

## Comparative analysis of retracted pre-print and peer-reviewed articles on COVID-19

Manraj Singh Sra<sup>1</sup>, Mehak Arora<sup>1&</sup>, Archisman Mazumder<sup>1&</sup>, Ritik Mahaveer Goyal<sup>1</sup>,

Giridara G Parameswaran<sup>2</sup>, Jitendra Kumar Meena<sup>3\*</sup>

<sup>1</sup> All India Institute of Medical Sciences, New Delhi

<sup>2</sup> Center for Disease Dynamics, Economics & Policy, New Delhi

<sup>3</sup> Department of Preventive Oncology, National Cancer Institute (NCI) Jhajjar, All India

Institute of Medical Sciences, New Delhi

\*Corresponding author

& Mehak Arora and Archisman Mazumder have contributed equally

### Corresponding author:

Dr. Jitendra Kumar Meena

Department of Preventive Oncology,

National Cancer Institute (NCI) Jhajjar,

All India Institute of Medical Sciences, New Delhi

Email: [drmeenajk@gmail.com](mailto:drmeenajk@gmail.com)

**Keywords:** COVID-19, Retractions, Bibliometrics, Pre-prints

**Article Type:** Original Research

**Abstract word count:** 245

**Main text word count:** 1412

**Tables:** 1

**Figures:** 1

**Supplementary Material:** Table 4; Figure: 1

**Conflict of interest:** The authors have no competing interests to declare that are relevant to the content of this article.

**Funding source:** No funding was received for conducting this study

**Data availability statement:** Data used in this is available subject to a standard data use agreement from The Centre for Scientific Integrity, the parent non-profit organization of Retraction Watch. Web link: <http://retractiondatabase.org/RetractionSearch.aspx?>

**Contributors:** MSS, GGP, and JKM conceptualized the study. MSS, MA, AM, and JKM were involved in data collection and analysis. MSS, MA, AM, and RG wrote the first draft of the manuscript. GGP and JKM provided critical revisions to the manuscript and supervised the study.

## **Abstract**

### **Introduction**

Due to the accelerated pace and quantum of scientific publication during the COVID-19 pandemic, a large number of articles on COVID-19 have been retracted. Pre-prints though not peer-reviewed offer the advantage of rapid dissemination of new findings. In this study, we aim to systematically compare the article characteristics, time to retraction, social media attention, citations, and reasons for retraction between retracted pre-print and peer-reviewed articles on COVID-19.

### **Methods**

We utilized the Retraction Watch database to identify retracted articles on COVID-19 published from 1<sup>st</sup> January 2020 to 10<sup>th</sup> March 2022. The articles were reviewed and metadata such as article characteristics (type, category), time to retraction, reasons for retraction, and Altmetric Attention Score (AAS) and citation count were collected.

### **Results**

We identified 40 retracted pre-prints and 143 retracted peer-reviewed articles. The median (IQR) retraction time for pre-print and peer-reviewed articles was 29 (10-81.5) days and 139 (63-202) days ( $p = 0.0001$ ). Pre-prints and peer-reviewed article had median (IQR) AAS of 26.5 (4-1155) and 8 (1-38.5), respectively ( $p = 0.0082$ ). The median (IQR) citation count for pre-prints and peer-reviewed articles was 3 (0-14) and 3 (0-17), respectively ( $p = 0.5633$ ). The AAS and citation counts were correlated for both pre-prints ( $r = 0.5200$ ,  $p = 0.0006$ ) and peer-reviewed articles ( $r = 0.5909$ ,  $p = 0.0001$ ). The commonest reason for retraction for pre-prints and peer-reviewed articles concerns about data and results.

### **Conclusion**

The increased adoption of pre-prints results in faster identification of erroneous articles compared to the traditional peer-review process.

