 doi:10.20944/preprints202104.0448.v1

*Article*

# Did you Ask for Citations? An Insight into Preprint Citations *en route* to Open Science

**Mario Pagliaro**

Istituto per lo Studio dei Materiali Nanostrutturati, CNR, via U. La Malfa 153, 90146 Palermo, Italy; mario.pagliaro@cnr.it

**Abstract** Preprints are regularly cited in peer reviewed journal articles, books and conference paper. Are preprint citations somehow less important than citations to peer reviewed research papers? This study investigates citation patterns between 2017 and 2020 for preprints published in three preprint servers, one specializing in biology (bioRxiv), one in chemistry (ChemRxiv), and another hosting preprints in all disciplines (Research Square). As evaluation of scholarship continues to largely rely on citation-based metrics, this analysis and its outcomes will be useful to inform new research-based education in today's scholarly communication.

**Keywords:** preprint; citations; scholarly communication; open science; peer review; impact factor

## 1. Introduction

Following the invention of the world wide web in 1989, in the 1990s evidence started to emerge that the internet would have reshaped scholarly communication. Research papers would no longer be printed in paper journals but rather published in digital format on the web after peer review, or even before review of anonymous referees in the form of "preprints". Particle ("high-energy" in physics jargon) physicists were used to circulate paper-copy preprints since 1969 when a database management system aggregating preprints circulated between different institutions was established in Stanford, an urban area in California hosting a large particle accelerator. In 1991 the Stanford Physics Information Retrieval System became the first database accessible through the web [1]. In 2003 Harnad forecasted that "once the preprints and postprints of all 2 million articles appearing annually in the world's 20,000 peer-reviewed journals are openly accessible, research progress will become much more rapid and interactive" with "every article hyperlinked directly to each article it cites, and new scientometric search engines providing rich new measures and predictors of research usage, direction and impact" [2].

Nearly two decades later, the number of preprints uploaded online in 2019 nearly reached 227,000. Growth since 1991 was exponential with doubling in less than 10 years [3]. Still, the ratio of preprints to all scientific articles published in the first nine months of 2020 was only 6.4%, with only three basic science disciplines showing significant uptake: physics (close to 35% preprints to all-papers published in 2019), mathematics (about 30% in 2019), and biology (6.5% in 2020) [3].

Hosting close to 1,800,000 preprints by late 2020, the world's largest preprint server (arXiv), managed by the Library of Cornell University at https://arxiv.org, in 2020 reached a 14,861/month publication rate from 12,989/month in 2019 (+14%) [4]. For comparison, the number of papers published by the main biology preprint server (bioRxiv), a preprint repository managed by the USA-based Cold Springer Harbor Laboratory at https://biorxiv.org, in 2020 reached an average publication rate of 3,193/month [5].

From Advance to Zenodo, as of April 2021 a list of 68 preprint servers were found in the directory made openly accessible on the internet by Rittman [6], a manager at

 doi:10.20944/preprints202104.0448.v1



CrossRef, a not-for-profit organization based in the USA for scholarly publishing. For comparison, another comprehensive list of preprint servers in early 2020 found 57 preprint servers [7]. These online repositories host preprints either in specific research fields (EarthArXiv, bioRxiv, etc.) or in all scientific disciplines (SSRN, Preprints, Authorea, etc.).

"Not having been subject to peer review", reads a reputed manual for authors of research papers written in English published in 2017, "preprints are treated as unpublished material" [8]. Fifteen years before Perelman, a world's leading mathematician, published in arXiv the first of three papers appeared only in preprint form between 2002 and July 2003 in which he provided a proof of the geometrization conjecture including the Poincaré conjecture. Only the first paper [9] by early April 2021 has been cited 2,503 times [10].

This indicates that, starting from physicists and mathematicians (the original contributors to arXiv), the scientific community has never dealt with preprints as "unpublished material", but rather as scientific articles which are read, studied and cited, even though not having gone through the peer review process. Accordingly, preprints are regularly cited in peer reviewed journal articles, books and conference papers and presentations. For example, articles posted in arXiv between its launch on August 1991 and 2016 received 135,782 citations, of which 23,288 were received only in 2016 [11].

