Scientometrics (2016) 106:787–804
DOI 10.1007/s11192-015-1798-9

Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison

Anne-Wil Harzing[1] · **Satu Alakangas**[2]

Received: 7 September 2015 / Published online: 26 November 2015

Abstract This article aims to provide a systematic and comprehensive comparison of the coverage of the three major bibliometric databases: Google Scholar, Scopus and the Web of Science. Based on a sample of 146 senior academics in five broad disciplinary areas, we therefore provide both a longitudinal and a cross-disciplinary comparison of the three databases. Our longitudinal comparison of eight data points between 2013 and 2015 shows a consistent and reasonably stable quarterly growth for both publications and citations across the three databases. This suggests that all three databases provide sufficient stability of coverage to be used for more detailed cross-disciplinary comparisons. Our cross-disciplinary comparison of the three databases includes four key research metrics (publications, citations, *h*-index, and hI, annual, an annualised individual *h*-index) and five major disciplines (Humanities, Social Sciences, Engineering, Sciences and Life Sciences). We show that both the data source and the specific metrics used change the conclusions that can be drawn from cross-disciplinary comparisons.

Keywords Google Scholar · Scopus · Web of Science · *H*-index · hIa · Citation analysis · Research metrics

Introduction

In the last decade, the use of metrics for research evaluation has become an integral part of the academic landscape. The adverse impact of this "audit culture" is well documented (see e.g. Adler and Harzing 2009; Mingers and Willmott 2013). However, since the

✉ Anne-Wil Harzing
anne@harzing.com;
http://www.harzing.com

1 Middlesex University, The Burroughs, Hendon, London NW4 4BT, UK

2 University of Melbourne, Parkville Campus, Parkville, VIC 3010, Australia

reversal of this trend is unlikely, research into fairer and more inclusive ways of measuring research performance is gaining more and more momentum. This article is part of that movement. We investigate how the use of different data sources and the choice of different metrics influences cross-disciplinary comparisons of research performance. We do so by presenting a longitudinal analysis of publication and citation metrics drawn from Google Scholar, Scopus, and the Web of Science for a group of 146 academics across 37 different sub-disciplines, classified into five major disciplines: Humanities, Social Sciences, Engineering, Sciences, and Life Sciences.

We are not to first to compare research metrics across different databases. Prior studies compared Google Scholar and the Web of Science for specific disciplines, such as Business & Management (Amara and Landry 2012; Mingers and Lipitakis 2010), Earth Sciences (Mikki 2010), Computer Science (Franceschet 2010), and Astronomy, Environmental Science, Philosophy and Public Health (Wildgaard 2015). Other studies compared Scopus and the Web of Science for individual universities (Torres-Salinas et al. 2009; Vieira and Gomes 2009) and Scopus and Google Scholar for the Social Sciences in four Spanish universities (Etxebarria and Gomez-Uranga 2010).

Studies that focused on *all three* data-bases were typically fairly limited in scope, for instance looking at one specific academic publication (Bar-Ilan 2010), one individual (Jacso 2008), a single or a small group of journals in Oncology and Condensed Matter Physics (Bakkalbasi et al. 2006), Business & Economics (Levine-Clark and Gil 2008), General Medicine (Kulkarni et al. 2009), Social work (Bergman 2012), Psychology (Roales-Nieto and O'Neill 2012) and South African Environmental Science (Adriaanse and Rensleigh 2013), or a small group of researchers in specific disciplines such as Library & Information Systems (Meho and Yang 2007), Nursing (De Groote and Raszewski 2012), and Soil Research (Minasny et al. 2013). The study by Delgado-López-Cózar and Repiso-Caballero (2013) was the only study comparing a large group of journals (277) across the three databases, but the study was confined to a single discipline (Communication Studies).

Overall, the conclusion of nearly all of these studies is that Google Scholar provides broader coverage for most disciplines and that the Web of Science and Scopus provide fairly similar results. There are no studies, however, that provide large-scale and comprehensive cross-disciplinary comparisons between all three databases. Moreover, only two studies were published in the last 3 years and two-thirds of the studies were conducted at least 5 years ago. Given that coverage for both Google Scholar and Scopus has increased over the last couple of years, it would seem opportune to conduct an up-to-date study.

In addition, only two studies have taken an explicitly longitudinal approach to database comparisons. De Winter et al. (2014) compared the growth of citations to 56 classic research articles between Google Scholar and the Web of Science. On average, Google Scholar citations increased by 87.2 % between 2005 and 2013 and Web of Science citations by 63.9 %, i.e. a growth of approximately 1 % per month for Google Scholar and 0.7 % for the Web of Science. However, Google Scholar also showed a significant retroactive expansion of approximately 2.5 % per month, whereas Web of Science retroactive growth was negligible. Harzing (2014) provided a longitudinal study of the 2011–2013 growth of Google Scholar citation metrics for 20 Nobel Prize winners in Medicine, Chemistry, Physics, and Economics, hence covering a select group of academics in the Life Sciences, Sciences, and Social Sciences. She concluded that—after an earlier period of significant retroactive expansion for Chemistry and Physics—Google Scholar citations were increasing at a fairly stable rate of 1.5 % per month between 2012 and 2013.

Our study thus presents two major contributions. First, we provide a 2-year—July 2013 till July 2015—longitudinal comparison of the rate of growth of publications and citations across the three databases. We find that that all three databases show a consistent quarterly growth for both publications and citations, and can thus be used as alternative data sources. Second, rather than focusing on a single discipline, we provide a cross-disciplinary comparison, including 37 different sub-disciplines, classified into five major disciplines: Humanities, Social Sciences, Engineering, Sciences and Life Sciences. We find that the data source (Google Scholar, Scopus or Web of Science) and the specific metrics used change the conclusions that can be drawn from cross-disciplinary comparisons. We further argue that fair and inclusive cross-disciplinary comparisons *are* possible, provided we use Google Scholar or Scopus as a data source, and the recently introduced hI, annual—a *h*-index corrected for career length and co-authorship patterns—as the metric of choice (Harzing et al. 2014).

