Journal of Library and Information Studies 24:1 (June 2026) pp.1-16
https://doi.org/10.6182/jlis.202606_24(1).001

Open, but Uncounted: Citations to Open Access Preprints and Working Papers of Published Papers

Paul Donner[1]

Abstract

The citations to open access preprint and working paper versions of papers that are also formally published in journals are not regularly counted in global multidisciplinary proprietary citation index databases, nor are they considered in most bibliometric studies. They are, however, integrated into paper citation counts in other citation databases. It can be argued that open access preprint citations reflect a relevant part of scientific impact of a work as proxied by citation counts, particularly in scientific disciplines that make extensive use of preprints/working papers for rapid open communication. Here we present a large-scale study of these usually uncounted open access preprint citations for the Web of Science database to provide an exploratory overview of the magnitude of this phenomenon and its distribution across disciplines. To this end, we introduce a methodology to link cited reference strings referring to preprints/working papers to their matching published journal versions. We find that at least 390,000 WoS-indexed publications have 1.1 million preprint version citations which are not included in ordinary citation counts.

Keywords: Open Access; Preprints; Citation Analysis; Working Paper; Bibliometrics

1. Introduction

Many scientific publications are published as preprints before an eventual formal publication in a scientific journal. Rzayeva et al. (2025) have studied the adoption of preprinting since the introduction of the first electronic preprint server, arXiv, in 1991. They found that the share of published papers with preprint versions has increased to around 3% of peer-reviewed literature in recent years, as covered in Dimensions (Herzog et al., 2020), with a conspicuous second wave of adoption after 2017. While this means that the vast majority of published research is not available as preprints, there are striking differences in preprinting adoption across disciplines: the preprinting adoption rate is close to 10% in the physical sciences and technology domain, but only about 1% in the humanities (Rzayeva et al., 2025, p. 18). The domains of life and medical sciences and social and behavioral sciences show intermediate adoption levels of 3 to 5% in recent years. It is also these latter domains which account for much of the post-2017 adoption increase. In absolute terms, the number of preprints is large. Rzayeva et al. (2025) identified 4.1 million preprints, which, as they pointed out, is not a comprehensive estimate because Dimensions does not cover working papers not posted to repositories.

The practice of preprinting is closely linked to the open access (OA) movement (Suber, 2012). Digital preprinting and its precursor, hardcopy

[1] German Center for Higher Education Research and Science Studies, Berlin, Germany
E-mail: donner@dzhw.eu

臺灣大學學術期刊資料庫

preprint circulation culture (Youngen, 1998), preceded the OA movement which embraced depositing freely accessible versions of papers on repository servers as the "green road" to OA (Suber, 2012, Chapter 3). The green road encompasses the OA publication of both preprints, drafts before submission to or acceptance by journals, and postprints, author versions of accepted or published papers (Note 1).

Open access preprint and working paper versions of papers may be cited, but such preprint version citations are usually not counted in conventional citation analysis, because the policy of major citation-tracking scholarly databases is that only citations to formally published items are counted. For papers available first as preprints and later published in journals, there can be a considerable timeframe in which only the preprint version is accessible and citable. For instance, for mathematics publications, Wang et al. (2020) reported that it takes an average of 16 months from arXiv submission to online journal publication.

Whether it is appropriate to cite preprints has not been finally settled. Teixeira da Silva (2017) has argued that preprints should not be cited except to critically evaluate them, because they have not been subjected to peer review. This reasoning is supported by the common observation that not all preprints are eventually published as formal journal papers: Rzayeva et al. (2025) found that about half of preprints could be linked to a published peer-reviewed version. Respondents to a recent survey on attitudes about preprints expressed concerns about the reliability and credibility of preprints and were uncertain about whether and how to cite preprints (Ni & Waltman, 2024). Nonetheless, preprints, whether electronic or print, have been cited to some extent in the literature as studies using the WoS cited reference search have shown (Aman, 2015; Fu, 2003; Youngen, 1998).

On the other hand, preprint and peer-reviewed versions are often very similar or even indistinguishable in content. In those cases, concerns about reliability of preprints are either groundless, or equally applicable to the peer-reviewed versions. Those preprints which are later formally published in reputable journals are often unchanged or only little changed in content between preprint and final published version. Klein et al. (2019) studied the text similarities between arXiv and bioRxiv preprints and their published versions, and concluded: "Our results suggest that the contents of the vast majority of final published papers are largely indistinguishable from their preprint versions" (p. 336). Stephen (2022) conducted a qualitative analysis of peer review reports for 40 social science preprint-publication pairs. While final published versions were in general somewhat longer than preprint versions, the overall semantic content was changed only slightly to moderately. Moreover, many preprints are accepted versions of manuscripts without the publisher formatting (Ni & Waltman, 2024, p. 754). These arguments obviously do not apply to preprints never formally published in journals, which are not within the scope of this study. The evidence of very high similarity of preprints and published versions suggests that these versions of papers are typically not too different to be treated as different works in citation analysis of published papers. In other words, citing the preprint versions of these published papers is appropriate, as there are only minor differences between the versions. (Note 2).

