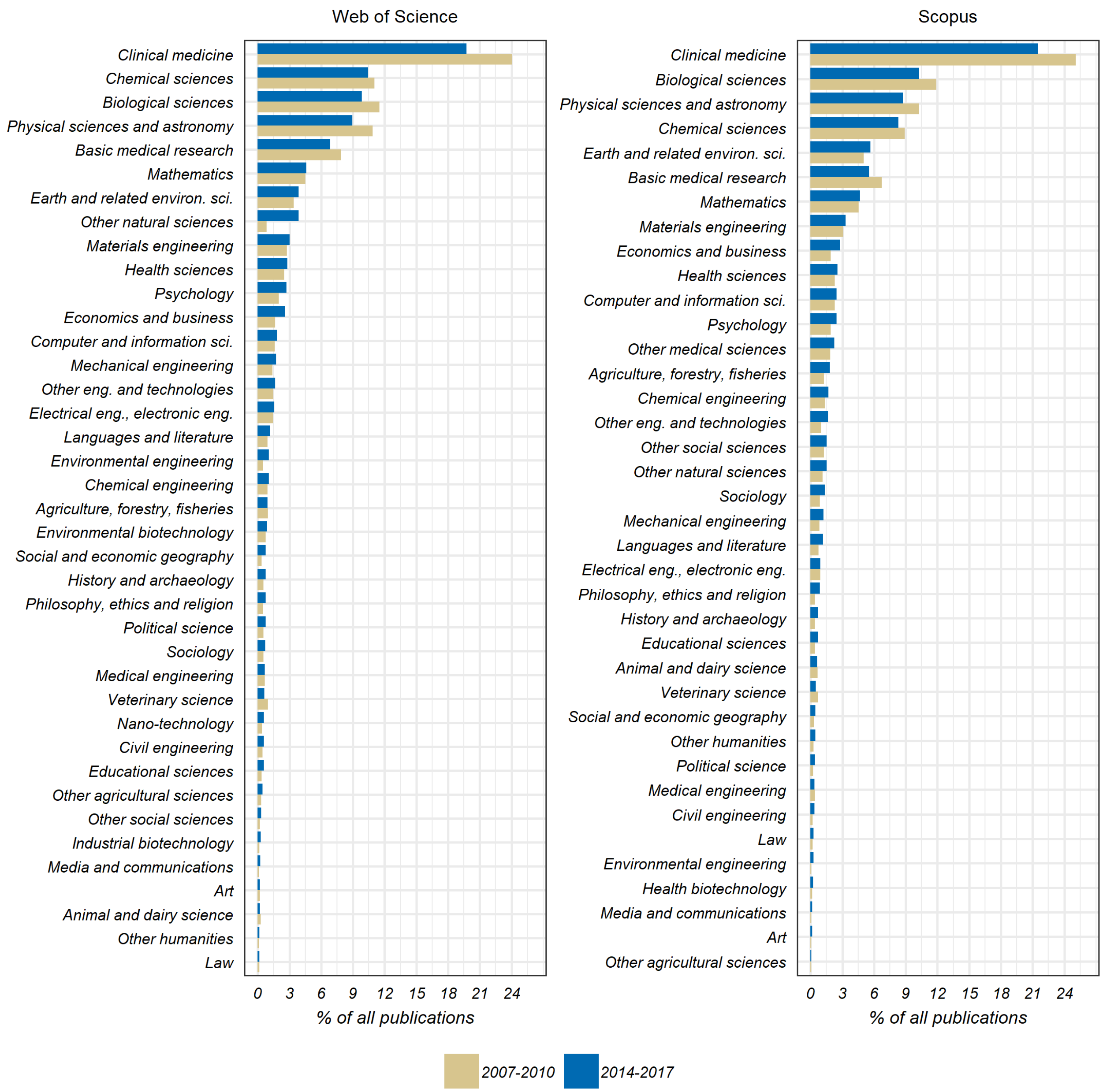


Figure 28 The percentage of total University publications accounted for by each OECD discipline, based on fractional counting, grouped years 2007-2010 and 2014-2017.

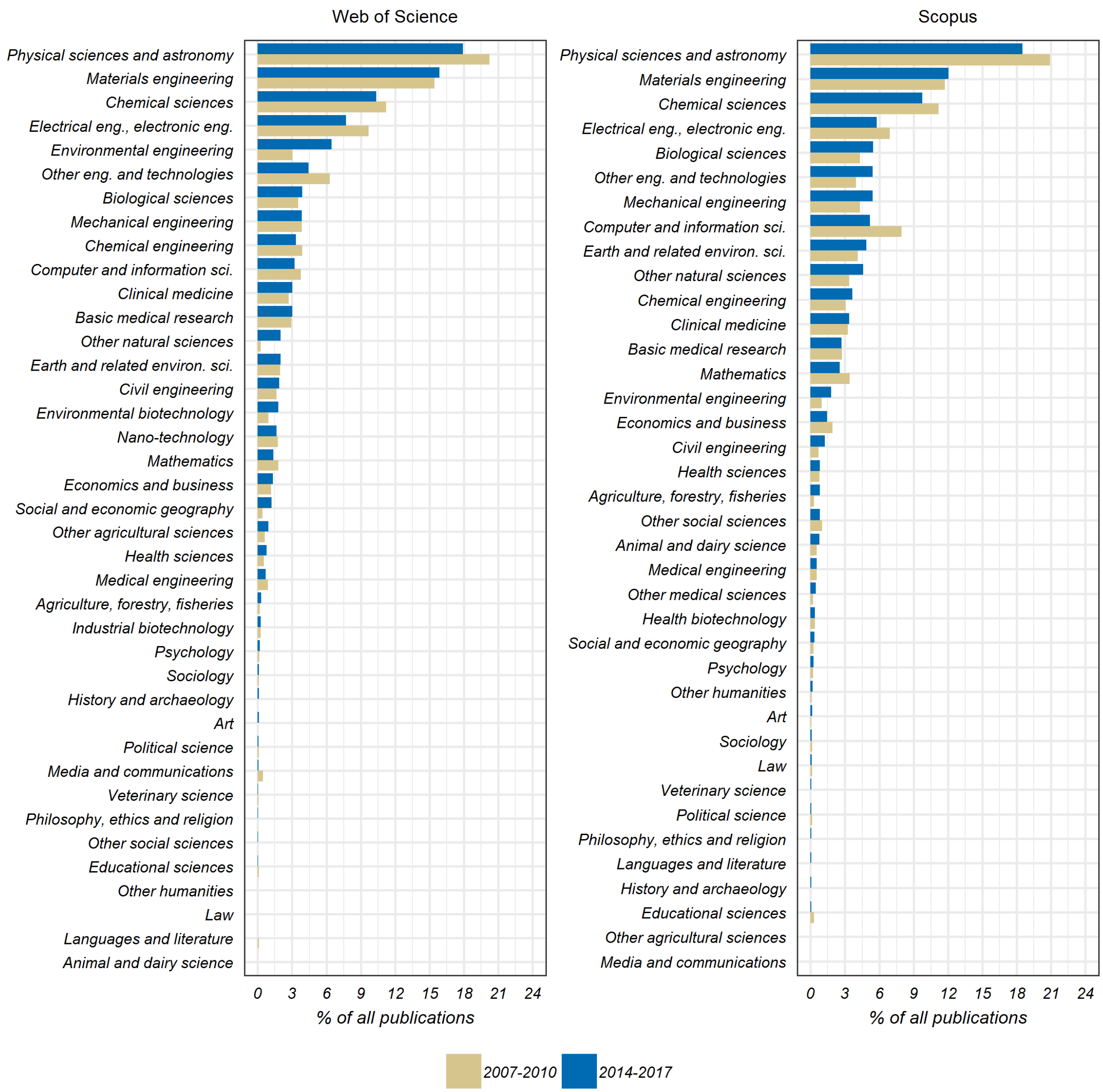


Figure 29 The percentage of total FhG publications accounted for by each OECD discipline, based on fractional counting, grouped years 2007-2010 and 2014-2017.

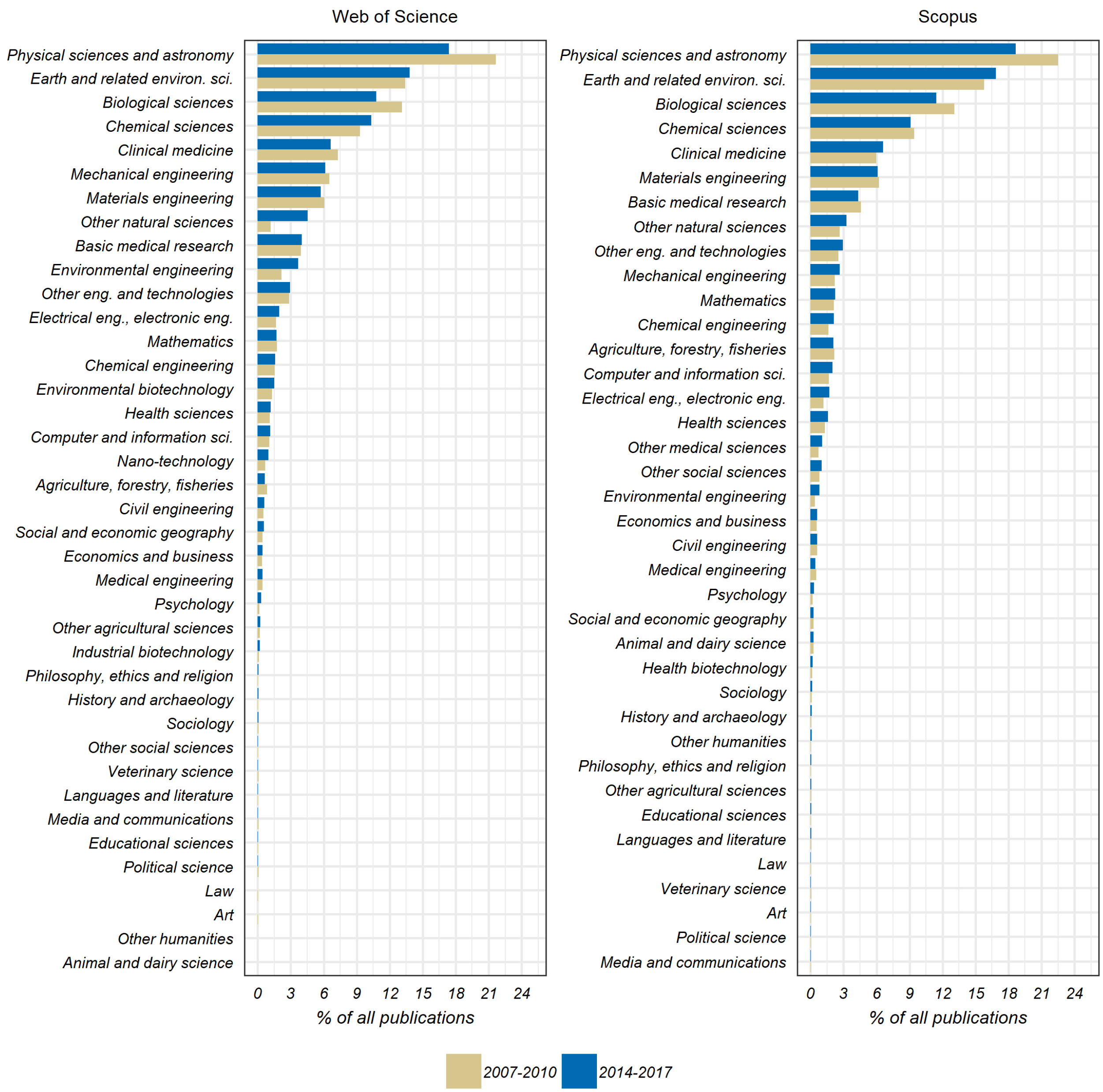


Figure 30 The percentage of total HGF publications accounted for by each OECD discipline, based on fractional counting, grouped years 2007-2010 and 2014-2017.

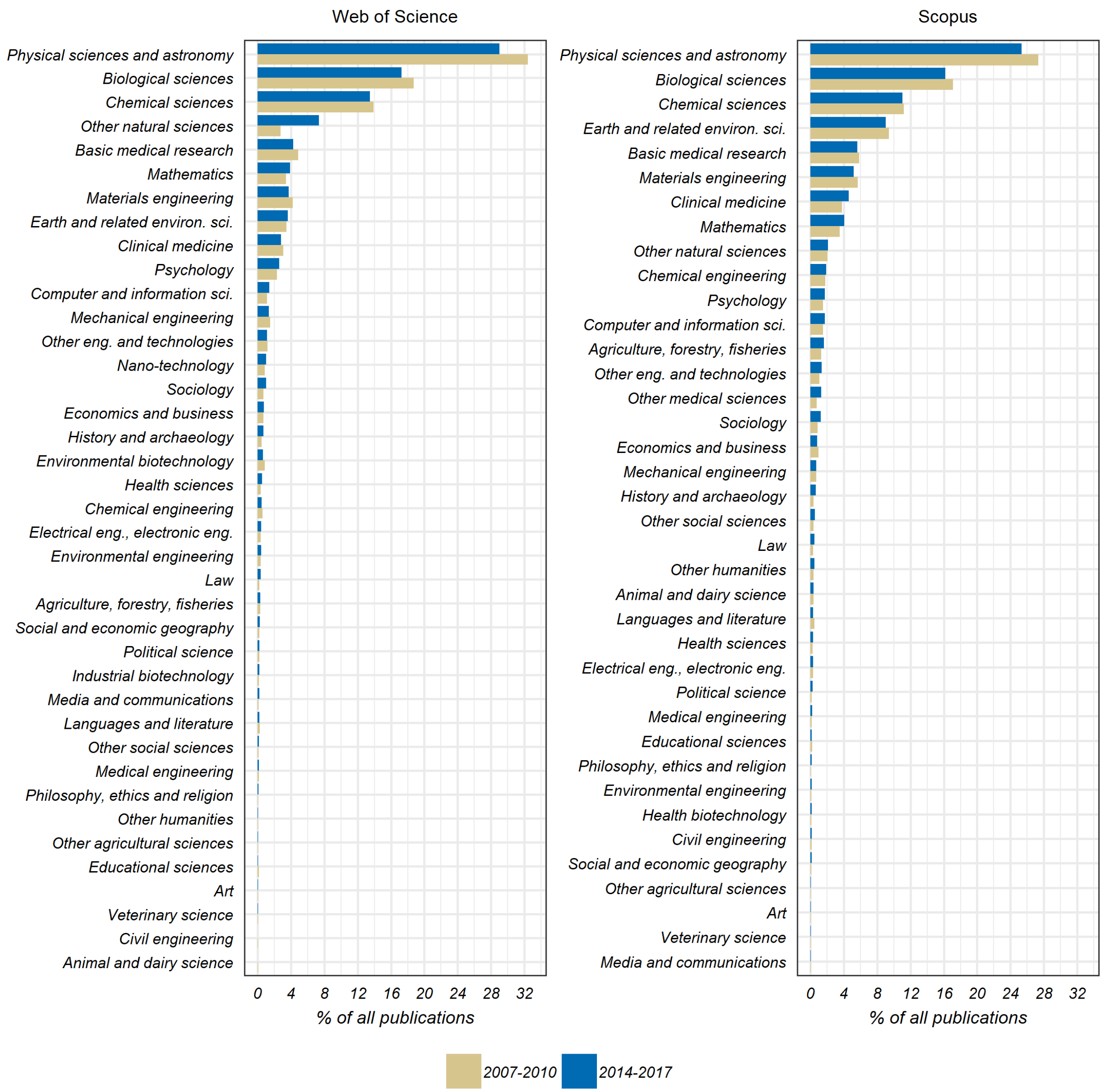


Figure 31 The percentage of total MPG publications accounted for by each OECD discipline, based on fractional counting, grouped years 2007-2010 and 2014-2017.

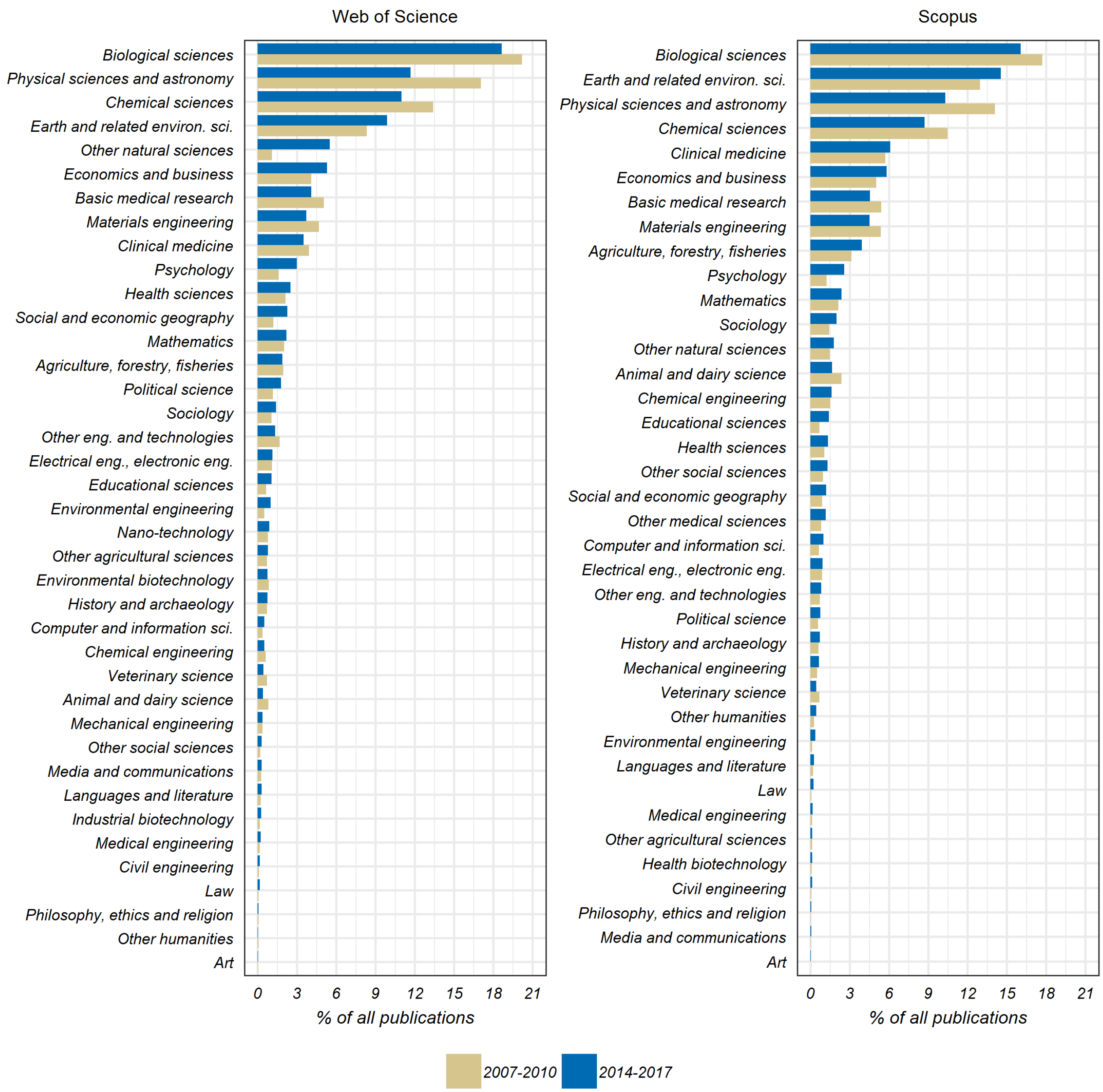


Figure 32 The percentage of total WGL publications accounted for by each OECD discipline, based on fractional counting, grouped years 2007-2010 and 2014-2017.

