



Correlating article citedness and journal impact: an empirical investigation by field on a large-scale dataset

Giovanni Abramo¹ · Ciriaco Andrea D'Angelo^{1,2} · Flavia Di Costa²

Received: 13 July 2022 / Accepted: 16 December 2022 / Published online: 9 January 2023
© The Author(s) 2022

Abstract

In spite of previous research demonstrating the risks involved, and counsel against the practice as early as 1997, some research evaluations continue to use journal impact alone as a surrogate of the number of citations of hosted articles to assess the latter's impact. Such usage is also taken up by research administrators and policy-makers, with very serious implications. The aim of this work is to investigate the correlation between the citedness of a publication and the impact of the host journal. We extend the analyses of previous literature to all STEM fields. Then we also aim to assess whether this correlation varies across fields and is stronger for highly cited authors than for lowly cited ones. Our dataset consists of a total of almost one million authorships of 2010–2019 publications authored by about 28,000 professors in 230 research fields. Results show a low correlation between the two indicators, more so for lowly cited authors as compared to highly cited ones, although differences occur across fields.

Keywords Article impact · Journal impact factor · WoS · Bibliometrics · Skewness · Scientific publishing · Italy

Introduction

In seeking publication, researchers aspire to highly prestigious (often referred to as high-impact or high-quality) scientific journals, aiming at the widest possible distribution of their new knowledge, the greatest access to their scientific community, and the best hopes of career progress and resulting continued scientific output. Considering the strict review processes and high rejection rates, the achievement of publication in

✉ Giovanni Abramo
giovanni.abramo@iasi.cnr.it
Ciriaco Andrea D'Angelo
dangelo@dii.uniroma2.it
Flavia Di Costa
flavia.di.costa@uniroma2.it

¹ Laboratory for Studies in Research Evaluation, Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy, Rome, Italy

² Department of Engineering and Management, University of Rome "Tor Vergata", Rome, Italy

such journals is indeed a great source of satisfaction. Reviewers are chosen from among the most qualified scholars in the relevant fields, and typically offer observations, suggestions and even criticisms that serve in improving the specific paper and as further stimuli to research. On the side of the scientific community as audience, readers look to prestigious journals with respect and high expectations in terms of the reference value of the contents. High-impact journals attract and select the better works, likely to attract more citations (Traag, 2021), and therefore achieve better diffusion of the knowledge.

In research organizations, it is not uncommon to come across more or less formalized incentive systems that encourage researchers to publish in prestigious journals, and include aspects of evaluation in consideration of the numbers of their articles in such journals. Irrespective of their contents, there can be a tendency to transfer the prestige of the journals to the articles and their authors, so that those who succeed in reaching publication in high-impact journals are looked upon with admiration, even a sort of envy.

Peer-reviewed journals serve in the verification of new knowledge, encoded in written form (articles), and then its transfer from producers (authors) to consumers (scholars, practitioners). The process is the same as that for other channels, serving in quality-control and transfer of products of all kinds from maker to consumer. However, just as consumers resist being fooled by fine displays (buyer beware!, all that glitters is not gold), and investors by a flashy prospectus (what's the bottom line?), we, as scientists, evaluators and research administrators, should exercise caution in considering that a fine journal guarantees the high impact of all the published products on future advances in knowledge. Quoting Nobel laureate Aaron Ciechanover (2013): "It doesn't matter where you publish; it matters what you publish". Eugene Garfield, the father of the journal impact factor (JIF) was also well aware of this issue of the true citedness of the published articles: "It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the JIF as a surrogate" (Garfield, 2001). Authoritative scholars, and even entire scientific associations and congresses, have continued to express similar positions (DORA, 2012; Hicks et al., 2015; Seglen, 1997; Wouters et al., 2015).

In the absence of any other information, the prestige of the journal of publication might serve as a kind of certification, or guarantee of quality: the higher it is, the more soundly one could possibly assume the articles published are of high quality. Moreover, in research evaluation exercises, the available citation time windows are often not long enough to guarantee the robustness of citation impact indicators, especially in certain disciplines: in such cases, in fact, compared to the early citations registered, the JIF can function as a better predictor of the future impact of an article (Abramo et al., 2010), or along with the citations can serve well as a co-predictor (Abramo et al., 2019).

Still, as we have noted, in evaluating the work of individual researchers and their institutions, caution is due in taking the JIF alone as a surrogate for citations, particularly in the case of articles above 2 years old, where its use can prove insidious (Abramo et al., 2010). It is also well known that the high prestige of a journal, as measured by the average citedness of the articles published in it, is usually determined by a small share of highly cited articles (Antonoyiannakis, 2020; Egghe, 2005; Kiesslich et al., 2021; Larivière & Sugimoto, 2019; Milojević et al., 2017; Seglen, 1989, 1992, 1994, 1997; Waltman, 2016; Zhang et al., 2017). Apart from these few articles, most others have a citedness that does not deviate much from that of articles published in less prestigious journals. It is also not uncommon to find articles published in high-impact journals, but which are never cited (Larivière et al., 2016; Lozano et al., 2012; Zhang et al., 2017).

It is therefore no surprise that, before us, bibliometricians have investigated the relationship between journal prestige and the citedness of hosted articles (de Oliveira Silva et al., 2021; Seglen, 1989, 1994, 1997; Zhang et al., 2017). In this paper we return to the issue, but extending the scale and disciplinary scope of the population under observation, and also deepening the investigation in several respects, by testing: (i) the correlation between the citedness of a publication and the impact of the host journal, not just overall, but also at field level (RQ1); (ii) whether such correlation varies among researchers within each field (RQ2); (iii) whether this JIF-citedness correlation is stronger for highly cited rather than for lowly cited authors (RQ3).

The work is rooted in the seminal articles by Seglen (1989, 1994, 1997), who analyzed the JIF-citedness relationship on a dataset of 16 Norwegian principal investigators in biomedical research. Twenty years later, Zhang et al. (2017) returned to the same methods, but on a larger sample of around 600 investigators and 18,000 publications, again from Norway and in the biomedical field, for the period 1992–2013, with the aim of verifying the earlier results of Seglen. In fact the findings were in line with Seglen's. At the level of the publication portfolio of individual authors, Zhang et al. found a moderate correlation on average, and in any case, characterized by extreme variability between authors. Finally, their analysis confirmed Seglen's conjecture that the relationship between article citedness and journal impact is much stronger for top-cited authors than for less cited colleagues. More recently, de Oliveira Silva et al. (2021) investigated a sample of 4022 sports sciences articles, showing that altmetric scores have a stronger relationship with number of citations than the JIF.

