

A comprehensive analysis of articles submitted to preprint servers from one laboratory (VKPrasad Lab at UCSF): Download statistics, rates of rejection, and reasons for rejection: Are preprint servers acting fairly or playing politics?

Alyson Haslam, PhD,¹ Vinay Prasad, MD, MPH¹

¹University of California San Francisco, 550 16th St, 2nd Fl, San Francisco, CA 94158

Running title: Outcomes of papers submitted to preprint servers

Corresponding author:

Alyson Haslam, PhD

Department of Epidemiology and Biostatistics

UCSF Mission Bay Campus | Mission Hall: Global Health & Clinical Sciences Building |

550 16th St, 2nd Fl, San Francisco, CA 94158

(E-mail: alyson.haslam@ucsf.edu)

Disclosure: Vinay Prasad's Disclosures. (Research funding) Arnold Ventures (Royalties) Johns Hopkins Press, Medscape, and MedPage (Honoraria) Grand Rounds/lectures from universities, medical centers, non-profits, and professional societies. (Consulting) UnitedHealthcare and OptumRX. (Other) Plenary Session podcast has Patreon backers, YouTube, and Substack. Alyson Haslam has no disclosures to report.

Funding: None

Word count: 1295

Number of tables: 1

Number of Figures: 0

Abstract

Introduction: Preprint servers have become an increasingly popular way to disseminate scientific information, in part because research articles can be published faster on these servers than via traditional peer-reviewed avenues. While there is no formal peer-review with preprint articles, preprint servers often have a vetting process for published articles, which lacks transparency.

Purpose: We sought to evaluate the submission process of preprint servers by assembling a comprehensive list of articles submitted to these servers and noting their fate.

Methods: We included all articles submitted to SSRN, medRxiv, and Zenodo and that arose from the VKPrasad Laboratory (www.vkprasadlab.com), a health policy and epidemiology lab at UCSF.

Results: Of 16 unique submissions, 6 (38%) resulted in articles being rejected or removed. 4 of those rejected were initially submitted to SSRN and two were initially submitted to medRxiv. All removed articles were on the topic of COVID. Three (50% of rejected/removed articles) were eventually accepted at another preprint server. The median number of downloads for a rejected/removed article that was later accepted by a different server was 4142. The median time from submission to acceptance was 2 days and 4 days for submission to decision of rejection.

Discussion: The submission and acceptance process for preprint servers appears to have inconsistent standards and be a non-transparent process. These servers appear to have a more stringent vetting process for articles on COVID topics, but because of the novelty of the virus, there are fewer absolutes about what is known, suggesting that a free exchange of scientific information is being stifled.

Introduction

Preprint servers provide researchers with an avenue to quickly disseminate study findings and research ideas. The publication of scientific information on preprint servers is now considered part of the mainstream publishing process,¹ especially as large medical journals integrate preprint posting as part of their journal's submission process. The utility and popularity of preprint servers was expedited during the COVID pandemic when the need for scientific understanding of a novel virus required faster communication than traditional avenues.²

Posted articles are not formally peer-reviewed, although they do go through a post publication peer-review, as readers can leave comments on the server and on social media, where these articles can be broadly shared. Yet, preprint websites have a vetting process, which lacks transparency, only after which articles can be posted.

We sought to evaluate the submission process of preprint servers by assembling a list of articles submitted to these servers and noting their fate. We focused on submissions by the VKPrasad Laboratory at the University of California San Francisco, as we had access to all communications. The lab is a productive health/oncology policy research lab with ~80 publications in peer reviewed journals the last two years. Full details available at www.vkprasadlab.com.

Methods

We built a set of all articles that our laboratory (vkprasadlab.com) has written and submitted to preprint servers. We included all articles submitted to SSRN, medRxiv, and Zenodo, regardless of date or topic. We searched the author centers of the preprint websites and historical work emails (July 24, 2023) for instances of original articles (including research, commentaries, and reviews) that were submitted to the preprint servers, noting dates, correspondence, and outcomes of these submissions. All included articles had to have Dr. Vinay Prasad as an author. From the preprint servers' websites, we searched for each published article and noted the number of views and downloads (as of July 24, 2023).

We conducted all analysis in Microsoft Excel. In accordance with 45 CFR §46.102(f), this study was not submitted for University of California, San Francisco institutional review board approval because it involved publicly available data and did not involve individual patient data.

Results

We found 16 instances of unique articles being initially submitted to one of these servers (7 to SSRN, and 9 to medRxiv). The table lists the characteristics of these submissions, including those with multiple submissions. 10 articles were on the topic of COVID, 5 were on the topic of cancer, and 1 was on the topic of monkeypox.

6 of these articles (38%) were removed from submission; 4 were initially submitted to SSRN and two were initially submitted to medRxiv. All removed articles were on the topic of COVID. Three of the removed articles were later submitted to and accepted at one of the other preprint servers – 1 to SSRN, 1 to med Rxiv, and 1 to Zenodo. For the 4 studies removed by SSRN, a general explanation of, "Given the need to be cautious about posting medical content, SSRN is selective on the papers we post" was provided. The explanations for article removal by medRxiv were "it is not a systematic evaluation with

reproducible methodology” and “medRxiv is intended for research papers, and our screening process determined that this manuscript fell short of that description.”