Methods

Sample

Our sample consists of 146 Associate Professors and Full Professors at the University of Melbourne, Australia. Constraining our sample to a single university allows us to control for extraneous variability and thus concentrate on the differences between the three databases. Moreover, the University of Melbourne displays excellence in most disciplines and was ranked in the top-40 in the 2014–2015 Times Higher Education ranking for all five disciplinary areas covered in this article. This makes it more likely that any cross-disciplinary differences found in our study are truly caused by differential coverage across databases rather than differences in academic performance. The ability to provide a reliable comparison across disciplines is further enhanced by the fact that this university has very formalised, standardised and centralised procedures for internal promotion. Two-thirds of the academics in our sample had gone through at least two internal promotions, whereas only 18 % had been appointed at their current level from outside the university.

We sampled two Associate Professors and two Professors[1] in all 37 disciplines represented at the University of Melbourne. The university in question is composed of ten distinct Faculties, one of which (Arts & Music) was excluded from our study, as most academics in these disciplines do not produce traditional research outputs. Despite representing a very full range of disciplines, the University of Melbourne has a relatively strong focus on the Life Sciences and Sciences and a traditional strength in Economics &

[1] Two professors in Law and Physics unfortunately had to be removed from the final sample, as their publication patterns were very uncharacteristic of their field. The Law professor (now a Professor at Stanford with a dual appointment in Medicine and Law, and an Honorary Professor in the Medical Faculty at Melbourne) specialised in Health Law and had published mainly in medical journals. His publication pattern (large number of publications and many co-authors) and citation pattern (very high level of citations) was very atypical for the field of Law, which is characteristic by very low citation levels. He was also a Federation Fellow, i.e. a very high performing academic. Including him in the Humanities sample would have completely changed our results; in the Web of Science he had four times as many citations on his own as all of the 19 other Humanities academics combined. The Physics professor was a Particle Physicist. Many of her publications had more than a thousand authors. As a result her publication and citation metrics far exceeded that of any of the other academics in our sample, thus distorting any comparisons. Moreover, as the number of co-authors resulted in very large datafiles, we experienced problems in exporting the data from Scopus for this author as well as difficulties in importing the WoS datafiles into Publish or Perish, thus making data collection impossible in some months. Unfortunately, by the time we realised these problems, it was too late to select a replacement.

Business. Hence these disciplines might have a stronger presence in our sample than for instance Social & Political Sciences, which aggregates Sociology, Anthropology, Geography, and Political Science. We are aware that grouping sub-disciplines into major disciplinary areas is not unambiguous. However, our study included a large variety of sub-disciplines in each major disciplinary area, thus increasing confidence in our results.

The 37 disciplines were subsequently grouped into five major disciplinary fields:

- *Humanities* Architecture; Building & Planning; Culture & Communication; History; Languages & Linguistics, Law (19 observations),
- *Social Sciences* Accounting & Finance; Economics; Education; Management & Marketing; Psychology; Social & Political Sciences (24 observations),
- *Engineering* Chemical & Biomolecular Engineering; Computing & Information Systems; Electrical & Electronic Engineering; Infrastructure Engineering; Mechanical Engineering (20 observations),
- *Sciences* Botany; Chemistry; Earth Sciences; Genetics; Land & Environment; Mathematics; Optometry; Physics; Veterinary Sciences; Zoology (44 observations),
- *Life Sciences* Anatomy and Neurosciece; Audiology; Biochemistry & Molecular Biology; Dentistry; Obstetrics & Gynaecology; Ophthalmology; Microbiology; Pathology; Physiology; Population Health (39 observations).

Individual academics were selected randomly within each sub-discipline, although individuals with very common names were avoided to mitigate problems with author disambiguation. We aimed to select one male and one female academic at both Associate and Full Professor level for each discipline, but had to compromise for some disciplines that had a paucity of female academics at senior levels. Overall, 56.2 % of our sample was male. Table 1 presents the descriptives for our sample. On average academics had been

Table 1 Descriptive statistics: years active, # of papers and citations, *h*-index and hIa index for 146 academics

	N	Minimum	Maximum	Mean	SD
WoS years active	146	3	47	23.84	9.016
Scopus years active	146	5	46	23.69	8.969
GS years active	146	8	46	25.64	8.086
WoS total # of papers	146	3	309	77.25	64.346
Scopus total # of papers	146	3	309	86.37	68.304
GS total # of papers	146	22	519	147.46	97.799
WoS total # of citations	146	0	11,287	1871.68	2238.092
Scopus total # of citations	146	0	11,740	1978.27	2179.222
GS total # of citations	146	58	16,507	3290.88	3122.853
WoS *h*-index	146	0	54	18.91	13.188
Scopus *h*-index	146	0	48	16.92	10.920
GS *h*-index	146	3	65	26.06	13.185
WoS hIa index	146	0.00	1.07	0.3623	0.18991
Scopus hIa index	146	0.00	1.11	0.4075	0.19075
GS hIa index	146	0.05	1.75	0.5757	0.26238
Valid N (listwise)	146				

publishing for 22 years at the Associate Professor level and for 29 years at level of Full Professor. The number of publications and citations varied widely across our sample, as did the *h*-index and hIa index, and included academics that had no citations in the Web of Science or Scopus (all academics had citations in Google Scholar).

Data Sources and procedures

The data sources used in this article are the Web of Science, Scopus, and Google Scholar. Thomson Reuter's Web of Science has long been considered the "gold standard" for citation analysis and until 2004 was the only data source available. Elsevier's Scopus is now a well-established alternative to the Web of Science and is used in many international rankings of universities such as the Times Higher Education ranking. Hence, the use of these two databases does not need further justification. Google Scholar, however, is not without its critics. In particular, Jacsó's many studies (see e.g. Jacsó 2010) have documented serious doubts about the level of accuracy of its citation counts. Hence our choice of Google Scholar as a third database in this article deserves further explanation.

Unfortunately, although there are many studies using Google Scholar as their datasource, so far large-scale investigations of Google Scholar accuracy are rare. However, even 4 years after its introduction, Vaughan and Shaw (2008) found that 92 % of Google Scholar citations in the field of library and information science represented what they called "intellectual impact", and that most citations came from journal articles. In the largest published verification project to date, the London School of Economics project on impact in the Social Sciences (2011), Google Scholar citations were collated for all traceable publications of a sample of 120 academics spread across five social science disciplines. Subsequently, both publications listed and their citing sources were verified and manually cleaned to remove duplicate entries, unacknowledged citations, publishers' publicity materials etc. The correlation between the original scores and the cleaned scores was 0.95. More recently, Harzing's (2013, 2014) studies, using a sample of 20 Nobel Prize winners in Economics, Physics, Chemistry and Medicine, showed that Google Scholar displayed stability over time, presented comprehensive coverage, and provided non-biased comparisons across disciplines.