4.3.4 Co-publication analysis for universities, Fachhochschulen and research associations

The following section discusses the co-publishing practices of the German universities, Fachhochschulen and research associations in terms of their international co-publications and co-publications with each other. The proportion of publications from each group that involved international collaboration are presented in Figure 33. The proportions and trends for each group were comparable between WoS and Scopus. The MPG had the highest proportion of co-publications, with three-quarters of its publications involving international collaboration in 2017. This is consistent with the MPG's focus on basic research specialising in the natural science disciplines examining research questions of international relevance.

Conversely, the Fachhochschulen and the FhG had the lowest proportions of internationally co-authored papers, similarly consistent with their missions of transferring practical knowledge.

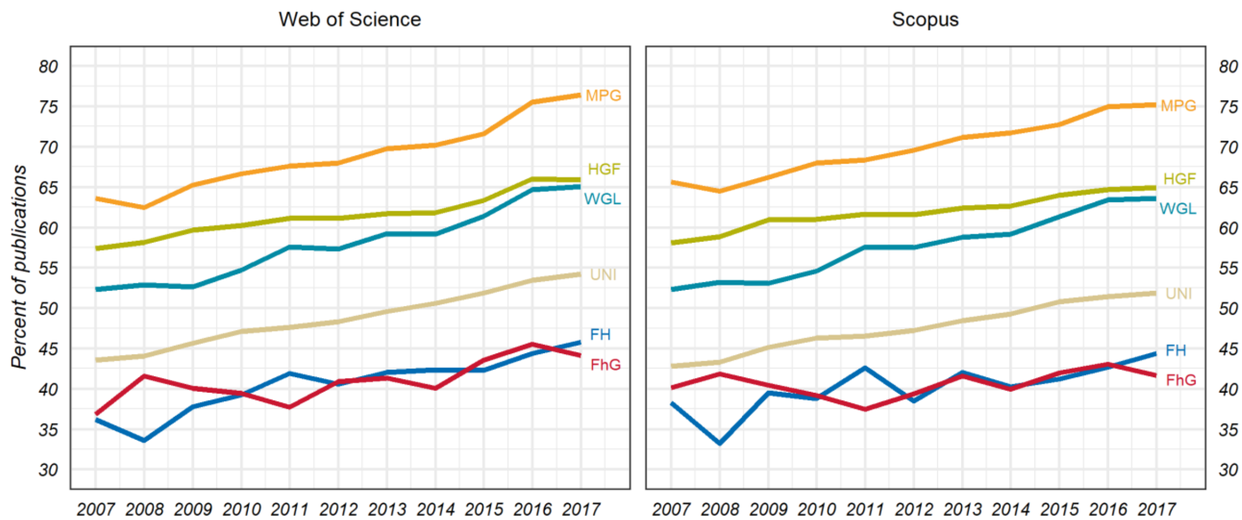


Figure 33 The percentage of publications from German universities, Fachhochschulen and research associations involving international co-publication, based on whole counts from Web of Science and Scopus

The proportion of each group's co-publications with every other group in 2007 and 2017 are presented in Tables 9 and 10. For example, in Table 9, 11.6% of the FhG's co-publications were with the HGF, but 1.1% of the HGF's co-publications were with the FhG. These proportions refer to papers with at least one author from each of the groups denoted, but may also include authors from additional groups or countries other than Germany.

The patterns of inter-group co-publications were similar between WoS and Scopus. Co-publications with the universities comprised the bulk of co-publication for all of the research associations and the Fachhochschulen: between 70-80% of each groups' co-publications were with the universities. Conversely, nearly half of the universities' co-publications were with the HGF, and they co-published least with the FhG and the Fachhochschulen. This was also true for the HGF, MPG and WGL, with less than 3% of each of these groups co-publications involving the FhG and the Fachhochschulen.

There were strong ties between the HGF and the MPG, with each accounting for a substantial proportion of each other's co-publications. The WGL also frequently co-published with the HGF, though the HGF's co-publications less often included the WGL. These patterns of co-publishing between groups were stable over time, but with a slight shift of all groups toward greater co-publishing with the Fachhochschulen.

Table 9 The proportion of co-publications between German universities, Fachhochschulen and research associations in 2007 and 2017, based on whole counts from Web of Science.

Year	Sector	FhG	HGF	MPG	WGL	Uni	FH
2007	FhG	0.0	1.08	0.92	1.06	4.2	2.4
2007	HGF	11.6	0.00	20.87	11.37	47.6	14.4
2007	MPG	6.4	13.48	0.00	10.22	27.5	4.2
2007	WGL	4.4	4.43	6.16	0.00	17.7	5.2
2007	Uni	75.9	80.03	71.61	76.43	0.0	73.8
2007	FH	1.7	0.99	0.44	0.92	3.0	0.0
2007	Total	100.0	100.00	100.00	100.00	100.0	100.0
2017	FhG	0.0	1.03	0.84	1.18	5.0	3.6

2017	HGF	9.6	0.00	21.54	12.42	48.7	15.5
2017	MPG	4.1	11.25	0.00	7.93	22.1	5.9
2017	WGL	4.5	5.05	6.16	0.00	18.9	4.4
2017	Uni	77.5	80.71	70.02	77.10	0.0	70.6
2017	FH	4.3	1.97	1.44	1.37	5.4	0.0
2017	Total	100.0	100.00	100.00	100.00	100.0	100.0

Table 10 The proportion of co-publications between German universities, Fachhochschulen and research associations in 2007 and 2017, based on whole counts from Scopus.

Year	Sector	FhG	HGF	MPG	WGL	Uni	FH
2007	FhG	0.0	1.12	1.05	0.67	4.2	2.6
2007	HGF	11.7	0.00	19.85	11.46	47.1	12.1
2007	MPG	7.2	12.89	0.00	10.30	27.4	3.9
2007	WGL	2.8	4.63	6.41	0.00	18.2	4.2
2007	Uni	76.3	80.52	72.27	76.86	0.0	77.1
2007	FH	1.9	0.83	0.42	0.71	3.1	0.0
2007	Total	100.0	100.00	100.00	100.00	100.0	100.0
2017	FhG	0.0	1.02	0.96	1.11	5.2	3.9
2017	HGF	9.3	0.00	22.15	12.76	49.1	15.7
2017	MPG	4.4	11.21	0.00	7.75	21.4	5.7
2017	WGL	4.1	5.11	6.14	0.00	18.8	4.2
2017	Uni	77.7	80.65	69.31	77.03	0.0	70.4
2017	FH	4.6	2.01	1.44	1.34	5.5	0.0
2017	Total	100.0	100.00	100.00	100.00	100.0	100.0

4.4 International co-publications

International collaboration, including the exchange and application of knowledge, sharing of resources and co-publishing of papers, is a key component of scientific research. International co-publications are publications that were produced by authors from two or more countries. By looking at patterns of co-publication between countries, we can identify the countries Germany most frequently partners with, and the disciplines in which Germany most or least actively collaborates as measured by co-publications. The data presented in this section is based on whole counts of publications. That is, each publication counts as one publication toward the count for each country with an author involved in its production.

4.4.1 German national and international co-publishing

Different perspectives of Germany's national publishing practices can be seen in Figures 34 and 35. Figure 34 depicts the number and proportion of Germany's total publications with authors from only Germany, including publications written by only one author. Figure 35 shows the number and proportion of Germany's total publications that involved collaboration between two or more authors from different German institutions.

The proportion of German publications with no international co-authorship decreased from 54% in 2007 to 45% in 2017 in Scopus, and from 53% in 2007 to 41% in 2017 in WoS placing Germany 10th amongst the countries compared in this report in terms of proportion of publications involving international co-authorship. Over the same time period, the proportion of publications that were collaborations between two or more authors from German institutions remained consistently between 15% and 17%. Combined, these figures indicate that German authors are co-publishing together at stable rates over the last decade, but they are co-publishing with authors outside of Germany more often than in previous years.

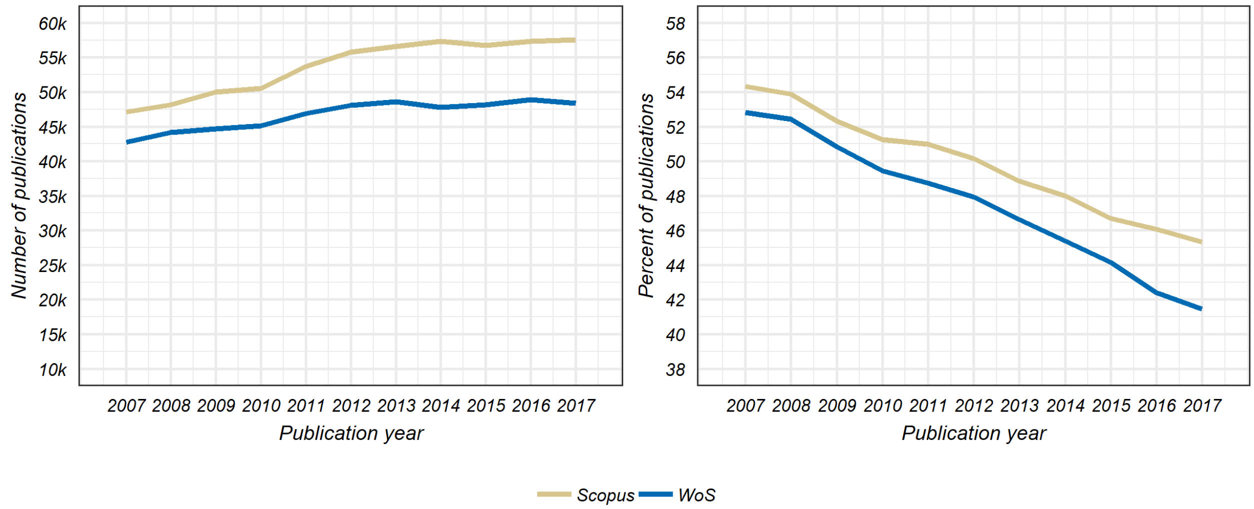


Figure 34 The number and percent of German publications without international co-authorship based on the whole count of publications, from Web of Science and Scopus, 2007-2017.

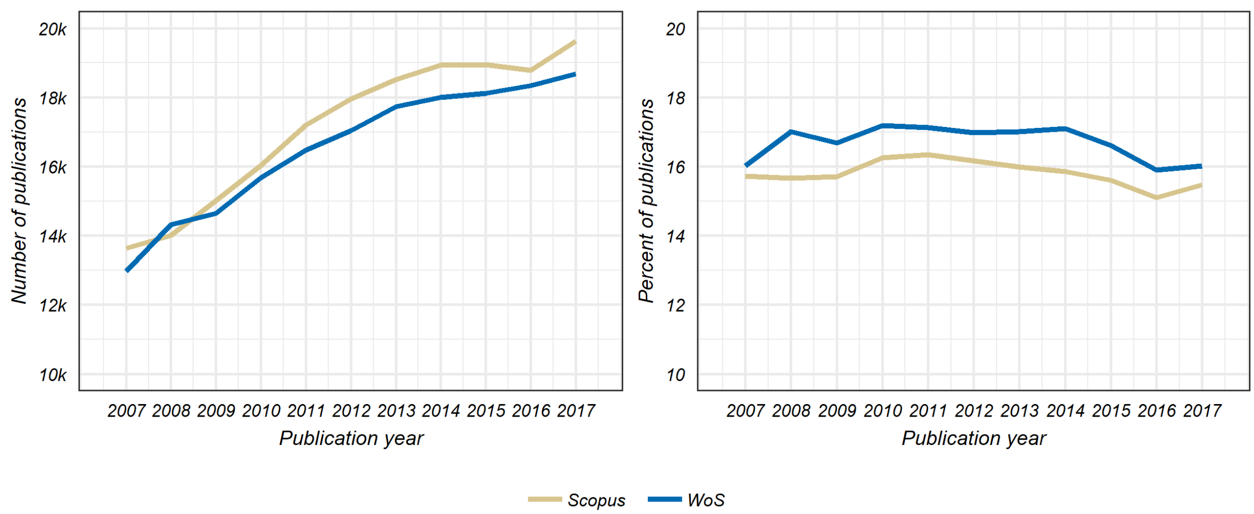


Figure 35 The number and percent of German publications with intra-Germany co-authorship based on the whole count of publications, from Web of Science and Scopus, 2007-2017.

This increase in international co-publications reflects a trend of higher rates of international collaboration across nearly all disciplines. The proportion of German publications involving international co-publishing in each discipline in 2007 and 2017 are presented in Figure 36. Authors in the hard sciences such as physical sciences and astronomy, earth and related environmental sciences, other natural sciences, and biological sciences exhibited the highest levels of co-publishing, with 65-75% of publications involving international co-publishing in 2017. With such high existing levels of co-publishing, these disciplines showed the least increase in co-publishing rates at between 15% and 23%, and co-publishing in ‘other natural sciences’ decreased by 3%. International co-publishing is likely strongest in these fields due to the need to share resources for large-scale projects and that such fields address research questions with worldwide applications.

In contrast, the largest increases in co-publishing were observed for disciplines primarily from the arts and humanities, such as philosophy, ethics and religion (up 225.5%), art (up 202.7%), other humanities (up

164.8%), political sciences (up 160.2%), and languages and literature (up 152.6%), although the proportion of publications from these disciplines which involved international co-authorship remained below 35% in 2017. Relatively less co-publishing is expected in these disciplines than the hard sciences as these disciplines often study regional issues. As such these changes may reflect an actual increase in the level of international co-publishing in the arts and humanities, but are also likely affected by the increased coverage in the databases of journals from these disciplines over time.

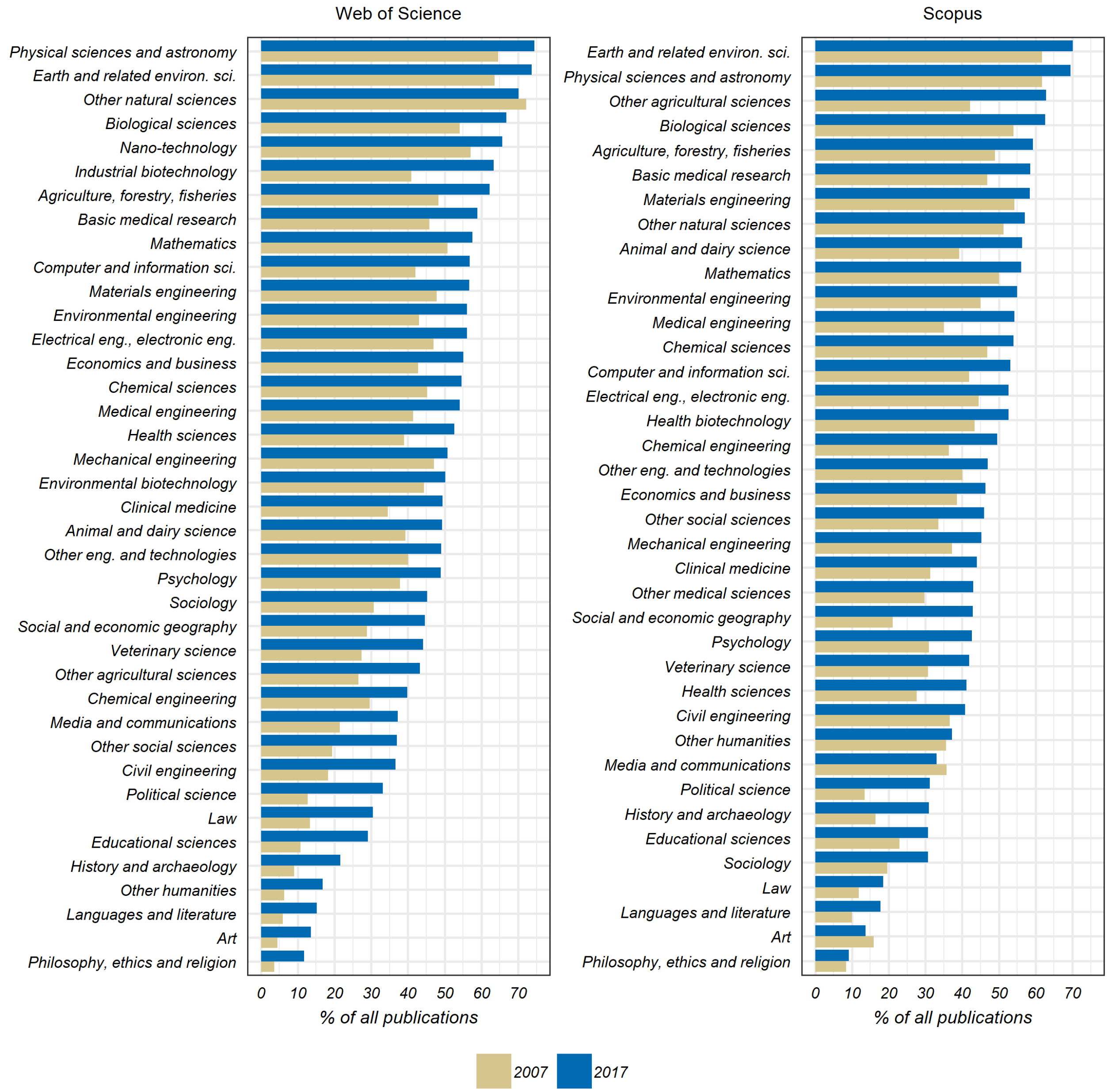


Figure 36 The percentage of whole counts of German publications with international co-authorship by OECD discipline in 2007 and 2017

4.4.2 Co-publication analysis for countries and groups of countries

The number of co-publications from each country and group of countries are presented in Figures 37 and 38, and the proportion of each country or group's publications that involved collaboration are presented in

Figures 39 and 40. All countries increased their number of internationally co-authored publications and consequently, the proportion of international collaborations between 2007 and 2017. Until 2013, Germany produced the third highest number of co-publications, behind the USA and the United Kingdom, until it was overtaken by China in 2014 to be pushed to 4th place where it continued to sit in 2017. These placements are at least partially due to the high rates of publishing from these countries and as such the proportion of publications which involved international co-authorship are a better indicator of co-publishing practices than numbers alone.

