



The citations chase: A random researcher's view

Psarftis, Harilaos N.

Published in:
ORMS Today

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Psarftis, H. N. (2021). The citations chase: A random researcher's view. *ORMS Today*, 6(7).
<https://pubsonline.informs.org/doi/10.1287/orms.2021.03.03/full/>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The Citations Chase¹

Bibliometrics – the most important publishing statistic?

By Harilaos N. Psaraftis

The importance attached to citations as an indicator of academic excellence, prominence, impact, influence and the like is on the rise. This urgency can be termed “the citations chase.”

Bibliometrics is the use of statistical methods to analyze publications or the science of citations, and activity in this area has grown considerably over the years. Citations statistics are produced, analyzed and discussed, sometimes ad nauseam. The dictum “publish or perish” has transcended into “be cited or perish.” Being published is simply not enough.

Why are Citations so Important?

Citations are important in hiring, promotion and tenure decisions, not to mention salary negotiations, funding proposals, etc. They are often considered a proxy for academic excellence, prominence, impact and influence. Certainly, they are not the only proxy, but they are becoming increasingly important.

The main trigger for this article was a number of recent bibliometric studies in which I was lucky to find my name included. Among them was a database known as the “100,000 highly cited researchers,” as described in Ioannidis et al. [1]. The actual number of scientists listed is 105,000. A colleague told me I was in that database, which to me was a pleasant surprise.

Some 6,880,389 scientists who have published five or more papers form the content of the database. Of those, the top 105,000 are selected and ranked using a composite index. Roughly 40 indices are collected for each author. The database is Scopus-based (the largest abstract and citation database of peer-reviewed literature) and breaks down scientific disciplines into 22 scientific fields and 176 subfields. Fields include broad areas such as engineering, economics and business, biology and more, and the subfields go into further detail, i.e., optoelectronics and photonics, urban and regional planning, operations research, logistics and transportation and many others. In that sense, a scientist can be ranked not only within the entire database, but also within the field and subfields in which they are associated.

Looking at the database and the accompanying paper, I was impressed that researchers in some areas get widely differing citation numbers than other areas. For instance, the 99 percentile of citations in “Nuclear and Particle Physics” is 32,708. However, the citation number for

¹ Published in *OR/MS Today*, June 2021. Link to article:
<https://pubsonline.informs.org/doi/10.1287/orms.2021.03.03/full/>

“Astronomy and Astrophysics” is 16,244, “Ecology” is 8,302 and “Drama & Theater” is only 288. For some subfields there are zero people in the top 100,000+ list. This means that it is difficult, or even futile, to compare scientists across different disciplines.

Still, digging deeper, some interesting “higher order” statistics can be calculated. Suppose a scientist has X citations. Then we may define the following:

(a) C_s , the percentage of X within the population of *that person’s citations*, that is ascribed to papers in which he or she is sole author.

(b) P_s , the percentage of X within the population of *that person’s papers*, that is ascribed to papers in which he or she is sole author.

These percentages are *second-order* citations statistics, being ratios of first-order statistics. They can also be defined in a broader sense as C_{sf} (or C_{sfl}), that is, percentages of X within the population of that person’s citations, that are ascribed to papers in which he or she is sole or first author (or sole, first or last author, respectively). In some disciplines the last author position is considered honorary. Similar definitions pertain to P_{sf} (or P_{sfl}).

Note that the values of C and P , as defined above, are not necessarily the same, because they are drawn from different populations. In fact, we can also define what we call the “claim to fame ratio” (for lack of a better name) as the ratio $R = C/P$ (for each of their three variants). R is a *third-order* citations statistic, being the ratio of two second-order statistics. A value of R higher than 1 means that an author’s prominence is more visible within the spectrum of their citations than within the range of their papers. Note that R_s is undefined if an author has zero sole-author papers, as in this case both C_s and P_s are zero. To my surprise, there are about 895 authors in the database who have never authored a paper by themselves.

Breaking it Down

To experiment with this, I calculated the three variants of C , P and R for those scientists in the database whose *primary subfields* are either: (a) Logistics & Transportation or (b) Operations Research. I chose these subfields because Logistics & Transportation is my own primary subfield and Operations Research is my secondary subfield.

Logistics & Transportation has a 99-percentile citation of 1,997. According to this criterion, I was disheartened to see it ranking lower than entomology, ornithology, nursing, veterinary sciences and applied ethics, among others. But it ranks higher than civil engineering, aeronautics and astronautics and general mathematics, among others. Of the 6,880,389 researchers, 15,386 list Logistics & Transportation as their primary subfield, and 84 of them are in the top 100,000+ list. Table 1 presents the top 50 scientists on this list, together with their three R variants. (C and P statistics and secondary subfields are available in the unabridged online version of the paper.)