Google Scholar's native interface is not very suitable for bibliometric analyses. We therefore used Publish or Perish (Harzing 2007) to collect citation data from Google Scholar. Publish or Perish is used primarily in conjunction with Google Scholar, but also offers extensive data import facilities, providing the ability to import amongst others Scopus and Web of Science data. There are currently nearly 1000 published papers referring to the Publish or Perish program, which provides further evidence that—in spite of its limitations—Google Scholar is perceived to be a useful source of bibliometric data.

In a 4-month trial phase before starting the main data collection, search queries for individual authors were refined on an iterative basis through a detailed comparison of the results for the three databases. For Google Scholar this involved creating fine-tuned exclusion strings (e.g. *NOT "A* Tordesillas" faiblesse Socrate Planton* d'aristote Homemmedida hablar "AM Nuevo" Aristote sofisti platon* fishing hake sophistes politiquee lieux monde perelman kairos Vaquero pueblos muerte todos toutes sofista*) and including year limitations to ensure we excluded similarly-named academics. In the Web of Science, we frequently had to use discipline limitations and affiliation to uniquely identify individuals [e.g. *au* = (*tong s**) *and* (*sh* = (*life sciences biomedicine*) *or wc* = (*multi-disciplinary sciences*)] *refined by: web of science categories* = (*endocrinology metabolism or neurosciences or obstetrics gynecology or biochemistry molecular* *biology or pediatrics or cell biology or genetics heredity or reproductive biology or developmental biology*) *and organizations-enhanced = (monash ivf or monash med ctr or monash university or university of melbourne)*. In Scopus, searching was slightly more straightforward as author IDs have been allocated to all authors. However, many authors had multiple author IDs and thus creating search queries was by no means simple [e.g. (*“Rundell, John” 26036632700) OR AU-ID*(*“Rundell, John” 7003579659) OR AU-ID(“Rundell, John” 36885412000) OR AU-ID(“Rundell, John” 37084458800)*]

Searches for Google Scholar were defined in the Publish or Perish multi-query centre, making the running of monthly data collection quick and easy to execute. For Scopus and the Web of Science, a more time-consuming process of repeating individual queries was needed. Hence data were collected on a quarterly basis only. Author Identifiers did change occasionally in Scopus, thus needing detailed verification checks in every data collection round. Searches for Scopus and the Web of Science were conducted in their native interfaces, exported and subsequently imported into Publish or Perish to allow for calculation of the various citation metrics.

Metrics

Definition of metrics

- *Publications* total number of publications per academic
- *Citations* total number of citations per academic
- *h*-index an academic with an index of *h* has published *h* papers each of which has been cited in other papers at least *h* times
- *hIa* hI $_{norm}$/academic age, where:
  - *hI $_{norm}$* normalize the number of citations for each paper by dividing the number of citations by the number of authors for that paper, and then calculate the *h*-index of the *normalized* citation counts
  - *Academic age* number of years elapsed since first publication
- *Growth rate* the growth rate of papers and citations over time was first calculated for each academic individually and then averaged over the 146 academics. This ensured that each academic was given equal weight

hIa: a new metric to allow for cross-disciplinary comparisons

Publish or Perish automatically calculates a wide range of metrics for every academic in our sample. In this paper, we only report the most commonly used bibliometric metrics such as the total number of papers, the total number of citations, and the *h*-index. However, as we want to ensure comparability across disciplines and levels of appointment, we also include a relatively new metric that corrects for differences in career stage and discipline: hI, annual or hIa for short, introduced by Harzing et al. (2014). The hIa provides an indication of the average number of impactful single-author equivalent publications an academic publishes per year. We expect its typical value to be well below 1.0 for most academics (Harzing and Mijnhardt 2015).

To illustrate the substantial differences in co-authorship patterns across disciplines, please note that the Humanities and Social Sciences academics in our sample on average published papers with only 2–2.5 authors. Engineering papers averaged at around 4 authors, whereas papers in the Sciences and Life Sciences averaged around 5 and 6.5 authors respectively. Overall, the average number of authors per paper in our sample ranged from 1 for a Humanities academic (i.e. only single-authored work) to 25 for a Life Sciences academic.

Results

Longitudinal comparisons

We first present a longitudinal comparison of the growth of the three databases in terms of the number of papers and citations. As is shown in Fig. 1, all three databases showed a consistent quarterly growth in terms of the number of papers published, ranging from 1.5 to 3.5 %. This is to be expected as our sample consists of high-performing academics that could be expected to increase their publication output over time.

However, growth in publications could also be due to expansion of the databases in question. Unfortunately, separating the two is impossible without a detailed verification process of the nearly 50,000 publications in our sample. However, it is likely that the stronger growth for Scopus in the second half of data collection period is due to the fact that Scopus has recently made a firm commitment to further expand its coverage of pre-1996 publications and citations (Chrysomallis 2014).

Figure 2 shows that there is also a consistent quarterly growth rate for citations ranging from just below 3–6 %. Overall, the Web of Science shows the lowest variability in its quarterly growth rates, both for papers and citations. There are no strong outliers for Google Scholar or Scopus either. However, for citations Google Scholar seems to experience its most substantial growth in the 1st quarter and its most modest growth in the 3rd quarter, whereas for Scopus the 4th quarter shows the largest growth. Finally, it is

