UNIVERSITY OF COPENHAGEN

Supplement to ACUMEN deliverable 5.4a: Description and comparison of indicators in Google Scholar and Web of Science

Wildgaard, Lorna Elizabeth; Larsen, Birger; Schneider, Jesper

Publication date:
2014

Document version
Early version, also known as pre-print

ACUMEN

academic careers understood through measurement and norms

FP7 Grant Agreement	266632
Dissemination level	Public (PU)
Work Package	WP5-Bibliometric Indicators
Version	1.0
Release Date	24. April 2014
Author(s)	Lorna Wildgaard Birger Larsen Jesper W Schneider
Project Website	http://research-acumen.eu/

European Commission

7th Framework Programme

SP4 - Capacities

Science in Society 2010

Grant Agreement: 266632

EUROPEAN COMMISSION

SEVENTH FRAMEWORK PROGRAMME

Supplement to ACUMEN Deliverable 5.4a:
Description and comparision of indicators in Google Scholar and Web of Science

Lorna Wildgaard[a] Jesper W Schneider[b] Birger Larsen[c]

[a] Royal School of Library and Information Science, Birketinget 6, 2300 Copenhagen, Denmark

[b] Institut for Statskundskab - Dansk Center for Forskningsanalyse, Bartholins Allé 7, 8000 Aarhus C, Denmark

[c] Department of Communication and Psychology, Aalborg University Copenhagen, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark

Contents

Introduction

We collected publication and citation data in two databases to investigate the extent performance of author-level indicators are effected by choice of database, the stability of indicators across databases and ultimately to illustrate how differences in the computed indicators change our perception of individual researchers. In this report we begin by comparing database coverage, coverage at seniority and gender-level and then the performance of four basic indicators computed in both databases. In the main deliverable 5.4a, we investigate in a cluster analysis the performance of our previously identified 108 indicators of author-level impact. Understanding the effect of the database used to source the data and the demographics of the researchers in our sample, will enable us to put the results of our cluster analysis in perspective and direct future studies.

Coverage

Out of the ACUMEN shared data set of 2154 researchers, 750 were identified as unique scholars having a working link to their curriculum vitae including/and a publication list. Publication and citation data was retrieved from Web of Science (Wos) and from Google Scholar (GS). A direct comparison between the two databases showed that WoS has about the same coverage for researchers as Google Scholar, Table 1.

Table 1. Overall coverage of Scholars in WoS and GS

Researchers with CV and publication list	Researchers covered in Web of Science	Researchers covered in Google Scholar
750	741	748
Difference to CV	9	2
Coverage	98%	99%

The researchers listed in total 62046 publications on their CVs and publication lists. Overall GS retrieved 41613 unique records more than WoS. Wos covered 50% of the records reported on CVs and publication lists, while GS covered 116%, Table 2. In both databases records that could be claimed by the searched researcher but not written on the CV or publication list were included. This is because CVs and publication lists sometimes only report selected papers or are not completely up-to-date.

Table 2. Overall coverage of publications in WoS and GS

Number of publications on CV	Number of records in WOS	Number of records in Google Scholar
62046	30967	72580
Difference to CV	31079	+10534
coverage	50%	116%

Researcher coverage differs only slightly from discipline to discipline in the two databases, Table 3. However the depth of coverage in the databases differs greatly between WoS and GS, which is of great importance for individual assessment. Further disciplinary coverage within WoS varies as well, Table 4. In Wos Astronomy has a 58% coverage, while GS found more papers resulting in 132%

coverage. Environmental Science has 46% coverage in WoS and 104% in GS, Philosophy 23% in WoS and 97% in GS and Public Health 80% in WoS and 136% in GS.

Table 3. Coverage of researchers in WoS and GS

Discipline	Researchers with CV & Publication list	Number in Wos	Difference	Coverage	Number in Google Scholar	Difference	Coverage
Astronomy	203	192	11	94%	193	10	95%
Environmental Science	203	195	8	96%	195	8	96%
Philosophy	250	222	28	88%	229	21	91%
Public Health	137	132	5	96%	132	5	96%

Table 4. Disciplinary coverage in Wos and GS

Discipline	Number of publications on CV	Number in WoS	Difference CV	Coverage	Number in Google Scholar	Difference CV	Coverage
Astronomy	21169	12359	8810	58%	28127	+6958	132%
Environmental Science	16720	7820	8900	46%	17453	+733	104%
Philosophy	15090	3494	11596	23%	14708	382	97%
Public Health	9067	7294	1773	80%	12387	+3320	136%

Effect of database on author-level indicators

Raw citation count alone is not an indicator of impact; citation counts need to be benchmarked or normalized to similar research. Citation patterns differ greatly between sub-disciplines and the types of publications a researcher publishes. Also citations accumulate over time, so the year of publication must be taken into account. Four common indicators computed in Web of Science and Google Scholar were compared, Table 5.