This study investigates citation patterns between 2017 and 2020 for preprints published in two specialized preprint servers publishing research articles in chemistry (ChemRxiv) and in biology (bioRxiv), and between 2018 and 2020 in one multidisciplinary preprint repository (Research Square). Previous studies have investigated citations for preprints in arXiv [11,12,13,14]. Studying 1,495 mathematics cited preprints posted in arXiv, scientometrics scholars have lately unveiled three important findings: *i*) 71.8% of cited preprints are cited *before* journal publication; *ii*) about 50% of all preprint citations are received within 24 months since publication, after which the average citation rate of preprints decreases due to journal online publication (authors prefer to cite the journal version rather than the preprint when both are available); and, *iii*) the preprint version and the journal version of the same papers have different readerships with the preprint server reaching a much *broader* readership as shown by the fact that tor the preprint versions, 27.5% of the total citations (vs. 12.5% for the journal versions) originates from papers assigned to another discipline beyond mathematics (physics) [13].

As evaluation of scholarship continues to rely on citation-based metrics [15], this analysis and its outcomes will be useful to inform new research-based education in today's scholarly communication [16].

## 2. Methods

An online search for total citations of research papers published in ChemRxiv and bioRxiv between 2017 and 2020 was conducted in Scopus on April 7, 2021. A similar search was carried out in Dimensions for papers posted in Research Square between 2018 and 2020. Launched in 2004 by Elsevier, Scopus (https://www.scopus.com) is a scholarly database indexing scientific journals, books, and conference proceedings that by late 2019 included 23,452 active journal titles, about 120,000 conferences and 206,000 books from more than 5,000 publishers, adding some 3 million records every year [17]. Launched in 2018, Dimensions (https://www.dimensions.ai) is the most comprehensive bibliographics database, with 109 million publications indexed and about 1.1 billion citations as of September 2019. As of May 2020 Dimensions indexed more than 74,000 journals [17]. Data on cumulative preprints were obtained from different sources: [18] for biorXiv, [19] for ChemRxiv, and [20] for Research Square.

## 3. Results

Results in Table 1 show the large difference in the uptake of preprints between chemists and biologists. By the end of 2020 bioRxiv, launched on November 2013, hosted

 doi:10.20944/preprints202104.0448.v1



107,518 preprints whereas ChemRxiv, launched on August 2017, hosted 6,127 preprints. Amid scholars in the basic sciences, indeed, chemists are those with the lowest uptake of preprint [21] as well as of open access (OA) [22] publishing models.

**Table 1**. Preprint citations and number of preprints of selected preprint servers, 2017-2020.

| Preprint server | Year | Citations[a,b] | Cumulative preprints[c] |
|---|---|---|---|
| bioRxiv | 2020 | 23,820 | 107,518 |
| | 2019 | 11,280 | 68,801 |
| | 2018 | 5,880 | 39,620 |
| | 2017 | 2,643 | 18,837 |
| ChemRxiv | 2020 | 1,432 | 6,127 |
| | 2019 | 437 | 2,289 |
| | 2018 | 85 | 1,053 |
| | 2017 | 7 | 350 |
| Research Square | 2020 | 3,208 | 38,448 |
| | 2019 | 45 | 5,666 |
| | 2018 | 8 | 2 |
| | 2017 | - | - |

[a]Source: Scopus, 2021; [b]Source: Dimensions, 2021; [c]Source: Ref.s 18, 19 and 20

In one year only between 2019 and 2020 Research Square went from 5,666 preprints hosted in 2019, to 38,448 in 2020 (+580%). Since then, as the time of writing (April 2021), the website added another 32,251 preprints approaching the 80,000 (77,641) threshold [23]. Growth was driven by preprints dealing with COVID-19 which led Research Square to rapidly become one of the top three preprint servers by volume for research related to the disease [24]. Researchers likely opted to publish findings and reviews on the disease in a platform whose majority owner (Springer Nature) publishes some of the leading medicine journals. The platform, however, publishes preprints in all scientific disciplines including language studies, philosophy and law, with submissions rapidly increasing in many other fields. By April 2021, for example, the server hosted 636 preprints in the chemical sciences.