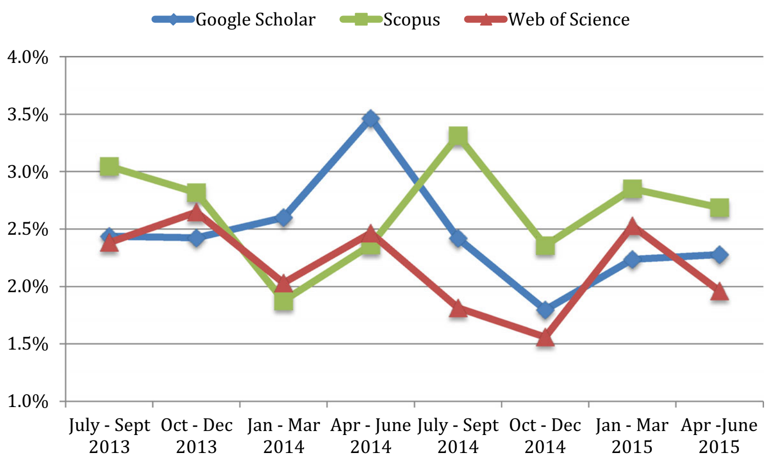


Fig. 1 Quarterly % increase in the number of papers July 2013–July 2015 (average growth of publications per academic)

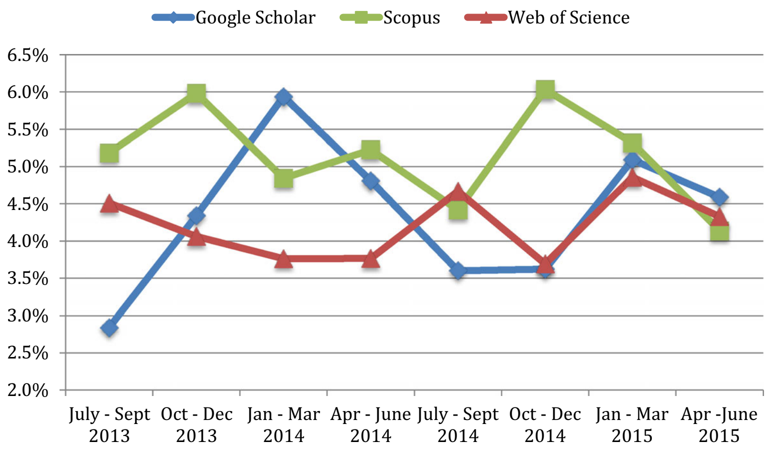


Fig. 2 Quarterly % increase in the number of citations July 2013–July 2015 (average growth of citations per academic)

Table 2 Average quarterly increase in papers, citations and *h*-index per academic across the three databases, July 2013–July 2015

Database	Percentage increase in metric		
	Papers (%)	Citations (%)	*h*-index (%)
Google Scholar	2.5	4.4	2.1
Scopus	2.7	5.1	2.9
Web of Science	2.2	4.2	2.1

The average quarterly increase was calculated by averaging the 8 quarterly data-points, which in themselves represent the average growth per academic for the relevant metric

important to note that none of the databases showed a decline in coverage, thus fulfilling the requirement of stability.

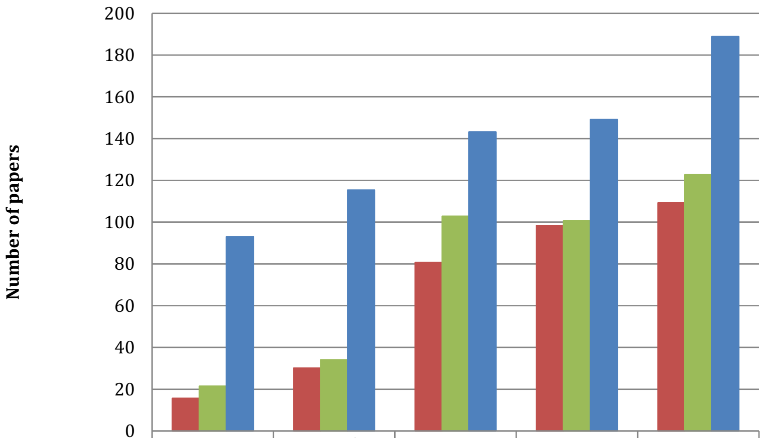
As Table 2 shows, the average quarterly increase in the number of papers per academic is lowest for the Web of Science at 2.2 %, with Google Scholar and Scopus presenting higher growth rates at 2.5 and 2.7 % respectively. With regard to citations and the *h*-index, average growth rates for Google Scholar and the Web of Science are very similar, with Scopus showing a stronger increase for both. The average quarterly increase of 4.4 % for Google Scholar citations is very similar to the 4.6 % increase that Harzing (2014) found for the April 2012 to April 2013 period in her study of 20 Nobel Prize winners. It appears there might be some regularity to Google Scholar growth for research active academics, regardless of whether they are Nobel Prize winners or "mere mortals".

Cross-disciplinary comparisons

We now turn to a comparison of the three databases (Google Scholar, Scopus and the Web of Science) across the five main disciplinary fields in our study: Humanities, Social Sciences, Engineering, Sciences, and Life Sciences. For this comparison, we use the most recent data available, i.e. the data we collected in the first week of July 2015. A comparison of the average number of papers per academic across disciplines (see Fig. 3) shows that Scopus reports a higher number of papers than the Web of Science for all disciplines, but that the difference is largest for Engineering. The number of papers in Google Scholar is substantially higher than both the Web of Science and Scopus for every discipline. However, the differences are particularly large for the Social Sciences and the Humanities, where Google Scholar reports 3–4 times as many papers as the two other databases.

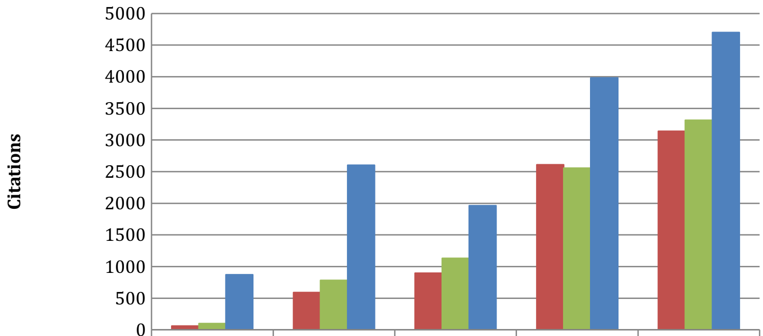
A fair number of the additional papers found by Google Scholar are what are normally called "stray citations", where minor variations in referencing lead to duplicate records for the same paper. This is especially the case in disciplines with publications that do not take the traditional journal article format, such as books, software, and conference papers. Referencing norms are less clear-cut for these types of publications, thus often leading to multiple records for the same publication. Hence, unless individual academics' records are manually cleaned and stray citations merged, we should not attach too much significance to the actual number of papers in Google Scholar as many of these "papers" might be duplicate records with just one or two citations.