臺灣大學學術期刊資料庫

The convention of citation index databases to only count citations to formally published papers but not citations to their preprint versions could potentially disadvantage authors and institutions committed to sharing their findings early, openly, and widely, by depositing their papers as preprint OA versions via preprint platforms or as working papers on personal or institutional websites.

Many studies have tried to identify the citation advantage of OA, see Langham-Putrow et al. (2021) for a review. The typical working hypothesis of these studies is that the reach of OA papers is greater than that of toll access papers, as the toll access may exclude researchers who cannot afford subscription or access fees. More readers would directly lead to greater citation counts (Lawrence, 2001). As far as these studies have not included citations to preprint versions of publications, their results would be questionable, as we will argue below.

Unlike the common approach, here we investigate for the first time the possibility of an OA citation disadvantage, specifically the phenomenon of undercounting of the total citations to a publication by omission of citations to preprint OA versions. It is currently not known if this phenomenon exists at any significant scale and what this scale may be. Therefore, the primary objective of this paper is to investigate if a preprint OA citation disadvantage does exist in reality, can be identified, and can be quantified.

Citations to preprints in general are not a quantitatively marginal phenomenon. For example, Aman (2015) found that for publications in 1991–2013, there were about 900,000 WoS-indexed references to arXiv preprints alone, which is enough to motivate an updated analysis including the wider preprint literature.

The argument for including preprint citations in citation analysis of published journal papers is as follows. Not counting citations to preprint versions is difficult to defend as a methodological choice. If citation counts are treated as an indicator for a latent construct of 'scientific impact' (Aksnes et al., 2019), there is no justification for excluding preprint citations because they provide just as much information on use and influence on later scientific research as citations to published versions. This is based on the assumption that the preprint and final published version of a publication are identical or similar in content that they can be considered as belonging to one publication family. Certainly, there are also cases in which preprints are thoroughly developed and modified such that the formally published versions must be considered as different works. It is an important open research question for which fraction of preprints this is true, and if such changes can perhaps be detected from metadata. However, this is beyond the scope of the present study and is left for future work. In summary, preprint version citations are just as relevant to be used for citation impact measurements as published version citations of the same publications, with the caveat that the publications should be recognized as 'the same.'

At present, the indexing policy of preprints and citations to preprint versions in the two major commercial scholarly citation platforms, Scopus and Web of Science (WoS), is as follows. Scopus has included preprints as listed documents on author profiles since 2021, but citations to preprint versions of published research are not included in

臺灣大學學術期刊資料庫

their citation counts (Note 3). In 2023, Clarivate Analytics launched the Preprint Citation Index, covering documents from initially five platforms. This index provides citation counts tallying references from preprints to preprints, and links cited references to the WoS Core Collection (https://webofscience.help.clarivate.com/en-us/Content/pre-prints/pre-prints.htm). However, citation counts to preprint and final published versions of the same item are not unified in WoS. Preprint version citations of works later published in journals are not counted in WoS or Scopus.

Given that preprints are frequently cited, that many preprints are developed into formally published articles, and that widely used commercial citation index databases do not unify preprint and journal version citations, the present study addresses the following research questions:

1. How many cited references to OA preprints/working papers in WoS can be matched to records of published versions?
2. How many WoS-indexed papers have uncounted preprint citations from other papers in WoS?
3. Which disciplines are most affected by uncounted preprint citations?

The remainder of the paper proceeds as follows. In section 2, we review the literature on the OA citation advantage with a focus on inclusion of preprint/working paper citations as a further motivation for this study. Section 3 describes the data and methods. The results are presented in section 4 and discussed in section 5. Section 6 concludes the paper with interpretation and recommendations.

2. Literature Review

Over the past decades, OA has been established as a common practice in scientific publishing in its various different incarnations (Suber, 2012) and it has also become positioned lately as a core component of the wider open science movement. Within the context of citation analysis, some interesting results about OA publishing have been reported. One advantage of OA publishing is the acceleration of scientific communication. Preprint versions may draw researchers' attention to potentially forthcoming papers, which they may thus read, incorporate in their research, and cite earlier than research that was not available in preprint form. Yu and Lin (2016) have examined the first-citation speed of papers in library and information science journals as a case study. Within journals publishing both OA and non-OA papers, they found that OA papers were first cited considerably earlier than their toll access counterparts. However, this study did not look into possible effects of preprints.