The list of the top five journals citing preprints in bioRxiv and ChemRxiv (Table 2) shows that the publications most frequently citing preprints posted in these repositories are reputed multidisciplinary or specialized chemistry and biology journals. Preprints in bioRxiv and ChemRxiv are chiefly cited in journal original research and review articles, accounting in both cases for >93% of the citations. Preprints in bioRxiv are cited more three times more frequently in conference papers when compared to preprints in ChemRxiv.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2021 doi:10.20944/preprints202104.0448.v1



**Table 2**. Top five journals citing preprints in bioRxiv and ChemRxiv, and type of citing documents.[a]

| Rank | Journal | Citations | Document type | Documents citing (share) |
|---|---|---|---|---|
| 1 | *Scientific Reports* | 1,116 | Article | 37,694 (70%) |
| 2 | *eLife* | 954 | Review | 12,567 (23.5%) |
| 3 | *Plos ONE* | 923 | Conference paper | 1,730 (3.2%) |
| 4 | *Nature Communications* | 787 | Book chapter | 939 (1.75%) |
| 5 | *International Journal of Molecular Sciences* | 685 | Note | 584 (1.1%) |
| 1 | *Organic Letters* | 90 | Article | 1,788 (71.6%) |
| 2 | *Journal of the American Chemical Society* | 70 | Review | 616 (24.7%) |
| 3 | *Angewandte Chemie International Edition* | 66 | Book chapter | 35 (1.4%) |
| 4 | *Journal of Biomolecular Structure and Dynamics* | 45 | Conference paper | 33 (1.3%) |
| 5 | *Journal of Chemical Information and Modeling* | 44 | Note | 26 (1%) |

[a]Source: Scopus, 2021

Contrary to preprints in chemistry posted at ChemRxiv, preprints in biology posted at bioRxiv and Research Square are chiefly cited by OA journals. This is not surprising, considering that chemists are the researchers with the lowest fraction of published OA papers [21]. On the other hand, the top three journals citing ChemRxiv preprints (*Organic Letters*, *Journal of the American Chemical Society* and *Angewandte Chemie International Edition*) are amid the most reputed chemistry journals. The remaining two (*Journal of Biomolecular Structure and Dynamics* and *Journal of Chemical Information and Modeling*) are renown structural chemistry journals.

It is relevant, that the second subject area citing preprints in ChemRxiv (Table 3) is "Biochemistry, Genetics and Molecular Biology", which became in the course of 2020 the main subject area of preprints posted in bioRxiv. Until late 2018, the main subject area for preprints posted on bioRixv was neuroscience which by October 2018 had become the first of bioRxiv collection to cross the 6,000 preprint threshold [18].

**Table 3**. Top five subject areas citing preprints in bioRxiv and ChemRxiv.[a]

| Rank | Subject area | Preprints |
|---|---|---|
| | bioRxiv | |
| 1 | Biochemistry, Genetics and Molecular Biology | 26,488 |
| 2 | Medicine | 16,875 |
| 3 | Agricultural and Biological Sciences | 11,021 |
| 4 | Immunology and Microbiology | 8,119 |
| 5 | Neuroscience | 7.710 |
| | ChemRxiv | |
| 1 | Chemistry | 1,421 |
| 2 | Biochemistry, Genetics and Molecular Biology | 780 |
| 3 | Chemical Engineering | 595 |
| 4 | Materials Science | 438 |
| 5 | Medicine | 326 |

[a]Source: Scopus, 2021

These findings confirm that also in the case of chemistry and biology, preprints reach a much broader readership. Preprints in bioRxiv are vastly cited, for instance, by researchers in agricultural sciences, whereas preprints in ChemRxiv are twice more frequently cited by researchers in biochemistry than by scientists working in materials science (Table 3).

Table 4 shows that preprints in the field of medical microbiology posted at Research Square are cited at significantly higher rate when compared, for instance, to preprints in oncology (entry 5 in Table 4) [20]. There are no clear trends. For example, by early April 2021 the 101 preprints in analytical chemistry (not shown in Table 4) have received 20 citations, with a high (0.2) citation/preprint ratio, whereas the 37 preprints in nanotechnology on Research Square by early April 2021 had never been cited.

**Table 4**. Top 10 research fields and number of citations for preprints in Research Square as of April 7, 2021.

| Entry | Research category | Preprints | Citations | Citation/preprint |
|---|---|---|---|---|
| 1 | Medical and Health Sciences | 33,396 | 3,546 | 0.11 |
| 2 | Clinical Sciences | 12,024 | 1,062 | 0.09 |
| 3 | Public Health and Health Services | 11,930 | 1,345 | 0.11 |
| 4 | Biological Sciences | 7,123 | 690 | 0.10 |
| 5 | Oncology and Carcinogenesis | 4,740 | 130 | 0.03 |
| 6 | Genetics | 3,941 | 233 | 0.06 |
| 7 | Biochemistry and Cell Biology | 2,375 | 397 | 0.17 |
| 8 | Cardiorespiratory Medicine and Haematology | 1,880 | 153 | 0.08 |
| 9 | Medical Microbiology | 1,653 | 681 | 0.41 |
| 10 | Neurosciences | 1,499 | 95 | 0.06 |

In general, between 2019 and 2020 the growth in citation rate for preprints at Research Square (+7,000%) has been one order of magnitude higher than the increase in number of preprint published (+580%). In detail, citations went from 45 (for 5,666 preprints) in 2019 to 3,208 in 2020 (for 38,848 preprints) in 2020 [20].