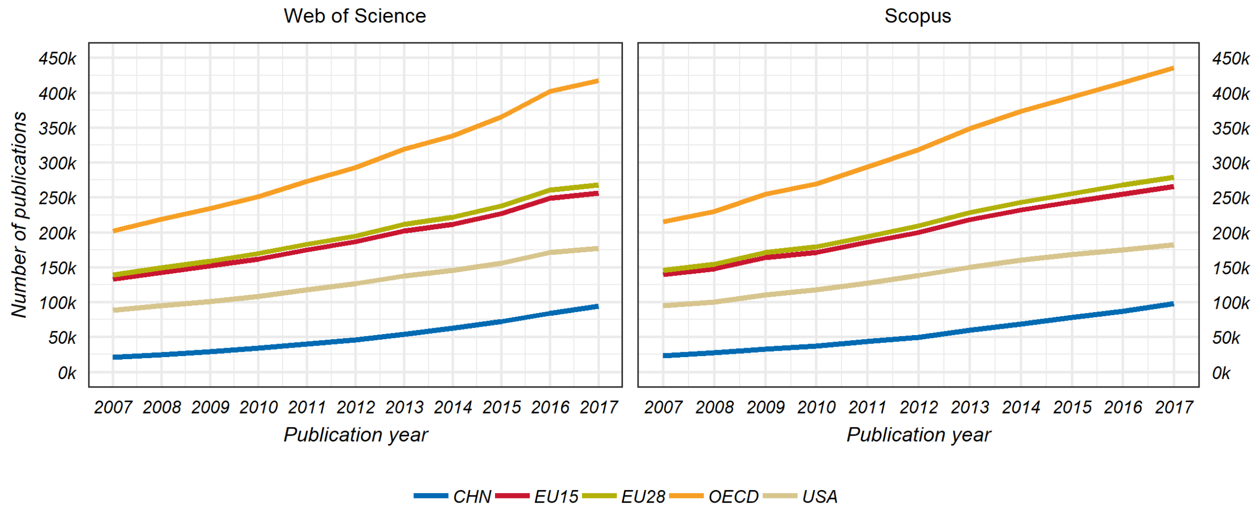


Figure 37 The whole count of publications from China, USA, and the EU15, EU28 and OECD countries which involved international co-authorship, from Web of Science and Scopus, 2007-2017.

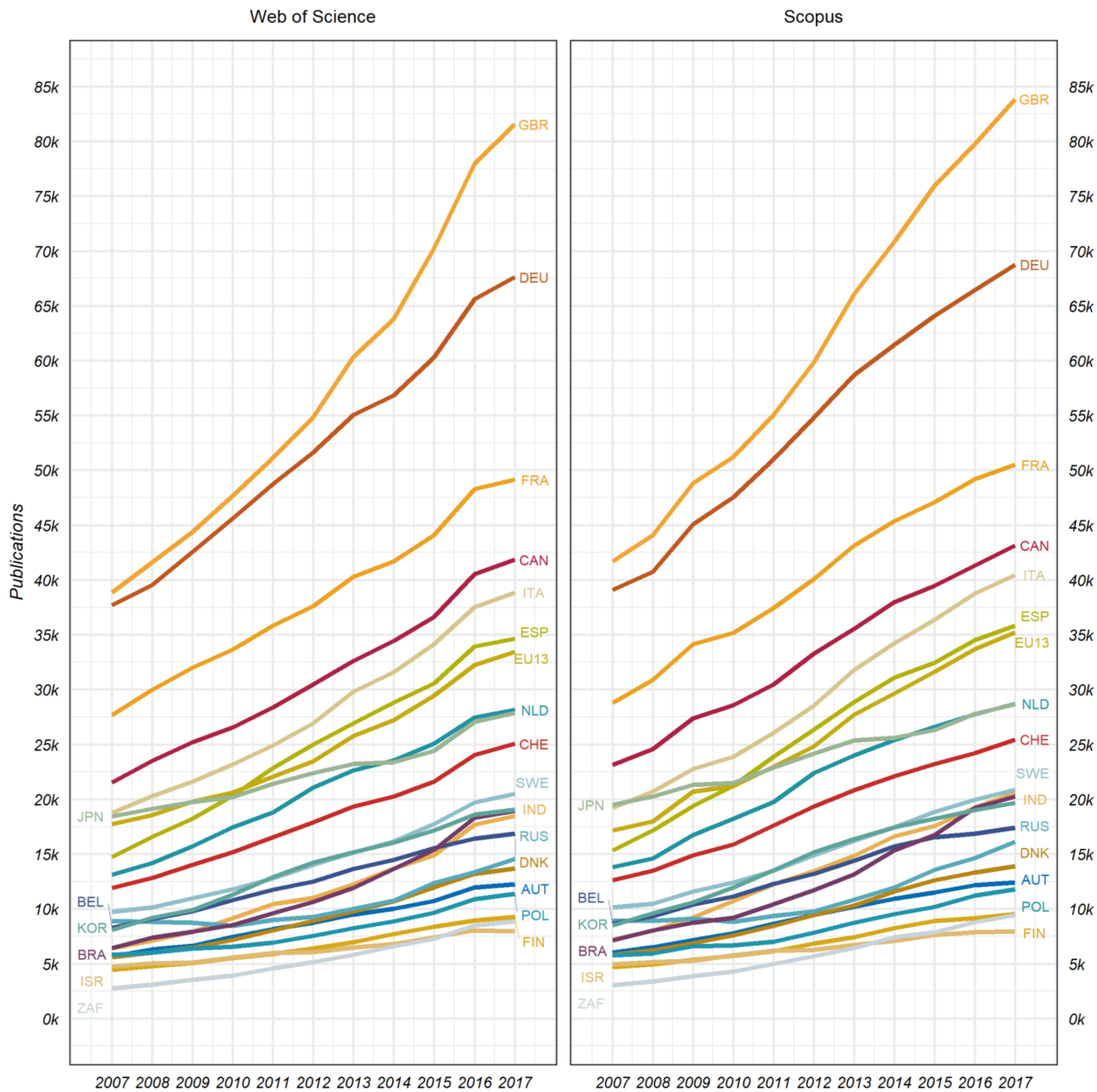


Figure 38 The whole count of publications from selected countries which involved international co-authorship between 2007 and 2017

When proportions are considered, Germany was 10th of the individual countries with 57.9% of its 2017 publications involving international co-authorship. Switzerland, Austria and Belgium had the highest levels of internationally co-authored papers with more than 70% of their publications involving other countries. Conversely, India, China, Japan and Korea had the lowest proportions of internationally co-authored papers at less than 35%. Despite being the highest producer of publications, the USA remained amongst the countries with the lowest proportion of internationally co-authored papers with 41% in 2017.

Scopus consistently recorded fewer internationally co-authored publications throughout the reference period than did WoS. The proportion of collaborations for most countries was approximately 1-3 percentage points lower in Scopus than WoS, although that difference was increased for countries such as

Poland (up to 5 points lower), India (up to 6 points lower), and China, Russia and South Africa (up to 8 points lower).

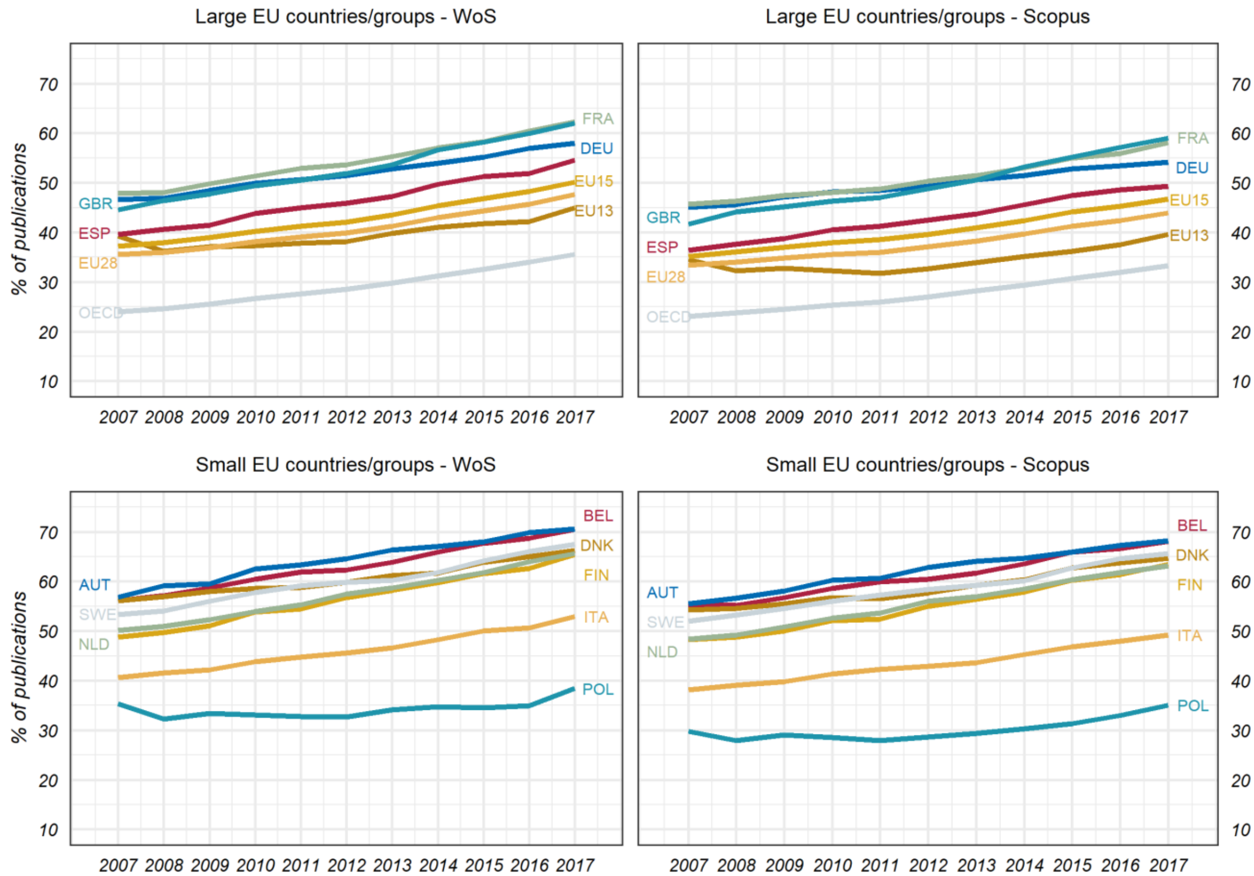


Figure 39 The percentage of publications from EU and OECD countries and groups which involved international co-authorship, based on whole counts, from Web of Science and Scopus, 2007-2017.

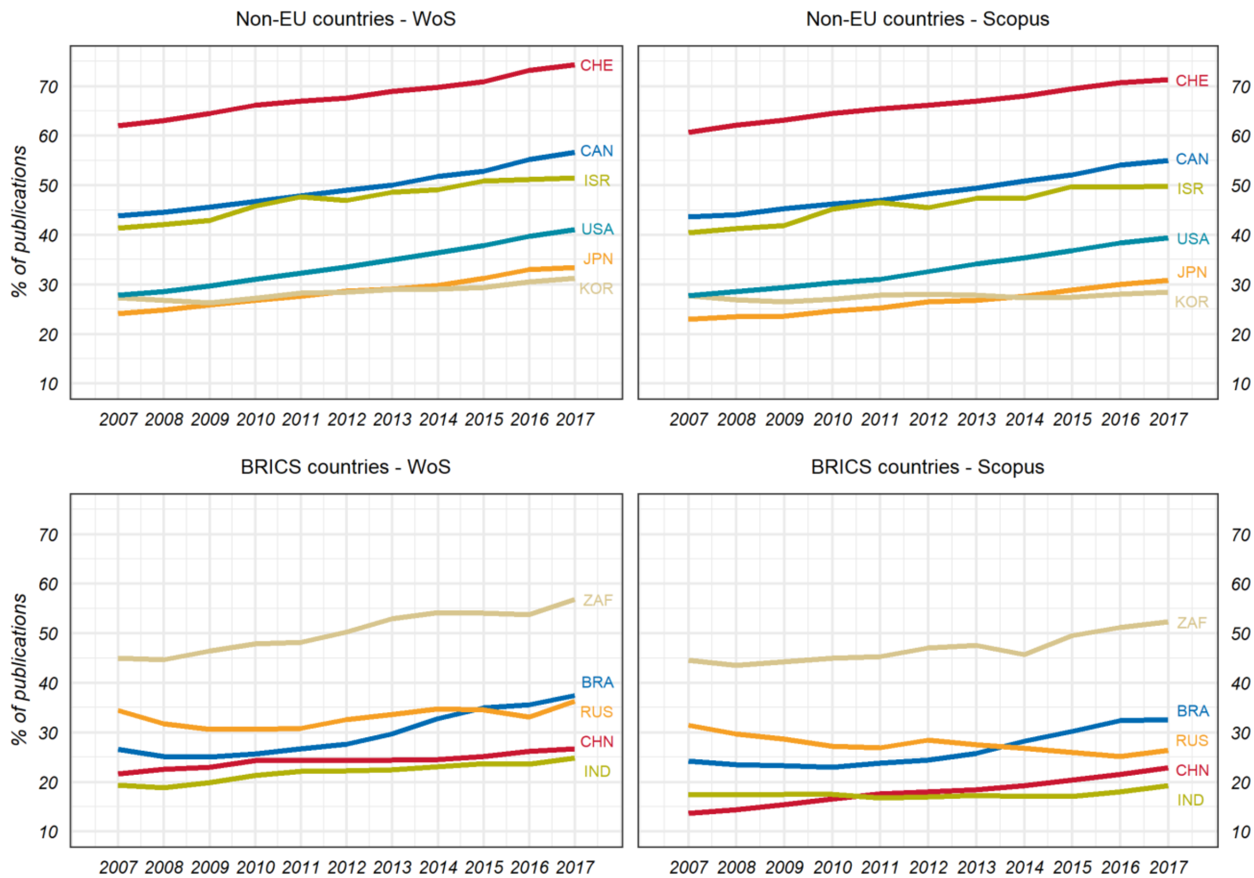


Figure 40 The percentage of publications from non-EU and BRICS countries and groups which involved international co-authorship, based on whole counts, from Web of Science and Scopus, 2007-2017.

4.4.3 Key international co-authoring countries with Germany and selected countries

The 20 countries which most often co-published papers with Germany, the USA, China, the United Kingdom and the Netherlands are identified in Figures 41 to 45 based on the proportion of publications co-published with each country. Germany's profile of countries it co-authors with is similar in both WoS and Scopus: Germany published most often with the USA, with a stable 30% of German publications co-published with the USA throughout the reference period. Other frequent partners included the United Kingdom (20% of German publications were co-published with the United Kingdom), France (approximately 13%), and Switzerland and Italy (both around 12%).

The most frequent co-authoring countries for the USA, China, the United Kingdom, and the Netherlands are similar to those for Germany. The USA is the most common co-authoring country for China, the United Kingdom, and the Netherlands. In particular, 46% of China's 2017 publications were co-published with authors from the USA. This finding is partially due to the USA's large number of publications, in addition to its strong scientific reputation. The USA co-published approximately 25% of its 2017 publications with authors in China (up from 9% in 2007), followed by 13% with UK authors and 11% with German authors. Germany is one of the five most frequent co-publishing countries for China, the USA, the United Kingdom, and the Netherlands, while Canada, France and Italy also often appear in the top five co-publishers for each of the selected countries.

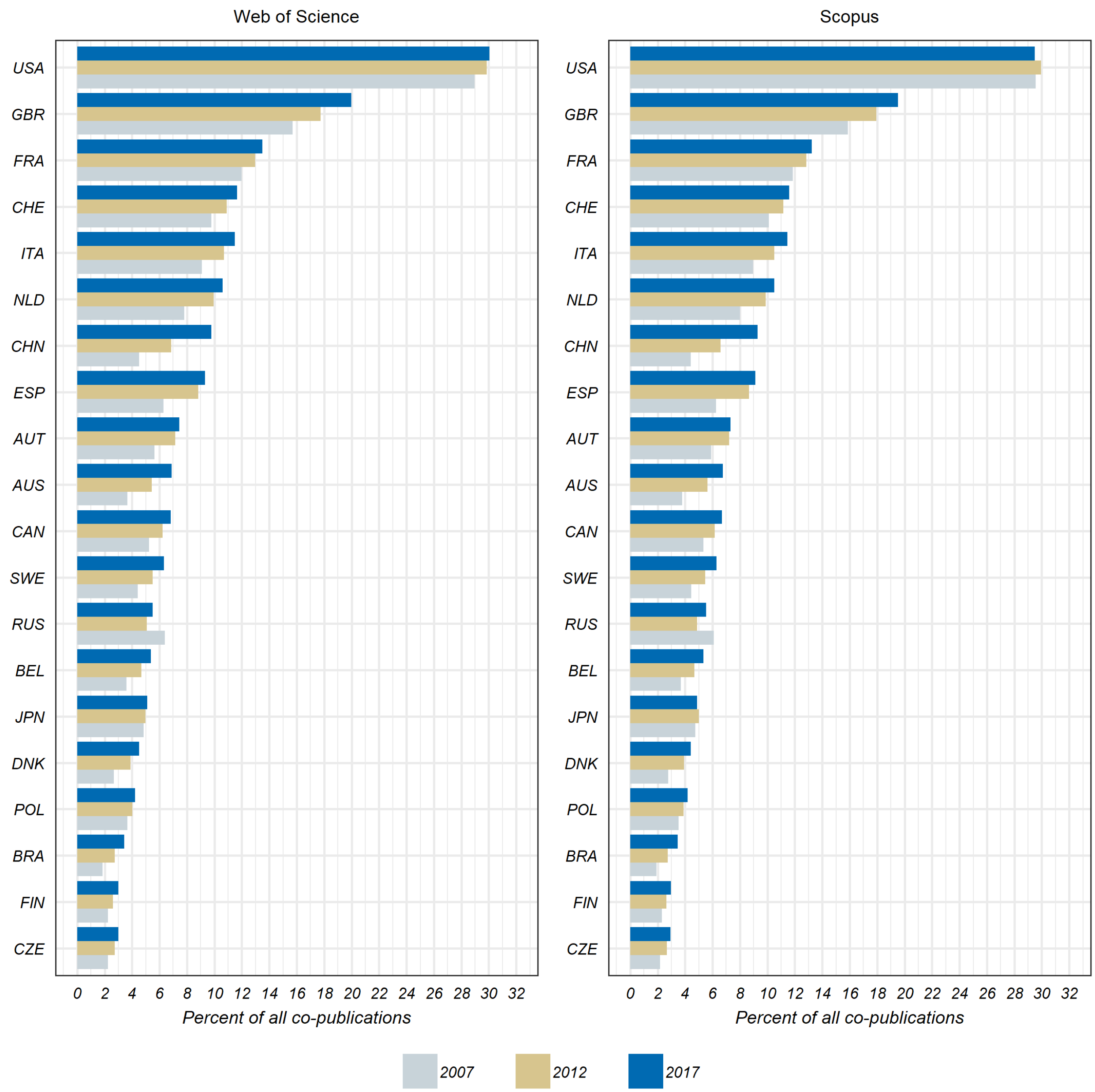


Figure 41 The 20 countries Germany most often co-publishes with based on the percentage of all German co-publications, whole counting, Web of Science and Scopus.