There are, to our knowledge, no prior studies directly addressing the postulated preprint OA citation disadvantage. However, many studies have investigated the related phenomenon of OA citation advantage, and the typical method of such studies is a motivation for the present paper, therefore the remainder of this section is about this literature. The majority of these studies have not included citations to OA preprint versions, which could be considered a methodological flaw. The basic setup of such studies is to compare the citation frequencies of two classes of papers, OA and non-OA papers. However, implicitly or explicitly, the theoretically relevant concept of interest in science studies is scholarly impact, that

臺灣大學學術期刊資料庫

is, the degree of influence that some research has had on subsequent research, for which citations serve as a proxy variable. Only OA papers, by definition, could have also been issued as a preprint OA version. A preprint version may have exerted influence on later research which could have manifested in citations. Such preprint citations do signify formally recorded scholarly influence and impact as much as citations to published papers, and should therefore be counted. This leads to systematic undercounting of citations to OA publications–a bias–vis-à-vis non-OA publications, and hence underestimation of their scholarly impact.

This section reviews the four OA citation advantage studies we could find which, in contrast to the majority of such studies, have explicitly included preprint citations. It also reviews some data on the citation data sources of the general OA citation advantage literature to understand which share of this literature may be affected by the methodological flaw of not including preprint citations.

Davis and Fromerth (2007) studied citation count differences between papers published only as formal journal versions and those also published as arXiv preprint versions in four mathematics journals. They were able to handle the issue of citations to preprint versions by relying on citation counts of a scholarly publication database specialized in mathematics:

> *MathSciNet citations were used instead of ISI citations since journal coverage in the latter is incomplete, and because MathSciNet tracks citations to preprints in the arXiv in addition to citations to the formal publication* (Davis & Fromerth, 2007, p. 205).

Using this data, they show that publications with arXiv preprints are in fact cited somewhat more frequently, likely because of authors' self-selection of papers for quality.

In the field of high energy physics (HEP), an OA citation advantage study of Gentil-Beccot et al. (2010) was also able to rely on a citation database that integrates preprint and journal version citations (SPIRES, Stanford Public Information Retrieval System) and clearly identified this property as the principal precondition for the validity of their study:

> *The HEP community relies heavily on communication through preprints, and therefore SPIRES counts citations to and from preprints. In SPIRES, citations to preprints are aggregated with the citations to the published versions, once available, treating the two versions as a single entity. This differs from most bibliometric approaches which only consider citations from published articles to published articles, and this feature is crucial for the conclusions of this study* (Gentil-Beccot et al., 2010, p. 347).

Brown and Zimmermann (2017) studied the history of a population economics journal, among other topics also with regard to citations to working paper versions and formally published versions of the same papers. They used Google Scholar citation counts, which unify citations to working paper versions, published OA, and journal versions. They found that papers with working paper version receive twice as many citations as other papers.

Finally, Henneken et al. (2006) studied citations to astronomy and physics papers either

臺灣大學學術期刊資料庫

published as preprints on arXiv or not. They used citation count data from the Astrophysics Data System (ADS) because this database consolidates references to either the arXiv or the journal version of one paper under one record, ensuring complete citation counts. They found that "papers that appear as e-prints are both cited more and read more than papers that do not appear as e-prints" (Henneken et al., 2006, section 4, para. 1).

These examples are atypical. More frequently, OA citation advantage studies, of which there are a lot, do not use a citation data source that integrates preprint and journal version citations into one complete citation count. A look at data from a recent meta-analysis of such studies can provide evidence for this claim. Langham-Putrow et al. (2021) conducted a comprehensive meta-analysis of studies on the OA citation advantage, analysing 134 studies. In their screening of the studies, they also recorded the citation data sources of the studies (in their Table 1). They found that most studies used either WoS (61%) or Scopus (28%), both of which do not unify preprint and journal version citations. 22% of studies used Google Scholar, which does unify these citations to different versions of a paper. 14% used a different citation data source, some of which perform the unification (SPIRES, NASA ADS, MathSciNet) while others presumably do not. Some studies (18%) used more than one citation data source. Thus, the majority of OA citation advantage studies have used data sources which may have unwittingly diminished the accuracy of their findings because these sources, WoS and Scopus, do not count citations accruing to preprint versions of papers. This puts papers published also as OA preprints at a disadvantage that only they could receive citations as preprint versions.