## 4. Discussion

A citation to a preprint server is a citation that will not be counted by the scholarly database used by the commercial USA-based company (Clarivate Analytics) publishing every year the journal impact factor (JIF) of indexed journals. This may explain the “concern” found in the 2018 report of the international scientific publishers association that:

> “preprints (which can be brought up to date) may become a go-to place for the version of record, undermining publisher business models” as well as “concerns… over the loss of citations from journals to preprints servers, with well over 8,000 citations to bioRxiv reported on Web of Science” [25].

Being the outcome of an highly skewed distribution for which typically 15% of the papers in a journal account for *half* the total citations [26], the impact factor is a misleading statistical indicator because the vast majority (typically 85%) of the journal’s articles will actually get *fewer* citations than indicated by the JIF. Due to this simple fact, namely that the JIF is a statistically poor indicator, it should not be used to evaluate research [27]. As noted by Curry in 2012, “if you are judging grant or promotion applications and find yourself scanning the applicant’s publications, checking off the impact factors, you are statistically illiterate” [28].

In a seminal study published three years later investigating the fundamental cause of the ongoing “impact factor mania”, Casavedell and Fang unveiled that:

> “The impact factor mania persists because it confers significant benefits to individual scientists and journals. Impact factor mania is a variation of the economic theory known as the ‘tragedy of the commons,’ in which scientists act

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2021 doi:10.20944/preprints202104.0448.v1



> rationally in their own self-interests despite the detrimental consequences of their actions on the overall scientific enterprise [29]".

This in its turn explains why "elite" (*i.e.*, high JIF) journals do not expand to accommodate all meritorious articles, continuing in what Casavedell and Fang call "artificial restrictions on journal size serve to perpetuate the current wasteful system that requires authors to cascade serial submissions from one journal to another" [29]. Put simply, by expanding the journal size beyond a certain threshold, the denominator in the equation affording the JIF would rapidly become too large and the JIF would dramatically drop regardless of increase in citations in the JIF equation's numerator. This is what happened, for example, in the case of so-called "mega journals" (large-volume OA journals accepting also replication studies and negative results) whose JIF rapidly decreased a few years since inception [30], due precisely to the skewedness of citation distribution identified by Seglen in 1992 [26].

As put it by Davis, a citation "is the basis for a system of rewarding those who make significant contributions to public science" [31]. This is the original idea that led Garfield to introduce the JIF in 1955 [32]: scientific quality is associated with citations from peers. Hence, there is no logical basis for promotion and tenure committee, or for a research funding agency, to evaluate differently a citation to a preprint or to a peer reviewed paper. Accordingly, the world's largest research funding agency (the National Institutes of Health, USA) as early as of 2017 recommended applicants and awardees to include in applications, proposals and reports the citation of any "interim research product" [33].

In brief, purposeful evaluation of scholarship today includes also the evaluation of preprints [34], and will take into account the number of preprint citations. I also recommend scholars to include next to the number of preprint citations also the value of altmetrics indicators (such as those provided by Altmetric.com, Mendeley and PlumX) which track and report altmetrics data measuring the online impact of the preprint's research [35].

## 5. Outlook and Conclusions

Along with enhanced visibility, collaboration, professional and funding opportunities [36], the benefits of preprints for scholars include an higher number of citations directly in terms of preprint citations. If the preprint is subsequently published in a peer reviewed journal, the journal article can be merged with the preprint, allowing the database to sum the citations (easily done for instance on Google Scholar by selecting the preprint and the journal article and choosing "Merge" from the "Actions" menu).