## **Main Text**

### **Introduction**

Since the start of the COVID-19 pandemic, a large number of articles on epidemiology, diagnostics, clinical management, and other aspects of the disease have flooded the scientific publication space creating doubts over the authenticity and scientific rigor. (Boschiero et al., 2021) It is estimated that 4% of the world's research output in 2020 on the Dimensions database was COVID-19 related. (Nicholas Fraser & Bianca Kramer, 2021) Due to the growing doubts about research quality, the publication industry put forth retractions of a large number of COVID-19-related articles, although the proportion of articles retracted was nearly the same as other research. (Peterson et al., 2022)

Pre-publication review systems, although effective and efficient, might fail to prevent the publication of invalid literature. There have been instances where retractions have occurred even from high-impact journals and such retractions have happened over months while the typical pre-pandemic delay in retraction was around 3 years. (Else, 2020). During the pandemic, there was a large increase in the number of studies released as pre-prints. Pre-print though not peer-reviewed has the advantage of facilitating rapid dissemination of the findings, but this practice also has the danger of promoting the quick propagation of faulty research. (King, 2020)

Previously descriptive studies on retracted articles related to COVID-19 have been carried out, but comprehensive comparison between pre-print and peer-reviewed articles has not been done. (Shimray, 2021) In this study, we aim to compare the article characteristics, time to retraction, social media attention, citations, and reasons for retraction of peer-print and peer-reviewed articles on COVID-19.

## **Materials and methods**

### **Data Acquisition**

In this study, we utilized the Retraction Watch Database maintained by the Center of Scientific Integrity to identify retracted articles on COVID-19/SARS-COV2. (*Retraction Watch Database*, n.d.) This database was accessed on 10<sup>th</sup> March 2022 under a data use agreement. Using the PubMed ID or DOI of the retracted articles we determined the Altmetric Attention Scores (AAS) and citations on 24<sup>th</sup> March 2022 from the Altmetric and Dimensions database, respectively. (*Dimensions*, n.d.; *Discover the Attention Surrounding Your Research – Altmetric*, n.d.)

### **Search Strategy**

Retracted studies on COVID-19/SARS-COV2 were identified by searching for the keywords “COVID-19”, “COVID”, ”SARS-COV-2”, ”2019-nCOV” and “Coronavirus” in the title of all pre-prints and peer-reviewed articles in the Retraction Watch Database published from 1<sup>st</sup> January 2020 to 10<sup>th</sup> March 2022.

### **Variables**

For our analysis, we collected the name of the journal/pre-print server, country of the first author, article type, reason of retraction, date of publication, date of retraction, AAS, and citations received. Articles were classified as pre-prints if they were ever published on any of the following preprint platforms: medRxiv, bioRxiv, arXiv, Social Science Research Network (SSRN), OSF Preprints, and Research Square. Articles that were published in peer-reviewed journals and ever never posted on a pre-print platform were classified as peer-reviewed articles. Articles were classified into case reports, clinical studies, commentary/editorials, conference abstracts, letters, meta-analyses, research articles, and review articles. Article types and reasons for retraction were based on the classification used in the Retraction Watch Database. (Supplementary Table 1 and 2). AAS was considered to be

zero for all studies with no Altmetric records, as assumed in previous studies. (Serghiou et al., 2021) We also examined if notice of retraction had been displayed on the pre-print platforms and if the full text of the retracted pre-prints was still available.

### **Statistical Analysis**

All data were entered into Microsoft Office Excel and statistical analysis was done in R. (*R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>, n.d.*) Categorical variables were summarized as counts and proportions. Continuous variables were checked for normal distribution using the Shapiro-Wilk test. Continuous variables were summarized as mean (standard deviation) or median (inter-quartile range), based on the results from the Shapiro-Wilk test. Categorical variables were compared using Fisher's exact test and continuous variables were compared using paired t-test or Kruskal-Wallis rank-sum test. Log-rank test was used to compare the time to retraction between pre-print and peer-reviewed articles. All p-values were two-tailed and p-values <0.05 were considered to be significant. Spearman rank correlation test was used to examine the correlation between continuous variables. UpSet plot was created using the UpSetR package to show the intersection of sets for the reason of retraction. (Conway et al., 2017).