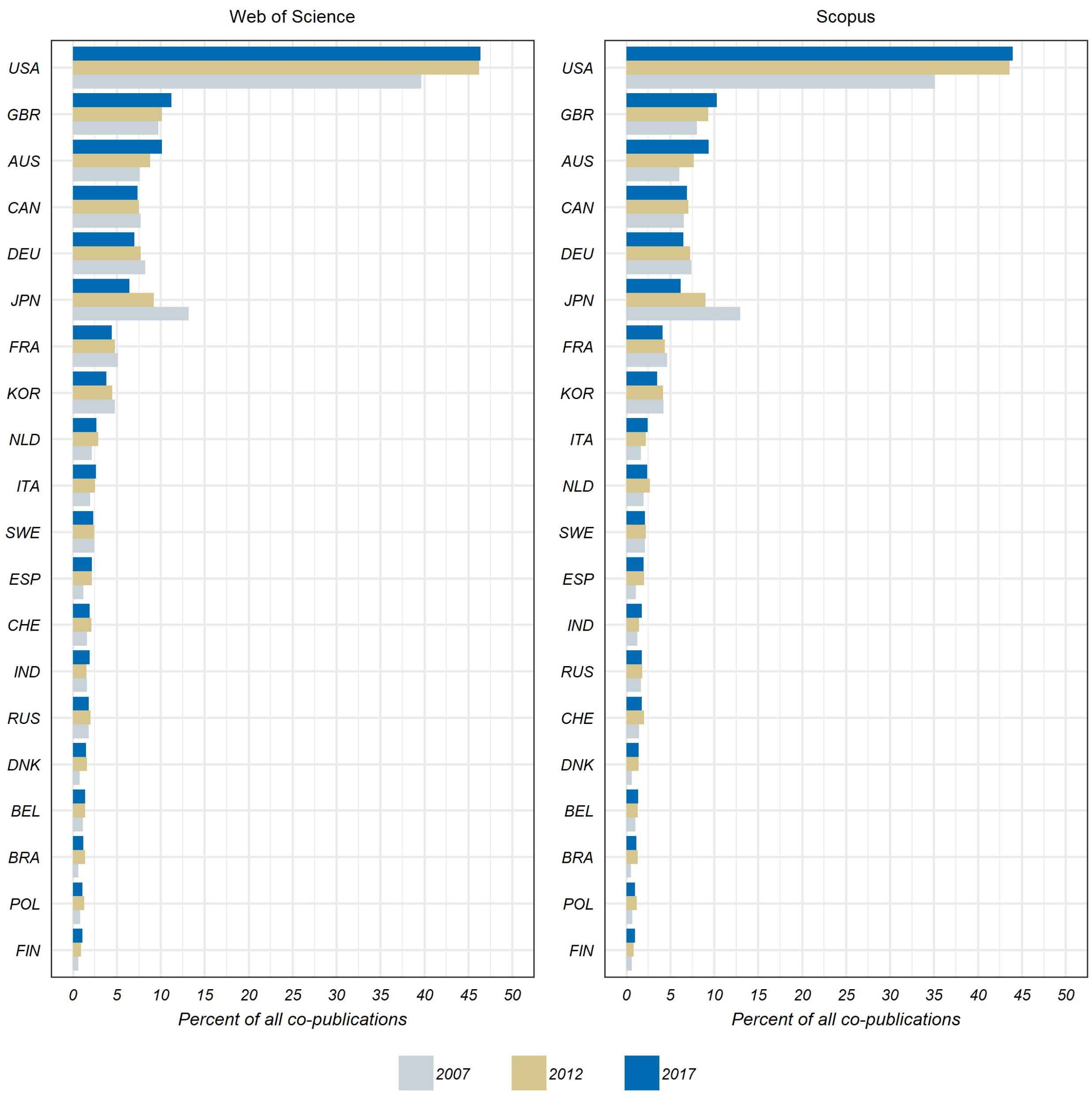


Figure 42 The 20 countries China most often co-publishes with based on the percentage of all Chinese co-publications, whole counting, Web of Science and Scopus.

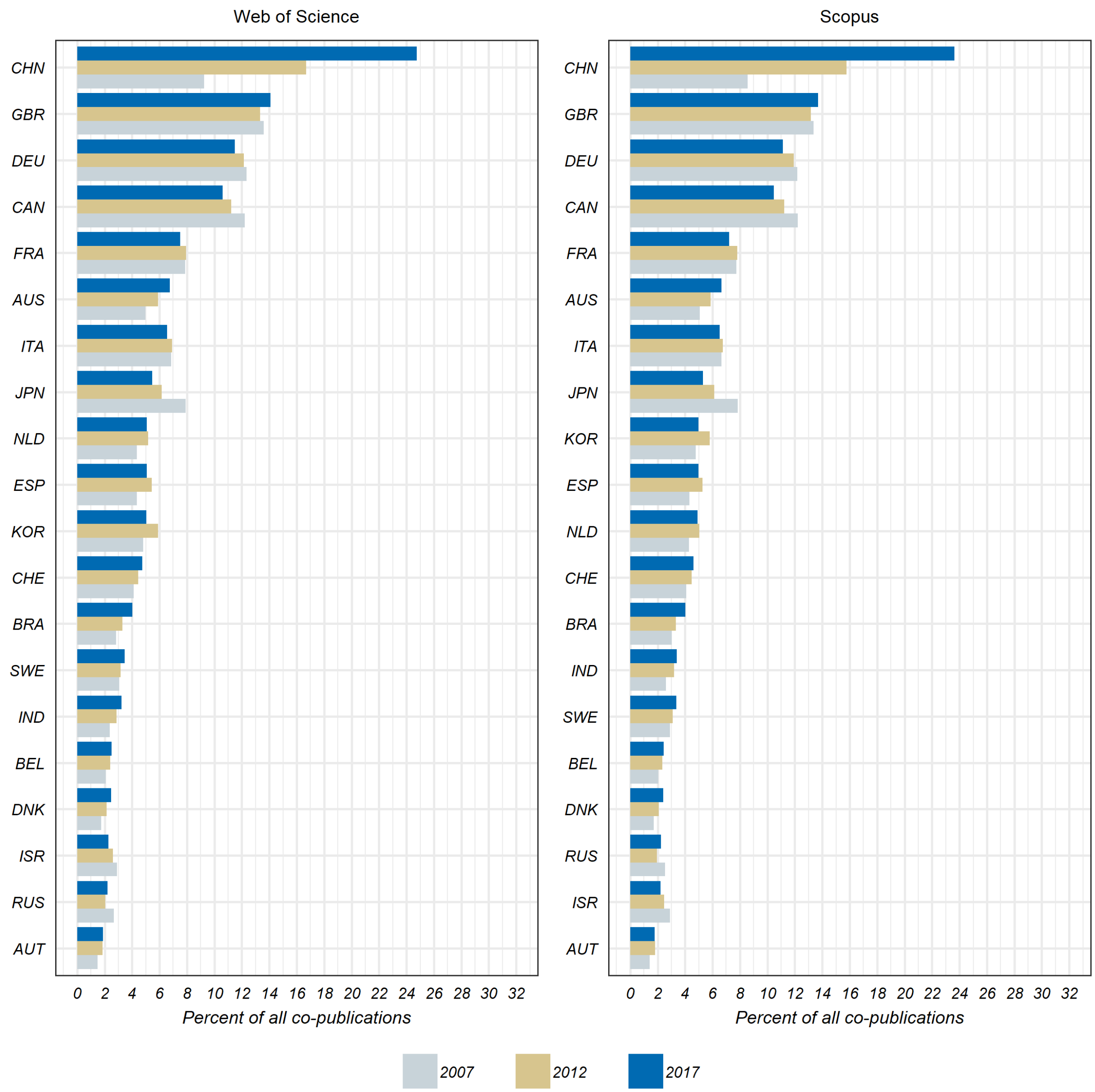


Figure 43 The 20 countries the USA most often co-publishes with based on the percentage of all American co-publications, whole counting, Web of Science and Scopus.

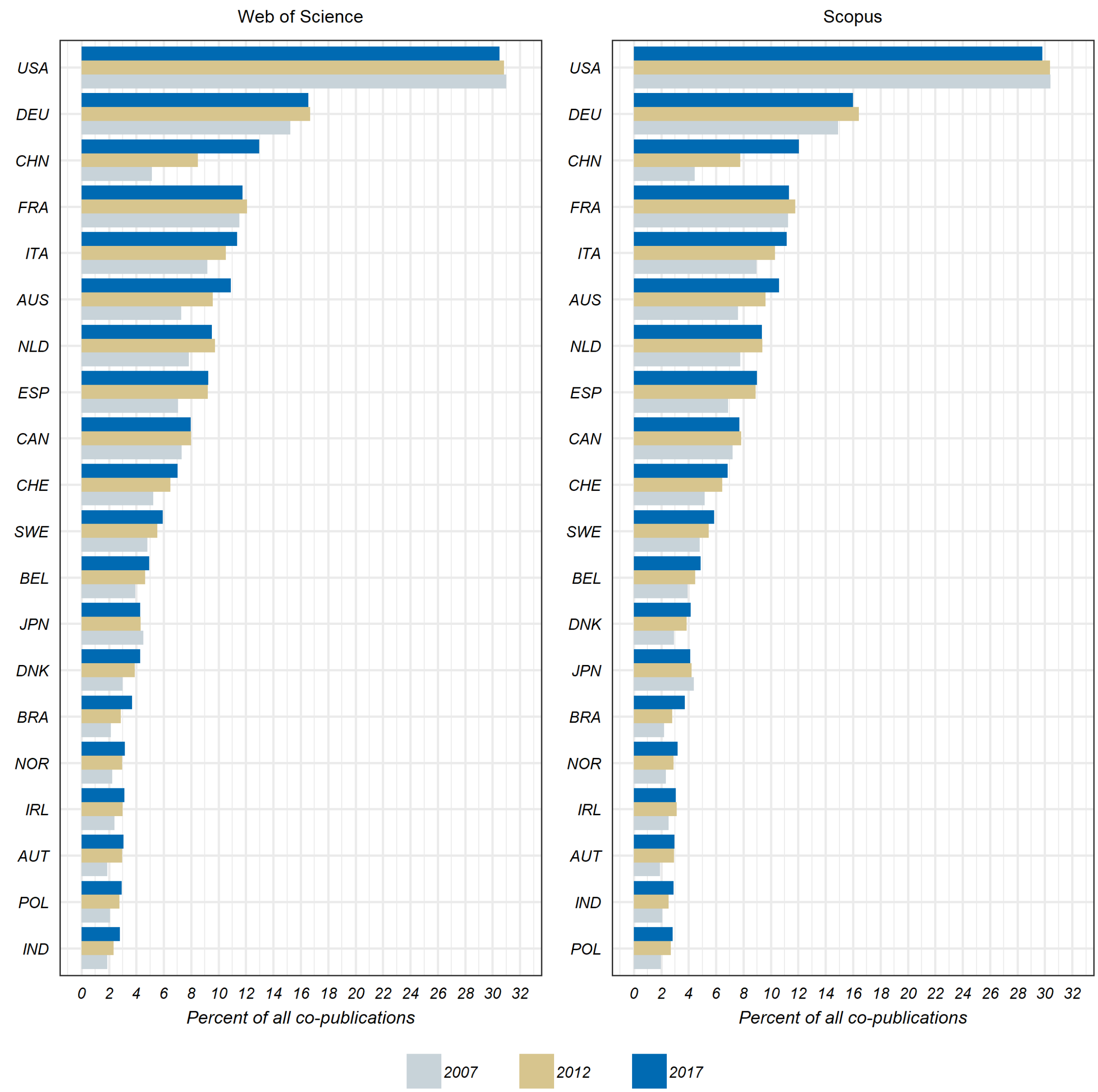


Figure 44 The 20 countries the UK most often co-publishes with based on the percentage of all UK co-publications, whole counting, Web of Science and Scopus.

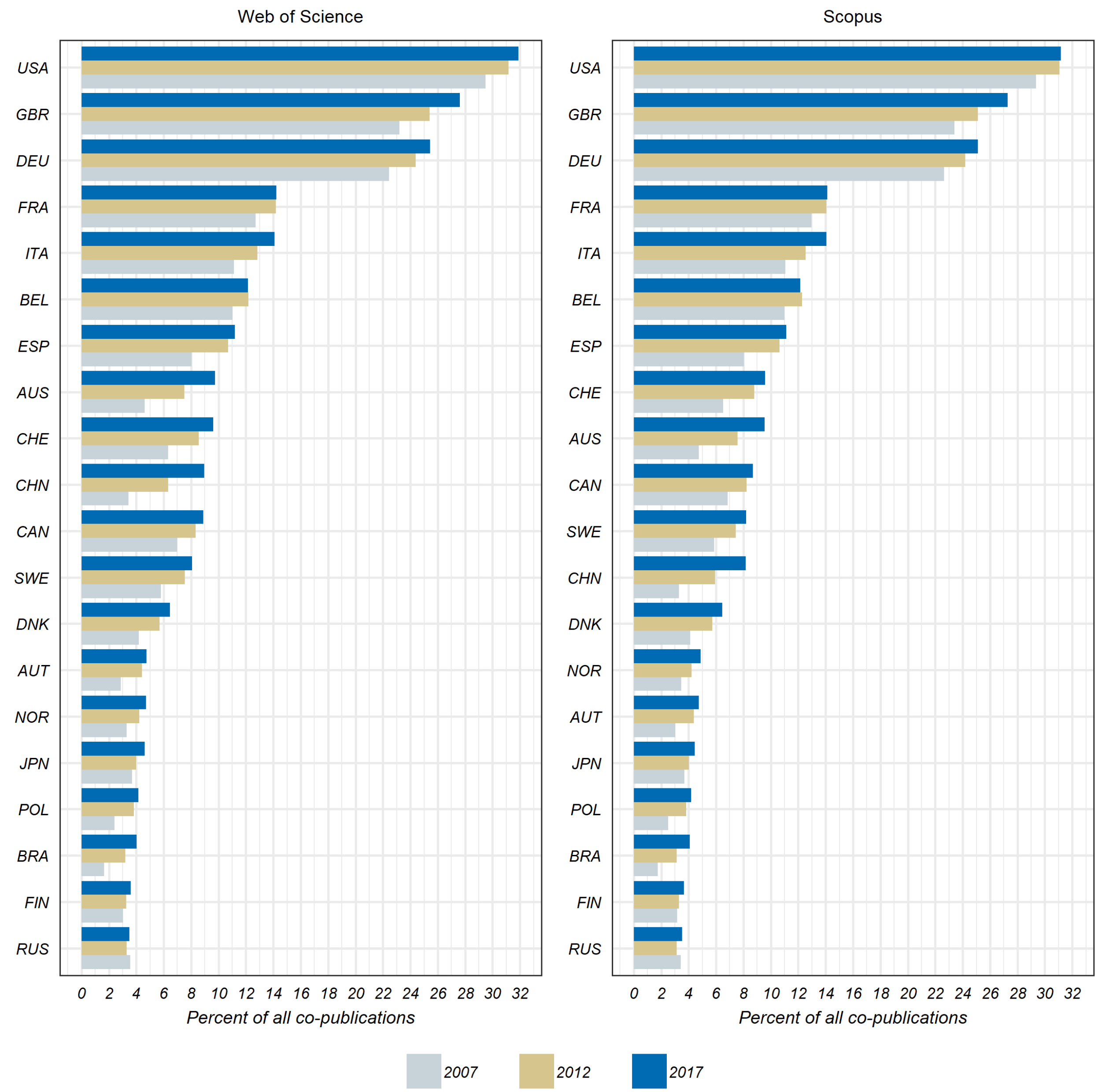
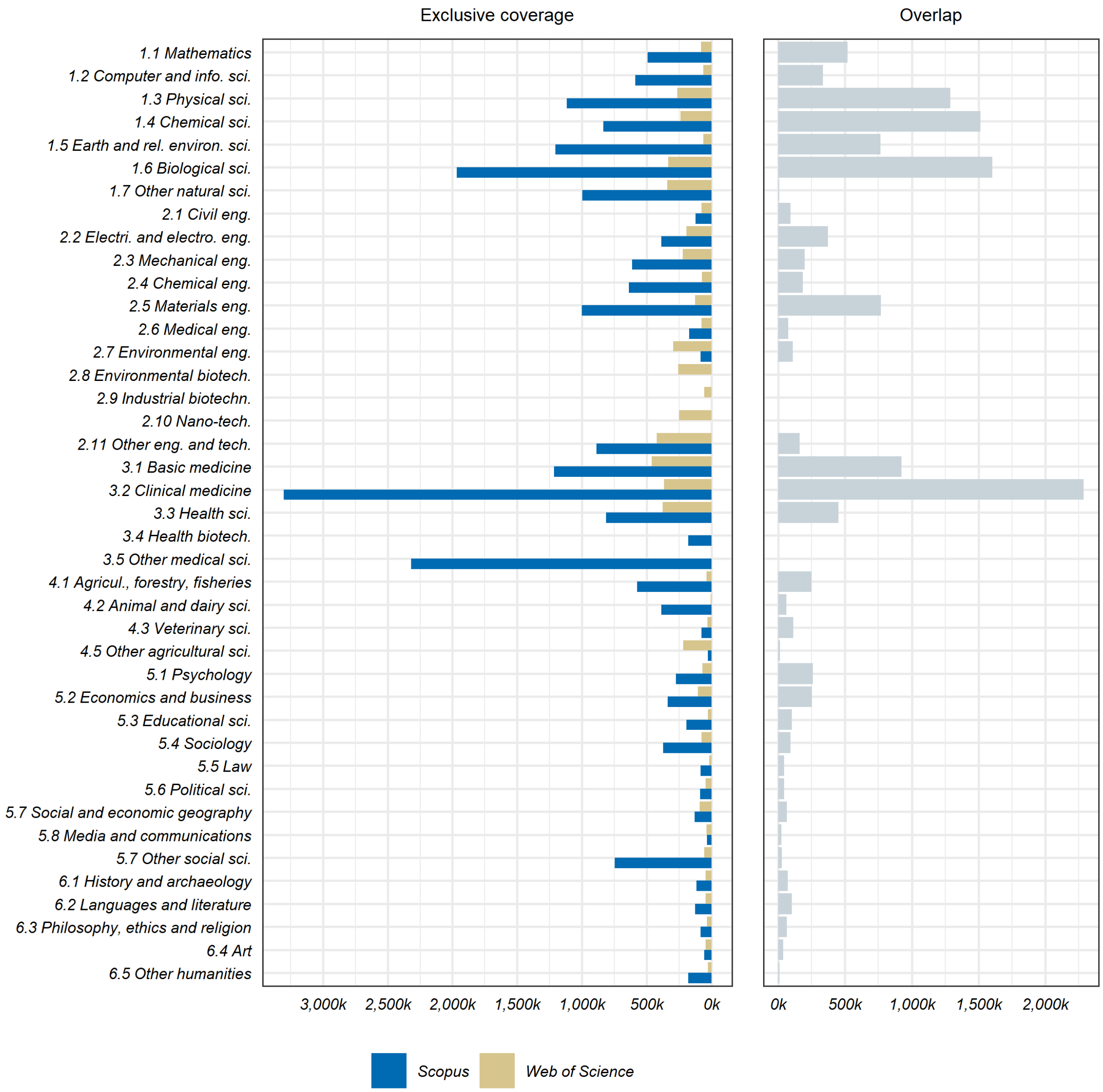


Figure 45 The 20 countries the Netherlands most often co-publishes with based on the percentage of all Dutch co-publications, whole counting, Web of Science and Scopus.