To win skepticism amid young scholars under the "publish-or-perish" pressure [37], I recommend in teaching open science the use of practical examples as case studies. For instance, in 2018 Kievit and co-workers published in *PeerJ Preprints* a preprint [38] on raincloud plots (a new data visualization tool providing bias-free statistical information while preserving the desired inference at a glance nature of other plot visualization devices). In the second version of preprint posted in *Wellcome Open Research* eight months later, the team inserted a new section illustrating the benefits derived from posting a preprint, including 18 citations, which is worth reading:

> "Firstly, posting the manuscript as preprint has vastly widened the reach. To date (March 2019) our preprint was viewed 9803 times, with 6,309 downloads. However, views and downloads alone don't necessarily entail engagement. Since publication the preprint alone has already been cited 18 times. Moreover, in depth engagement has gone well beyond mere citations. Several individuals have created their own useful tutorials, summarizing our paper and asking useful questions, posted constructive criticism, discussed raincloud plots as part of various plotting alternatives, created a shiny app, wrote an accessible tutorial using native R datasets, a

 doi:10.20944/preprints202104.0448.v1



> new package, creating various animated interactive visualisations... Our codebase itself received feedback through various avenues including formal pull requests on github, comments on the preprint, twitter replies and email. In this new version of our paper we have tried our best to integrate all these suggestions and comments, which without fail have improved the usability of our code.
>
> "Social media, specifically twitter, provided the central hub where all these benefits coalesced. The paper has been tweeted at least 750 times, with an estimated reach of up to 1,500,000 total followers, and as such is the principal driver for the engagement our preprint has received. This engagement has yielded invaluable feedback, comments, and suggestions, and were even lucky enough to track down the first instance of an early precursor of the raincloud plot (Ellison, 2018). Moreover, the paper itself was inspired by a twitter discussion, and brings together co-authors who have never met in person. Together, these interactions illustrate the fundamentally two-way street of new publishing models, which facilitate access without paywalls and allow for near instantaneous improvements to ongoing work." [39]

Making possible the aforementioned "two-way street of new publishing models" with nearly immediate "improvements to ongoing work", the preprint allows to reap the benefits of open science enhancing the number of citations and related citation-based metrics which continue to be used for appraising scientists [15,34].

It is enough to review the citations of preprints posted in arXiv to realize that physicists and mathematicians never made a distinction between citation of a study deposited in arXiv or citation of a study in peer reviewed journal. In other words, they never added the "preprint" word in the reference as required by certain research funding agencies and by certain journals. Scholars in physics know for instance that reputed journals such as *Physical Review* until 1960 used peer review only for half of the papers received, and even in that case peer review consisted in the editor asking to one referee an opinion on a manuscript for which the editor needed advice [40]. From Krebs' 1937 work on the citric acid cycle rejected by *Nature* and published in *Experientia*, to Ernst's 1966 work on nuclear magnetic resonance spectroscopy rejected twice by the *Journal of Chemical Physics* to be finally published in the *Review of Scientific Instruments*, up to Mullis' polymerase chain reaction discovery rejected by *Science* and published in *Methods in Enzymology*, the fact that several discoveries leading to major scientific progress (and to Noble Prizes) were actually rejected shows evidence that the peer review system has significantly delayed scientific progress and perhaps also *suppressed* it [41].

Finding that nearly 50% of all research papers published in 2020 were openly accessible, Dudley has recently unveiled the main drivers and obstacles to open access publishing [42]. The causal loop diagram in Figure 1 allows to visualize how the dominance of the pay-to-publish OA model and its reinforcing factors led to mandates for OA reform which continued to focus on promoting OA via pay-to-publish. The latter model has built a financial barrier which in its turn created demand for sub-standard journals from authors in low income countries, reinforcing the dichotomy between scholars in wealthier and those in poorer countries. Indeed, in 2017 the largest number of researchers who published in "predatory" journals were from India, Nigeria, Turkey, and other economically developing countries [43].

The outcome of this situation is that in wealthier countries the article processing charge (APC) of "elite" OA journals exceeds $3,000 and at times even the $5,000 threshold (Table 5), while in poorer countries numerous sub-standard ("predatory") OA journals emerged charging low APCs.