Table 5. Average difference between indicators computed in Google Scholar and Web of Science

Discipline	Difference in mean academic age GS : WoS	Difference in mean CPP GS:WoS	Difference in mean H-index GS:WoS	Difference in mean m-quotient GS:WoS	Difference in mean g-index GS:WoS
Astronomy	+3 years	-4.5 CPP	+3.6h	0	+8.7g
Environmental Science	+4 years	-0.3 CPP	+2.7h	+0.7	+5.3g
Philosophy	+6 years	+2.9 CPP	+4.6h	+0.17	+9.3g
Public Health	+3 years	+1.4 CPP	+3.5h	+0.1	+7.8g

Across all disciplines the academic age of researchers are on average 4 years older in Google Scholar than Web of Science. Academic age is the number of years since the first publication for the researcher recorded in the database. This information is used to adjust many indicators to the length of a researcher's career to enable comparability. The average number of citations per paper is however only 0.7 citations between the two databases and the m-quotient is similar as well, with only a difference of 0.2; the h-index is on average 3.7 h higher in Google Scholar than Web of Science and likewise the g-index is also higher by 8.1. However, the performance of indicators of individual impact should not be compared across disciplines. Within disciplinary analysis reveals larger differences that favour Google Scholar as it produces the higher numbers, however data collection proved more reliable in Web of Science and as such we assume the reliability of the indicators to represent the actual publications and reception of the individual scholar is more accurate in WoS, Table 5. Interestingly the m-quotient is very similar on average per researcher in both databases. The m-quotient makes the h-index comparable, as it divides h by the number of years since the researcher's first publication recorded in the database thus enabling the comparison of researchers with different length of career.

Age and seniority

Early career researchers are defined as PhD and Post Docs, middle career are Assistant professors and senior researchers are associate professors. In this report we call professors "established researchers". As expected early career researchers are not as highly cited as researchers who have had a longer career. This is not an indication of quality, but simply that during their short career the work of these early career researchers has not had enough time to accumulate citations. Comparing their citations to field norm is uninformative. However, comparing their citations per paper to the expected number of citations of the articles in journals they publish in (CWTS indicator *average mjs mcs*) can be an indication of impact. In the WoS data set 396 researchers performed under the *average mjs mcs* (Sample A) and 345 researchers performed better than *average mjs mcs*, (Sample B). Normally field benchmarks are computed using the average number of citations per paper for a WoS subject category which may or may not represent the sub-specialty of the researcher. However, as *average mjs mcs* is calculated with a two year citation window, the junior researcher needs to have been published for two years to allow fair comparison, Table 6. This indicator is only comparable as an expected performance benchmark to the number of citations received to articles and reviews retrieved from WoS. The Table shows that publications written by senior and established staff are only performing marginally better than junior or middle career researchers. Seniority is not a classification of academic age, a Post Doc can for example have 6 or 15 yearlong publishing history. Apart from age, gender and nationality can have an effect on researchers' career paths and research output.

Table 6: Summary of actual citations to expected seniority performance (WoS)

Seniority	Average mjs mcs	Number of researchers	Number of researchers performing better	% achieving ≥expected
Astronomy				
PHD	7,583046907	15	9	60%
Post Doc	12,4729792	48	21	43%
Assis Prof	12,54805936	26	11	42%
Assoc Prof	16,36060726	66	29	43%
Full Professor	18,64497503	37	17	45%
Environmental Science				
PHD	11,54813557	3	0	0
Post Doc	4,932046506	17	8	47%
Assis Prof	8,275902941	39	14	35%
Assoc Prof	10,08383101	85	37	43%
Full Professor	12,4342212	51	25	49%
Philosophy				
PHD	1,237678971	8	2	25%
Post Doc	2,110023794	22	6	27%
Assis Prof	4,261891167	44	8	18%
Assoc Prof	3,826703308	73	18	24%
Full Professor	5,019210551	75	22	29%
Public Health				
PHD	6,30695831	9	4	44%
Post Doc	8,843720756	14	6	42%
Assis Prof	9,154821404	30	14	46%
Assoc Prof	12,69529504	50	26	52%
Full Professor	14,6056222	29	15	51%