A more meaningful comparison is therefore to contrast the average number of citations per academic across disciplines (Fig. 4). In terms of citations, Scopus reports higher levels than the Web of Science for all disciplines except the Sciences, where its citation levels are marginally lower. Its coverage of pre-1996 publications and citations is still lower than the Web of Science, but as indicated above Elsevier has started a large-scale expansion program in this respect. Google Scholar metrics exceed both Scopus and the Web of Science, with 4.5 and 14 times as many citations in Google Scholar than the Web of Science for the Social Sciences and the Humanities. For Engineering, using Google Scholar roughly doubles citations, whereas even for the Sciences and the Life Sciences, the use of Google Scholar on average still increases citations by 50 %.



	Humanities	Social Sciences	Engineering	Sciences	Life Sciences
Web of Science	16	30	81	98	109
Scopus	21	34	103	101	123
Google Scholar	93	115	143	149	189

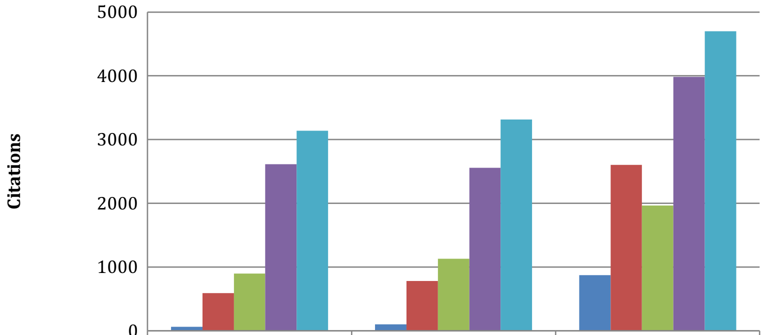
Fig. 3 Average number of papers per academic across five disciplines and three databases, July 2015



	Humanities	Social Sciences	Engineering	Sciences	Life Sciences
Web of Science	61	591	897	2612	3139
Scopus	100	782	1132	2558	3313
Google Scholar	871	2604	1964	3984	4699

Fig. 4 Average number of citations per academic across five disciplines and three databases, July 2015

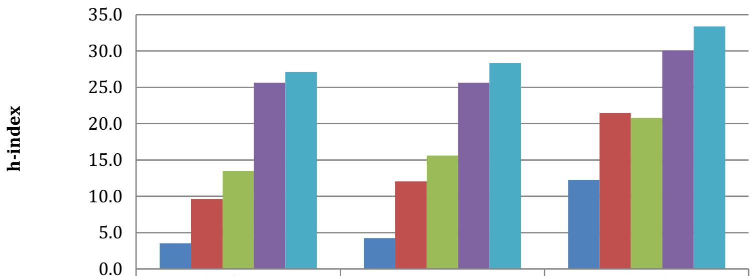
Figure 5 displays citation levels from another perspective and shows how the choice of database affects a disciplinary comparison. In the Web of Science and Scopus disciplinary patterns are very similar, with the Sciences and Life Sciences towering over the three other disciplinary areas. In the Web of Science, the average Life Science academic has more than 50 times as many citations as the average Humanities scholar and 3.5–5 times as many citations as the average scholar in Engineering and the Social Sciences.



	Web of Science	Scopus	Google Scholar
Humanities	61	100	871
Social Sciences	591	782	2604
Engineering	897	1132	1964
Sciences	2612	2558	3984
Life Sciences	3139	3313	4699

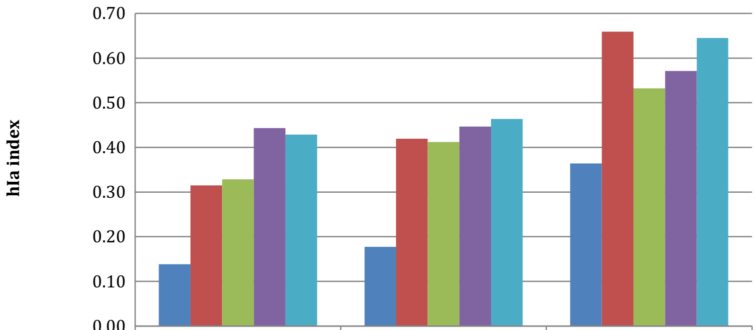
Fig. 5 Average number of citations per academic for five different disciplines in three different databases, July 2015

In Google Scholar, however, disciplinary differences in citation levels are far less pronounced. An average Life Science academic only has 5 times as many citations as an average Humanities academic and 1.8–2.4 times as many citations as an average academic in the Social Sciences or Engineering. A similar pattern is found if we look at the *h*-index (Fig. 6) instead of the total number of citations. In the Web of Science, the *h*-index of the average Life Sciences academic is nearly 8 times as high as for the average Humanities academic and nearly 3 times as high as for the average Social Scientist. In Google Scholar these differences are reduced to 2.7 times as high for Humanities and only 1.5 times as high for the Social Sciences.



	Web of Science	Scopus	Google Scholar
Humanities	3.5	4.3	12.3
Social Sciences	9.6	12.0	21.5
Engineering	13.5	15.6	20.8
Sciences	25.6	25.6	30.1
Life Sciences	27.1	28.3	33.4

Fig. 6 Average *h*-index per academic for five different disciplines in three different databases, July 2015



	Web of Science	Scopus	Google Scholar
Humanities	0.14	0.18	0.36
Social Sciences	0.32	0.42	0.66
Engineering	0.33	0.41	0.53
Sciences	0.44	0.45	0.57
Life Sciences	0.43	0.46	0.65

Fig. 7 Average hIa per academic for five different disciplines in three different databases, July 2015

An even more striking picture appears when, instead of the regular *h*-index, we compare disciplines using the annualised individual *h*-index (hIa for short), which corrects the *h*-index for differences in career length and co-authorship patterns (Fig. 7). Using this metric dramatically reduces the differences between disciplines for any database. Even in the Web of Science the difference between Science and Life Science academics on the one hand and Social Science and Engineering academics on the other hand is now relatively small, the latter only showing 30 % higher metrics. Only the Humanities show significantly lower scores. In Scopus, four of the five disciplines now have very similar scores, whereas in Google Scholar the average for the Social Sciences is marginally higher than the Life Sciences average, and substantially higher than the average for Engineering and the Sciences. Using the hIa and Google Scholar, even Humanities scholars have scores that make the application of metrics for this discipline appear feasible.