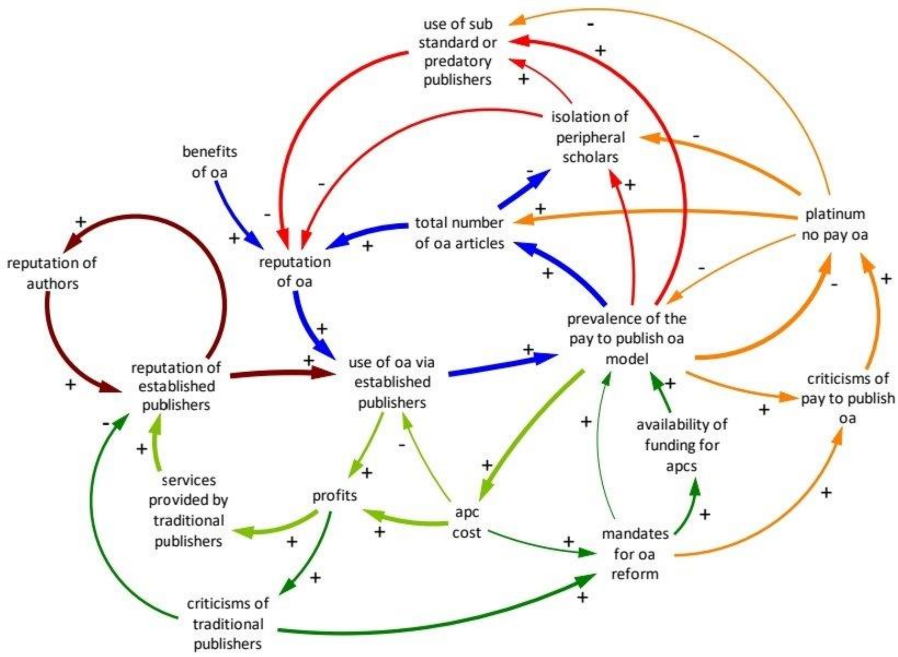


**Figure 1**. Dudley's closed loop diagram describing the current (early 2021) situation in the scientific publishing system. Arrows (causal links), indicate how a change in the causal variable affects change in the second variable. Change in the same direction is indicated with a plus sign. Change in the opposite direction, with a minus sign. [Reproduced from Ref.42, Creative Commons CC BY 4.0 license]

**Table 5**. APC for selected elite scientific journals as of April 2021.

| **Journal** | **Publisher** | **APC** (USD) |
|---|---|---|
| *Nature Communications* | Springer Nature | 5,560 |
| *Advanced Science* | Wiley | 5,000 |
| *JACS Au*[a] | American Chemical Society Publishing | 5,000 |
| *Science Advances* | American Association for the Advancement of Science | 4,500 |
| *PLOS Biology*[b] | Public Library of Science | 4,000 |
| *eLife* | eLife Sciences Publications | 3,000 |
| *Cell Reports* | Elsevier | 5,200 |

[a]In case of (CC-BY) license, otherwise in case of (CC- BY-NC-ND) license, $4,000; [b]For discovery reports, $3,350

To escape this vicious cycle, scholars may first freely publish the outcomes of their work in preprint form, and then publish the preprint either in low cost or free ("platinum") OA journals or in paywalled journals taking care to "green" self-archive their papers [2]. On the other hand, Laakso in 2013 found that in 2010 a share exceeding 80% (a figure approaching to 1 million out of 1.1 million articles investigated) of all articles could be self-archived after 12 months of publication [44]. Yet, only 12% scholarly articles were found to be actually self-archived, with scholars in certain disciplines such as chemistry and chemical engineering not even reaching 10% [45]. Nearly ten years later, extending the analysis to all articles indexed in the Web of Science, a commercial bibliographic database, the share of green self-archived articles was found to be 4%, with another 7% made OA directly by journal editors (in a so called "bronze" model) [46]. This, once again, reveals the widespread need amid scholars in all disciplines for updated education of practical value in the field of open science [16].

In conclusion, gone are the days in which preprints, for example in bioRxiv, were highly read and shared online, but poorly cited [47]. Showing evidence that preprints are now regularly cited in peer reviewed journal articles, books and conference papers by investigating citation patterns for preprints published in ChemRxiv, Research Square and bioRxiv, this study further substantiates the value of open science also in relation to citation-based metrics on which evaluation of scholarship continues to rely on. As empirical evidence derived from analyzing measures of experimental and statistical rigor including data reported for crystallographic quality, effect sizes in gene-association studies, and use of statistics in neuroscience and psychology, reveals that "elite" (high JIF) scientific journals are those with the lowest reliability of published research [48], the outcomes of this study will further inform new research-based education in today's scholarly communication [16].