5 Micro-level comparison on diverging assessment between Web of Science and Scopus

5.1 Illustrating differences via duplicated and exclusive publications

Currently applied bibliometric indicators are mostly size-independent and implement this relative approach by relating a publication to a specific environment of similar publications. Due to differences in coverage, WoS and Scopus apply different environments to appraise the same publication, which results in the macro-level differences presented in this study. In order to analyse the impact of different environments on the valuation this section will analyse publications covered in both databases and compare the database-specific valuation of these publications. I.e. we investigate how the same content is evaluated by both databases to provide an understanding for the macro-level differences described before.



This procedure is motivated by the observation that the core differences between both databases consist of the respectively exclusive content. Consequently any differences in the valuation of the same content result from differences in the respective environment, i.e. the exclusive content. Hence a comparison of the diverging valuation of the same content does not inform on the content itself, but on the exclusive content causing any differences and therefore the databases themselves.

Figure 46 presents the overlap of articles and reviews[3] and the exclusive publications indexed in only one of the two databases by OECD disciplines[4]. As mentioned before, a strong focus on natural sciences and medical & health sciences might be observed. The share of more application-oriented engineering and technology sciences is lower, still, they obtain a substantial share of indexed publications. On the contrary few publications are indexed for agricultural sciences, social sciences or humanities.

Comparing the distribution of duplicated publications with the exclusive publications, we observe few differences. In terms of publication numbers also the natural sciences and medical & health sciences have by far the highest share of exclusive publications, the engineering and technology sciences obtain a lower share and agricultural sciences, social sciences and humanities the lowest share. Furthermore the share of additionally indexed publications in Scopus exceeds in nearly every discipline the exclusive share of the Web of Science and does so by a wide margin. While for example around 2.25 million articles and reviews are consistently assigned to the discipline *Clinical medicine* in Web of Science and Scopus for the years 2007 to 2016, Scopus lists another 3.25 million records in this category, while Web of Science adds less than 0.5 million additional records to this particular category[5]. In addition Scopus indexes another 2.25 million records in the category *Other medical sciences*, which does not exist in the Web of Science mapping. Consequently, Scopus holds a much larger number of articles and reviews in the OECD field *Medicine & health sciences*, a higher-level agglomeration of all medical and health-related OECD disciplines, than Web of Science and this observation also holds for all other OECD fields.

In general the additional share of publications indexed in Scopus mimics the distribution of the overlap, adding proportionally more publications by discipline. The numbers of exclusive publications of the Web of Science are smaller in magnitude and at least among the natural sciences, medical & health sciences and engineering & technology more evenly distributed. Exclusive publications in agriculutural sciences, social sciences and humanities are hardly detectable. Exceptions to these observations consist mostly of the residual classes *Other ** and occasions in which mappings do not make uniform use of a certain OECD discipline. In general the corpus of Web of Science indexed publications hardly diverges much from the corpus of duplicates, while Scopus presents a much larger corpus, which deviates proportionally from the overlap.

---

[3] Publications indexed in WoS and Scopus are identified in the German *Competence Centre for Bibliometrics* by comparing hash values on a subset of the available metadata strings.

[4] The database providers' official mappings were applied to convert the respective database-specific classification to the OECD FOS classification.

[5] Due to differences in the mappings of the database providers' classifications to the OECD classification, some duplicated publications might be assigned to different disciplines and are consequently included in the discipline-specific share of exclusive publications.

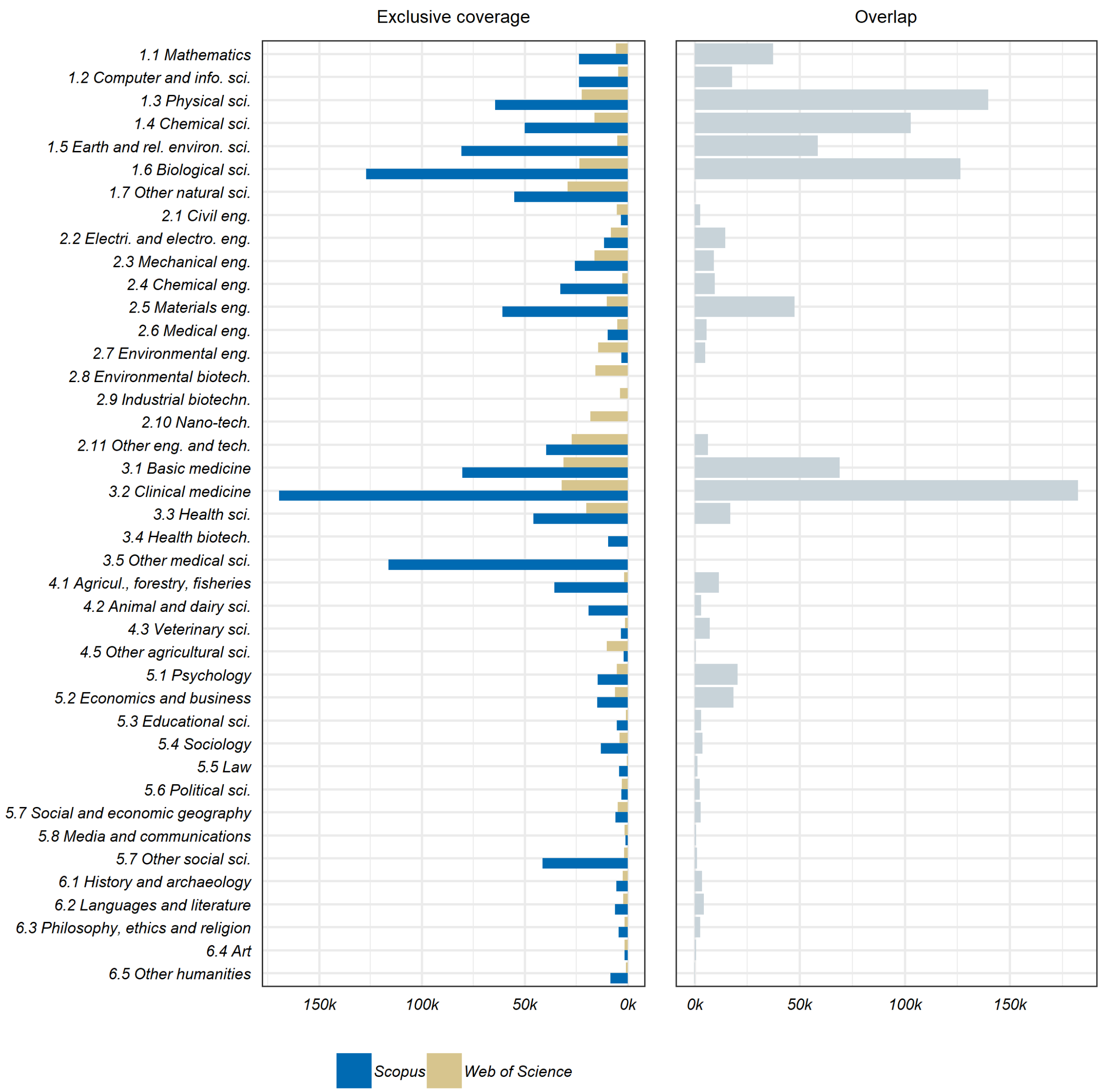


Figure 47 Exclusive and overlapping German articles and reviews in Web of Science (SCIE, SSCI, AHCI) and Scopus in 2007 to 2016 according to OECD Revised FOS disciplines.

Likewise observations might also be made by observing the number of exclusive and duplicated German publications in Figure 47. The so-called “hard” sciences and their applications in engineering and technology are responsible for most German duplicated and exclusive publications, while the respective share in the “soft” sciences is much smaller.

5.2 Resulting micro variation in disciplines and on publications’ valuation

Given the initial motivation and this study’s focus on German publications, we utilize the duplicates of German publications on the right side of Figure 47 to analyse the database-specific exclusive shares in generally indexed publications depicted on the left side of Figure 46. Therefore we will normalise every duplicated German publication in the two environments defined by the exclusive publications. In detail we compute for every duplicated article i affiliated to a German address in 2009 the ratio

$$\frac{obtained\ citations_i^{(s)}}{expected\ citations_i^{(s)}}, \qquad (1)$$

where s denotes the source of citation and expected citations counts, i.e. Web of Science or Scopus. Expected citations are computed as the mean number of citations received by articles from 2009 in the three-year post-publication period 2009-2011 assigned to the same OECD discipline[6].

By varying s we obtain a separate citation-based valuation of a duplicated German publication i. Hereby the variation in obtained citations and expected citations could differ, because the exclusive share of publications by each source s might have a varying effect on the general citation level in a discipline expressed in the expected citation counts and the particular impact of a publication i in that context, which might be stronger or weaker than the change in expected citations. Indeed differences in the change of obtained and expected citations faciliate changes in the valuation of publication i, as a uniform increase (or decrease) in obtained and expected citations would not change the ratio and the resulting valuation of the respective publication would stay constant.

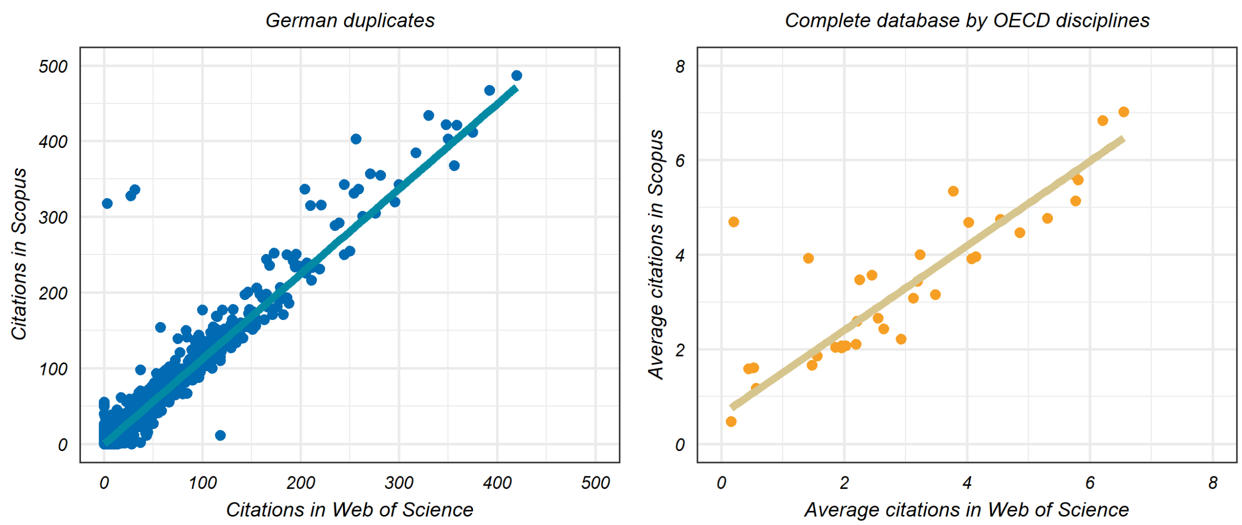


Figure 48 Scatterplot of citations obtained by duplicated 2009 German publications (left) and scatterplot of global average citation counts by OECD discipline (right). Outlier not shown.

Given equation (1) we firstly compare absolute citation counts and expected citation counts from the Web of Science and Scopus in Figure 48. In both graphs we observe a strong positive correlation underlining findings in previous sections, that in general both databases present a fairly similar picture of national bibliometric evaluation. However some publications obtain more (or less) citations in Scopus than postulated by the otherwise linear relation. Also the expected citations of the mapped OECD disciplines follow a clearly positive correlation, although most disciplines possess a higher expected citation count in either Web of Science or Scopus than postulated by the linear relationship. Especially these deviations in

[6] The national Excellence Rate illustrated earlier counts how many publications exceed in their citations the 90% quantile of citations in the particular discipline and consequently presents a robust, but single number instead of a distribution of normalised counts, complicating further analysis.

the obtained or expected citation counts from the mean line might cause a non-uniform change in the ratio of equation (1) and consequently a different valuation of German publications in the respective setting.

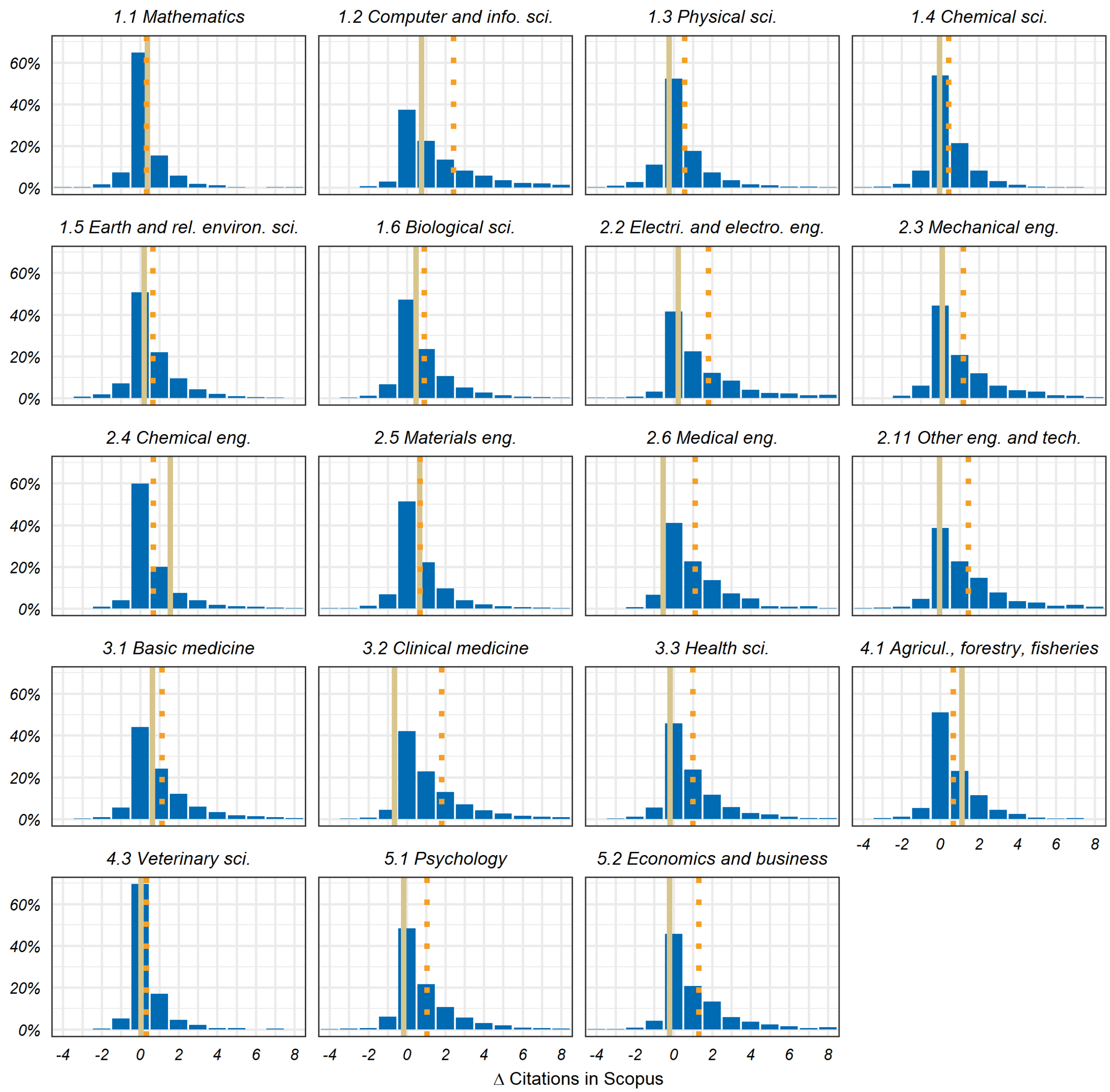
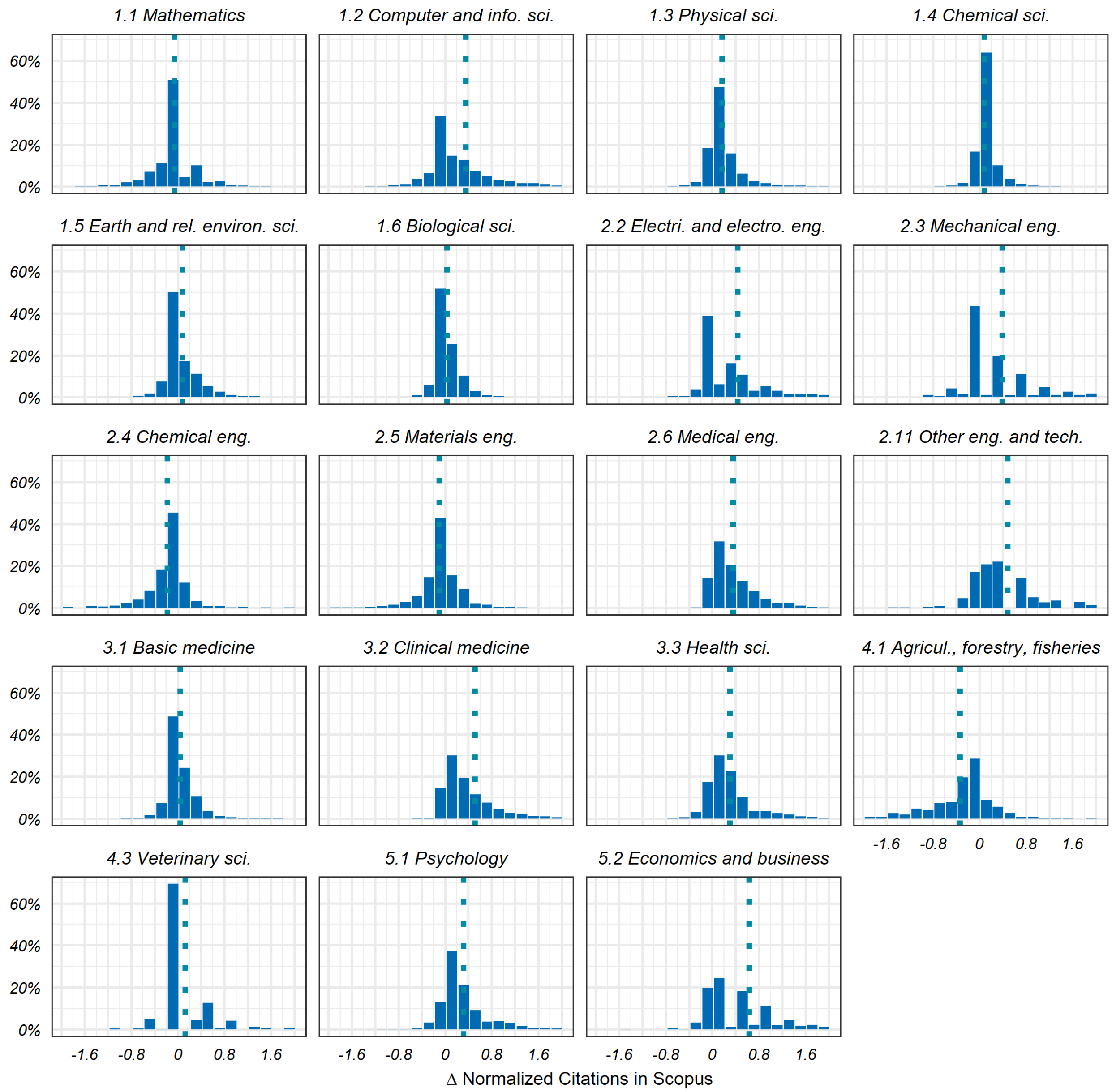


Figure 49 Histogram of different citation counts to duplicated German 2009 publications (blue bars), their average (orange dotted line) and differences in database-wide expected citations counts (dark yellow line) by OECD disciplines.