The current work extends the analysis to all STEM fields, verifying the results of our predecessors on a broader and more diversified population, and above all the questions of potential differences across fields. The dataset of analysis consists of 28,000 Italian professors and their scientific production indexed in Web of Science (WoS) in the period 2010–2019, totaling nearly one million authorships. The Italian case is convenient for the analysis, because of the uniquely fine-grained field classification of all professors (370 fields) and the availability of a highly accurate authorships disambiguation algorithm.

The next section of the paper reviews the literature relevant to the current study. “**Methods**” section describes the data and methods designed to respond to the research questions. “**Results**” section provides the results of the analysis and “**Discussion and conclusions**” section concludes with comments on the main findings and certain of their implications.

Literature review

In the bibliometric field, the oldest and most widely used indicators are at journal level. The introduction of the JIF some 40 years ago, by the Institute for Scientific Information (ISI) of Eugene Garfield, was intended to provide US university librarians with quantitative, objective methods for the selection of journals. Beginning as the main indicator of the Journal Citation Report (JCR), the JIF then spread unstoppably, assuming new purposes in research evaluation across all fields of the hard sciences (Garfield, 2001), in the meantime transforming the publishing industry and significantly influencing recruitment practices, as well as resource allocation and even the directions of research activities (Archambault & Larivière, 2009). Individual researchers use the JIF not only as an indicator of journal quality, but also for other purposes, such as in choosing a narrower

selection of papers out of the huge offer, for view, reading and potential citation (De Rijcke et al., 2016; Rushforth & De Rijcke, 2015).

Its success has fueled the proliferation of multiple variants, including the ones coined by the publishers of all the different bibliographic repertories. Paradoxically, a considerable number of studies have illustrated critical aspects of the indicator: the asymmetry between the numerator (citations of citable items) and the denominator (citable and non-citable items); the many differences between disciplines in habits of publication and citing; the problems of insufficient citation window; the asymmetry of the citation distributions; journal self-citations; the lack of transparency on its calculation, which casts doubt on the results (Lariviere & Sugimoto, 2019; Mingers & Leydesdorff, 2015; Seglen, 1997). Added to these are the weakening relationships between JIF and citedness of papers due to the advent of digital repositories (Lozano et al., 2012), and the increasing recourse to new channels of communication on research results, i.e. via social media, which can impact on the citations to an article more than the prestige of the publishing journal (de Oliveira Silva et al., 2021). Last but not least, the JIF is related to the mean of the distribution of citations of hosted publications, which is notoriously highly skewed (Bornmann & Leydesdorff, 2017; Glänzel & Moed, 2002; Leydesdorff, 2008; Kiesslich et al., 2021; Radner, 1998). For these reasons, most scholars have urged caution when using journal indicators for individual evaluation (Brito & Rodríguez-Navarro, 2019; Jarwal et al., 2009; Larivière et al., 2016; Marx & Bornmann, 2013; Moed, 2020; Moed & van Leeuwen, 1996; Paulus et al., 2018; van Leeuwen & Moed, 2005).

The pioneering investigations of Seglen (1989, 1992, 1994, 1997) on the relationship between article citedness and JIF have been particularly influential in the scientific community, as regards the effects of such indicators in evaluation processes, and have inspired several important initiatives, such as the San Francisco Declaration on Research Assessment (DORA), the Leiden Manifesto for research metrics, and The Metric Tide review on the role of metrics in research assessment and management (DORA, 2012; Hicks et al., 2015; Wouters et al., 2015).

In a 1997 article, Seglen summarized four criticalities: “(i) Use of JIF conceals the difference in article citation rates (articles in the most cited half of articles in a journal are cited 10 times as often as the least cited half); (ii) JIF is determined by technicalities unrelated to the scientific quality of their articles; (iii) JIF depends on the research field: high JIFs are likely in journals covering large areas of basic research with a rapidly expanding but short-lived literature that use many references per article; (iv) Article citation rates determine the JIF, not vice versa”. Traag (2021), however, argued that simply removing consideration of JIF does not negate the influence of journals on the citedness of hosted articles.

Recently, several studies have returned to favoring the use of journal-based indicators under certain conditions (Abramo et al., 2010, 2019; Bonaccorsi, 2020; Waltman & Traag, 2021). Waltman and Traag (2021) also demonstrated through computer simulations that “depending on the assumptions that are made, the JIF may be a more accurate indicator of the value of an article than the number of citations of the article”. Abramo et al. (2019) showed that with very short time windows (0 to 2 years) JIF together with early citations can play a useful role in predicting future impact of publications, even though for longer ones, the weight of early citations is dominating and the JIF is not informative any longer.

Decades after Seglen first raised his criticisms and began investigations, the global research community, and that of research assessment in particular, continues to debate and study his original questions. Callaway (2016), for example, reviews some of the main

complaints, and reports anew the cases of Nature and Science journals, where 75% of 2013–2014 hosted articles garnered fewer citations than the 2015 JIF of the journals.

Kim et al. (2020) effectively illustrate the methodological difficulties in estimating journal influence on article performance. The approach proposed by the authors consists of using a proxy for individual article quality, unrelated to the publishing journal reputation/impact: i.e. number of citations to the preprint, posted on arXiv.org. Using the title and digital object identifier (DOI), all articles in arXiv on high energy physics, astrophysics, and condensed matter were linked to the corresponding published versions indexed in Microsoft Academic Graph (MAG). The authors showed that “estimates of the effect of journal reputation on an individual article’s impact (measured by citations) are likely inflated”, and found “little systematic evidence that the role of journal reputation on article performance has declined.”

Traag (2021), aiming to uncover the causal mechanism between journal impact and article citedness, also looked at citations to arXiv preprints, but applying a more sophisticated model for comparing these to the citations received by the full versions, as recorded in Scopus. The results show that journal impact “filtering” does not cancel out the influence of journals on the citedness of hosted articles, and that therefore: (i) articles that attract more citations are more likely to be selected and published in high-impact journals; (ii) articles in high-impact journals will be cited even more frequently because of the publication venue.

Zhang et al. (2017) adopted a different approach for verification of the pioneering studies by Seglen (1992, 1994, 1997), some 20 years later and on a larger sample of scientists. In particular, the authors examined: (i) the skewness of article citedness; (ii) the correlation between article citedness and impact of the hosting journal; (iii) the real benefit for scholars, in terms of received citations, of publishing in journals with higher impact. Whereas Seglen’s work was based on the study of 16 biomedical principal investigators affiliated with only one Norwegian institution, who authored 907 publications, Zhang et al. (2017) consider all scientists in biomedical research working at Norwegian institutions (approximately 600) and their production of approximately 18,000 publications between 1992 and 2013.