To understand the profound difference that different metrics and data sources can make for cross-disciplinary comparisons, Table 3 compares the *h*-index using the Web of Science as a data source with the hIa using the Scopus and Google Scholar as a data source. When looking at the latter two, comparisons across disciplines are much less fraught than when looking at a traditional metric, such as the *h*-index, and the most commonly used database, Thomson Reuters' Web of Science. Hence provided we use a metric corrected for co-authorships (the hIa) and a database with more comprehensive coverage across disciplines (i.e. Scopus or Google Scholar), four of the five main disciplines in our study in fact show very similar research metrics.

Detailed comparisons at the individual level

Thus far, our discussion has focused on the aggregate level. On average, we found that Scopus and Google Scholar provided more comprehensive coverage and a fairer comparison between disciplines. However, aggregate comparisons might conceal important differences at the individual level. If we were to use Scopus or Google Scholar instead of the Web of Science, what would the impact be on individual academics?

As shown in Table 4, nearly all academics in our sample had higher metrics in Google Scholar than in the Web of Science. Two of the three cases where the number of papers in the Web of Science exceeded the number of publications in Google Scholar concerned academics where in the Web of Science all chapters of their authored books were (erroneously) added as individual publications. The third case was an academic for whom we

Table 3 Comparing the average Web of Science *h*-index per academic with the average Scopus and Google Scholar hIa per academic across five disciplines, July 2015

Discipline	Web of Science *h*-index	Life Sciences = 100	Scopus hIa	Life Sciences = 100	Google Scholar hIa	Life Sciences = 100
Humanities	3.5	13	0.18	38	0.36	56
Social sciences	9.6	36	0.42	91	0.66	102
Engineering	13.5	50	0.41	89	0.53	82
Sciences	25.6	95	0.45	96	0.57	89
Life sciences	27.1	100	0.46	100	0.65	100

Table 4 Individual comparisons of Google Scholar and Scopus coverage with Web of Science coverage for publications, citations, *h*-index and hIa, July 2015

	Number of academics (out of 146) for whom the metric in question is higher or lower than the corresponding metric in the WoS					Affected academics
	Higher than WoS	<5 % Lower	5–10 % Lower	10–25 % Lower	>25 % Lower	
GS publications	143	2	0	0	1	None; differences are caused by Web of Science errors + one mega-authored paper
GS citations	145	0	0	1	0	
GS *h*-index	145	1	0	0	0	
GS hIa	146	0	0	0	0	
Scopus publications	133	3	5	4	1	Older academics
Scopus citations	110	6	7	15	8	Social Sciences 13 %*
Scopus *h*-index	115	9	8	11	3	Humanities 21 % Life Sciences 28 %
Scopus hIa	113	3	10	17	3	Sciences 43 %

* 13 % of the academics in our sample working in the Social Sciences had a lower number of citations in Scopus than in the Web of Science

had missed two publications in Google Scholar through searching with her full given name, a strategy that was needed in her case for author disambiguation purposes. These missing publications didn't have a high level of citations and her citation count in Google Scholar was substantially higher than in the Web of Science.

There is only one author for which we found fewer citations in Google Scholar than in the Web of Science. This author's most cited publication was in fact missing in both Google Scholar and Scopus. However, the likely reason for this omission was that the paper in question had well over 1000 authors and that our academic was not one of the key authors. Finally, the only case in which the Google Scholar *h*-index was lower than the Web of Science *h*-index was one where the Google Scholar *h*-index was 37 and the Web of Science *h*-index 38. Overall, we therefore argue that, amongst the 146 academics in our sample, there are no cases where the Google Scholar metrics are substantively lower than the Web of Science metrics.

In Scopus, there are a larger number of individual academics that show lower research metrics than in the Web of Science. However, of the 13 academics with fewer publications in Scopus than the Web of Science, 10 miss only 3–7 papers in Scopus; the three remaining academics miss 10, 12 and 24 publications respectively. In two of the 13 cases, the Web of Science inappropriately over-reported publications as it listed all book chapters in an authored book as individual publications. Taking these Web of Science errors into account, Scopus reports a higher number of publications than the Web of Science for more than 92 % of the academics in our sample. Where publications are missing, by and large these are publications that were published before 1996.

Scopus' performance in terms of citations is a little less impressive. Even though three quarters of our sample of academics show a higher citation count in Scopus than in the Web of Science, there are 36 academics that have a lower number of citations in Scopus. Two-thirds of these academics have published a substantial number of papers in the 1970s and/or 1980s. Even though in many cases these publications were included in Scopus, their citation levels in Scopus were typically lower than in the Web of Science, most likely because not all of the pre-1996 publications that cite these papers have so far been added in Scopus. That said, the citation difference between Scopus and the Web of Science in 2015 (a year after the start of the Scopus expansion initiative) was already much smaller than in 2013; the difference had been reduced by more than 40 %. As a result the average number of citations of citations per academic in 2015 is roughly 100 higher for Scopus than for Web of Science, whereas in 2013 it was roughly 50 lower.

The *h*-index and hIa metrics for Scopus show a similar pattern, with well over three quarters of our academics having a higher *h*-index and hIa with Scopus data than with the Web of Science data. Of the 31 academics with a lower *h*-index, only 14 showed a difference that was larger than 1. These were all older academics working in the Sciences or Life Sciences; on average they had been active for 41 years (against the sample average of 26). With an average *h*-index of 31 (against a sample average of 20), they also had a relatively high Scopus *h*-index, in spite of the missing publications and citations.

Overall, Scopus still has some catching up to do for pre-1996 publications and citations. However, Scopus already reports higher metrics than the Web of Science for more than three quarters of our sample. Furthermore, the Scopus expansion program is still in progress, and with the passing of time the proportion of academics with a substantial number of pre-1996 publications and citations will decline naturally.