Table 7: Overall performance of researchers compared to disciplinary benchmark (WoS)

Discipline	Number of researchers	Number in WoS	% researchers performing better than expected citation score
Astronomy	192	12359	45%
Environmental Science	195	7820	43%
Philosophy	222	3494	25%
Public Health	132	7294	49%

Gender

In the WoS data set there are 580 male researchers and 161 female researchers. Overall 44% of the female researchers perform better than expected, while 47% of the male researchers perform better than expected. Performance on a disciplinary level is shown in Table 8.

Table 8. Gender performance better than expected on a disciplinary level (WoS)

	Number of researchers	Number of publications	% of researchers performing better than expected	Citations per paper Sample A	Citations per paper Sample B
Astronomy					
Male	162	11163	59%	14.1	29.8
Female	30	1196	80%	15.7	29.5
Environmental Science					
Male	160	6874	46%	11.1	16.6
Female	35	946	60%	7.5	20.8
Philosophy					
Male	179	2889	32%	3.2	8.2
Female	43	605	20%	2.9	14.3
Public Health					
Male	79	4458	55%	13.1	19.4
Female	53	2836	32%	14.7	17.0

The average academic age in Sample A and Sample B are the same, 14 years. However Sample B, the high performing group, have on a greater amount of citations to a smaller amount of papers than Sample A, resulting in a higher rate of Citations Per Paper. Even though they produce fewer papers the female researchers' publications are achieving on average a higher impact than their male counterparts in all disciplines except Public Health.

Nationality

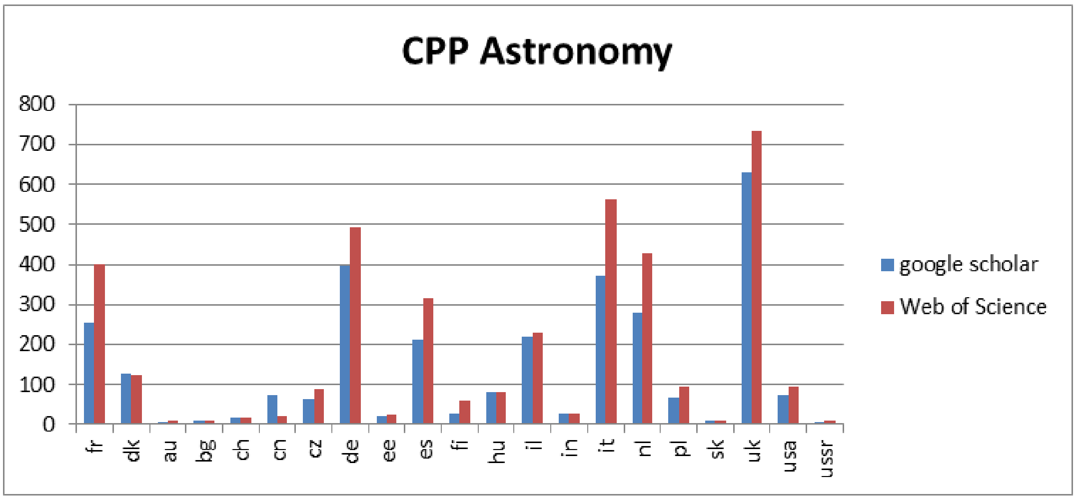
Nationality can also have an effect on researcher output and reception of their work. The researchers in our sample of researchers that are covered in GS and WoS are primarily western European, Table 9.