Fraser et al. (2020) offer a good example for the importance of preprint citations to OA advantage studies, using Scopus and augmented Scopus citation counts to papers also deposited as bioRxiv preprints in the following way:

> *Since preprints are not themselves indexed in Scopus, we utilized the Scopus raw reference data, which includes a "SOURCETITLE" field including the location of the cited object. We queried the SOURCETITLE for entries containing the string "biorxiv" (case-insensitive, partial matches), and retrieved 4,826 references together with the metadata of their Scopus-indexed citing articles. References were matched to preprints via fuzzy matching of titles and direct matching of DOIs, although DOIs were only provided in a minority of cases. In total 4,387 references (90.9%) could be matched to a bioRxiv preprint* (p. 622).

They seemingly did not unify preprint and journal version citations. They report total Scopus citation counts for bioRxiv-preprinted papers of 45,121 citations and 27,658 citations to matched control articles. They found that, "2,021 citations were made to preprints that were subsequently published as journal articles" (p. 629). Thus, the raw OA citation advantage is about 70% ((45,121 + 2,021) / 27,658), and about 10% of the advantage can be attributed to preprint citations.

臺灣大學學術期刊資料庫

3. Data and Methods

3.1 Data source

We carried out this study using an October 2024 licenced snapshot of the WoS (Birkle et al., 2020). This snapshot consists of the WoS Core Collection, publication years from 1980 to present, which includes the Science, Social Science, and Arts and Humanities Citation Indexes (CI) and the two Proceedings Indexes, but not the regional/national CIs, Book CI, Emerging Sources CI, or the new Preprint CI. We have used WoS data in the "KB" data infrastructure, whose processing details are reported in Schmidt et al. (2025).

3.2 Matching procedure

Cited reference strings were text-indexed within a relational database environment for convenient and fast search queries. We restricted the candidate pool of cited references in WoS to those which were not linked by WoS to any indexed source item, and which had some author name information present. We stored all cited reference strings which contained either a text pattern that matches any of 10 preprint platform names, or any of the strings "working paper," "discussion paper" or "preprint" in the cited source title field. The 10 preprint platforms were arXiv, medRxiv, bioRxiv, OSF, PsyArXiv, SSRN, EconStor, PhilArchive, viXra, ChemRxiv. Missing cited publication years were in some cases automatically replaced with identified year-like patterns in the cited source title field when present. Reference strings with cited title containing "reply," "comment," "corrigendum," "correction" and "vol" were removed, as these were found to lead to false matches in the following steps. The remaining candidate preprint cited reference strings were matched to WoS-indexed source items with the following conditions:

- exact match on author family name and given name (or initials),
- publication year of cited reference earlier or equal to indexed source item's publication year,
- cited reference title and indexed source item title similar with Jaro-Winkler similarity > 0.9 and Levenshtein similarity > 0.88.

These values were experimentally found to have very high precision (practically no false positives), but the false negative rate is unknown. We do not claim that these methodological choices are optimal in terms of precision-recall trade-off, but only that they are reasonable for the purpose of investigating the approximate size of the preprint citation disadvantage phenomenon. More sophisticated methods with verified accuracy should be developed for an actual application in practice.

4. Results

Research question 1: How many cited references to OA preprints/working papers in WoS can be matched to records of published versions?

The search for preprint/working paper cited reference strings in WoS identified a candidate pool of about 4.4 million references. Of these, 1,121,172 references could be matched to WoS-indexed items (about 25%). These citations were not included and counted as citations by WoS and would be overlooked in conventional citation analysis. In the preliminary candidate pool, there were 2.8 million references to arXiv preprints, which can be compared to the result of Aman

臺灣大學學術期刊資料庫

(2015) who found 900,000 arXiv references in WoS ten years earlier.

Research question 2: How many WoS-indexed papers have uncounted preprint citations from other papers in WoS?

The 1.1 million identified citations refer to 393,705 WoS items, which is about 2.8 preprint citations per 'preprint-cited' paper on average. The papers with any identified preprint citations had accrued about 18.9 million regular WoS citations. 14% of these items, about 56,000 papers, attracted more preprint citations than regular citations. This specific subset of most affected publications received on average 1.0 regular WoS citations and 3.6 preprint citations, overlooked in ordinary citation analysis using WoS data.

In some cases, the number of preprint citations can take on extremely high values. Table 1 shows the 10 publications with the highest identified preprint citation counts, with their regular WoS citation counts and basic bibliographic data. These are all exceptionally highly cited recent papers, mostly from computer science. Nevertheless, their ordinary WoS citation counts, as visible to users, can be an underestimation by hundreds, occasionally by thousands, of citations.