Discussion

Based on a sample of 146 senior academics, we provided a longitudinal and cross-disciplinary comparison of three major bibliometric databases: Google Scholar, Scopus and the Web of Science. First, we presented a longitudinal comparison of the rate of growth of publications and citations across the three databases and showed a consistent quarterly growth across all three databases. Second, we provided a cross-disciplinary comparison for four key research metrics: publications, citations, *h*-index, and hI, annual, an annualised individual *h*-index (see Harzing et al. 2014).

Our sample included 37 different sub-disciplines, classified into five major disciplines: Humanities, Social Sciences, Engineering, Sciences and Life Sciences. We found that the data source and the specific metrics used change the conclusions that can be drawn from cross-disciplinary comparisons. More specifically, we found that when using the *h*-index as a metric and the Web of Science as a data source, the average academic in the Life Science and Sciences had an *h*-index that was nearly eight times as high as their counterpart in the Humanities, and two to three times as high as their counterparts in Engineering and the Social Sciences respectively. However, when using the hI, annual and Google Scholar or Scopus as a data source, the average academic in the Life Sciences, Sciences, Engineering and the Social Sciences shows a very similar research performance; whereas the average Humanities academic has a hI, annual that is half to two-thirds as high as the other disciplines.

Drilling down to the level of individual academics to compare research metrics across the three databases, we found that Google Scholar provides a broader coverage and thus higher research metrics than the Web of Science for all academics in our sample. For Scopus the same was true for more than 90 % of the academics in terms of publications and for more than three quarters of the academics in terms of citations. Most of the missing publications and citations concerned pre-1996 material. As Scopus' pre-1996 expansion program is still ongoing, we expect its coverage to match that of the Web of Science for most, if not all, academics in the near future.

Comparing our longitudinal findings with the two earlier longitudinal studies (De Winter et al. 2014; Harzing 2014) we find that the number of publications and citations grow at a fairly stable rate in all three databases, ranging from 0.75 to 0.88 % per month for publications and 1.42 to 1.71 % per month for citations. Although we are unable to separate "natural" growth from retroactive growth through database expansion, the fairly low rate of growth in our current study seems to suggest that—in comparison to De Winter et al.'s (2014) study—retroactive growth is modest in our sample. De Winter's study compared 2013 data for WoS and Google Scholar with 2005 data, the latter being only a year after Google Scholar's introduction. Likewise, Harzing (2013) documented that Google Scholar's retroactive growth between 2011 and 2012 was still fairly high for Chemistry and Physics Nobelists, mainly due to improvement in previously weak coverage for some important publishers in this field. However, in a follow-up study (Harzing 2014), the growth rate for Google Scholar between 2012 and 2013 was very similar to the growth rate found in our current study for 2013–2015. Hence, this suggests that Google Scholar's retroactive expansion has now stabilised. As documented in our current paper, we do still see some retroactive expansion for Scopus through its pre-1996 expansion program (Chrysomallis 2014). As a result, both Google Scholar and Scopus have in our view become credible alternatives to the Web of Science.

Our results with regard to coverage in the three databases confirm previous studies in that Google Scholar provides the most comprehensive coverage and that coverage for the Web of Science and Scopus is similar. However, our study improved on these studies in a number of ways. First, we included a relatively large sample of academics. Second, our sample covered all major disciplines and was thus able to provide a detailed comparison of coverage of the three databases across disciplines. Third, in addition to our cross-disciplinary comparison, we also compared coverage for every individual in our study and were thus able to combine broad disciplinary comparisons with a fine-grained individual analysis. Finally, our comparison across the databases did not just include publications and citations, but also the *h*-index and the newly introduced hIa index. As a result, our paper presents a much more systematic and comprehensive comparison of the three major sources of citation data than previous studies.

Limitations and suggestions for further research

Although our study is the first to present a comprehensive longitudinal and cross-disciplinary comparison across three major sources of citation data, it is not without limitations. The first limitation is related to our sample. Our focus on a single university allowed us to control for extraneous variation. Its universal excellence across disciplines and formalised, standardised and centralised promotion procedures also enabled us to concentrate on inherent cross-disciplinary differences across the three databases and different research metrics. However, this came at the price of external generalisability. Hence, it would be useful to complement our study with a random sample of academics in a range of different universities.

A second limitation is related to our focus on senior academics only. Again, we made this choice purposefully as it allowed us to counteract differences in career structures across disciplines. In the (Life) Sciences academics often complete one or more postdocs before being appointed as Lecturer or Senior Lecturer. This is less common in the Humanities and Social Sciences, and in "shortage" sub-disciplines such as Accounting & Finance academics are often appointed as Senior Lecturer even before finishing their PhD. Hence comparing Lecturers or Senior Lecturers across disciplines might mean comparing apples and oranges. Moreover, many junior academics only have a limited number of publications and citations and hence idiosyncratic results are likely. At senior level academics are more comparable and have also gone through the homogenising effect of internal promotions. However, again this means we cannot generalise our finding to junior academics and thus a complementary study for this group might bring additional insights.

A final limitation is linked to one of our two preferred databases: Google Scholar. Unlike the Web of Science and Scopus, Google Scholar doesn't have a strong quality control process and simply crawls any information that is available on academic related websites. Although most of Google Scholar's results come from publisher websites, its coverage does include low quality "publications" such as blogs or magazine articles. However, previous studies have shown that the majority of citing references presents intellectual content. A second drawback of using Google Scholar is the large number of duplicate papers found by Google Scholar. These are normally called "stray citations", where minor variations in referencing lead to duplicate records for the same paper. It is possible, though very time-consuming, to manually clean every academic's record by merging stray citations and removing non-academic publications. Doing so for the first author, a fairly extreme case as there are many stray citations to for instance the Publish or Perish software, reduces her number of papers in Google Scholar from 244 to 106. However, this is still about twice as much as her number of publications in the WoS (47) and Scopus (58) as Google Scholar includes books, book chapters, software, and publications in journals not included the WoS and Scopus. It should also be noted that the problem of "stray citations" is not unique to Google Scholar. If one conducted a "cited reference" search[2] for the WoS one would find many stray citations, especially for academics in the Social Sciences and Humanities. Without submitting monthly data change requests to Thomson Reuters, the first author's "number of publications" in the WoS cited reference search would far exceed that of Google Scholar. Likewise, Scopus lists 340 secondary documents for the first author, 77 to the Journal Quality List—a web publication aggregating various journal ranking lists—alone. Google Scholar thus seems to have better aggregating mechanisms than both the Web of Science and Scopus. However, all three databases are continuously evolving and hence we would recommend periodic monitoring of changes in both coverage and level of accuracy.