Table 9. Nationality of researchers

Nationality	nResearchers	% sample A	% sample B	Nationality	nResearchers	% sample A	% sample B
British	105	74	26	Finnish	14	85	15
Italian	78	78	12	Estonian	8	100	0
German	54	64	36	American	5	20	80
Spanish	46	80	20	Slovakian	4	100	0
Dutch	42	73	27	Bulgarian	2	100	0
French	33	54	46	Indian	2	100	0
Danish	27	92	8	Australian	1	0	100
Chzec	24	87	13	Chinese	1	0	100
Israelian	24	87	13	Greek	1	100	0
Polish	21	85	15	Russian	1	100	0
Hungarian	18	100	0	Swiss	1	0	100

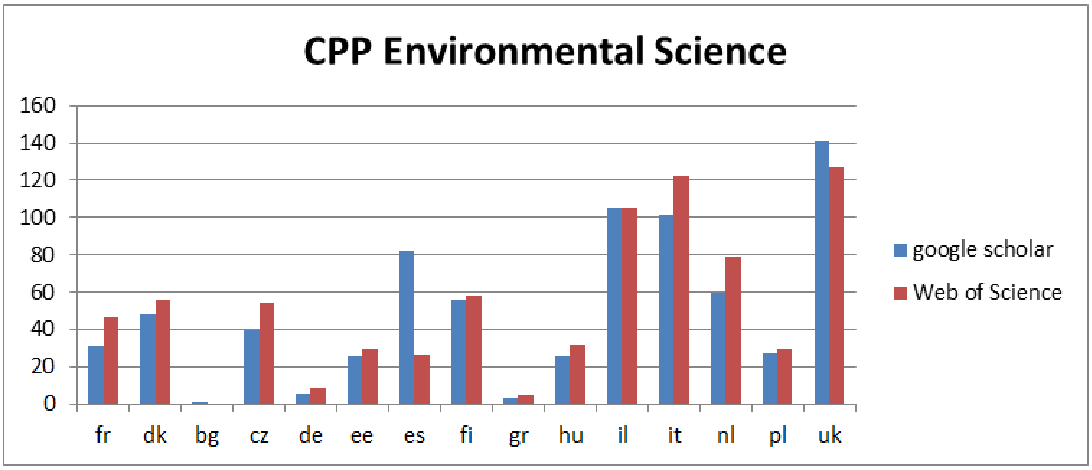
There is no clear grouping of nationalities in Sample A and Sample B. However, there is definite advantage for scholars of certain nationalities and disciplines to find citations in Google Scholar rather than WoS, Tables 10, 11 , 12, 13.

Table 10. Citations per paper in Astronomy



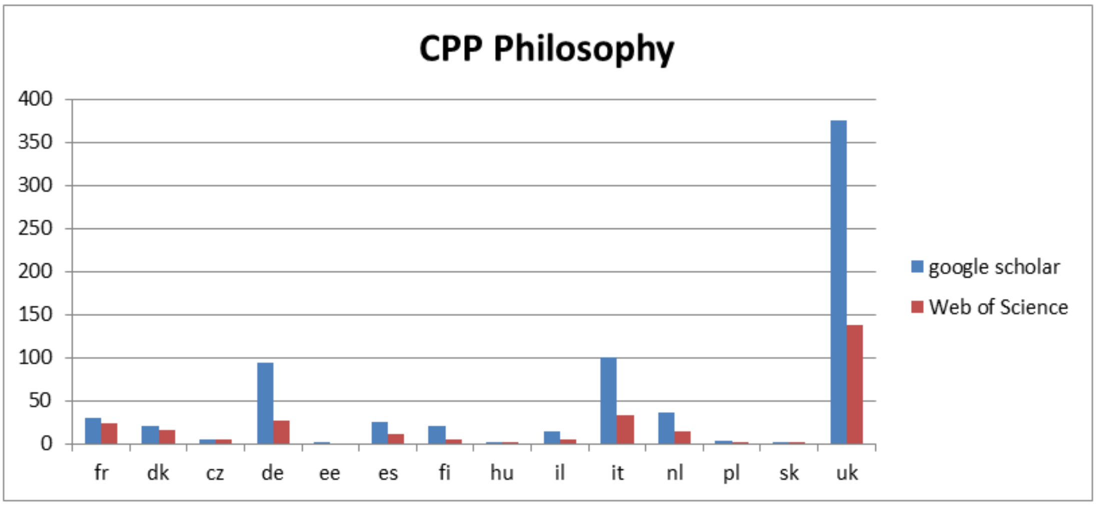
Conference papers are an important publication type for Astronomers, and as we experienced in our data-collection these were not available in our version of Web of Science and seriously reduced the amount of publications and citations per researcher. However, Web of Science still results in higher CPP for all researchers than Google Scholar.