## Results

We included 183 articles in the final analysis, out of which 40 were pre-prints and 143 were peer-reviewed articles. (Supplementary Figure 1) Research articles accounted for 62.8% (n=115) of all retracted articles. Among peer-reviewed articles, 6 journals accounted for 34.7% of the retractions. Most of the retracted pre-prints were posted on medRxiv (n=23, 57.5%) followed by the SSRN (n=8, 20%) and bioRxiv (n=7, 17.5%). Characteristics of studies included in the analysis have been summarized in Table 1. All pre-print servers displayed a notice of retraction for the pre-prints included in this study. The full text was available for 77.5% (n=31) of the articles posted, these articles had been posted on medRxiv, bioRxiv, or Research Square. The full text was unavailable for articles posted on SSRN or OSF Preprints.

The time to retraction of pre-print article was significantly shorter than peer-reviewed article. (Figure 1 (A); median (IQR): 29 (10-81.5) days vs 139 (63-202) days,  $p = 0.0001$ ). For pre-print article there was no correlation between AAS and days to retraction ( $r = -0.0124$ ,  $p = 0.9392$ ), citations and days to retraction ( $r = -0.1797$ ,  $p = 0.2673$ ), but AAS and citations were positively correlated ( $r = 0.5200$ ,  $p = 0.0006$ ). Similarly peer reviewed articles also did not have any correlation between AAS and days to retraction ( $r = -0.0918$ ,  $p = 0.2756$ ) and between citations and days to retraction ( $r = 0.9000$ ,  $p = 0.2850$ ). But there was a significant correlation between AAS and citations ( $r = 0.5909$ ,  $p = 0.0001$ ).

Figure 1 (B) and (C) show that concerns/issues about data and results were the commonest cause for retraction of both pre-print as well as peer-reviewed articles. Reasons for retraction of pre-prints and peer-reviewed articles by article type have been detailed in Supplementary Tables 3 and 4, respectively.

## Discussion

The surge in spurious articles on COVID-19 has made it imperative to develop methods for the rapid identification and retraction of such studies. The key finding of our study was that the pre-print articles had a statistically significant smaller delay in retraction compared to peer-reviewed articles. Preprint servers have increased the speed with which scientific information can be submitted, disseminated, and critiqued. Since pre-prints are not paywalled they allow researchers can get valuable feedback from a vast community of peers. (Sarabipour et al., 2019) Even for peer-reviewed articles, previous studies have shown that more retractions happen in open access journals and because of post-publication scrutiny by the scientific community. (Avissar-Whiting, 2022) Additionally, because pre-printed articles are not peer-reviewed, people reading the articles are more critical of the results, methodology, and conclusions which inflates the chances of noticing errors. However, further studies are needed to examine this hypothesis as some have argued that during the pandemic, non-peer-reviewed scientific information was used without caution and scrutiny. (Ravinetto et al., 2021) A study on all biology and medicine-related papers showed that downstream retraction of pre-printed articles is faster. (Avissar-Whiting, 2022) Considering that the quality of the article doesn't improve much between a preprint and the final peer-reviewed article, it isn't surprising that earlier and easier access to the results for a critical appraisal can lead to faster retractions. (Carneiro et al., 2020) In our study, we found no correlation between the time to retraction and the AAS and citations. Previously it has been shown that retracted articles receive greater social media attention than similar non-retracted articles. (Serghiou et al., 2021) But non-scientific audiences and bots contribute to a large portion of the social media discussions of retracted articles, thus explaining the lack of a correlation between the time to retraction and the AAS. (Abhari et al., 2022)



Another interesting point to note is the difference in the causes of retraction between preprints and peer-reviewed articles. Some peer-reviewed articles have been retracted for reasons like fake peer review and duplication of articles, which exposes the faults of the peer-review process. A non-peer-reviewed or duplicate article that is vetted as an original peer-reviewed article can spread scientific misinformation, potentially easier than preprints that are mentioned to be non-peer-reviewed. On the other hand, concerns/issues with the results were the most common cause of retraction of preprints.

Our study has some limitations. Firstly this was an exploratory study aimed to compare preprint and peer-reviewed retractions of COVID-19-related articles. However, our study design and scope are insufficient to conclude causality, and further research is needed to understand the causes leading to a delay in retractions. Also, we did not have longitudinal data of the AAS and citations to compare pre and post-retraction differences. Possibly some retracted articles might have gotten attention after retraction leading to an increase in AAS post retraction.