We explore any accruing discipline-specific deviations from the linear trend in Figure 49. For all 19 OECD disciplines with more than 400 duplicated articles affiliated with a German institution in 2009 we depict the difference in raw citation counts (blue bar), the mean difference (orange dotted line) and any difference in the expected citation counts. Hence more than 40% of such publications in the OECD discipline *Economics and business* obtain exactly the same number of citations in Web of Science and in Scopus. 20% of German duplicates in *Economics and business* receive one additional citation in Scopus. While in all disciplines the largest share of publications obtains exactly the same number of publications, we also consistently find a right-skewed distribution, in which aforementioned German publications receive in general more citations in Scopus than in Web of Science. The orange dotted lines depict the mean of these differences and consequently summarise the distributional effects in the blue bars in a single number. By comparing disciplines it might be noted that Scopus favours German publications in raw citation counts in some disciplines more than in others. Duplicated German publications in *Clinical medicine*, *Economics and business* or *Computer and info. sci.* seem to benefit especially from Scopus, while duplicated German publications in *Veterinary sc.* obtain few additional citations in Scopus.



Any difference in the raw citation count of a duplicated publication represents an altered standing of these duplicated publications in the different environments. General changes in the environments are expressed via differences in the expected citation counts, which are depicted via the dark yellow line in Figure 49. In general these changes in the expected citation counts seem less pronounced and are sometimes even negative. In any case the additional, exclusive publications by Scopus alter the expected counts because they differ structually, e.g. in citations over time or between disciplines, citations to non-indexed publications or different citation potentials due to longer or shorter reference lists, from the overlapping set of publications. Via these general changes in expected counts, as well via changes to the raw citation counts of duplicated German publications, the database specific exclusive content depicted on the left side of Figure 46 defines differences in the national bibliometric evaluation of Germany.

5.3 Observing structural differences via normalised citations

Given equation (1) every duplicated 2009 German article is individually contrasted with its corresponding expected citation count and the resulting distribution in differences between both databases is depicted in Figure 50. The blue bins indicate the percentage of duplicated German articles with difference Δ in normalised citations $norm.\ cit.$, i.e.

$$\Delta\ norm.\ cit. = \frac{obtained\ citations_i^{(Scopus)}}{expected\ citations_i^{(Scopus)}} - \frac{obtained\ citations_i^{(WOS)}}{expected\ citations_i^{(WoS)}}.$$

Accordingly more than 60% of German duplicated articles in *Veterinary sci.* find their normalised citation impact altered in the range of -0.2 to 0. The blue dotted line indicates the mean of this distribution of differences. As to be expected by Figure 49 the mean difference is positive for most disciplines. Duplicated German articles in the discipline *Agricul., forestry, fisheries* observe on average the most severe negative effect on their citation-based impact stemming from the use of Scopus and its set of exclusive publications. This might also be seen in Figure 49, where the average in additional raw citations for publications in this discipline is outrun by the change in expected citation counts. The same observation also holds for *Mathematics* and *Chemical eng.*, while articles in most other disciplines observe a positive average effect in line with the higher shares in the German Excellence Rate reported for Scopus than Web of Science in Figure 8.

The publication level rationale for the increase in the Excellence Rate for Germany might be observed in Figure 51. The top panel shows the distribution of $\Delta norm.cit._{(i)}$ for every duplicated 2009 German article. The 40%-60% quintile, depicted in the darkest blue shade, starts right above the zero line of no effect and almost reaches up to the 0.1 line. Given that a normalised citation count of 1 for a publication is commonly interpretated as exactly reaching the discipline-specific citation expectation an increase of 0.1 might be interpretated as 10% increase in this indicator. Consequently the 20% most strongly affected German duplicates improve on their normalised citation impact by over 25% in Scopus, while the 20%-40% quintile of affected German publications might note reductions on their normalised citation impact of -5% to zero. Consequently the distribution is skewed to the right allowing duplicated German publications to obtain on average a higher normalised citation impact in Scopus than the Web of Science.

The lower panel of Figure 51 subdivides the duplicated 2009 German articles by sectors based on the institution encoding of the German *Competence Centre for Bibliometrics*. In general all sectors show similar distributions to the national one. The distribution of the Higher Education Institutions sector, which is dominated by German universities, but also includes universities of applied sciences (Fachhochschulen) or specialised schools/colleges, resembles most closely the national distribution, as these institutions are responsible for the largest share of German publications. Surprisingly the MPG obtains the weakest effect on increased normalised citations of all sectors, although its Excellence Rate clearly outperforms all other sectors. On the contrary the strongest effect might be observed for the publications affiliated with the German business sector (denoted by “Economy”) and the FhG, which according to their Excellence Rate, find themselves on the lower half of citation-based impact assessment.

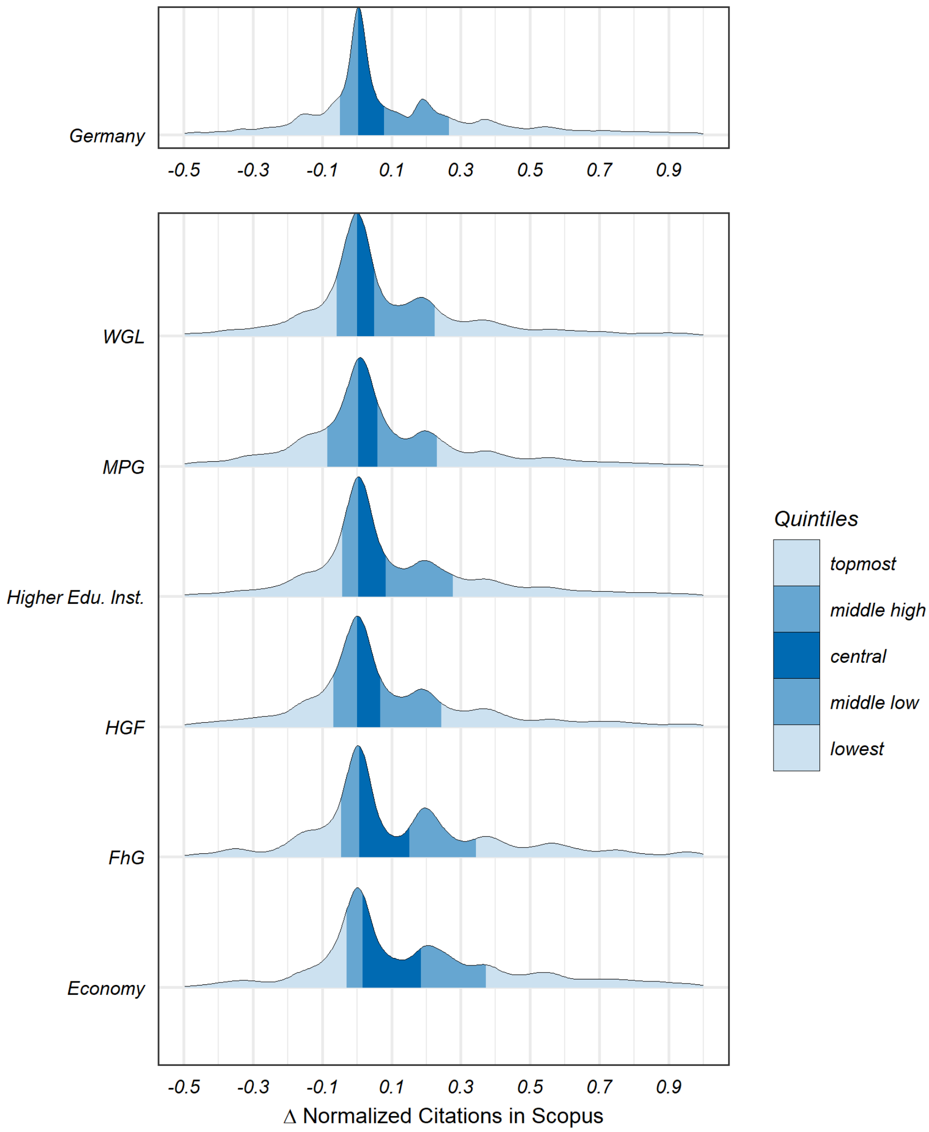
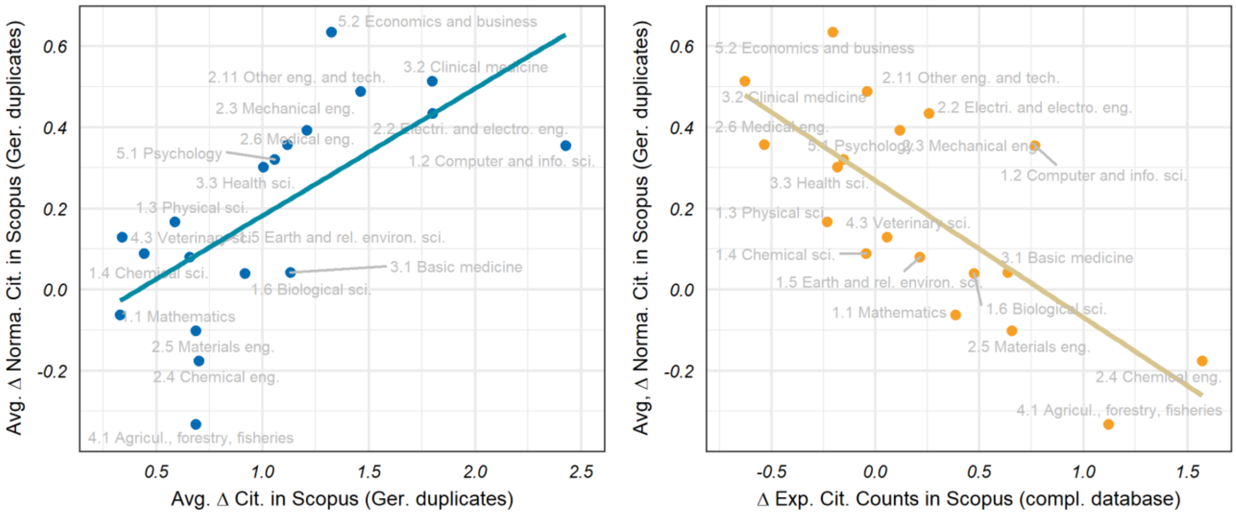


Figure 51 Density and quintiles of differences in normalized citations on all duplicated 2009 German publications and by sector.

Discipline-specific differences are given in Figure 52. In the left panel the average of differences in citations for duplicated 2009 German articles by discipline are contrasted with the discipline-specific average of differences in the respective normalised citations. The pronounced positive relations show that on average any additional gain in the raw citation count translates to the normalised citations count implicating that the expected citation counts adapt on average perfectly to the increase in raw citations. Consequently and as it has been observed in the macro-level comparison in the sections above, the Web of Science and Scopus both consistently report the same general standing of Germany in the global science system. Although raw citation counts are higher in Scopus the expected citation counts adapt accordingly resulting in the same overall picture.



However the right panel of Figure 52 indicates at the same time a discipline-specific difference between Web of Science and Scopus. We observe that the average of differences in normalised citations is negatively correlated with increases in the expected citation counts in the respective disciplines. Also, given the relatively small number of OECD disciplines available and any potential distortions by the database-specific mapping, duplicated 2009 German articles perform less well in disciplines in which the exclusive content of Scopus strongly alters expected citations counts. The substantial shake-up of these disciplines implied by the additional publications indexed exclusively in Scopus apparently penalises the valuation of German duplicates and might be analysed via network analysis disclosing the position and role these exclusive publications take on and thereby alter the valuation of German publications.

6 The growing influence of Chinese publication on Germany's bibliometric impact

6.1 Chinese publications as a non-marginal effect

Since 2016 China counts as the biggest single national "producer" of publications[7], only ten years after reaching the second rank in 2006, and only 16 years after outpacing other BRICS countries in 2000[8]. This increase of Chinese publications is extensive and thus beyond what could be considered marginal. It rather affects the whole database and every country participating in the global science system. Furthermore it is stimulated by economic growth and political factors (Zhou and Leydesdorff 2016) rendering it an external macro-level intervention on the science system. Consequently, the question arises, how this unprecedented growth of contributions from a single country influences bibliometric indicators. Bibliometric impact measures might be especially affected, as they relate a country's publications to the general publication universe nowadays strongly influenced by the Chinese publication growth.

To assess this effect we firstly apply author affiliation data to identify Chinese publications. We build our analysis upon the classical assumption, that any listed author has contributed an essential part to a publication and consequently assign any publication to the set of Chinese publications with at least one listed author affiliated to an institution residing in China.

Based upon this definition we quantify the effect of the Chinese publication increase on bibliometric impact measures by inferring what would have happened without them. Consequently, we constructed a counterfactual bibliometric world without Chinese publications and contrasted this with the actual bibliometric universe allowing for an assessment of the effect Chinese publications exhibit on the bibliometric universe.

This approach borrows from the treatment effect literature in Economics (Rubin 1974; Imbens and Wooldridge 2009). Based upon observational data, treatment effects models infer if and how a treatment causally affects a target audience. Ideally these models compare the same observational units with and without the treatment on some outcome variable and declare any difference to denote a causal effect of the treatment. Obviously any unit can either be exposed or not be exposed to the treatment and a direct comparison on the same unit is infeasible (Holland 1986). Consequently, treatment effects models apply carefully constructed substitute comparisons exploiting the untreated units of the population. However, as the Chinese publications affect the whole Web of Science publications universe, no unaffected units are available, but have to be constructed artificially.

Hence, we recount citation links after excluding the aforementioned set of Chinese publications. Next, these counts are applied to re-compute the Web of Science Subject Categories based citation statistics for the Mean Normalized Citation Score (MNCS)[9] and Excellence Rate[10]. We subsequently compare each

---

[7] according to fractionally counted articles and reviews indexed in Scopus, q.v. country shares in section *Bibliometric reporting comparing Web of Science and Scopus*

[8] This section is based on an elaborated country-level analysis (Stahlschmidt and Hinze 2018), which is extended to a specific analysis for Germany and its sectors. In line with this report to the EFI we also apply fractional counting instead of whole counting.

[9] The MNCS normalises the citations of every publication by comparing them to the average number of citations by other publications of the same discipline in the same publication year. This procedure facilitates a comparison between different disciplines and time periods which usually differ in their usage of citations foiling a comparison on absolute values.

non-Chinese publication to these statistics to obtain counterfactual national impact statistics $Impact_{counterfactual}$.

Finally, these values are contrasted with the actual national impact statistics $Impact_{actual}$ for sector i via

$$\Delta Impact^{(i)} = Impact^{(i)}_{actual} - Impact^{(i)}_{counterfactual}$$

where $\Delta Impact^{(i)}$ quantifies how national bibliometric impact indicators are affected by Chinese publications.

6.2 Quantifying the effect

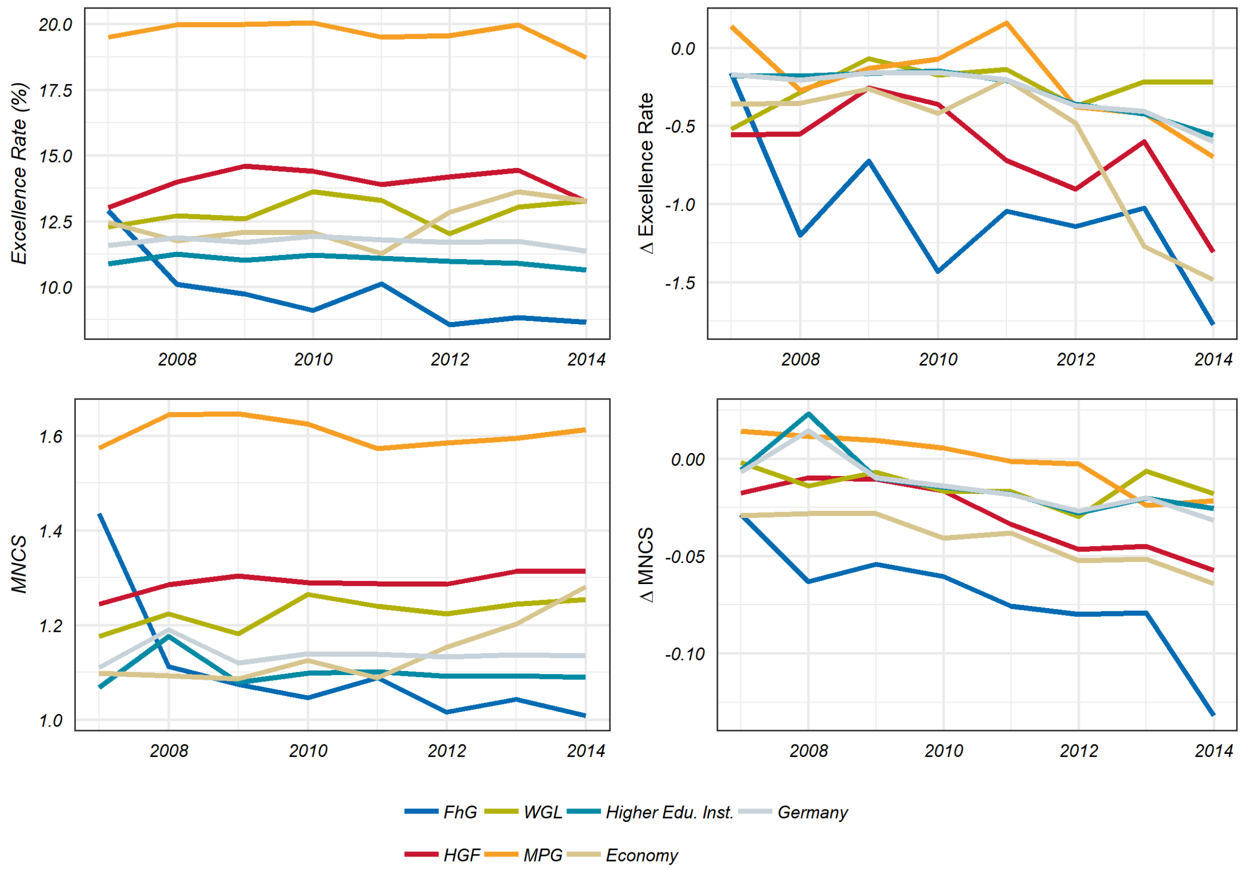


Figure 53 Upper panel: Actual Excellence Rate by sector (left) and changes induced by Chinese publications (right); lower panel: Actual MNCS by sector (left) and changes induced by Chinese publications (right).

The evolution of sectoral impact statistics like the Excellence Rate and MNCS are depicted on the left of Figure 53, while the graphs on the right detail the effect China exhibits as they illustrate the resulting differences $\Delta Impact^{(i)}$. In general all sectors apart from the Fraunhofer Society (FhG) find their share of highly cited publications to exceed the expectation of 10%. Germany as a country holds a stable share of more than 11%, while the Max-Planck-Society (MPG) excels in a rate of nearly 20%. Similar findings might also be observed by inspecting the sectoral MNCS. Every sector finds itself above the global

[10] q.v. section *Bibliometric reporting comparing Web of Science and Scopus*

expectation of 1, while the FhG performs on this inner research oriented metric less well and the MPG outperforms all other sectors clearly. Germany as a country exceeds the global expectation by more than 11%.