Table 1: Top 50 scientists who list Logistics & Transportation as their primary subfield, together with their three R variants.

No.	Scientist	Rank in database	R_s	R_{fs}	R_{sn}
1	Daganzo, Carlos F.	2,689	1.549	1.314	1.062
2	Hensher, David A.	3,391	0.659	1.069	1.038
3	Cervero, Robert	4,391	0.419	0.768	0.938
4	Bhat, Chandra R.	6,078	1.938	1.458	1.152
5	Flyvbjerg, Bent	6,369	1.347	1.263	1.131
6	Train, Kenneth	6,564	2.142	1.120	1.048
7	Yang, Hai	9,212	1.859	1.973	1.178
8	Handy, Susan	11,556	0.953	1.515	1.072
9	Mokhtarian, Patricia L.	12,712	0.878	1.140	1.020
10	Banister, David	16,770	0.949	0.983	0.822
11	Ewing, Reid	16,929	0.336	1.071	0.885
12	Williams, Allan F.	17,807	1.053	0.810	0.983
13	Bell, Michael G.H.	19,015	1.456	1.428	1.234
14	Rietveld, Piet	19,482	0.490	0.726	1.000
15	Papageorgiou, Markos	20,557	0.815	1.767	0.998
16	Elvik, Rune	23,067	1.072	1.076	1.041
17	Abdel-Aty, Mohamed	24,607	2.086	1.540	1.306
18	Mannering, Fred L.	26,805	0.481	0.728	1.039
19	Levinson, David	33,196	1.779	1.383	0.939
20	van Wee, Bert	33,776	0.833	0.711	0.961
21	Mahmassani, Hani S.	35,371	1.181	1.381	1.215
22	Notteboom, Theo	35,534	1.264	1.483	1.170
23	Sheu, Jih Biing	36,011	1.473	1.258	1.121
24	Pucher, John	36,211	0.194	1.053	1.005
25	Evans, Leonard	37,374	1.064	1.047	1.019
26	Viano, David C.	37,396	0.492	0.789	0.730
27	Timmermans, Harry J.P.	38,336	0.704	1.194	1.026
28	Noland, Robert B.	39,947	0.945	0.989	1.120
29	Golob, Thomas F.	40,344	2.599	1.093	1.011
30	Zhang, H. Michael	41,740	2.750	1.914	1.050
31	Summala, Heikki	42,596	0.925	1.000	1.062
32	Kockelman, Kara M.	43,320	1.130	1.100	1.166
33	Lo, Hong K.	44,538	2.916	1.714	1.063
34	Verhoef, Erik T.	45,217	1.179	1.251	1.105
35	Shinar, David	45,813	0.831	1.171	1.043
36	Wong, Sze Chun	46,880	1.112	1.095	1.004
37	Ben-Akiva, Moshe	48,874	0.190	1.363	0.993
38	Airey, Gordon D.	53,301	5.196	2.777	1.372
39	Psaraftis, Harilaos N.	53,647	1.671	1.366	0.992
40	Axhausen, Kay W.	53,857	0.714	1.157	0.895
41	Masad, Eyad	55,018	3.205	1.753	1.193
42	Hooendoorn, Serge P.	57,979	1.106	2.387	1.018

43	Nagel, Kai	58,002	1.721	2.255	0.835
44	Lord, Dominique	58,415	1.271	2.111	1.145
45	Huang, Hai Jun	59,806	1.875	1.368	1.009
46	Karlaftis, Matthew G.	60,094	0.958	1.145	1.029
47	Hall, Randolph W.	60,575	0.898	0.789	0.810
48	Shope, Jean T.	60,791	0.958	1.055	1.020
49	Quddus, Mohammed A.	61,882	1.510	1.977	1.582
50	Hauer, Ezra	63,248	0.912	1.108	1.072

The database also lists 319 scientists with operations research as their primary subfield, drawn from a group of 20,758 scientists among the 6,880,389. Operations research has a 99-percentile citation of 3,435. According to this criterion, it ranks higher than Logistics & Transportation but lower than economics, gerontology, sports sciences, optics and dentistry, among others. Table 2 lists the top 50 operations research scientists, together with the three variants of their R values. (Note. You will see many INFORMS members on either table.)

Table 2: Top 50 scientists who list operations research as their primary subfield, together with their three R variants.