OA preprinting has recently been growing rapidly, so it is interesting to see if this trend is also reproduced in citations. To test this, we have calculated the shares of WoS publications with any preprint citations by year. For the years up to 2018, this is below 1% annually, and between 2019 and 2024 the share is between 1.0 and 1.6%. While in both absolute and relative terms, not many publications are affected by uncounted

Table 1. The Ten Publications with Highest Preprint Citation Count as of October 2024

Preprint citations	WoS citations	First author	Year	Title
10,666	28,357	Devlin, Jacob	2019	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
4,842	2,901	Liu, Yinhan	2019	RoBERTa: A Robustly Optimized BERT Pretraining Approach
3,063	94,907	He, Kaiming	2016	Deep Residual Learning for Image Recognition
2,938	52,971	Vaswani, Ashish	2017	Attention Is All You Need
2,498	8,990	Ioffe, Sergey	2015	Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift
1,614	3,531	Brown, Tom B.	2020	Language Models are Few-Shot Learners
1,399	5,862	Goodfellow, Ian	2020	Generative Adversarial Networks
1,207	52,386	Bates, Douglas	2015	Fitting Linear Mixed-Effects Models Using lme4
970	385	Kirillov, Alexander	2023	Segment Anything
909	49,901	Ronneberger, Olaf	2015	U-Net: Convolutional Networks for Biomedical Image Segmentation

臺灣大學學術期刊資料庫

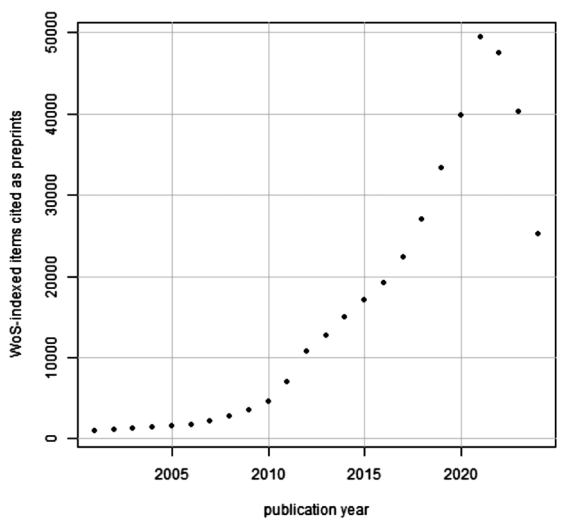
preprint citations, the share of these publications is increasing in more recent years.

This growth trend of absolute numbers of WoS-indexed items with any preprint citations is shown in Figure 1 in the period from 2001 to 2024. The chart confirms that few papers from the 2000s are affected and there is a steady increase over time. At the time of analysis, there were already about 50,000 papers from both 2021 and 2022 which had uncounted preprint citations. These numbers are dynamic, as still more preprints might become cited, especially for the most recent years.

Another question about time dynamics is relevant to look at: When were these preprint citations made? This is visualized in Figure 2, showing the total number of annually citations to any preprint version of papers indexed in WoS, regardless of the publication year of the cited document. In contrast to the previous analysis, these numbers are static, as the cited references in each year are fixed once all papers published in a given year are indexed. If we disregard the citing year 2024, for which data was incomplete at the time of analysis, we can notice a break in the growth from 2019 to 2020, perhaps due to the impact of Covid, and a decrease after 2021.

Citing preprint versions becomes obsolete when a final journal version is published, which typically happens shortly after a preprint was issued. It then seems likely that citations to preprint versions would be made in a brief window of time when only the preprint version of paper was available. Figure 3 displays the difference of the preprint publication year and the year in which it was cited as percentages. For example, 29% of all preprint citations were made in the same year when the preprint was issued, namely year 0 in the figure. The highest share of preprint citations are made in the year following preprint publication, about 34%. Very few citations to preprints are made 5 or more years after preprint publication, only about 6% of the total.