The results confirm Seglen’s findings that “there is no consistent positive relationship between individual article citedness and the JIF of the journal in which the article is published”. Citation distributions are skewed across journals: less cited articles and highly cited articles can appear in any journal. However, while most articles are rarely cited, the few highly cited articles appear more often in high-impact journals. The correlation between the two indicators referring to the scientific portfolios of individual authors is moderate on average and characterized by extreme variability. Finally, the lack of consistent positive relationships is more evident among the majority of researchers with average or lower overall citedness.

Methods

Data

The dataset for the current study consists of the scientific publications by Italian academics over the period 2010–2019 that are indexed in WoS. In the Italian university system, all academics are classified in one and only one field, named scientific disciplinary sector

(SDS), 370 in all. The SDSs are grouped into 14 disciplines, named university disciplinary areas (UDAs). The current analysis is limited to the SDSs for which the WoS coverage is acceptable for bibliometric assessments: 230 out of the total 370 (62%), comprised within a total of 11 UDAs.¹

The data concerning professors were extracted from the database of Italian university personnel maintained by the Ministry of Universities and Research (MUR). For each professor, this database provides information on their gender, affiliation, field classification, and academic rank at the end of each year.² In addition to the limitation on disciplinary sectors, a further selection was made at the level of the academics, limiting these to professors tenured over the entire 2010–2019 period, totaling 28,040.

Data on output and relevant citations are extracted from the Italian Observatory of Public Research, a database developed and maintained by the authors, derived under license from the Clarivate Analytics WoS Core Collection. Beginning with the raw data of the WoS, and applying a complex algorithm to reconcile the authors' affiliations and disambiguation of the true identity of the authors (D'Angelo et al., 2011), each publication (articles, letters and reviews hosted by journals with JIF) is attributed to the university professor who produced it.³

Table 1 shows the size of the working dataset in aggregate form. The 28,040 professors under review authored a total of 424,309 unique publications. In terms of sizes per UDA, Medicine leads the ranking in number of professors, ahead of Industrial and information engineering, and Biology leads in total publications. As regards the average number of publications per professor, four of the UDAs show values above overall average (28.0): Medicine (42.0), Chemistry (41.3), Physics (33.8), and Biology (29.3).

Indicators and data analysis plan

To pursue our objectives, we correlate two indicators for all publications in the dataset:

- the number of citations received, normalized to the average citations of all WoS cited publications, of the same year and WoS subject category, referred to as AII (article impact index),⁴
- the JIF of the host journal, normalized to the average JIF of the journals of the same year and subject category, referred to as JII (journal impact index).

Note that the use of normalized indicators is an absolute necessity in comparing citations/impact factors of publications from different fields and different years (Waltman, 2016).

For the first indicator, the citation count is taken on 31/12/2021. For the second, the JIF is the one at year of publication. The two indicators are by definition correlated, but because of the high skewness of citation distributions within journals, strong correlation is not expected.

¹ 1—Mathematics and Computer Science; 2—Physics; 3—Chemistry; 4—Earth Sciences; 5—Biology; 6—Medicine; 7—Agricultural and veterinary sciences; 8—Civil Engineering; 9—Industrial and Information Engineering; 10—Psychology; 11—Economics and Statistics.

² <http://cercauniversita.cineca.it/php5/docenti/cerca.php>, last accessed on 12 December 2022.

³ The harmonic average of precision and recall (F-measure) of authorships, as disambiguated by the algorithm, is around 97% (2% margin of error, 98% confidence interval).

⁴ AII is quite similar to the Category Normalised Citation Impact (CNCI) available in WoS or its variant available in SCOPUS, known as Field-Weighted Citation Impact (FWCI).

Table 1 Dataset of the analysis

Discipline	# Fields	# Professors	# Authorships	Average publications per professor
1—Mathematics and computer science	10	2229	31,909	14.3
2—Physics	8	1525	51,605	33.8
3—Chemistry	11	2174	89,873	41.3
4—Earth sciences	12	775	18,437	23.8
5—Biology	19	3486	102,284	29.3
6—Medicine	50	6476	272,278	42.0
7—Agricultural and veterinary sciences	30	2339	58,801	25.1
8—Civil engineering	18	1471	23,516	16.0
9—Industrial and information engineering	42	3903	97,713	25.0
10—Psychology	12	976	19,200	19.7
11—Economics	18	2686	20,797	7.7
Total	230	28,040	786,413	28.0

In the following section we investigate this correlation: (i) at overall and field levels; (ii) referring to the publication portfolio of each professor in the dataset; (iii) distinguishing between the top (10%) cited and bottom (10%) cited professors. The data analysis, which uses scatter plots and Pearson ρ correlations between the two indicators, is performed using the STATA 12 statistical software package. Outliers were not removed except where specifically reported in the text.

Results

We present the results taking two perspectives: from a first analysis with the article as the fundamental unit of investigation, next with the author as fundamental unit. The first analysis aims at answering RQ1, the second at answering RQ2 and RQ3.

Analysis at the article level

Overall (424,309 dataset publications), the Pearson ρ correlation between AII and JII is 0.308. Excluding the 17 publications with AII greater than 100, the correlation coefficient rises to 0.345. The data scatterplot (Fig. 1) confirms a weak association between the two dimensions under analysis.

The correlation between AII and JII can also be represented by generating a heat plot matrix, in which each cell indicates the relative frequency of the AII-JII row-column combination. For brevity, rather than presenting the entire distribution of absolute values for the two indicators, Table 2 shows their percentiles with respect to the reference distribution (world publications of the same year and SC),⁵ also highlighted by grey

⁵ Relative frequencies in the matrix sum up to 100%.

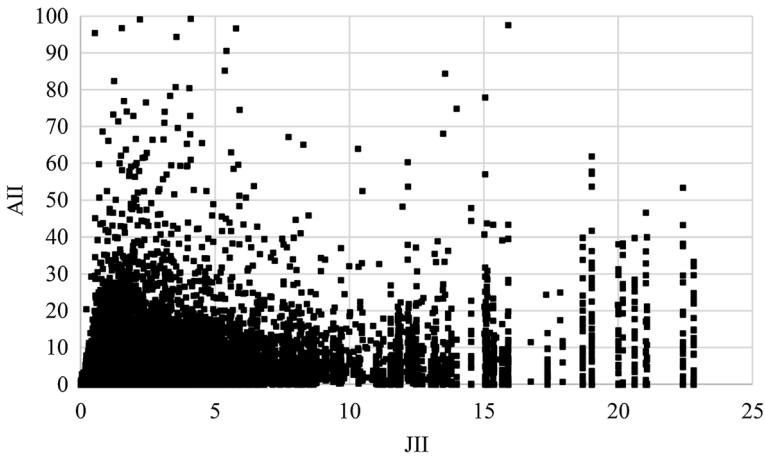


Fig. 1 Distribution of publications in the dataset, by citedness (AII) and journal impact (JII)