Table 11. Citations per paper in Environmental Science



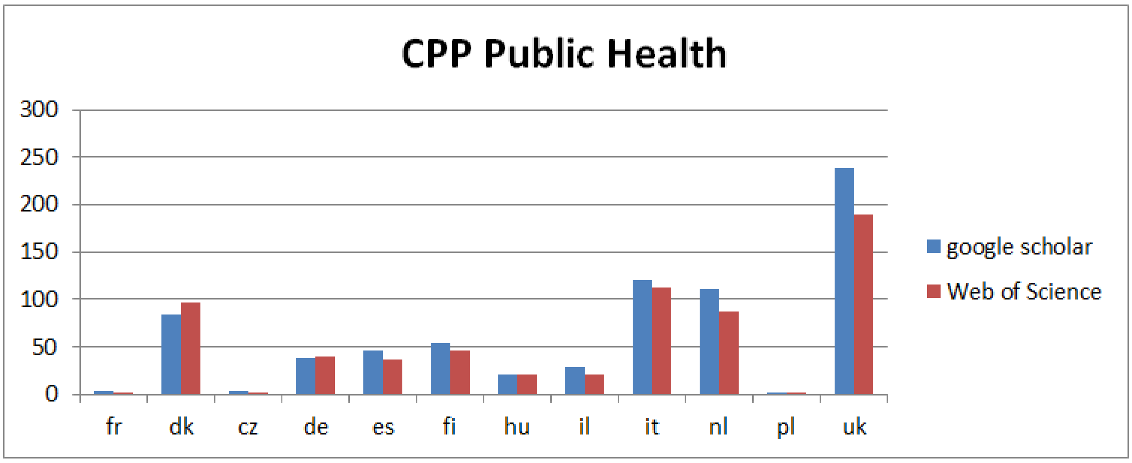
CPP is slightly improved in Web of Science across all nationalities apart from a noticeable improvement in Google Scholar for Spanish researchers.

Table 12. Citations per paper in Philosophy



Google Scholar clearly out performs Web of Science in indicating CPP for researchers in Philosophy, whereas for Public Health the resulting CPP is only slightly higher.

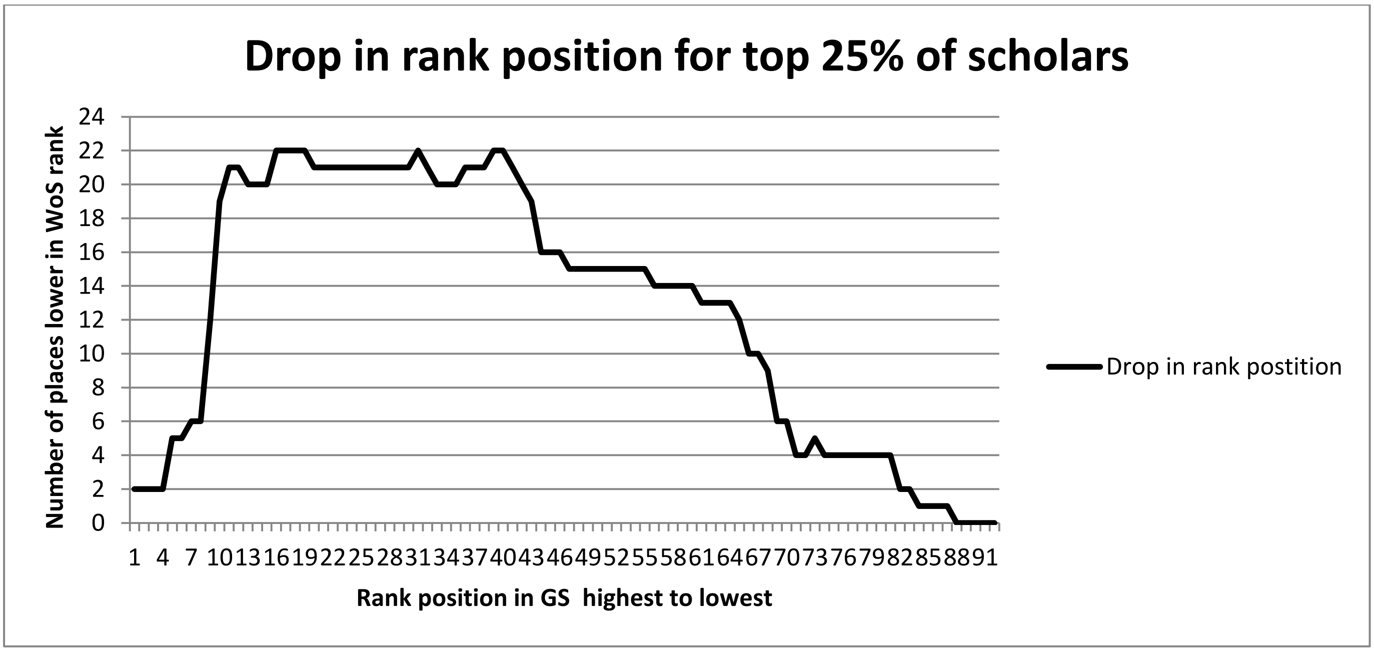
Table 13. Citations per paper in Public Health