This stability is not to be expected given the increasingly negative effect Chinese publications impose on German publications according to the right side of Figure 53. In general Germany as a country but also the individual sectors observe negative $\Delta Impact^{(i)}$ values, i.e. their impact in the counterfactual world without Chinese publications is higher than their actual impact values. The FhG is affected most severely losing more than 1.5 % points on their Excellence Rate in 2014, while the effect on the MPG is less pronounced. Overall Germany as a country is clearly negatively affected and this effect is increasing over time. This combined observation allows for a new perspective on the stable actual impact, as the German science system apparently manages to hold on to its position although the general environment evolves against it.

6.3 Explaining the effect

The evolution of the environment is defined by changes in the ratio of obtained citations to expected citations (MNCS), respectively obtained citations evaluated against thresholds (Excellence Rate). In this respect Stahlschmidt and Hinze (2018) show that in general terms the obtained citations and any field-specific citation statistics increase due to the growing share of Chinese publications.

These changes in the obtained, respectively field-specific citation statistics represent empirical symptoms of underlying mechanisms, which can be exemplified via a highly stylized bibliometric toy model. We omit for a moment any citations across time or disciplines and any influence non-source items might have. In such a perfectly encapsulated setting citations are distributed as a zero-sum game from reference lists and the expected citation count equals the average reference list length (Garfield 1979). Consequently, any expansion in terms of additional publications might only increase (decrease) the field-specific citation statistics if these additional publications include more (less) than usual references.

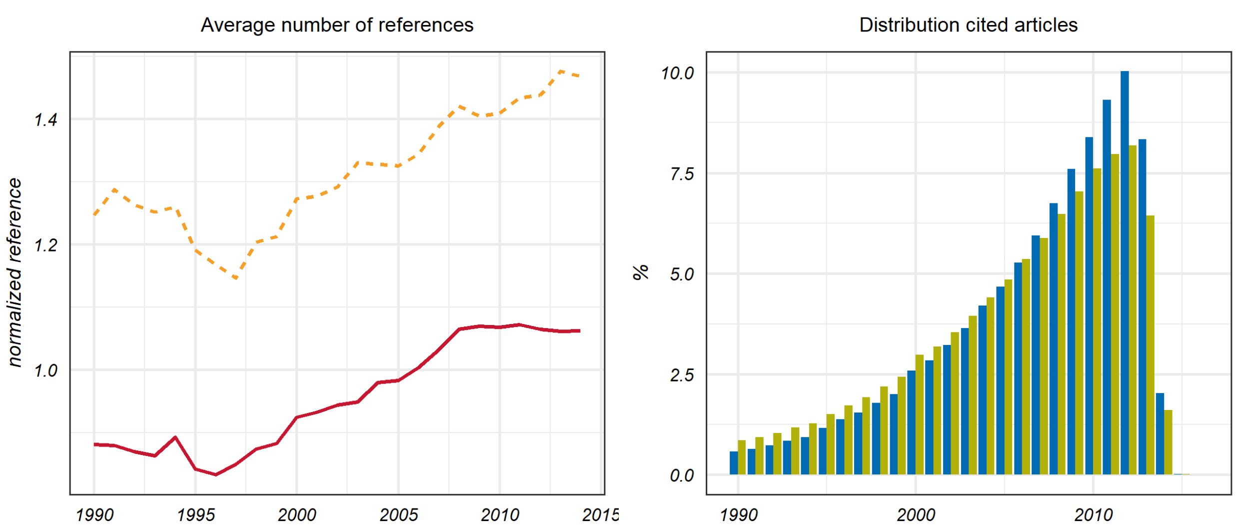


Figure 54 Left panel: Normalized length of Chinese reference lists (solid red line) and normalized length of Chinese reference lists within citation window (dashed orange line); right panel: Distribution of cited article over time from citing articles for Chinese (blue) and non-Chinese (green) publications from 2014.

In Figure 54 we depict the average of the normalized reference list lengths of Chinese publications (solid red line). These lists have been shorter, i.e. below 1, than the global average in the beginning of the observation period, but due to substantial growth around the millennium change the average normalized

Chinese reference list has exceeded the world average since 2006. However, the observed increase in field-specific citation statistics begins much earlier and can be observed uniformly across all Subject Categories. Therefore Chinese publication seem unlikely to differ strongly in their use of citations across disciplines. On the contrary citations crossing the yearly time periods are common and the varying time focus of Chinese and non-Chinese publications can exemplarily be observed for citing publications from 2014 in the right panel of Figure 54. Accordingly Chinese publications focus more strongly on more recent publications, as their share of references to publications from 2007 to 2014 surpasses the non-Chinese shares and trails them for all preceding years.

But the applied three-year citation window curtails the count of relevant citations to cited publications not older than two years and consequently favours the Chinese focus on more recent literature. Accordingly the dashed line in the left panel in Figure 54 restricts the count to references within the three-year citation window. Thus Chinese reference lists exceed the non-Chinese ones in terms of citations utilized in the whole observation period and along the way raise field-specific citation statistics.

In general any country will benefit from Chinese publications if the additional citations received will outweigh the rise in field-specific counts. Two factors might influence to what extent other countries will receive additional citations from Chinese publications. First an outward-looking China, which cites foreign papers relatively more often than national publications. Second a non-uniform spread of these citations among countries.

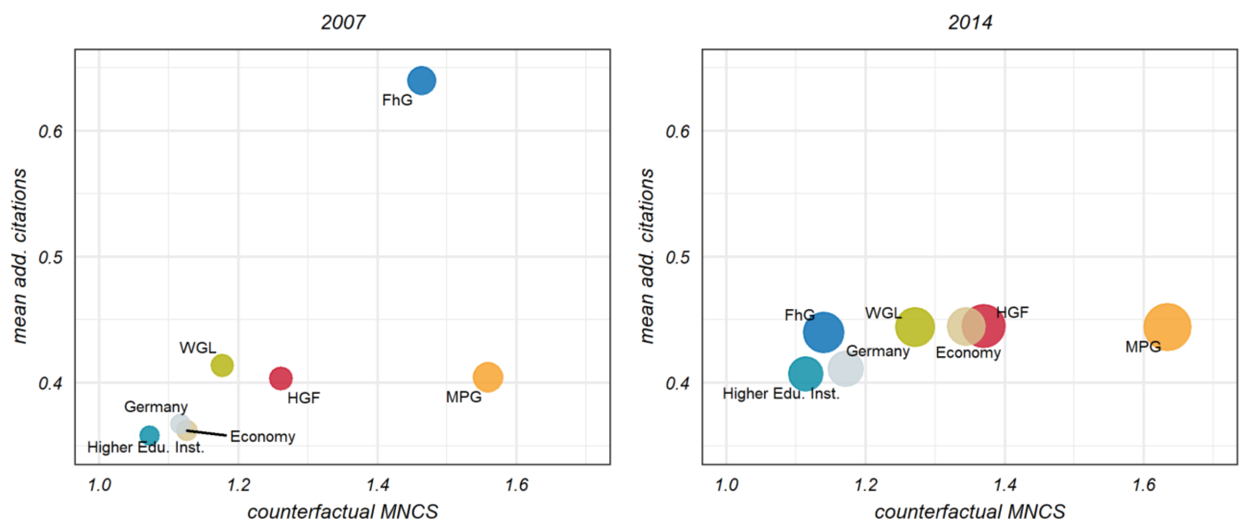


Figure 55 Sectoral difference between counterfactual national impact and average, normalized, additional citations from China in 2007 and 2014. Point size indicates the share of national publications becoming cited by China.

As the share of Chinese national self-citations stays constant (Stahlschmidt and Hinze 2018), we analyse how outward citations from Chinese publications are distributed among German sectors. Figure 55 graphically relates the additional Chinese citations received by sectors in 2007 and 2014 to their concurrent general scientific impact expressed via the counterfactual MNCS (x-axis). In detail the y-axis describes how many additional normalized citations a paper receives from Chinese publications on average once it is cited by a Chinese publication and the point size denotes the share of national publications being cited by Chinese publications. If Chinese publications were to adopt the common citation practice, we would expect a positive relation between a sector's global scientific impact without China and the additional citations received from Chinese publications.

Comparing the two graphs we note at first the growth in point size between 2007 and 2014, i.e. the increase in the national share of papers being cited by Chinese publications. This observation, caused by the rising number of Chinese publications, facilitates the increase in the size of the effect Chinese publications exhibit on Germany and other countries' impact statistics. Another noteworthy change affects the FhG, which not only decreases in its counterfactual MNCS, but also in the number of obtained citations from Chinese publications.

While this valuation of Chinese publications of FhG publications aligns consistently with the decrease of the FhG in the counterfactual setting, the share of additional citations obtained by the MPG does not align with its counterfactually evaluated impact. In 2007 as well as in 2014 the MPG obtains relatively fewer additional citations than postulated by its counterfactual citation impact and finds itself only on the same level as the Helmholtz Association (HGF) and Leibniz Association (WGL) although the globally evaluated counterfactual impact clearly exceeds the impact by the HGF and WGL. This difference in relevance attribution by Chinese publications has also been observed for Germany and other OECD countries (Stahlschmidt and Hinze 2018), as these countries gain fewer citations than postulated by their counterfactual standing.

6.4 Conclusions

Contrary to the counterfactual status quo Chinese publications exhibit a strong citation focus on Asian countries (Stahlschmidt and Hinze 2018) and due to their non-marginal size affect the citation-based impact of Germany and other OECD countries negatively, i.e. while Germany and its sectors obtain more citations than before, the concurrent increase in field-specific citation statistics exceeds the increase in obtained citations and causes a negative effect on the measured impact. While any normative appraisement of the diverging perspective on relevance by Chinese publications might be hard to make, the observed mechanism highlights that current citation-based impact statistics not only mirror the national performance, but also changes in the global environment which might be accounted for in interpreting citation-based impact statistics.

7 Appendix A: Conference Proceedings in the Web of Science

The number of conference proceedings published by each country between 2007 and 2017 indexed in WoS are shown in Figures 56 and 57, and the proportion of worldwide proceedings these account for are in Table 11. The decrease in numbers between 2016 and 2017 is likely a processing artefact as other document types, such as articles and reviews, take precedence in processing over conference proceedings. Although China and the USA published similar numbers of proceedings in 2007, since then China has published between 28% and 55% more conference proceedings per year than the USA. Of note in Figure 57 is that despite the United Kingdom producing the third highest number of articles and reviews, it drops to sixth place after India, Russia, Germany, Japan and Italy in relation to proceedings, suggesting the disciplines in which the United Kingdom predominantly publishes may not be conference-oriented disciplines. As with articles and reviews, the number of Indian proceedings published has dramatically increased over time - a 561% increase between 2007 and 2016 - which increased its share of worldwide publications from 1.6% in 2007 to 7.4% in 2016. Russian proceedings have also risen, increasing in number by 379% since 2007 to account for 4.1% of worldwide proceedings in 2016. Germany's proceedings remained relatively stable, with between 12,400 and 15,600 proceedings published per year betwen 2007 and 2016 accounting for 4-5% of worldwide proceedings, placing Germany consistently fourth behind China, the USA, India and Japan, having only been overtaken by Russia in 2017.

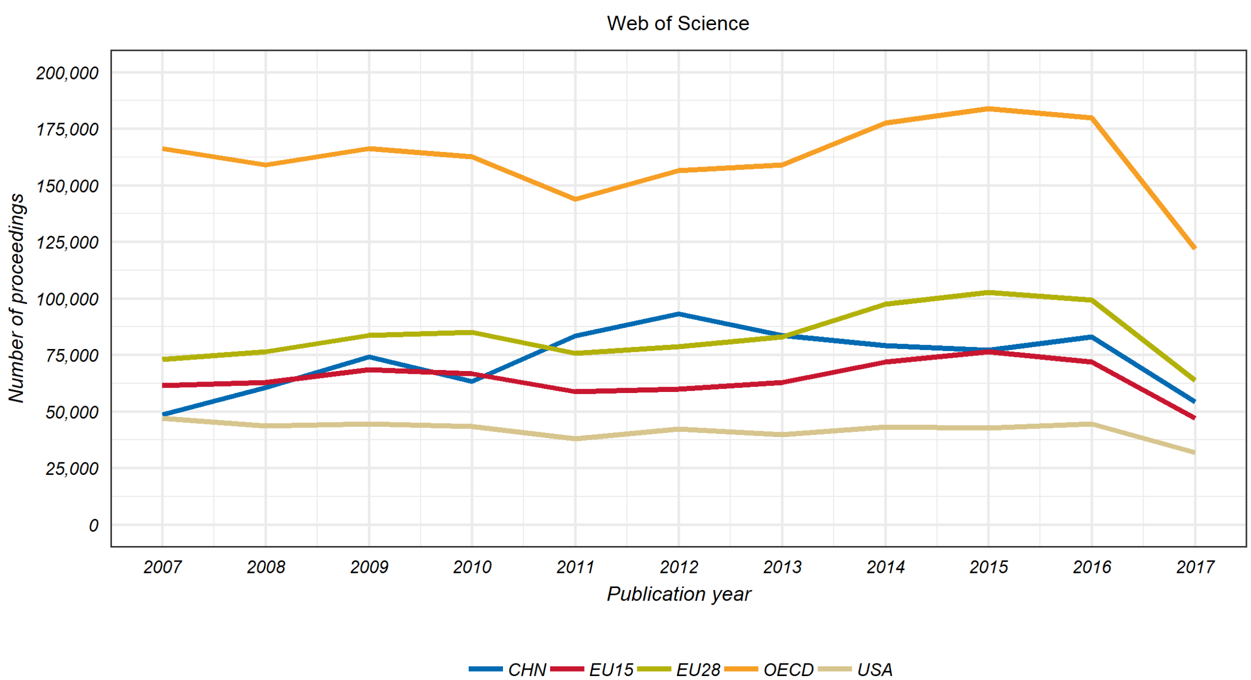


Figure 56 The fractional count of conference proceedings from China, USA, and the EU15, EU28 and OECD countries between 2007 and 2017 from Web of Science.

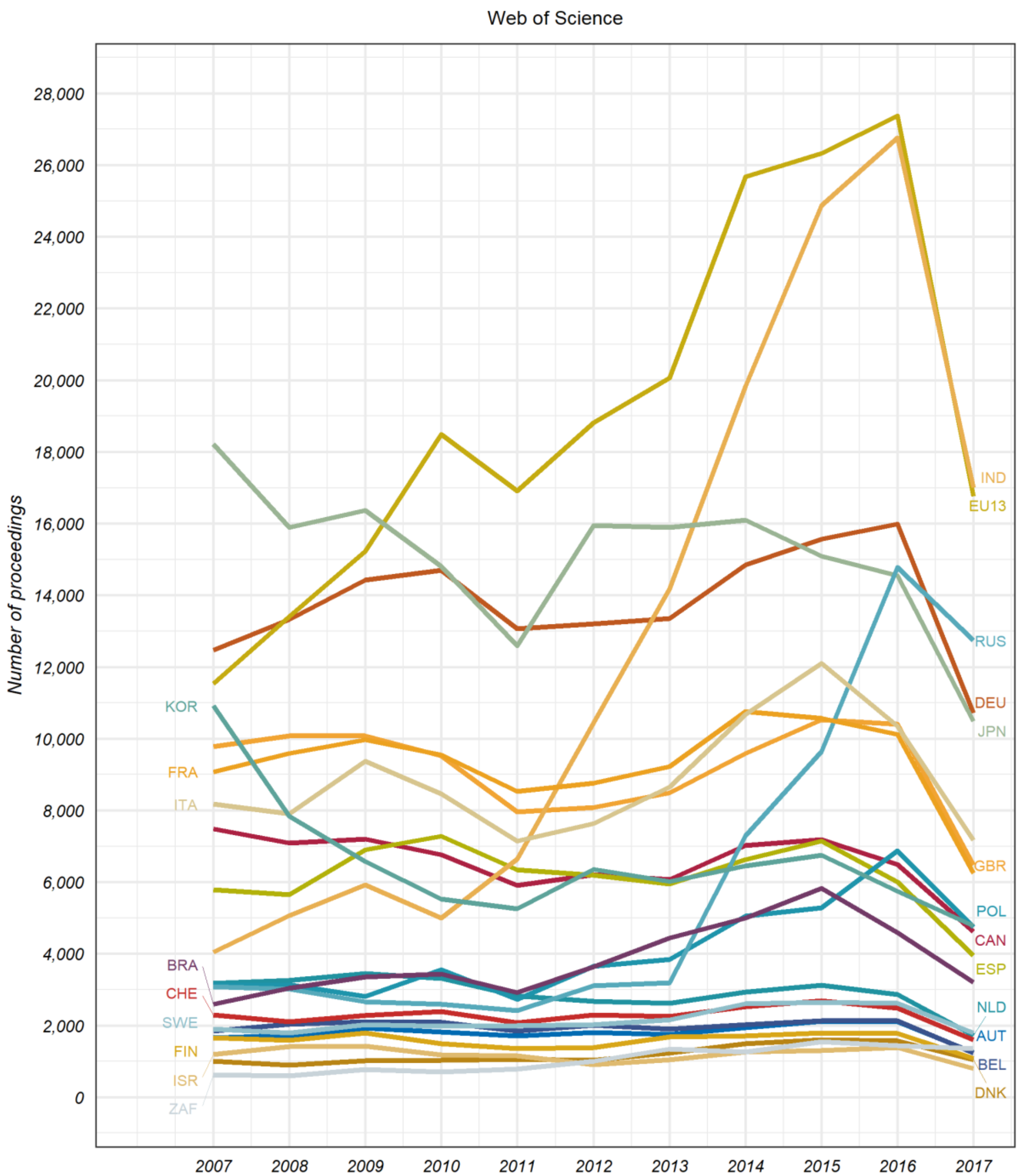


Figure 57 The fractional count of conference proceedings from selected countries and groups between 2007 and 2017 from Web of Science.

Table 11 The percentage of world conference proceedings accounted for by selected countries and groups between 2007 and 2017, based on fractional counting, from Web of Science.