Table 2 Heat plot of journal impact (JII decile, 1 = max, 10 = min) and average article citedness (AII decile, 1 = max, 10 = min) for publications in the dataset

		AII (decile)									
		1	2	3	4	5	6	7	8	9	10
JII (decile)	1	5.2%	3.2%	2.2%	1.5%	1.1%	0.7%	0.5%	0.4%	0.2%	0.3%
	2	3.3%	3.7%	3.4%	2.8%	2.2%	1.7%	1.2%	0.8%	0.4%	0.4%
	3	2.0%	2.6%	2.6%	2.4%	2.0%	1.7%	1.3%	0.9%	0.5%	0.4%
	4	1.2%	1.7%	1.8%	1.8%	1.6%	1.4%	1.1%	0.8%	0.5%	0.5%
	5	0.7%	1.1%	1.3%	1.3%	1.3%	1.2%	1.0%	0.8%	0.5%	0.5%
	6	0.5%	0.8%	1.0%	1.1%	1.1%	1.0%	0.9%	0.7%	0.4%	0.5%
	7	0.3%	0.5%	0.7%	0.7%	0.8%	0.8%	0.8%	0.7%	0.5%	0.5%
	8	0.2%	0.4%	0.6%	0.7%	0.8%	0.8%	0.7%	0.6%	0.5%	0.9%
	9	0.1%	0.2%	0.2%	0.3%	0.4%	0.4%	0.5%	0.5%	0.4%	0.6%
	10	0.0%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.3%	0.3%	0.5%

gradation. If there is a strong correlation between the two dimensions of analysis, we expect an equally strong concentration of high relative frequencies on the main diagonal. Table 2 shows no such concentration, except in the upper left part of the matrix, the part characterized by the “top” deciles of both AII and JII. It is no surprise that the highest relative frequency (5.2%) concerns precisely the top publications, in terms of both indicators. This is due to the relative positioning of Italian scientific production, which features significantly higher average impact than the world average (Consiglio Nazionale delle Ricerche, 2019).

Note that the entire matrix is populated, even cell AII=1–JII=10, albeit with only 138 publications out of 424,309. Scrolling down column AII=1, we observe decreasing frequencies, but still greater than 1 up to the fourth decile of JII. Likewise, scrolling across row JII=1, frequencies again decrease but remain greater than 1 up to the fifth decile.

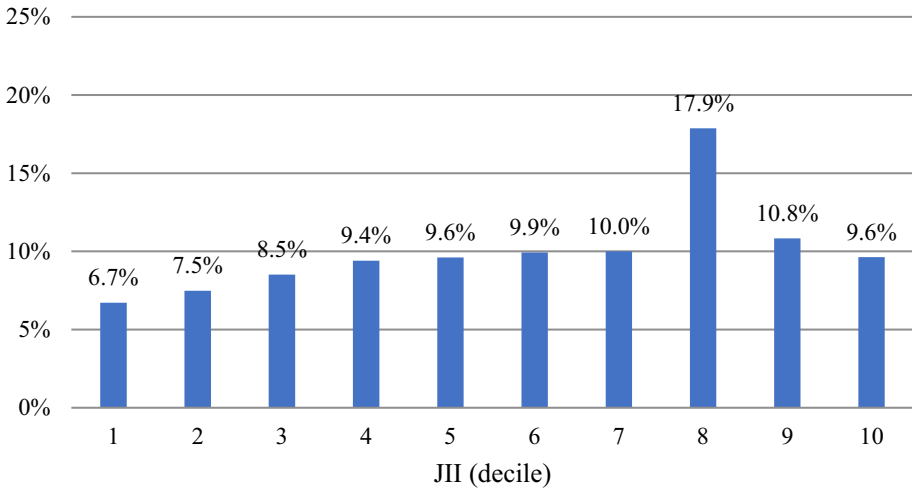


Fig. 2 Share of publications in the AII D10, by JII decile (1 = max, 10 = min)

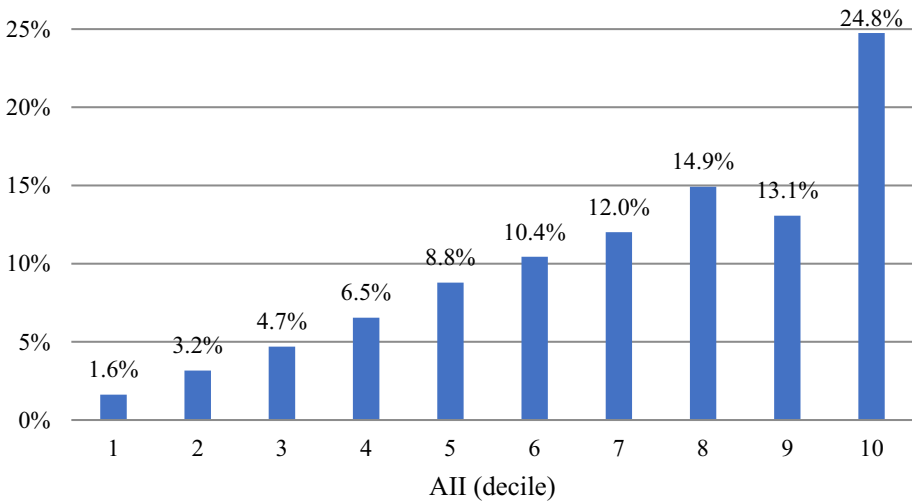


Fig. 3 Share of publications in the JII D10, by AII decile (1 = max, 10 = min)

It may also be interesting to analyze the distribution of JII of publications in the tenth decile (D10) by AII, or AII of publications in D10 by JII (Figs. 2, 3).

In D10 by AII (Fig. 2) we have 21,857 publications⁶ (5.2% of total 424,309), of which 6.7% (1467 publications) are hosted in journals falling in D1, and a full 41.7% in journals with JII greater than or equal to the median.

⁶ Of which 18,014 never cited.

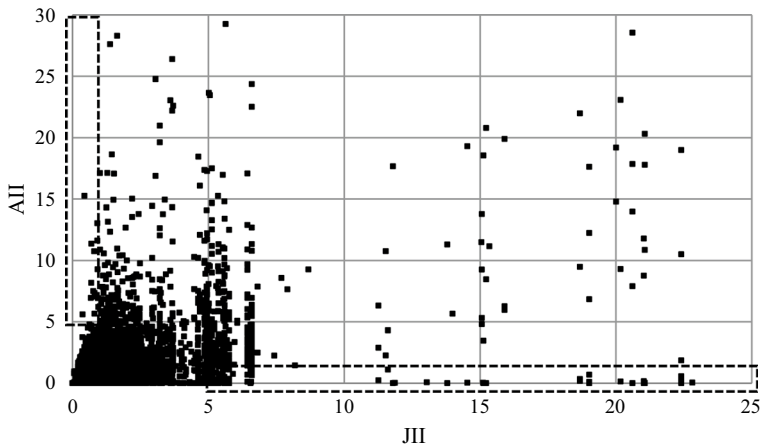


Fig. 4 Distribution of publications of Italian “Cardiovascular diseases” professors, by citedness (AII) and journal impact (JII). For better visualization, 13 outlier publications (AII higher than 30) are removed

In D10 by JII (Fig. 3) we find 8508 publications (2.0%), from which only 1.6% (i.e. 138 papers) place in D1 by AII, and only 24.8% above median.