## **Conclusion**

The present study shows that there is another benefit to encouraging people to put up their work on preprint servers in the form of earlier recognition of errors and open review leading to prompt retractions. As we go forward, the scientific community should promote preprints as they provide wider and faster reach, more attention, and rapid peer feedback. Importantly, future studies should aim at developing methods for the rapid identification of studies requiring retractions.

## References:

- Abhari, R., Vincent, N., Dambanemuya, H. K., Bodon, H., & Horvát, E.-Á. (2022). *Twitter Engagement with Retracted Articles: Who, When, and How?* (arXiv:2203.04228). arXiv. <https://doi.org/10.48550/arXiv.2203.04228>
- Avissar-Whiting, M. (2022). Downstream retraction of preprinted research in the life and medical sciences. *PLOS ONE*, *17*(5), e0267971. <https://doi.org/10.1371/journal.pone.0267971>
- Boschiero, M. N., Carvalho, T. A., & Marson, F. A. de L. (2021). Retraction in the era of COVID-19 and its influence on evidence-based medicine: Is science in jeopardy? *Pulmonology*, *27*(2), 97–106. <https://doi.org/10.1016/j.pulmoe.2020.10.011>
- Carneiro, C. F. D., Queiroz, V. G. S., Moulin, T. C., Carvalho, C. A. M., Haas, C. B., Rayê, D., Henshall, D. E., De-souza, E. A., Amorim, F. E., Boos, F. Z., Guercio, G. D., Costa, I. R., Hajdu, K. L., Van Egmond, L., Modrák, M., Tan, P. B., Abdill, R. J., Burgess, S. J., Guerra, S. F. S., ... Amaral, O. B. (2020). Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Research Integrity and Peer Review*, *5*(1). <https://doi.org/10.1186/s41073-020-00101-3>
- Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>
- Dimensions*. (n.d.). Retrieved April 19, 2022, from <https://app.dimensions.ai/discover/publication>
- Discover the attention surrounding your research – Altmetric*. (n.d.). Retrieved January 31, 2022, from <https://www.altmetric.com/>

- Else, H. (2020). How a torrent of COVID science changed research publishing—In seven charts. *Nature*, 588(7839), 553–553. <https://doi.org/10.1038/d41586-020-03564-y>
- King, A. (2020). Fast news or fake news? *EMBO Reports*, 21(6), e50817. <https://doi.org/10.15252/embr.202050817>
- Nicholas Fraser & Bianca Kramer. (2021). *Covid19\_preprints*. <https://doi.org/10.6084/m9.figshare.12033672.v58>
- Peterson, C. J., Alexander, R., & Nugent, K. (2022). COVID-19 article retractions in journals indexed in PubMed. *The American Journal of the Medical Sciences*, 364(1), 127–128. <https://doi.org/10.1016/j.amjms.2022.01.014>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. (n.d.).
- Ravinetto, R., Caillet, C., Zaman, M. H., Singh, J. A., Guerin, P. J., Ahmad, A., Durán, C. E., Jesani, A., Palmero, A., Merson, L., Horby, P. W., Bottieau, E., Hoffmann, T., & Newton, P. N. (2021). Preprints in times of COVID19: The time is ripe for agreeing on terminology and good practices. *BMC Medical Ethics*, 22(1), 106. <https://doi.org/10.1186/s12910-021-00667-7>
- Retraction Watch Database. (n.d.). Retrieved January 31, 2022, from <http://retractiondatabase.org/RetractionSearch.aspx?AspxAutoDetectCookieSupport=1>
- Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., & Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLOS Biology*, 17(2), e3000151. <https://doi.org/10.1371/journal.pbio.3000151>
- Serghiou, S., Marton, R. M., & Ioannidis, J. P. A. (2021). Media and social media attention to retracted articles according to Altmetric. *PLOS ONE*, 16(5), e0248625. <https://doi.org/10.1371/journal.pone.0248625>