Figure 1. Number of WoS-indexed Items with Any Preprint Citations by Publication Year



臺灣大學學術期刊資料庫

Figure 2. Number of Citations to Preprint Versions of WoS-indexed Items by Citing Year

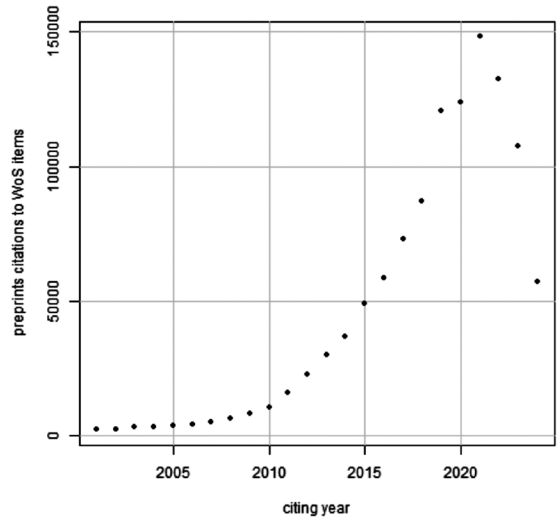
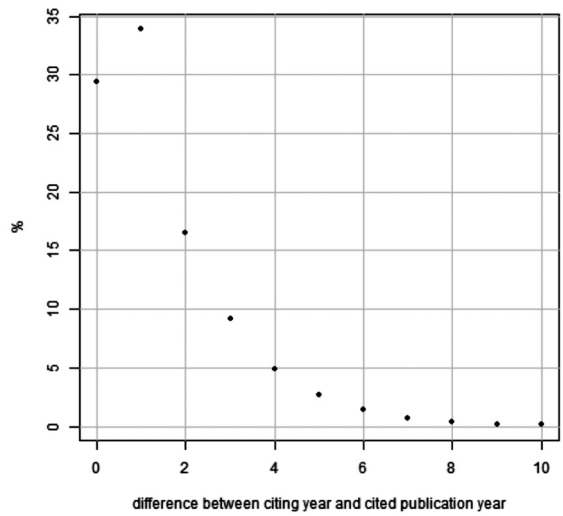


Figure 3. Difference between Preprint Publication Year and the Year of Citation as Percentages of Total Preprint Citations



Research question 3: Which disciplines are most affected by uncounted preprint citations?

Some scientific disciplines have a more developed culture of using preprints, including physics and astronomy, mathematics, computer science, and economics. It is plausible that such disciplines are also disproportionately affected by preprint citation differences which create

臺灣大學學術期刊資料庫

an OA preprint citation disadvantage. To test which disciplines are most affected by uncounted preprint citations, we calculated the number of papers with any preprint citations by WoS Subject Category. Because the OA preprint citations phenomenon is accelerating in recent years, we have restricted this analysis to papers published in 2020 to 2023. Table 2 shows the five Subject Categories with the highest percentage of papers with any preprint citations. These disciplines belong to the domain of physics and mathematics, and their rates of papers with preprint citations are over 10%. Economics has a rate of about 9% (not included in the table). There are many WoS Subject Categories with very few or no preprint citation papers, primarily in the humanities.

In summary, these results mean that a preprint citation disadvantage clearly exists and that overlooked preprint citations can be automatically identified, counted, and analyzed. The phenomenon of overlooked preprint citations is increasing in recent years, and is concentrated in a modest number of scientific disciplines with a strong preprint culture.

5. Discussion

This is an exploratory study with several limitations. The proposed method of matching preprint citations to their published version papers, while very similar to the method of Fraser et al. (2020), is experimental. As mentioned, we have used relatively strict matching procedures to rule out false positives as much as possible. This means the study is missing an unknown number of additional citations to preprint versions, for example, cases in which the first author was changed or the title was modified more than our decision rule allowed for. Further research on preprint citation matching should explore systematically which matching conditions lead to optimal results in terms of precision-recall trade-off. This issue cannot be detached from the underlying question: how to draw the distinction between the degree of difference in metadata of publication content that still indicates different versions of the same publication as opposed to indicating two different documents which may still share some common origin, but are too

Table 2. Five WoS Subject Categories with Highest Proportion of Papers with Preprint Citations (2020–2023)

WoS subject category	Papers with preprint citations	Total papers	Percentage of papers with preprint citations
Physics, Particles & Fields	9,587	62,120	15.43
Statistics & Probability	6,605	49,321	13.39
Social Sciences, Mathematical Methods	1,653	13,045	12.67
Mathematics	17,406	143,754	12.11
Quantum Science & Technology	1,904	15,990	11.91

臺灣大學學術期刊資料庫

different to be considered identical (Note 4). For this reason, the presented figures are inaccurate and conservative estimates of the real numbers to some extent, but the degree of inaccuracy could not be quantified.

While the number of papers with uncounted preprint citations is not large, the OA preprint citation disadvantage is increasing over time as OA preprints are becoming more common in many disciplines (Rzayeva et al., 2025) and can be expected to worsen further in the future. The average share of overlooked preprint citations for papers is modest. A lower limit for the preprint OA citation penalty is 5% of total citations for those publications with any preprint citations. However, as is common in scientometric variables, these overlooked citations are very unevenly distributed over publications. More than 14% of published papers with any preprint citations have more overlooked citations than regular WoS citations.