Conclusion

Our comparative study of publications, citations, *h*-index and hIa across 146 academics from five major disciplines was the first to present a comprehensive longitudinal and cross-disciplinary comparison across three major sources of citation data: Web of Science, Scopus and Google Scholar. Our longitudinal analysis showed a consistent and reasonably stable quarterly growth for both publications and citations across the three databases. This suggests that all three databases provide sufficient stability of coverage to be used for more detailed cross-disciplinary comparisons.

Our cross-disciplinary comparison of four key research metrics (publications, citations, *h*-index, and hI, annual—an annualised individual *h*-index) across five major disciplines (Humanities, Social Sciences, Engineering, Sciences and Life Sciences) showed that both

[2] This obviously begs the question why we didn't use the WoS Cited Reference search or the Scopus secondary documents in our comparison. For a comprehensive review of the many reasons why this is not a feasible option, please see Harzing (2013, pp. 1064–1065).

the data source and the specific metrics used change the conclusions that can be drawn from cross-disciplinary comparisons. We thus argue that a fair and inclusive cross-disciplinary comparison of research performance *is* possible, provided we use Google Scholar or Scopus as a data source, and the recently introduced hI, annual—a *h*-index corrected for career length and co-authorship patterns—as the metric of choice.

References

Adler, N., & Harzing, A. W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. *The Academy of Management Learning & Education, 8*(1), 72–95.

Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. *The Electronic Library, 31*(6), 727–744.

Amara, N., & Landry, R. (2012). Counting citations in the field of business and management: Why use Google Scholar rather than the Web of Science. *Scientometrics, 93*(3), 553–581.

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical digital libraries, 3*(1), 7.

Bar-Ilan, J. (2010). Citations to the "Introduction to informetrics" indexed by WOS. *Scopus and Google Scholar. Scientometrics, 82*(3), 495–506.

Bergman, E. M. L. (2012). Finding citations to social work literature: The relative benefits of using Web of Science, Scopus, or Google Scholar. *The Journal of Academic Librarianship, 38*(6), 370–379.

Chrysomallis, M. (2014, December 8). Scopus continues to add pre-1996 citations [Web log post]. http://blog.scopus.com/posts/scopus-continues-to-add-pre-1996-citations

De Groote, S. L., & Raszewski, R. (2012). Coverage of Google Scholar, Scopus, and Web of Science: A case study of the *h*-index in nursing. *Nursing Outlook, 60*(6), 391–400.

De Winter, J. C., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: A longitudinal study. *Scientometrics, 98*(2), 1547–1565.

Delgado-López-Cózar, E., & Repiso-Caballero, R. (2013). El impacto de las revistas de comunicación: Comparando Google Scholar Metrics, Web of Science y Scopus. *Comunicar: Revista Científica de Comunicación y Educación, 21*(41), 45–52.

Etxebarria, G., & Gomez-Uranga, M. (2010). Use of Scopus and Google Scholar to measure social sciences production in four major Spanish universities. *Scientometrics, 82*(2), 333–349.

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics, 83*(1), 243–258.

Harzing, A. W. (2007) *Publish or Perish*. http://www.harzing.com/pop.htm

Harzing, A. W. (2013). A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel Prize winners. *Scientometrics, 93*(3), 1057–1075.

Harzing, A. W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics, 98*(1), 565–575.

Harzing, A. W., Alakangas, S., & Adams, D. (2014). hIa: An individual annual *h*-index to accommodate disciplinary and career length differences. *Scientometrics, 99*(3), 811–821.

Harzing, A. W., & Mijnhardt, W. (2015). Proof over promise: Towards a more inclusive ranking of Dutch academics in Economics & Business. *Scientometrics, 102*(1), 727–749.

Jacso, P. (2008). Testing the calculation of a realistic *h*-index in Google Scholar, Scopus, and Web of Science for FW Lancaster. *Library Trends, 56*(4), 784–815.

Jacsó, P. (2010). Metadata mega mess in Google Scholar. *Online Information Review, 34*(1), 175–191.

Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA, 302*(10), 1092–1096.

Levine-Clark, M., & Gil, E. L. (2008). A comparative citation analysis of Web of Science, Scopus, and Google Scholar. *Journal of Business & Finance Librarianship, 14*(1), 32–46.

London School of Economics and Political Science. (2011). *Maximizing the impacts of your research: A handbook for social scientists*. http://www2.lse.ac.uk/government/research/resgroups/LSEPublicPolicy/Docs/LSE_Impact_Handbook_April_2011.pdf

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105–2125.

Mikki, S. (2010). Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics, 82*(2), 321–331.

Minasny, B., Hartemink, A. E., McBratney, A., & Jang, H. J. (2013). Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. *PeerJ, 1*, e183.

Mingers, J., & Lipitakis, E. (2010). Counting the citations: A comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics, 85*(2), 613–625.

Mingers, J., & Willmott, H. (2013). Taylorizing business school research: On the 'one best way'performative effects of journal ranking lists. *Human Relations, 66*(8), 1051–1073.

Roales-Nieto, J. G., & O'Neill, B. (2012). A comparative study of journals quality based on web of science, scopus and google scholar: A case study with IJP&PT. *International Journal of Psychology and Psychological Therapy, 12*(3), 453–480.

Torres-Salinas, D., Lopez-Cózar, E. D., & Jiménez-Contreras, E. (2009). Ranking of departments and researchers within a university using two different databases: Web of Science versus Scopus. *Scientometrics, 80*(3), 761–774.

Vaughan, L., & Shaw, D. (2008). A new look at evidence of scholarly citations in citation indexes and from web sources. *Scientometrics, 74*(2), 317–330.

Vieira, E., & Gomes, J. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics, 81*(2), 587–600.

Wildgaard, L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. *Scientometrics, 104*(3), 1–34.