**Funding**: This research received no external funding

**Data Availability Statement**: All data are available by contacting the author

**Conflicts of Interest**: The author declares no conflict of interest

## References

1. A. Gentil-Beccot, S. Mele, T.C. Brooks, Citing and Reading Behaviours in HighEnergy Physics: How a Community Stopped Worrying about Journals and Learned to Love Repositories, *arXiv* **2009**, 0906.5418. http://arxiv.org/ftp/arxiv/papers/0906/0906.5418.pdf
2. S. Harnad, Electronic Preprints and Postprints In *Encyclopedia of Library and Information Science*, Marcel Dekker, New York: **2003**. http://www.ecs.soton.ac.uk/~harnad/Temp/eprints.htm.
3. B. Xie, Z. Shen, K. Wang, Is preprint the future of science? A thirty year journey of online preprint services, *arXiv* **2021**, 2102.09066. https://arxiv.org/abs/2102.09066
4. Cornell University, arXiv Submission Rate Statistics, **2021**. Available online: https://arxiv.org/help/stats/2020_by_area/index (accessed 6 April 2021).
5. Rxivist, Site-Wide Metrics. **2021**. Available online: https://rxivist.org/stats (accessed 6 April 2021).
6. M. Rittman, Research Preprints: Server List, **2021**. https://docs.google.com/spreadsheets/d/17RgfuQcGJHKSsSJwZZn0oiXAnimZu2sZsWp8Z6ZaYYo/ (accessed 6 April 2021).
7. ASAPBio, Directory of preprint server policies and practices, **2020.** Available online: https://asapbio.org/preprint-servers (accessed 6 April 2021).
8. *The Chicago Manual of Style*, 17th edition, The University of Chicago Press, **2017**; Section 14.173: Journal article preprints.
9. G. Perelman, The entropy formula for the Ricci flow and its geometric applications, *arXiv*, **2002**, math/0211159, https://arxiv.org/abs/math/0211159
10. Google Scholar, accessed at https://scholar.google.com/ on April 7, 2021.
11. A. Noruzi, ArXiv popularity from a citation analysis point of view, *Webology* **2016**, *13* (2), 22. http://www.webology.org/2016/v13n2/editorial22.pdf
12. A. Ferrer-Sapena, R. Aleixandre-Benavent, F. Peset, E.A. Sanchez-Perez, Citations to arXiv preprints by indexed journals and their impact on research evaluation, *J. Inf. Sci. Theory Pract.* **2018**, *6*, 6-16. https://doi.org/10.1633/JISTaP.2018.6.4.1
13. Z. Wang, Y. Chen, W. Glänzel, Preprints as accelerator of scholarly communication: An empirical analysis in Mathematics, *J. Informetr.* **2020**, *14*, 101097. https://doi.org/10.1016/j.joi.2020.101097
14. V.A. Traag, Inferring the causal effect of journals on citations, *Quant. Sci. Stud.* **2021**, https://doi.org/10.1162/qss_a_00128
15. J.P.A. Ioannidis, K.W. Boyack, Citation metrics for appraising scientists: misuse, gaming and proper use, *Med. J. Aust.* **2020**, *212*, 247-249. https://doi.org/10.5694/mja2.50493
16. M. Pagliaro, Publishing scientific articles in the digital era, *Open Sci. J.* **2020**, *5*, 3. https://doi.org/10.23954/osj.v5i3.2617
17. V.K. Singh, P. Singh, M. Karmakar, J. Leta, P. Mayr, The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis, *Scientometrics* **2021**, https://doi.org/10.1007/s11192-021-03948-5
18. http://api.biorxiv.org/reports/content_summary (accessed 7 April 2021).
19. GitHub, ChemRxiv Dashboard, **2021**. Available online: https://chemrxiv-dashboard.github.io (accessed 7 April 2021).
20. https://app.dimensions.ai/ (accessed 7 April 2021).
21. H. Piwowar, J. Priem, V. Larivière, J. Alperin, P.L. Matthias, B. Norlander, A. Farley, J. West, S. Haustein, The state of OA: A large-scale analysis of the prevalence and impact of open access articles, *Peer J.* **2018**, *6*, e4375, https://doi.org/10.7717/peerj.4375
22. M. Pagliaro, Preprints in Chemistry: An Exploratory Analysis of Differences with Journal Articles, *Publications* **2021**, *9*, 5. https://doi.org/10.3390/publications9010005

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2021 doi:10.20944/preprints202104.0448.v1