No.	Scientist	Rank in database	R _s	R _{ts}	R _{sff}
1	Saaty, Thomas L.	1,543	1.569	1.093	1.037
2	Laporte, Gilbert	1,633	2.095	1.332	1.006
3	Sarkis, Joseph	2,043	1.057	0.963	0.940
4	Gunasekaran, Angappa	4,041	0.703	1.467	1.154
5	Glover, Fred	4,050	3.310	1.387	1.102
6	Cheng, T.C.E.	4,132	0.469	0.982	0.991
7	Banker, Rajiv D.	4,138	1.630	1.184	1.104
8	Lee, Hau L.	4,740	1.614	1.489	1.134
9	Kusiak, Andrew	4,907	0.654	1.130	1.048
10	Bertsekas, Dimitri P.	5,881	0.776	0.879	0.932
11	Goyal, Suresh	6,518	0.877	1.077	1.023
12	Mangasarian, Olvi L.	6,645	0.544	0.764	1.018
13	Whitt, Ward	6,973	1.209	1.170	1.013
14	Chan, Felix T.S.	7,002	1.295	1.414	1.270
15	Beasley, J. E.	10,269	1.432	1.417	1.013
16	Van Wassenhove, Luk	10,627	7.698	1.277	1.081
17	Mingers, John	10,638	1.181	1.197	1.055
18	Cachon, Gérard P.	10,996	1.020	1.050	1.067
19	Lee, Chung Yee	11,140	2.262	1.257	1.046
20	Tseng, Paul	11,492	1.294	1.107	1.005
21	Fisher, Marshall L.	12,125	1.108	0.740	1.052
22	Gendreau, Michel	13,168	0.522	1.473	0.986
23	Zhu, Joe	13,178	1.734	1.671	1.041

24	Towill, Denis R.	13,377	0.492	0.537	1.068
25	Ngai, Eric W.T.	13,386	1.495	1.461	1.077
26	Bertsimas, Dimitris	13,577	0.467	1.087	1.025
27	Combettes, Patrick L.	14,623	1.340	1.130	1.064
28	Sherali, Hanif D.	15,623	0.336	0.608	0.654
29	Bard, Jonathan F.	15,884	0.946	1.082	1.015
30	Cooper, William W.	15,943	0.038	0.353	0.782
31	Fukushima, Masao	17,304	1.577	1.475	1.056
32	Kleijnen, Jack P.C.	17,417	1.101	1.123	1.048
33	Pang, Jong Shi	17,578	0.648	0.846	1.087
34	Qi, Li qun	18,029	1.320	1.741	1.102
35	Brucker, Peter	18,283	1.347	0.918	1.091
36	Wright, Stephen J.	18,288	0.438	0.794	1.061
37	Tang, Christopher S.	19,440	1.903	1.495	1.098
38	Nesterov, Yurii	19,568	1.737	1.319	1.129
39	Hochbaum, Dorit S.	20,229	0.799	1.152	1.092
40	Hansen, Pierre	20,425	0.322	1.001	1.092
41	Keeney, Ralph L.	20,722	1.059	0.968	0.964
42	L'Ecuyer, Pierre	20,845	1.711	1.373	1.087
43	Taillard, Eric D.	20,995	1.832	1.292	1.001
44	Kao, Chiang	21,067	0.856	0.997	1.010
45	Goldfarb, Donald	21,228	1.785	0.704	0.731
46	Lasserre, Jean Bernard	21,246	1.228	1.037	1.031
47	ReVelle, Charles	21,426	1.021	1.443	1.176
48	Cordeau, Jean François	21,601	3.320	1.990	1.720
49	Guide, Jr. V. Daniel R.	21,985	1.685	1.226	0.983
50	Dekker, Rommert	22,047	3.669	1.457	0.836

I could find no discernible pattern in either table, for instance, regarding a possible correlation between the R values vis-à-vis the rank of the scientists in the database. Both tables confirm the wide diversity in citations statistics, even across classes of scientists who seem to have similar profiles. (Incidentally, I can name at least one person in either table who has passed away.)

Conclusion

In my opinion, even though bibliometrics is an interesting sport, its importance is overblown. What I think is more important is whether a paper's content is sound, whether it has improved upon the state of the art, or whether its results are useful to science, industry or society. In addition, spending time with industry may not be reflected at all in any citation metric, even though this may be just as important professionally. (I spent 5.5 years as CEO of the Port of Piraeus, during which time I wrote zero papers.)

Editor's note. An unabridged version of this article with more detailed considerations on this subject and with expanded tables (including all C and P values and secondary subfields) is available online².

Harilaos N. Psarftis is a professor in the Department of Technology, Management, and Economics at the Technical University of Denmark.

Reference

1. Ioannidis, J.P.A., Baas, J., Klavans, R., Boyack, K.W., 2019, "A standardized citation metrics author database annotated for scientific field," *PLoS Biology*, Vol. 17, No. 8, Article no. e3000384, <https://doi.org/10.1371/journal.pbio.3000384>.

² Link to unabridged version:

<https://pubsonline.informs.org/doi/10.1287/orms.2021.03.03/full/THE.CITATIONS.CHASE.pdf>