Given the shape of distributions in Figs. 2 and 3, very briefly: regardless of the high impact factor of the host journal, a publication may be scarcely or never cited; conversely, journals with a very low impact factor are much more likely to host publications that are scarcely or never cited.

Moving to disciplinary level, Fig. 4 shows an example SDS, in this case the scatterplot of the 8319 publications authored by professors in MED/11 (cardiovascular diseases).⁷ The Pearson’s correlation for AII and JII is 0.319. The graph shows 15 publications with JII values below one and AII greater than 5 (top left box), but also 107 publications with JII greater than 5 (bottom right box) and AII values less than 1. The size of this second set is substantial, amounting to slightly less than a quarter of the total 433 publications with JII greater than 5. Thus, the presence of publications hosted by journals with very high JIF (five times expected value) but never or scarcely cited is by no means marginal.

Figure 5 shows the boxplot of Pearson distribution between AII and JII, for the same analysis repeated with all 230 SDSs. The highest correlation is seen in GEO/12 (Oceanography and atmospheric physics, 364 publications), where the Pearson coefficient is 0.594. The second outlier is ING-IND/24 (Principles of chemical engineering), with Pearson ρ for the 1816 publications by professors of that SDS being 0.573.

The average Pearson ρ for the 230 SDSs is 0.334 (median 0.329), with variability between maximum 0.594 (as noted, in GEO/12) and minimum 0.107 (in SECS-S/05, Social statistics), and a rather small interquartile range of 0.116.

With reference to RQ1, we can conclude that the correlation between citedness and journal impact is rather weak at the overall level. At the field level, although varying substantially, it is never strong.

⁷ A publication co-authored by professors belonging to different fields will be multiple counted, i.e. considered in the set of publications from each of those fields.

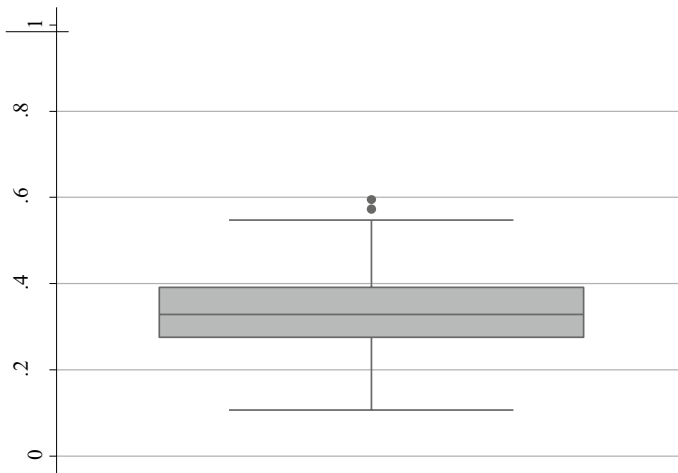


Fig. 5 Boxplot of Pearson correlations—citedness (AII) vs journal impact (JII) for 230 investigated SDSs

Analysis at author level

In this section, aiming to answer RQ2 and RQ3, we investigate the relation between article citedness and hosting-journal impact factor for the publication portfolios of all professors. As an example, Fig. 6 shows the scatterplot of data referring to the 292 publications authored in the period 2010–2019 by Gianluca Gaidano, Full Professor in MED/15 (Blood diseases) at the University of Piemonte Orientale. For this professor, there has been a significant correlation between AII and JII (Pearson $\rho = 0.675$). Figure 7 instead shows the case of Francesco Violi, Full Professor in MED/09 (Internal medicine) at the University of Rome ‘Sapienza’. In this case, the two indicators measured on his 240 publications seem to be completely unrelated, with the five publications hosted on the journals with the highest impact ($JII > 13$) all showing low or no citedness (AII between 0 and 0.4).

By grouping the professors of the dataset by SDS, we obtain distributions of correlation indices that may reveal the presence or absence of elements of differentiation among fields, useful for the response to RQ2. To ensure robustness of results, we consider only professors with a portfolio of at least 10 publications.

Taking the example of professors in the eight SDSs of the Physics UDA, Fig. 8 shows the box plots of the correlation distributions, and Table 3 the descriptive statistics. We note significant variability in the data among these SDSs, with median values ranging from minimum 0.184 in FIS/06 to maximum 0.433 in FIS/03. Many portfolios (around 10% of total) show negative AII vs JII correlation, the most of all in FIS/06, where this occurs for a full 30.4% of professors (14 in number). However, the confidence intervals all show a positive lower bound. The dispersion within SDSs is quite marked, with standard deviation values of the same order of magnitude as the averages, and coefficients of variation above 1 in FIS/02, FIS/05, and FIS/06.

The last column of Table 3 shows the Pearson correlation between the average AII and average JII of each of the eight Physics SDSs. With the exception of FIS/04 (Nuclear and subnuclear physics), this is always higher than mean ρ at individual level (column 4): for instance, in FIS/01 (Experimental physics) the correlation is 0.503, against 0.363 of the average ρ registered at individual level. Although the strength of the relationship varies

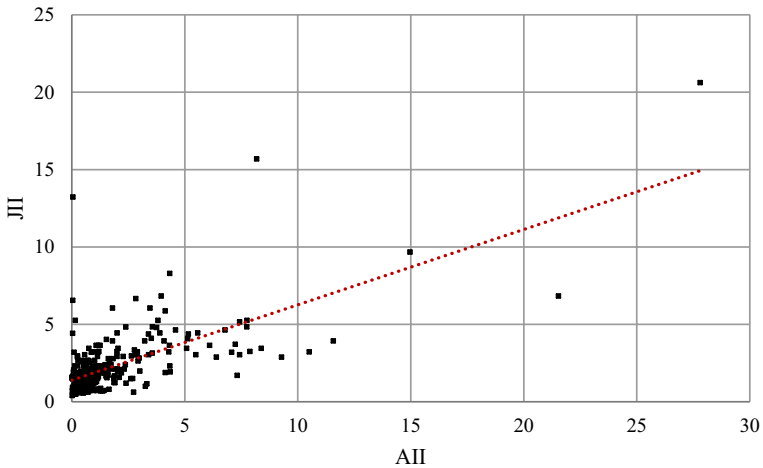


Fig. 6 Scatterplot—citedness (AII) and journal impact (JII) of publications of a professor in MED/15 (Blood diseases)