COUNTRY	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUT	0.7	0.7	0.7	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.5
BEL	0.7	0.8	0.7	0.8	0.7	0.7	0.6	0.6	0.6	0.6	0.5
BRA	1.0	1.2	1.2	1.3	1.1	1.2	1.4	1.5	1.6	1.3	1.3
CAN	3.0	2.7	2.5	2.5	2.2	2.0	2.0	2.1	2.0	1.8	1.8
CHE	0.9	0.8	0.8	0.9	0.8	0.8	0.7	0.7	0.8	0.7	0.6
CHN	19.4	23.3	26.1	23.8	31.1	30.6	27.2	23.3	21.7	22.9	21.2
DEU	5.0	5.1	5.1	5.5	4.9	4.3	4.3	4.4	4.4	4.4	4.2
DNK	0.4	0.3	0.4	0.4	0.4	0.3	0.4	0.4	0.5	0.4	0.4
ESP	2.3	2.2	2.4	2.7	2.4	2.0	1.9	2.0	2.0	1.7	1.5
FIN	0.7	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.4
FRA	3.6	3.7	3.5	3.6	3.2	2.9	3.0	3.2	3.0	2.8	2.4
GBR	3.9	3.9	3.5	3.6	3.0	2.7	2.8	2.8	3.0	2.9	2.5
IND	1.6	1.9	2.1	1.9	2.5	3.4	4.6	5.8	7.0	7.4	6.7
ISR	0.5	0.5	0.5	0.4	0.4	0.3	0.3	0.4	0.4	0.4	0.3
ITA	3.3	3.0	3.3	3.2	2.7	2.5	2.8	3.1	3.4	2.9	2.8
JPN	7.3	6.1	5.7	5.6	4.7	5.2	5.2	4.7	4.2	4.0	4.1
KOR	4.4	3.0	2.3	2.1	2.0	2.1	2.0	1.9	1.9	1.6	1.9
NLD	1.3	1.3	1.2	1.2	1.1	0.9	0.9	0.9	0.9	0.8	0.7
POL	1.2	1.2	1.0	1.3	1.0	1.2	1.2	1.5	1.5	1.9	1.9
RUS	1.2	1.2	0.9	1.0	0.9	1.0	1.0	2.1	2.7	4.1	5.0
SWE	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7
USA	18.8	16.8	15.7	16.3	14.1	13.9	12.9	12.7	12.0	12.3	12.5
ZAF	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5
OECD	66.3	61.1	58.4	61.2	53.6	51.4	51.7	52.2	51.8	49.7	47.8
EU13	4.6	5.1	5.4	7.0	6.3	6.2	6.5	7.6	7.4	7.6	6.6
EU15	24.5	24.2	24.1	25.1	21.9	19.7	20.4	21.1	21.5	19.9	18.4
EU28	29.1	29.3	29.4	32.0	28.2	25.8	26.9	28.7	28.9	27.5	25.0
WORLD	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

8 Appendix B: Country code list

Table 12 The country name, code and whether it is part of the EU13, EU15, EU28 or OECD for the countries included in this report.

Country	Code	EU13	EU15	EU28	OECD
Australia	AUS	-	-	-	Yes
Austria	AUT	-	Yes	Yes	Yes
Belgium	BEL	-	Yes	Yes	Yes
Brazil	BRA	-	-	-	-
Bulgaria	BGR	Yes	-	Yes	-
Canada	CAN	-	-	-	Yes
Chile	CHL	-	-	-	Yes
China	CHN	-	-	-	-
Croatia	HRV	Yes	-	Yes	-
Cyprus	CYP	Yes	-	Yes	-
Czech Republic	CZE	Yes	-	Yes	Yes
Denmark	DNK	-	Yes	Yes	Yes
Estonia	EST	Yes	-	Yes	Yes
Finland	FIN	-	Yes	Yes	Yes
France	FRA	-	Yes	Yes	Yes
Germany	DEU	-	Yes	Yes	Yes
Greece	GRC	-	Yes	Yes	Yes
Hungary	HUN	Yes	-	Yes	Yes
Iceland	ISL	-	-	-	Yes
India	IND	-	-	-	-
Ireland	IRL	-	Yes	Yes	Yes
Israel	ISR	-	-	-	Yes
Italy	ITA	-	Yes	Yes	Yes
Japan	JPN	-	-	-	Yes
Latvia	LVA	Yes	-	Yes	Yes
Lithuania	LTU	Yes	-	Yes	Yes
Luxembourg	LUX	-	Yes	Yes	Yes
Malta	MLT	Yes	-	Yes	-
Mexico	MEX	-	-	-	Yes
New Zealand	NZL	-	-	-	Yes
Norway	NOR	-	-	-	Yes
Poland	POL	Yes	-	Yes	Yes
Portugal	PRT	-	Yes	Yes	Yes
Romania	ROU	Yes	-	Yes	-
Russia	RUS	-	-	-	-
Slovak Republic	SVK	Yes	-	Yes	Yes
Slovenia	SVN	Yes	-	Yes	Yes
South Africa	ZAF	-	-	-	-
South Korea	KOR	-	-	-	Yes
Spain	ESP	-	Yes	Yes	Yes
Sweden	SWE	-	Yes	Yes	Yes
Switzerland	CHE	-	-	-	Yes
The Netherlands	NLD	-	Yes	Yes	Yes
Turkey	TUR	-	-	-	Yes
United Kingdom	GBR	-	Yes	Yes	Yes
United States	USA	-	-	-	Yes

9 Appendix C: Methodological details

This appendix discusses key methodological details to be considered in the interpretation of data from this report. This report is predominantly based on document types ‘articles’ and ‘reviews’ from the publication type ‘journal’. Data are extracted from the Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI) from the Web of Science (WoS), and from the entire Scopus database. Fractional counting is used for all data, except when examining co-publications when whole counting is used. Publications are examined for the period 2007 to 2017, and citation data are examined for the period 2007 to 2015. The following are key points to note in relation to the methodology used in compiling this report and the interpretation of its data.

9.1 Whole versus fractional counting

There are two methods for counting publications which have more than one author – whole counting or fractional counting. Consider an example of one paper with two authors, one from Germany and one from France. Whole counting assigns a whole count of the paper to each author so one paper is considered as one contribution from each author and the country with which they are affiliated. In our example, the one paper would count as one paper each for Germany and France, for a total count of two papers. Evidently this method of counting inflates the overall number of publications. A means of dealing with this inflation is using fractional counting in which each author is awarded a proportion of the paper. In this case, the count of publications for Germany and France would be 0.5 each and the total number of publications remains at one.

This report primarily uses fractional counting. In calculating the fractional counts for countries, the fractionalisation is applied at the level of the author. That is, each of the countries affiliated with the authors in our example receive 0.5 of the paper, and the proportions of all papers are summed for each country. If an author has an affiliation with two countries, say our author from Germany is also affiliated with the Netherlands, their proportion is further fractionalised and our resultant counts for the countries would be 0.5 for France and 0.25 each for Germany and the Netherlands. The same method is applied when calculating the fractional counts for the German universities and non-university research institutions, except that the fractional proportions based on authors are summed to the institution rather than the country. A criticism of fractional counting is that countries or institutions which frequently collaborate receive lower counts of publications are proportions are attributed away from them, and this should be considered in interpretting the data here.

The analysis of international collaboration in this report uses whole counting instead of fractional counting. International co-publications are items on which two or more authors from two or more countries collaborated. Counts of publications are calculated using the whole counting method, such that each paper counts as one co-publication per co-authoring country.

9.2 Conference proceedings

The number of conference proceedings for the reference countries and their share of worldwide conference proceedings based on WoS data were presented in the beginning of the appendix. These data are based on the document type ‘proceedings paper’ from the Conference Proceedings Citation Index and Book Citation Index for the Sciences and Social Sciences. There are two points to note regarding conference proceedings.

Firstly, conference proceedings may be classified as either document type ‘proceedings paper’, ‘article’, or both. To avoid double-counting conference proceedings which have been assigned to both document types, in the current report the data presented for conference proceedings refers only to items classified as ‘proceedings papers’. Conference proceedings which were assigned to ‘article’ are included in the bulk of the report which is based on document types ‘article’ and ‘review’.

Secondly, since the previous presentation of data relating to conference proceedings in the “Performance and Structures of the German Science System 2014” report, the recording of affiliation data in WoS has changed. Previously, WoS recorded only one author per proceedings paper, however now all authors of proceedings papers are recorded. As such, where previously each paper would be assigned to a country based only on one author, the assignment of proceedings papers to countries in the current report used the author-level fractionalisation method described above for all authors. As fractional counting is used, this change will not inflate the number of proceedings papers compared to previous data, however differences in shares of worldwide publications may be observed due to the introduction of more author data and fractionalisation at the author level.

9.3 Disciplines classification

Data disaggregated by scientific disciplines are presented using the OECD’s Fields of Science and Technology (FOS) classification. Both Elsevier and Clarivate Analytics have concordances mapping their classification structures - Scopus’ All Science Journal Classification (ASJC) and WoS’ ‘traditional’ Subject Categories classification scheme - to the FOS structure which was used in this report. However, neither database’s classification uses all of the disciplines in the FOS, nor does the FOS account for all categories in the databases’ classifications.

The FOS disciplines ‘nano-technology’, ‘environmental biotechnology’, and ‘industrial biotechnology’ are not used in Scopus. For WoS data, the FOS disciplines ‘health biotechnology’ and ‘other medical sciences’ are not used, and neither database uses the ‘agricultural biotechnology’ discipline. These differences in mapping need to be considered when directly comparing data disaggregated by discipline between WoS and Scopus in that, for example, items which were mapped to ‘nano-technology’ in WoS will have been classified into a different FOS discipline for Scopus data. Further, items in attributed to the ‘multidisciplinary’ category of Scopus’ ASJC have been excluded from presentation here as this category is not mapped to any discipline in the FOS classification. Also, a small proportion of items were not classified to any category in WoS and so were also excluded here as they were not able to be attributed to an FOS discipline.

In both databases, publications can be classified to more than one discipline. As such, the publications counts have also been fractionalised based on the number of classifications they were assigned to within their native structure and then aggregated to the FOS categories. The field-normalisation of indicators occurs on the classifications assigned in WoS and Scopus, and data is then presented by FOS discipline.

9.4 Citation window

While counts of publications can be reliably calculated as early as the following year, a period of time must elapse during which publications are disseminated, read and accumulated citations before counts of the publications’ citations can be calculated. As such, it is typical in bibliometric studies to analyse citations in a window of 3 to 5 years after the publishing year. Wang (2013) determined that 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. For this reason this report uses a 3 year citation window, which also maintains the relevancy of the data better than the longer 5 year window. As such, counts of citations and all indicators using citations are include all citations received within the publication year and the subsequent two years. Consequently, items published at the end of the year have a slight disadvantage in that they have slightly shorter window in which to accrue citations.

9.5 Self-citations

Self-citations can either be included or excluded from citations counts. Self-citations have been retained in the data for this report on this basis that, first, self-citation is a standard means of scientific communication and of building upon one's own previous body of work, and secondly that the patterns of self-citation are likely to be similar with fields so will not present an advantage or disadvantage due to differing citation practices after field normalisation (Bornmann et al. 2014).

9.6 Scientific Regard

The Scientific Regard indicator shows whether a country's publications are cited more or less than average compared to other publications from the same journals. The SR is calculated by comparing the observed rate of citations for a country's papers against the citation rate which would be expected considering the average citation rate of publications in the journals the country published in, and then transforming the scale to range between -100 and 100. Values of 0 indicate the country's publications are cited at the average rate of citations from those journals, while values over 0 indicate a higher citation rate than average, and values less than 0 indicate a lower than average citation rate compared to other publications from the same journals. The SR value for a country is calculated as:

$$SR_k = 100\ tanh\ ln\ (OBS_k/EXP_k)$$

where OBS_k is the observed rate of citations of country k's publications, and EXP_k is the expected citation rate based on the average citation rate of the journals in which country k publishes.

9.7 International Alignment

International Alignment values indicate if a country publishes in journals which are of greater or lesser international visibility as measured by citation rates. Positive IA values for a country indicate the country's papers were published in journals that are cited more frequently than average and so have higher visibility and impact. Conversely, negative IA values indicate the journals in which the country published were cited less frequently than the world average, and values of 0 indicate the citation rate was the world average. The IA value for a country is calculated as:

$$IA_k = 100\ tanh\ ln\ (EXP_k/OBS_w)$$

where OBS_w is the observed citation rate of all publications in the world, and EXP_k is the expected citation rate for country k based on the citation rates of the journals it publishes in.

9.8 Excellence Rate

The Excellence Rate identifies the proportion of a country's publications which are in the 10% most highly cited publications from each discipline and which could be considered of excellent quality on this basis. This report employed the method described by Waltman and Schreiber (2013) to calculate the 10% most frequently cited publications. In this method, the publications with citations above the 90th percentile are identified, however there may be a number of publications with citations on the threshold of the 10th percentile which, if included, would exceed the 10% required. As a secondary step then, the publications on the 90th percentile threshold are proportionally assigned to achieve exactly the top 10%. When interpretting Excellence Rates, 10% would be the expected value, with higher values indicating a higher proportion of publications in the subset of 'excellent' publications and thus better performance.

9.9 Differences from previous years' reports

As this report is the latest in a series, there are some differences in content and computational methods between this report and previous reports which should be noted.

First, the fractional counting method used in this report is applied at the level of the authors and aggregated upward, as described in this appendix, whereas it was applied at the organisational level in previous reports. This produces different counts between the reports, however the author-level counts provide more fine-grained counts of each country's publications. For instance, Table 13 shows an example of a publication's authorship with four authors from 3 institutions in 2 countries. When fractional counting is applied to the authors, each author receives 0.25 of the publication which aggregates to 0.75 for country 1 and 0.25 for country 2. When the fractional counting is applied at the organisational level, each organisation receives 0.33 of the publication which aggregates to 0.66 for country 1 and 0.33 for country 2. In this way, fractional counts of countries' publications will differ between this and previous reports. For further information, Waltman and Eck (2015) provides a useful discussion on counting methods and their impacts on field-normalised indicators.

Table 13 An example of a publication's authorship

Author	Institution	Country
Author 1	Institution 1	Country 1
Author 2	Institution 1	Country 1
Author 3	Institution 2	Country 1
Author 4	Institution 3	Country 2

Finally, the publications examined here were 'articles' and 'reviews', while previous reports included these but also 'letters' and 'notes'. This report also included the Arts and Humanities Citation Index to enable a more appropriate comparison with Scopus, although this index has not been included in previous reports. These differences in methodology and content should be considered when comparing between reports.

10 References

Aman, V. 2016. "Regional Coverage of Authorship in Wos and Scopus: Report Prepared for Thomson Reuters' Tender on 'Web of Science Vs. Scopus Comparison'." Berlin: Institute for Research Information; Quality Assurance.

Bar-Ilan, J. 2017. "Bibliometrics of 'Information Retrieval'–A Tale of Three Databases." In *Proceedings of the 2nd Joint Workshop on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries (Birndl)*, edited by P. Mayr, M. K. Chandrasekaran, and K. Jaidka, 83–90. Tokyo, Japan: CEUR Workshop Proceedings.

Bornmann, L., Bowman B. F., Bauer J., Marx W., Schier H., and M. Palzenberger. 2014. "Bibliometric Standards for Evaluating Research Institutes in the Natural Sciences." In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, edited by Cronin B. and C. R. Sugimoto, 201–23. Cambridge: The MIT Press.

Clarivate Analytics. 2017. "Evaluation Criteria for Web of Science Core Collection Journals." https://clarivate.com/essays/evaluation-criteria-web-science-core-collection-journals/.

———. 2018a. "Web of Science Core Collection: Introduction." http://clarivate.libguides.com/woscc/basics.

———. 2018b. "Web of Science: Summary of Coverage." https://clarivate.libguides.com/webofscienceplatform/coverage.

Donner, P. 2016. "WoS Vs Scopus – Subject Area Coverage: Report Prepared for Thomson Reuters for the Tender 'Web of Science Vs Scopus Comparison'." Berlin: Institute for Research Information; Quality Assurance.

Elsevier. 2017. "Scopus: Content Coverage Guide." https://www.elsevier.com/__data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf.

———. n.d. "Content Policy and Selection." https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection.

Garfield, E. 1979. "Is Citation Analysis a Legitimate Evaluation Tool?" *Scientometrics* 1 (4): 359–75.

Holland, P.W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 946–60.

Huang, Y., D. Zhu, Q. Lv, A. L. Porter, D. K. R. Robinson, and X. Wang. 2017. "Early Insights on the Emerging Sources Citation Index (Esci): An Overlay Map-Based Bibliometric Study." *Scientometrics* 111: 2041–57.

Imbens, G., and J. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Jacso, P. 2005. "As We May Search — Comparison of Major Features of the Web of Science, Scopus, and Google Scholar Citation-Based and Citation-Enhanced Databases." *Current Science* 89 (9). Current Science Association: 1537–47. http://www.jstor.org/stable/24110924.

Leydesdorff, L., and L. Bornmann. 2016. "The Operationalization of 'Fields' as Wos Subject Categories Wcs in Evaluative Bibliometrics: The Cases of 'Library and Information Science' and 'Science &

Technology Studies'." *J. Assoc. Inf. Sci. Technol.* 67 (3). New York, NY, USA: John Wiley & Sons, Inc.: 707–14. doi:10.1002/asi.23408.

Moed, H. F. 2005. "Citation Analysis in Research Evaluation." In, 119–36. Dordrecht, The Netherlands: Springer.

Mongeon, P., and A. Paul-Hus. 2016. "The Journal Coverage of Web of Science and Scopus: A Comparative Analysis." *Scientometrics* 106 (1): 213–28. https://EconPapers.repec.org/RePEc:spr:scient:v:106:y:2016:i:1:d:10.1007_s11192-015-1765-5.

Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.

Shah, S. R. U., K. Mahmood, and A. Hameed. 2017. "Review of Google Scholar, Web of Science, and Scopus Search Results: The Case of Inclusive Education Research." *Library Philosophy and Practice* 5-1-2017. http://digitalcommons.unl.edu/libphilprac/1544.

Somoza-Fernández, M., J-M. Rodríguez-Gairín, and C. Urbano. 2018. "Journal Coverage of the Emerging Sources Citation Index." *Learned Publishing* 31 (3): 199–204.

Stahlschmidt, S., and S. Hinze. 2018. "The Dynamically Changing Publication Universe as a Reference Point in National Impact Evaluation: A Counterfactual Case Study on the Chinese Publication Growth." *Frontiers in Research Metrics and Analytics* 3 (30).

Taşkın, Z., G. Doğan, S. Akça, I. Şencan, and M. & Akbulut. 2015. "Does Scopus Put Its Own Journal Selection Criteria into Practice?" In *15th International Conference of the International Society for Scientometrics and Informetrics*. Istanbul, Turkey.

Testa, J. 2018. "Journal Selection Process." https://clarivate.com/essays/journal-selection-process/.

Torres-Salinas, D., N. Robinson-García, Campanario J. M., and E. Delgado López-Cózar. 2014. "Coverage, Field Specialization and Impact of Scientific Publishers Indexed in the 'Book Citation Index'." *Online Information Review* 38 (1): 24–42.

Waltman, L., and N. J. van Eck. 2015. "Field-Normalized Citation Impact Indicators and the Choice of Appropriate Counting Method." *Journal of Informetrics* 9 (4): 872–94.

Waltman, L., and M. Schreiber. 2013. "On the Calculation of Percentile-Based Bibliometric Indicators." *Journal of the American Society for Information Science and Technology* 64 (2): 372–79. doi:10.1002/asi.22775.

Wang, J. 2013. "Citation Time Window Choice for Research Impact Evaluation." *Scientometrics* 94 (3): 851–72. doi:10.1007/s11192-012-0775-9.

Wang, Q., and L. Waltman. 2016. "Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus." *Journal of Informetrics* 10 (2): 347–64. doi:10.1016/j.joi.2016.02.005.

Zhou, P., and L. Leydesdorff. 2016. "A Comparative Study of the Citation Impact of Chinese Journals with Government Priority Support." *Frontiers in Research Metrics and Analytics* 1 (3). doi:10.3389/frma.2016.00003.