23. https://www.researchsquare.com/browse (accessed 8 April 2021).
24. H. Else, How a torrent of COVID science changed research publishing - in seven charts, *Nature* **2020**, *588*, 553.
25. International Association of Scientific, Technical and Medical Publishers, *The STM Report*, *5th edition*, The Hague: **2018**; p.10.
26. P. O. Seglen, The skewness of science, *J. Assoc. Inf. Sci. Technol.* **1992**, *43*, 628-638. https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0
27. P. O. Seglen, Why the impact factor of journals should not be used for evaluating research, *Bmj* **1997**, *314*, 497. https://dx.doi.org/10.1136%2Fbmj.314.7079.497
28. S. Curry, Sick of Impact Factors, 13 August 2012. Available at: http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors (accessed 7 April 2021).
29. A. Casadevall, F. C. Fang, Causes for the persistence of impact factor mania, *mBio* **2014**, *5*, e00064-14. https://doi.org/10.1128/mbio.00064-14
30. J. Brainard, Open-access megajournals lose momentum as the publishing model matures, *Science* **2019**, https://doi.org/10.1126/science.aaz4585
31. P. Davis, Journals lose citations to preprint servers, *The Scholarly Kitchen*, 21 May 2018. Available at: https://scholarlykitchen.sspnet.org/2018/05/21/journals-lose-citations-preprint-servers-repositories
32. E. Garfield, The history and meaning of the journal impact factor, *JAMA* **2006**, *295*, 90-93. https://doi.org/10.1001/jama.295.1.90
33. National Institutes of Health, Reporting Preprints and Other Interim Research Products, Notice Number: NOT-OD-17-050, 24 March 2017. Available at: https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html (accessed 7 April 2021).
34. M. Pagliaro, Purposeful evaluation of scholarship in the open science era, *Challenges* **2021**, *12*, 6. https://doi.org/10.3390/challe12010006
35. J. Bar-Ilan, G. Halevi, S. Milojević, Differences between Altmetric Data Sources - A Case Study, *J. Altmetr.* **2019**, *2*, 1. https://doi.org/10.29024/joa.4
36. I. Puebla, J. Polka, O. Rieger, Preprints: Their Evolving Role in Science Communication, *MetaArXiv* **2021**, https://doi.org/10.31222/osf.io/ezfsk
37. M. Tian, Y. Su, X Ru, Perish or Publish in China: Pressures on Young Chinese Scholars to Publish in Internationally Indexed Journals, *Publications* **2016**, *4*, 9. https://doi.org/10.3390/publications4020009
38. M. Allen, D. Poggiali, K. Whitaker, T.R. Marshall, R. Kievit Raincloud plots: a multi-platform tool for robust data visualization, *PeerJ Preprints* **2018**, *6*, e27137v1. https://doi.org/10.7287/peerj.preprints.27137v1
39. M. Allen, D. Poggiali, K. Whitaker, T.R. Marshall, R. Kievit, Raincloud plots: a multi-platform tool for robust data visualization, *Wellcome Open Res.* **2019**, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2
40. R. Lalli, 'Dirty work', but someone has to do it: Howard P. Robertson and the refereeing practices of *Physical Review* in the 1930s, *Notes Rec.* **2016**, *70*, 151-174. https://doi.org/10.1098/rsnr.2015.0022
41. D.F. Horrobin, The Philosophical Basis of Peer Review and the Suppression of Innovation, *JAMA* **1990**, *263*, 1438-1441. https://doi.org/10.1001/jama.1990.03440100162024
42. R. D. Dudley, The Changing Landscape of Open Access Publishing: Can Open Access Publishing Make the Scholarly World More Equitable and Productive?, *J. Librariansh. Schol. Commun.* **2021**, *9*, eP2345. http://doi.org/10.7710/2162-3309.2345
43. S. Besir Demir, Predatory journals: Who publishes in them and why?, *J. Informetr.* **2018**, *12*, 1296-1311. https://doi.org/10.1016/j.joi.2018.10.008
44. M. Laakso, Green open access policies of scholarly journal publishers: A study of what, when, and where self-archiving is allowed, *Scientometrics* **2014**, *99*, 475-494. http://doi.org/10.1007/s11192-013-1205-3
45. B.C. Björk, M. Laakso, P. Welling, P. Paetau, Anatomy of green open access, *J. Assn. Inf. Sci. Tec.* **2014**, *65*, 237-250. https://doi.org/10.1002/asi.22963
46. A. Maddi, Measuring open access publications: a novel normalized open access indicator, *Scientometrics* **2020**, *124*, 379–398. https://doi.org/10.1007/s11192-020-03470-0
47. S. Serghiou, J. P. A. Ioannidis, Altmetric Scores, Citations, and Publication of Studies Posted as Preprints, *JAMA* **2018**, *319*, 402-404. https://doi.org/10.1001/jama.2017.21168
48. B. Brembs, Prestigious Science Journals Struggle to Reach Even Average Reliability, *Front. Hum. Neurosci.* **2018**, *12*, 37. https://doi.org/10.3389/fnhum.2018.00037