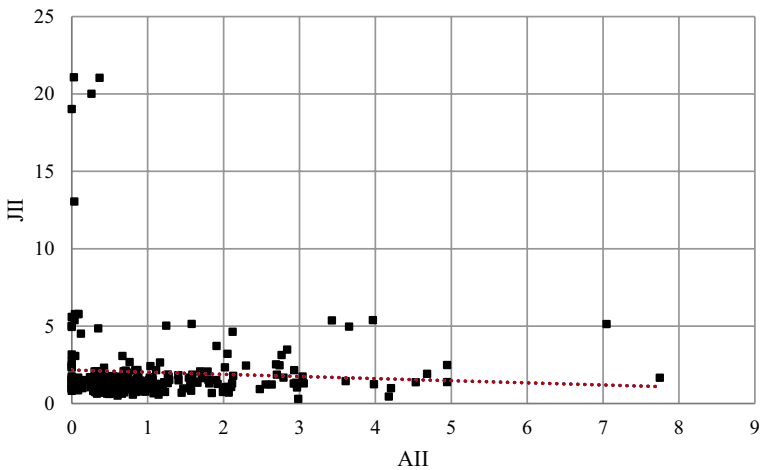


Fig. 7 Scatterplot—citedness (AII) and journal impact (JII) of publications of a professor in MED/09 (Internal medicine)

across SDSs, it is quite clear that the authors who average more citations also publish in higher impact journals.

From Table 4, for the 232 SDSs, the highest average Pearson ρ coefficients are in “Electrical converters, machines and switches” (ING-IND/32), then “Blood diseases” (MED/15) and “Electrical energy systems” (ING-IND/33); the lowest in “Physics for earth and atmospheric sciences” (FIS/06), “Nuclear plants” (ING-IND/19), and “Plastic surgery” (MED/19).

Table 5 shows the descriptive statistics for the entire dataset, disaggregated by ADU. The average values of the correlation indices range between 0.264 (UDA

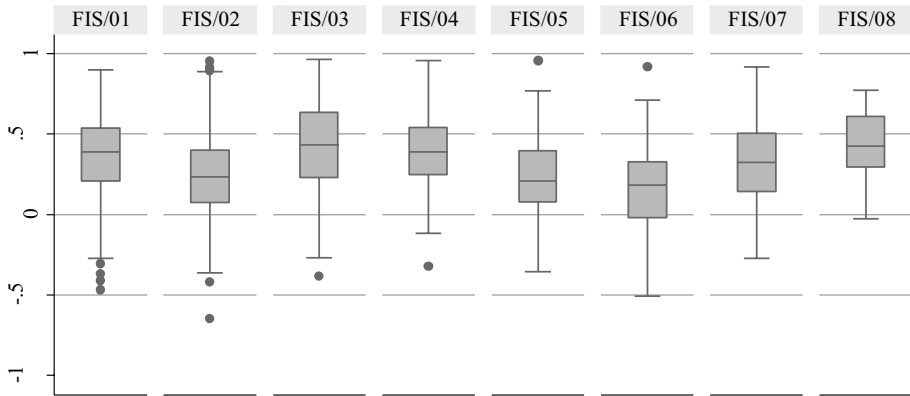


Fig. 8 Boxplot, AII vs JII Pearson correlation for professors in SDSs of Physics

Table 3 Descriptive statistics, AII vs JII Pearson correlation for professors in SDSs of Physics

SDS*	Obs	with negative ρ	Mean	Std. dev.	Min	Max	[95% Conf. Interval]		Avg AII vs Avg JII ρ
FIS/01	398	32 (8.0%)	0.363	0.251	-0.467	0.899	0.339	0.388	0.503
FIS/02	177	22 (12.4%)	0.260	0.279	-0.647	0.952	0.219	0.302	0.372
FIS/03	285	15 (5.3%)	0.427	0.264	-0.385	0.965	0.396	0.458	0.461
FIS/04	62	5 (8.1%)	0.389	0.247	-0.322	0.957	0.326	0.452	0.357
FIS/05	100	13 (13.0%)	0.247	0.255	-0.354	0.957	0.196	0.297	0.378
FIS/06	46	14 (30.4%)	0.165	0.293	-0.508	0.917	0.078	0.252	0.469
FIS/07	201	22 (10.9%)	0.323	0.257	-0.273	0.917	0.287	0.358	0.414
FIS/08	5	1 (20.0%)	0.415	0.307	-0.028	0.773	0.033	0.796	0.893
Total	1274	124 (9.7%)	0.342	0.269	-0.647	0.965	0.327	0.357	0.459

*FIS/01, Experimental physics; FIS/02, Theoretical physics, Mathematical models and methods; FIS/03, Physics of matter; FIS/04, Nuclear and subnuclear physics; FIS/05, Astronomy and astrophysics; FIS/06, Physics for earth and atmospheric sciences; FIS/07, Applied physics; FIS/08, Didactics and history of physics

1—Mathematics) and 0.382 (UDA 3—Chemistry). Mathematics presents distinctive characteristics: it is the UDA with the highest dispersion (std dev. 0.294), and the highest number of observations with negative correlation: for 249 out of total 1288 professors (19.3%). Cases of negative correlation are seen through all the UDAs, and overall concern 9% of the dataset. The maximum values of the correlation indices are never less than 0.94.

With reference to RQ2, on the AII-JII correlation for researchers within each field, we can certainly state that variation is considerable. While average values show a moderate relationship between AII and JII, the point values shift significantly from one author to another. In practically all SDSs, there is at least one professor with a decidedly negative correlation and at least one with a very strong positive correlation.