The OA preprint citation disadvantage of published research is primarily affecting publications in disciplines with a well established preprint publication culture such as mathematics, computer science, economics, physics and astronomy.

6. Conclusion

This study was able to show that formally published papers with OA preprint or working paper versions also receive significant numbers of citations to their preprint or working paper versions. As these preprint citations are not commonly counted in the citation of the final published versions of papers, those papers with OA preprint version are affected by a citation disadvantage.

In light of the presented results, professional bibliometricians, science policymakers, staff in charge of citation-based evaluation procedures and citation index producers are advised to consider this study as an early warning of a potentially increasingly relevant issue. Researchers who study OA citation advantages should always use a source of citation data that aggregates citations to preprint versions and formally published versions of articles, otherwise the validity of their study will be open to questioning on methodological grounds. We recommend that citation index database producers also link preprint version citations to indexed documents and offer users the option to include or exclude these citations. Unless the OA preprint citation disadvantage is addressed by improved citation analysis methodology, continuing ignorance of preprint citations might eventually be perceived as a stumbling block for wider OA adoption.

Notes

Note 1 It is critical to emphasize that the terms ‘preprint’ and ‘open access publication’ do not mean the same thing. While almost all formally deposited preprint publications are made available openly accessible, not all OA publications have available preprint versions. This present study focused explicitly on citations to OA preprint papers and working papers and their identified reviewed and formally published journal versions.

Note 2 This is not to suggest that the choice of citation is arbitrary when both preprint and published versions exist.

臺灣大學學術期刊資料庫

In those cases, the improvements that peer review can make obviously mean that the published version should be cited preferably.

Note 3 "*Preprints* are not integrated into any metrics in Scopus and do not influence assessment, including citation counts. The version-of-record (published, peer-reviewed articles) are the official representation of the research in Scopus." (Scopus, 2024)

Note 4 In several preprint services, users depositing preprints have the option of linking a formally published version of the uploaded document by either DOI or bibliographical reference. Such version linking information is not only a useful source for aggregating citation counts across versions, but may be also useful as gold standard data in determining similarity thresholds for fuzzy matching procedures that link cited reference strings of preprint versions to formally published documents indexed in citation databases.

References

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open, 9*(1), Article 2158244019829575. https://doi.org/10.1177/2158244019829575

Aman, V. (2015). Citing e-prints on arXiv. A study of cited references in WoS-indexed journals from 1991–2013. In A. A. Salah, Y. Tonta, A. A. Akdag Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics conference* (pp. 1107-1119). International Society of Scientometrics and Informetrics. https://www.issi-society.org/proceedings/issi_2015/1107.pdf

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies, 1*(1), 363-376. https://doi.org/10.1162/qss_a_00018

Brown, A. J. G., & Zimmermann, K. F. (2017). Three decades of publishing research in population economics. *Journal of Population Economics, 30*, 11-27. https://doi.org/10.1007/s00148-016-0620-1

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics, 71*, 203-215. https://doi.org/10.1007/s11192-007-1661-8

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies, 1*(2), 618-638. https://doi.org/10.1162/qss_a_00043

Fu, Y.-H. (2003). Citation patterns to traditional and electronic preprints. *Journal of Library & Information Studies, 1*(2), 81-94. https://doi.org/10.6182/jlis.2003.1(2).081 (in Chinese)

Gentil-Beccot, A., Mele, S., & Brooks, T. (2010). Citing and reading behaviours in high-energy physics. *Scientometrics, 84*(2), 345-355. https://doi.org/10.1007/s11192-009-0111-1

臺灣大學學術期刊資料庫

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., & Murray, S. S. (2006). Effect of e-printing on citation rates in astronomy and physics. *Journal of Electronic Publishing, 9*(2). https://doi.org/10.3998/3336451.0009.202

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies, 1*(1), 387-395. https://doi.org/10.1162/qss_a_00020

Klein, M., Broadwell, P., Farb, S. E., & Grappone, T. (2019). Comparing published scientific journal articles to their pre-print versions. *International Journal on Digital Libraries, 20*(4), 335-350. https://doi.org/10.1007/s00799-018-0234-1

Langham-Putrow, A., Bakker, C., & Riegelman, A. (2021). Is the open access citation advantage real? A systematic review of the citation of open access and subscription-based articles. *PloS ONE, 16*(6), Article e0253129. https://doi.org/10.1371/journal.pone.0253129