For accepted articles, the median time from submission to acceptance and posting was 2 days (range: 0-12 days). For removed articles, the median time from submission to rejection decision was 4 days (range: 1-27 days). The median number of downloads for a rejected/removed article that was later accepted by a different server was 4142 vs 300 for articles submitted and accepted without rejection or removal.

Discussion

While preprint servers do not have a formal peer-review process, it is unclear how posted articles are selected. MedRxiv’s vetting process is set up to identify articles that could cause harm, which is inherently subjective, particularly on topics (e.g., COVID19 policy) where there remains sizable uncertainty and important scientific debate.

The preprint server also flags articles that “might contradict widely accepted public-health advice.”³ Because of the novelty of COVID, public health advice has been mainly based on small, biased studies. For example, the CDC’s recommendation for mask mandates initially stemmed from a study of 2 masked hairdressers.^{4,5} Later, a Cochrane review of multiple randomized studies concluded that the evidence for face masks for COVID protection was uncertain.⁶ Thus, even widely promoted public health advice can be fallible, and repressing studies that contradict guidance based on weak evidence, even if generally accepted, is concerning. It is concerning that a preprint server would defer to governmental agencies, which may be captured by political processes and not performing the best possible science. Moreover, a preprint server can thus be used to quell opposing points of view.

According to medRxiv screening process, our article on Statistical Methodological errors was “not a systematic evaluation with reproducible methodology.” Yet, this article was posted on SSRN, where it has received almost 40,000 downloads to date. Moreover, a number of articles accepted by medRxiv appear to have similar methodology. For example, in one posted medRxiv article, a study sample included 9 COVID-positive people “who they [community healthcare workers] knew,”⁷ which is biased methodology and unlikely to be reproducible. This example not only shows the inconsistent standards for allowing articles to be posted, but the decision to post based on the quality of methodology should be determined through peer-review, by experts in the field.

Based on these observations, there appears to be greater vetting among articles on the topic of COVID, rather than other topics. Some have argued for the need to reign in misinformation during COVID, yet, ironically, a counter prevailing attitude is that because of the novelty of COVID, measures needed to be implemented before the topics could be fully studied. Understandably there has been equipoise about COVID issues, and the way to come to consensus on these issues is to have open dialogue in the scientific community, not to censor and subdue information.

Two instances of articles being rejected/removed concerned earlier versions of this analysis, and we have since updated our results. The reason this analysis was rejected from medRxiv is that the preprint server did not consider it scientific research, and SSRN provided the same reasoning as it did for other removed articles. Further explanation was not provided, but we would argue that our analysis is scientific research. Scientific research has been defined as, “the systematic collection, interpretation and

evaluation of data.”⁸ Our analysis 1) presents on a scientific-related topic (dissemination of scientific research) and 2) is a systematic evaluation of multiple observations, which is stronger methodology than case reports that are regularly reported in the literature. Moreover, medRxiv has allowed other analyses on similar topics to be posted.⁹ It is hard to escape the conclusion that preprint servers do not allow criticism of their own processes on their servers.

Our analysis has several limitations. First, these findings may not be generalizable to the experience of other researchers, as some may prefer to publish findings that align only with the current dogma and popular beliefs. Second, our experience was during the COVID pandemic, and experience during a non-pandemic time may be different. However, our findings suggest a concerning pattern of biased screening. Third, the justifications for removing articles were vague, and it is difficult to know the exact reasoning behind their decisions. When we did reach out for clarification, our efforts were unfruitful.

Conclusion

In conclusion, the submission and acceptance process for preprint servers appears to have inconsistent standards and be a non-transparent process. These servers appear to have a more stringent vetting process for articles on COVID topics, but because of the novelty of the virus, there are fewer absolutes about what is known, suggesting that a free exchange of scientific information is being stifled.

References

1. Hoy MB. Rise of the rxivs: how preprint servers are changing the publishing process. *Med Ref Serv Q*. 2020;39(1):84-89. doi:10.1080/02763869.2020.1704597
2. Fraser N, Brierley L, Dey G, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biol*. 2021;19(4):e3000959. doi:10.1371/journal.pbio.3000959
3. Kwon D. How swamped preprint servers are blocking bad coronavirus research. *Nature*. 2020;581(7807):130-131. doi:10.1038/d41586-020-01394-6
4. NETBURN D. Timeline: CDC mask guidelines during the COVID pandemic. *Los Angeles Times*. July 27, 2021.
5. Centers for Disease Control and Prevention. How to Protect Yourself & Others. July 13, 2020. Accessed June 26, 2023. https://web.archive.org/web/20200424080741/https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprepare%2Fprevention.html
6. Jefferson T, Dooley L, Ferroni E, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Libr*. January 30, 2023.
7. Bresser M, Erhardt RM, Shanaube K, et al. Evaluation of Covid-19 Ag-RDTs self-testing in Lesotho and Zambia. *medRxiv*. December 23, 2022. doi:10.1101/2022.12.21.22283827
8. Çaparlar CÖ, Dönmez A. What is Scientific Research and How Can it be Done? *Turk J Anaesthesiol Reanim*. 2016 Aug;44(4):212-218. doi: 10.5152/TJAR.2016.34711. Epub 2016 Aug 1. PMID: 27909596; PMCID: PMC5019873.
9. Raju V, Adakawa MI, Harinarayana NS, Yelemalla C. Indian COVID-19 Preprints Submissions in bioRxiv and medRxiv Preprint Servers. *medRxiv*. 2023:2023-02.