Citations per paper

In the previous section we exemplified database performance to nationalities using citaqtions per paper (CPP). CPP is considered a robust indicator of performance. But we wish to investigate if this indicator is database dependent or if it is database independent for the top performing researchers. It was possible to compute bibliometric indicators for 512 researchers in both WoS and GS. The number of CPP a researcher received in the Google Scholar data was compared to the Web of Science data. Even though there is a positive correlation between CPP in WoS and GS, r=0754, n=512, p=0.00, there is no correlation between the resulting ranks of the scholars. All scholars were ranked from highest to lowest CPP and there was no correlation between their rank position in Google Scholar and in Web of Science. The set was divided into quartiles to identify if the CPP was stabile as a rank across databases for the top 25% of CPP scholars, r=0.051, n=128 and p=0.566. By manually investing the change of rank position in this top set, we found that 72% of the scholars appear in the top set in both databases, however the remaining 28% of scholars are entirely different from Google Scholar to Web of Science. On average the rank of the researcher in Google scholar was 12 places higher than the ranking of the same researchers in Web of Science, figure 1.

Figure 1: Number of places a scholar drops when ranked using CPP in Google Scholar compared to Web of Science



In the WoS data set 396 researchers performed under the *average mjs mcs* (Sample A) and 345 researchers performed better than *average mjs mcs*, (Sample B). Continuing the investigation of the stability of CPP, we investigated if researchers' whose publications out-perform the expected benchmarks, were well represented in the top 25% CPP. Eighty-one out of the 128 highest ranking CPP researchers in WoS, 63%, were from Sample B, while 65 researchers from Sample B where ranked top 25% CPP in Google Scholar, making up 50% of this sample.

Conclusions and recommendations

Our main finding is that indicators are highly dependent on the database used to compute them and the resulting impact-rankings of researchers are different. As such it is of utmost importance that the database used to collect the publication and citation data is reported alongside the indicators. Researchers who compute their indicators using Web of Science data should not be compared with researchers who compute indicators in Google Scholar. Further, our own data collection showed that different versions of the same database can also produce different results.

- Even though Google Scholar provided more publications and citations on an individual level, the work needed to clean the data to ensure researchers are only attributed with works that they authored is time consuming and sometimes impossible due to name ambiguities.

- The data retrieved from Web of Science was reliable, but limited in its coverage of the individual, which was detrimental to the outcome of the computed indicators in some disciplines and for some nationalities.

- Disciplinary and national coverage of a database should be established before author-level indicators are computed, as coverage can limit fair indications of the impact of work. Based on our study, we would recommend Philosophers use Google Scholar, well aware that this recommendation incurs increased work in cleaning and importing the publication and citation data.

- Raw citation count alone is not an indicator of impact; citation counts need to be benchmarked or normalized to similar research. Generally indicators computed using Google Scholar data are higher than indicators computed using Web of Science data.

- The m-quotient provides an indication of impact adjusted to the academic age of the researcher, and proved comparable across Google Scholar and Web of Science.

- A benchmark of expected citations for the researcher's speciality was calculated using only the Web of Science data. This was used to compare the impact of the individuals' publications. The results showed that even though female researchers produce fewer papers, they have a higher impact on average in their specialty than male researchers.