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature, 411*(6837), 521-521. https://doi.org/10.1038/35079151

Ni, R., & Waltman, L. (2024). To preprint or not to preprint: A global researcher survey. *Journal of the Association for Information Science & Technology, 75*(6), 749-766. https://doi.org/10.1002/asi.24880

Rzayeva, N., Pinfield, S., & Waltman, L. (2025). *Adoption of preprinting across scientific disciplines and geographical regions (1991-2023)*. SocArXiv. https://doi.org/10.31235/osf.io/xdwc4_v2

Schmidt, M., Rimmert, C., Stephen, D., Lenke, C., Donner, P., Gärtner, S., Taubert, N., Bausenwein, T., & Stahlschmidt, S. (2025). The data infrastructure of the German *Kompetenznetzwerk Bibliometrie*: An enabling intermediary between raw data and analysis. *Quantitative Science Studies, 6*, 1129-1146. https://doi.org/10.1162/QSS.a.20

Stephen, D. (2022). Peer reviewers equally critique theory, method, and writing, with limited effect on the final content of accepted manuscripts. *Scientometrics, 127*(6), 3413-3435. https://doi.org/10.1007/s11192-022-04357-y

Scopus. (2024). *What are preprints?* Scopus Support Center. https://service.elsevier.com/app/answers/detail/a_id/31630/supporthub/scopus/kw/preprint/

Suber, P. (2012). *Open access*. MIT Press. https://doi.org/10.7551/mitpress/9286.001.0001

Teixeira da Silva, J. A. (2017). Preprints should not be cited. *Current Science, 113*(6), 1026-1027. http://www.jstor.org/stable/26494158

Wang, Z., Chen, Y., & Glänzel, W. (2020). Preprints as accelerator of scholarly communication: An empirical analysis in Mathematics. *Journal of Informetrics, 14*(4), Article 101097. https://doi.org/10.1016/j.joi.2020.101097

Youngen, G. K. (1998). Citation patterns to traditional and electronic preprints in the published literature. *College & Research Libraries, 59*(5), 448-456. https://doi.org/10.5860/crl.59.5.448

臺灣大學學術期刊資料庫

Yu, T.-C., & Lin, W.-Y. C. (2016). The scholarly communication speed of the library and information science open access journals as measured by first-citation. *Journal of Library & Information Studies, 14*(1), 151-179. https://doi.org/10.6182/jlis.2016.14(1).151 (in Chinese)

(Received: 2025/5/8; Accepted: 2025/11/4)

臺灣大學學術期刊資料庫

圖書資訊學刊　第24卷第1期　（2026.6）　頁1-16
https://doi.org/10.6182/jlis.202606_24(1).001

開放而未計：已出版論文的開放取用預行本與工作底稿之引用分析

Open, but Uncounted: Citations to Open Access Preprints and Working Papers of Published Papers

Paul Donner[1]

摘　要

在全球跨學科引文索引資料庫中，對已正式發表於期刊之論文的開放取用預行本和工作底稿的引用通常未被統計，且多數書目計量學研究也不會考慮這些引用；然而，在部分引文資料庫中，這些引用被納入論文的引用計算。對廣泛使用預行本或工作底稿進行快速開放交流的領域，開放取用預行本的引用實則反映了論文之科學影響力。故本文針對Web of Science資料庫中這些通常未被統計的開放取用預行本引用進行大規模研究，旨在探索此現象的規模及其在不同學科的分佈。本研究將引用預行本或工作底稿的參考文獻連結至相對應的已發表期刊版本，發現至少39萬篇WoS收錄的出版品擁有110萬次預行本的引用，且這些引用並未包含在常規的引用計算中。

關鍵字：開放取用、預行本、引文分析、工作底稿、書目計量學

[1] 德國柏林高等教育與科學研究中心
German Center for Higher Education Research and Science Studies, Berlin, Germany
E-mail: donner@dzhw.eu

註：本中文摘要由圖書資訊學刊編輯提供。
以APA格式引用本文：Donner, P. (2026). Open, but uncounted: Citations to open access preprints and working papers of published papers. *Journal of Library and Information Studies, 24*(1), 1-16. https://doi.org/10.6182/ jlis.202606_24(1).001
以Chicago格式引用本文：Donner, Paul. “Open, but Uncounted: Citations to Open Access Preprints and Working Papers of Published Papers.” *Journal of Library and Information Studies* 24, no. 1 (2026): 1-16. https://doi.org/10.6182/jlis.202606_24(1).001

臺灣大學學術期刊資料庫