Shimray, S. R. (2021). Research done wrong: A comprehensive investigation of retracted publications in COVID-19. *Accountability in Research*, 0(0), 1–14.  
<https://doi.org/10.1080/08989621.2021.2014327>

**Table 1:** Characteristics of retracted articles

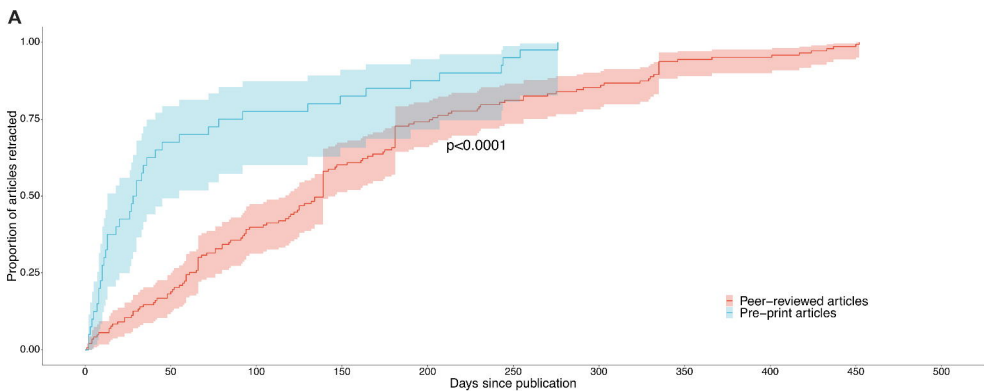
<b>Variables</b>	<b>Peer-reviewed (n=143)</b>	<b>articles</b>	<b>Pre-print Articles (n=40)</b>	<b>P-value</b>
<b>Year of publication</b>				
2020	91 (64.64)		28 (70.00)	0.860
2021	49 (34.27)		12 (30.00)	
2022	4 (2.10)		0 (0.00)	
<b>Country of 1<sup>st</sup> Author</b>				
China	26 (18.18)		5 (12.50)	0.054
United States	21 (14.69)		9 (22.50)	
Malta	27 (18.88)		0 (0.00)	
India	16 (11.19)		5 (12.50)	
Others	54 (37.76)		21 (52.5)	
<b>Article Type</b>				
Research Article	80 (55.94)		35 (87.50)	0.001
Review Article	23 (16.08)		0 (0.00)	
Conference Abstract/Paper	13 (9.09)		0 (0.00)	
Clinical Study	8 (5.59)		4 (10.00)	
Case Report	6 (4.20)		0 (0.00)	
Meta-Analysis	5 (3.50)		0 (0.00)	
Letter	5 (3.50)		0 (0.00)	
Commentary/Editorial	3 (2.10)		1 (2.50)	
<b>Citations</b>	3		3	0.563
<b>[median(IQR)]</b>	(0-17)		(0-14)	
<b>Altmetric Attention Score</b>	8		26.50	0.008
<b>[median(IQR)]</b>	(1-38.5)		(4-1155)	
<b>Number of days to retraction [median(IQR)]</b>	139 (63-202)		29 (10-81.5)	0.0001

Note: Data are shown as count (%) unless otherwise specified.

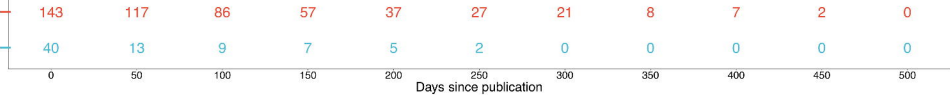
IQR: Inter-quartile range

## Figure Legend

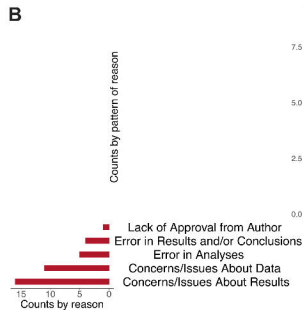
**Fig 1 (A)** Time to retraction of pre-print and peer-reviewed articles. Upset plot showing five commonest reasons for retraction for pre-print **(B)** and peer-reviewed **(C)** publications



Number of articles to be retracted



**B**



**C**