Table 4 Top and bottom five SDSs by AII vs JII Pearson ρ correlation

SDS	Obs ^a	UDA ^b	Avg Pearson ρ
ING-IND/32-Electrical converters, machines and switches	57	9	0.553
MED/15-Blood diseases	134	6	0.551
ING-IND/33-Electrical energy systems	65	9	0.512
AGR/07-Agrarian genetics	64	7	0.484
MED/25-Psychiatry	94	6	0.462
...			
M-PSI/07-Dynamic psychology	26	10	0.213
ING-IND/03-Flight mechanics	12	9	0.175
FIS/06-Physics for earth and atmospheric sciences	46	2	0.165
ING-IND/19-Nuclear plants	21	9	0.165
MED/19-Plastic surgery	53	6	0.122

^aWe list here only SDSs with at least 10 observations

^b2, Physics; 6, Medicine; 7, Agricultural and veterinary sciences; 9, Industrial and information engineering; 10, Psychology

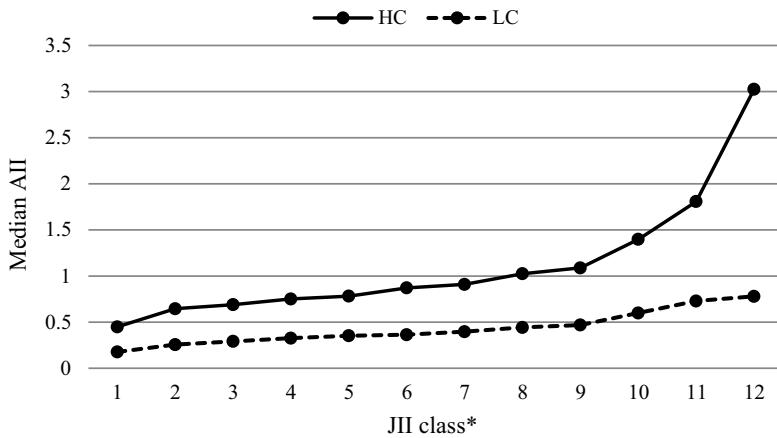
Table 5 Descriptive statistics, AII vs JII Pearson correlation for professors in the dataset, by UDA

UDA*	Obs	With negative ρ	Mean	Std dev.	Min	Max	[95% Conf. Interval]	
1	1288	249 (19.3%)	0.264	0.294	-0.599	0.948	0.248	0.280
2	1274	124 (9.7%)	0.342	0.269	-0.647	0.965	0.327	0.357
3	2037	106 (5.2%)	0.382	0.234	-0.586	0.983	0.372	0.392
4	620	51 (8.2%)	0.336	0.241	-0.555	0.940	0.317	0.355
5	2974	272 (9.1%)	0.347	0.259	-0.667	0.994	0.337	0.356
6	5258	422 (8.0%)	0.365	0.252	-0.536	0.986	0.358	0.371
7	1837	107 (5.8%)	0.369	0.226	-0.586	0.957	0.359	0.379
8	777	85 (10.9%)	0.294	0.257	-0.589	0.955	0.276	0.312
9	2960	224 (7.6%)	0.356	0.243	-0.697	0.979	0.347	0.365
10	600	82 (13.7%)	0.303	0.265	-0.553	0.943	0.282	0.324
11	718	103 (14.3%)	0.316	0.293	-0.637	0.972	0.294	0.337
Total	20,343	1825 (9.0%)	0.348	0.255	-0.697	0.994	0.344	0.351

*1, Mathematics and computer science; 2, Physics; 3, Chemistry; 4, Earth sciences; 5, Biology; 6, Medicine; 7, Agricultural and veterinary sciences; 8, Civil engineering; 9, Industrial and information engineering; 10, Psychology; 11, Economics

To answer RQ3, we now classify professors according to the total impact of their scientific production, measured as the sum of AII for all of their publications: professors in the top 10% by total impact in their SDS are tagged as “highly cited” (HC); those in the bottom 10% as “lowly cited” (LC).

For these two groups, we plot the median AII of their publications against “classes” of JII (Fig. 9). The data plot very clearly indicates a positive trend for citedness of a paper as a function of impact of the hosting journal. However, the curve for HCs is systematically above the curve for LCs, and also for high levels of journal impact (95th



* 1 = below 50th percentile; 2 = 50-55th; 3 = 55-60th; ... ; 11 = 95-98th; 12 = above 98th percentile

Fig. 9 Relation between journal impact (JII) and median article citedness (AII): comparison between highly cited (HC) and lowly cited (LC) authors

percentile and above), the difference in citedness for the former is significantly higher than for the latter.

Discussion and conclusions

Starting from the seminal works by Seglen (1989, 1994, 1997) and the update by Zhang et al. (2017), this work delves into questions of the correlation between citedness of a publication and the impact of the host journal, extending the analysis to all STEM fields in the Italian academic system. The analysis confirms the results of Zhang et al. (2017), but now also in fields beyond biomedical sciences. The data on the low correlation between the two indicators, however, tends to differ between fields: with Pearson ρ distribution ranging between 0.107 and 0.594, in any case quite concentrated around the central value (mean 0.334, st. dev. = 0.097).

Adopting individual publications as the unit of analysis, the correlation between citedness and journal impact is weak overall (Pearson $\rho = 0.308$; 0.345 excluding outliers). It is notable that the number of publications with little or no citation, even though hosted by journals in D1 for impact factor, is ten times that of publications that rank top-10% for citedness, but are hosted in journals with impact factor in the bottom 10%.

Taking professors as the unit of analysis, the correlation is extremely variable both between and within fields, in some SDSs spanning as much as -0.7 to +1. Aggregating professors and their relative publications by discipline, we see cases of negative AII-JII correlation in all the UDAs, overall affecting 9% of the professors.

Finally, the analyses confirm that the correlation between citedness and prestige of the hosting journal, although not strong, is certainly more significant for highly cited authors than lowly cited ones: i.e. although the overall data trend shows a positive correlation between citedness and prestige of the journal, the correlation is higher for the specific population of highly cited authors.

This difference in citedness, given the same JII, would evidently be attributable more to the quality of the authors and their articles than to the presence of a hypothetical “journal effect”. Interpretations that seek a causal effect of journal impact factors on citation impact could prove problematic. Indeed, as noted by Bornmann and Leydesdorff (2017) if a publication is of high quality, it could be accepted for publication in a journal with high JII and receive a high number of citations. In this case, the correlation between JII and AII is spurious, because the relationship between the two variables is mediated by a third variable, which in this case is the publication quality. In essence, it is not the journals that are cited, rather the quality articles they publish. This led Zhang et al. (2017) to observe that: “It is the highly cited authors that provide these articles, perhaps by being conscious about where they publish their more significant results”. In this case, one could hypothesize the presence of a self-selection bias: the highly cited authors would be more aware of their own value and inclined to submit their research products to journals with higher JII. The opposite could hold true for lowly cited authors.

Furthermore, the continued occurrence of a significant citedness gap between highly cited and lowly cited authors, even in journals with very high JII, leads to the hypothesis of phenomena of “scientist stratification”, i.e. a manifestation of the Matthew effect. For the lowly cited authors, even when they manage to publish in a high-impact journal, this does not seem to earn them premiums of citedness, although we could certainly imagine some gains in prestige among their scientific community.

The empirical evidence supports the conclusion that the most important criterion in seeking a journal for the publication of our manuscripts is the fit between the article content and the expectations of the journal’s target audience. Among the journals addressing the relevant “scientific public”, if we wish, we can then aim for the ones of greater prestige.