Table. Instances of articles submitted for posting on preprint servers at a productive health policy research lab at a top research university.

Title	Date submitted	Decision	Views/downloads (including full-text views)
SSRN			
Mask mandates and COVID-19: A Re-analysis of the Boston school mask study	7/19/2023	Removed – “Given the need to be cautious about posting medical content, SSRN is selective on the papers we post.”	
A comprehensive analysis of articles submitted to preprint servers from one laboratory (VKPrasad Lab at UCSF): Download statistics, rates of rejection, and reasons for rejection: Are preprint servers acting fairly or playing politics?	7/18/2023	Removed – “Given the need to be cautious about posting medical content, SSRN is selective on the papers we post.”	
Falsification Endpoints and the Pitfalls of Using Observational Studies in Reviews of Effectiveness of COVID-19 Vaccines: Critique of a Systematic Review and Meta-Analysis of BNT161b2 vaccination of 5-11-Year-Olds	4/19/2023	Removed – “Given the need to be cautious about posting medical content, SSRN is selective on the papers we post.”	
Interpretation of Wide Confidence Intervals in Meta-Analytic Estimates: Is the ‘Absence of Evidence’ ‘Evidence of Absence’?	4/13/2023	Removed – “Given the need to be cautious about posting medical content, SSRN is selective on the papers we post.”	
Statistical and Numerical Errors Made by the US Centers for Disease Control and Prevention During the COVID-19 Pandemic	3/14/2023	Accepted	40213/7959

COVID-19 vaccines: history of the pandemic's great scientific success & flawed policy implementation	11/14/2022	Removed – “Given the need to be cautious about posting medical content, SSRN is selective on the papers we post.”	3072/1028*
COVID-19 Vaccine Boosters for Young Adults: A Risk-Benefit Assessment and Five Ethical Arguments against Mandates at Universities	8/31/2022	Accepted	417488/90252
A Systematic Analysis of Post-Protocol Therapy in First Line Checkpoint Inhibitor Trials	10/29/2021	Submitted by journal	283/19
Why Is Research in Early-Stage Cancer Research so Low? A Re-Assessment of Budish, Roin and Williams	6/6/2017	Accepted	2454/239
medRxiv			
A comprehensive analysis of articles submitted to preprint servers from one laboratory (VKPrasad Lab at UCSF): Download statistics, rates of rejection, and reasons for rejection: Are preprint servers acting fairly or playing politics?	7/14/2023	Removed - "medRxiv is intended for research papers, and our screening process determined that this manuscript fell short of that description"	
Interpretation of wide confidence intervals in meta-analytic estimates: Is the 'Absence of Evidence' 'Evidence of Absence'?	7/11/2023	Accepted	277/52
Changes in Masking Policies in US Healthcare Facilities in the First Quarter of 2023: Do COVID-19 Cases, Hospitalizations, or Local Political Preferences Predict Loosening Restrictions?	7/11/2023	Accepted	502/435
Characteristics and quality of studies pertaining to masks published in the Morbidity and Mortality Weekly Report	7/11/2023	Accepted	2148/1797
Analysis of tweets discussing the risk of Mpox among children and young people in school (May-Oct 2022): Public health experts on Twitter consistently exaggerated risks and infrequently reported accurate information	5/16/2023	Accepted	1364/920

Cross-sectional analysis of Open payments for physicians at designated hemophilia centers in the US (2018-2020)	3/7/2023	Submitted by journal	659/92
Current landscape of disparity-focused research: a bibliometric analysis of 260 research articles	3/7/2023	Accepted	510/133
Statistical and numerical errors made by the US Centers for Disease Control During the COVID-19 Pandemic	3/6/2023	Removed – “it is not a systematic evaluation with reproducible methodology”	
An empirical analysis of lay media coverage on influenza prevention pre- and post-COVID 19: Mask recommendations were previously rare, now ubiquitous	2/14/2023	Accepted	1306/202
Estimation of time cost of anti-cancer drugs approved based on comparisons to best supportive care: a cross sectional analysis medRxiv	6/22/2022	Submitted by journal	1090/361
Zenodo			
COVID-19 vaccines: history of the pandemic's great scientific success & flawed policy implementation	12/6/2022	Accepted	6953/4142

*This was able to be viewed publicly for 18 hours before it was removed by SSRN.