Author contributions GA (Corresponding Author): Conceptualization; Methodology; Supervision; Validation; Writing—Review & Editing. CAD: Methodology; Data curation; Formal analysis; Investigation; Writing—original draft. FDC: Data curation; Formal analysis; Investigation; Visualization; Writing—original draft.

Funding The research project received no funding by third parties.

Data availability Web of Science raw data used in this study have been made available under license by Clarivate Analytics. The authors are not allowed to redistribute WoS data which therefore cannot be made available.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the manuscript. Giovanni Abramo and Ciriaco Andrea D’Angelo are members of the Editorial Board of the journal.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2010). Citations versus journal impact factor as proxy of quality: Could the latter ever be preferable? *Scientometrics*, *84*(3), 821–833. <https://doi.org/10.1007/s11192-010-0200-1>
- Abramo, G., D'Angelo, C. A., & Felici, G. (2019). Predicting long-term publication impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, *13*(1), 32–49. <https://doi.org/10.1016/j.joi.2018.11.003>
- Antonoyiannakis, M. (2020). Impact factor volatility due to a single paper: A comprehensive analysis. *Quantitative Science Studies*, *1*(2), 639–663.
- Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics*, *79*(3), 639–653.
- Bonaccorsi, A. (2020). Two decades of research assessment in Italy. *Addressing the criticisms. Scholarly Assessment Reports*, *2*(1), 17. <https://doi.org/10.29024/sar.28>
- Bornmann, L., & Leydesdorff, L. (2017). Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data. *Journal of Informetrics*, *11*(1), 164–175.
- Brito, R., & Rodríguez-Navarro, A. (2019). Evaluating research and researchers by the journal impact factor: Is it better than coin flipping? *Journal of Informetrics*, *13*(1), 314–324.
- Callaway, E. (2016). Publishing elite turns against impact factor. *Nature*, *535*(7611), 210–211.
- Ciechanover, A. (2013). A conversation with Aaron Ciechanover. Interview by Ushma S. Neill. *The Journal of Clinical Investigation*, *123*(10), 4093–4094. <https://doi.org/10.1172/JCI171859>
- Consiglio Nazionale delle Ricerche. (2019). *Relazione sulla ricerca e l'innovazione in Italia*. Cnr edizioni.
- D'Angelo, C. A., Giuffrida, C., & Abramo, G. (2011). A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. *Journal of the American Society for Information Science and Technology*, *62*(2), 257–269.
- de Oliveira Silva, D., Taborda, B., Pazzinato, M. F., Ardern, C. L., & Barton, C. J. (2021). The altmetric score has a stronger relationship with article citations than journal impact factor and open access status: A cross-sectional analysis of 4022 sport sciences articles. *Journal of Orthopaedic and Sports Physical Therapy*, *51*(11), 536–541. <https://doi.org/10.2519/jospt.2021.10598>
- De Rijcke, S., Wouters, P., Rushforth, A., Franssen, T., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, *25*(2), 161–169.
- DORA. (2012). *San Francisco declaration of research assessment*. Retrieved November 30, 2022, from <https://sfedora.org/read/>
- Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Elsevier.
- Garfield, E. (2001). Impact factors, and why they won't go away. *Nature*, *411*(6837), 522. <https://doi.org/10.1038/35079156>
- Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193.
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431.
- Jarwal, S. D., Brion, A. M., & King, M. L. (2009). Measuring research quality using the journal impact factor, citations and 'Ranked Journals': Blunt instruments or inspired metrics? *Journal of Higher Education Policy and Management*, *31*(4), 289–300.
- Kiesslich, T., Beyreis, M., Zimmermann, G., & Traweger, A. (2021). Citation inequality and the journal impact factor: Median, mean, (does it) matter? *Scientometrics*, *126*(2), 1249–1269.
- Kim, L., Portenoy, J. H., West, J. D., & Stovel, K. W. (2020). Scientific journals still matter in the era of academic search engines and preprint archives. *Journal of the Association for Information Science and Technology*, *71*(10), 1218–1226.
- Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S., & Curry, S. (2016). A simple proposal for the publication of journal citation distributions. *BioRxiv*. <https://doi.org/10.1101/062109>
- Larivière, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 3–24). Springer.
- Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journal evaluation. *Journal of the American Association for Information Science and Technology*, *59*(2), 278–287.
- Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology*, *63*(11), 2140–2145.

- Marx, W., & Bornmann, L. (2013). Journal impact factor: “The poor mans’ citation analysis” and alternative approaches. *European Science Editing*, 39(3), 62–63.
- Milojević, S., Radicchi, F., & Bar-Ilan, J. (2017). Citation success index—An intuitive pair-wise journal comparison metric. *Journal of Informetrics*, 11(1), 223–231.
- Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, 246(1), 1–19.
- Moed, H. F. (2020). Appropriate use of metrics in research assessment of autonomous academic institutions. *Scholarly Assessment Reports*, 2(1), 1.
- Moed, H. F., & van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature*, 381(6579), 186–186.
- Paulus, F. M., Cruz, N., & Krach, S. (2018). The impact factor fallacy. *Frontiers in Psychology*, 9, 1487.
- Radner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B Condensed Matter and Complex Systems*, 4(2), 131–134.
- Rushforth, A., & De Rijcke, S. (2015). Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva*, 53(2), 117–139.
- Seglen, P. O. (1989). From bad to worse: Evaluation by journal impact. *Trends in Biochemical Sciences*, 14(8), 326–327. [https://doi.org/10.1016/0968-0004\(89\)90163-1](https://doi.org/10.1016/0968-0004(89)90163-1)
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628–638.
- Seglen, P. O. (1994). Causal relationship between article citedness and journal impact. *Journal of the American Society for Information Science*, 45(1), 1–11.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079), 498–502.
- Traag, V. A. (2021). Inferring the causal effect of journals on citations. *Quantitative Science Studies*, 2(2), 496–504.
- Van Leeuwen, T. N., & Moed, H. F. (2005). Characteristics of journal impact factors. The effects of uncitedness and citation distribution on the understanding of journal impact factors. *Scientometrics*, 63(2), 357–371.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>
- Waltman, L., & Traag, V. A. (2021). Use of the journal impact factor for assessing individual articles: Statistically flawed or not? *F1000Research*, 9, 366.
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., & Franssen, T. (2015). *The metric tide: Literature review (Supplementary Report I to the independent review of the role of metrics in research assessment and management)*. HEFCE.
- Zhang, L., Rousseau, R., & Sivertsen, G. (2017). Science deserves to be judged by its contents, not by its wrapping: Revisiting Seglen’s work on journal impact and research evaluation. *PLoS ONE*, 12(3), e0174205.