



Pandemic scientific data sharing recommendations: examining and re-imagining pre-print servers after the end of the world-wide emergency

Shira Doron¹ o and Westyn Branch-Elliman^{2,3} o

¹Division of Geographic Medicine and Infectious Diseases, Department of Medicine, Tufts Medical Center, Boston, MA, USA, ²Department of Medicine, VA Boston Healthcare System, Boston, MA, USA and ³Harvard Medical School, Boston, MA, USA

Abstract

Early in the pandemic, pre-print servers sped rapid evidence sharing. A collaborative of major medical journals supported their use to ensure equitable access to scientific advancements. In the intervening three years, we have made major advancements in the prevention and treatment of COVID-19 and learned about the benefits and limitations of pre-prints as a mechanism for sharing and disseminating scientific knowledge.

Pre-prints increase attention, citations, and ultimately impact policy, often before findings are verified. Evidence suggests that pre-prints have more spin relative to peer-reviewed publications. Clinical trial findings posted on pre-print servers do not change substantially following peer-review, but other study types (e.g., modeling and observational studies) often undergo substantial revision or are never published.

Nuanced policies about sharing results are needed to balance rapid implementation of true and important advancements with accuracy. Policies recommending immediate posting of COVID-19-related research should be re-evaluated, and standards for evaluation and sharing of unverified studies should be developed. These may include specifications about what information is included in pre-prints and requirements for certain data quality standards (e.g., automated review of images and tables); requirements for code release and sharing; and limiting early postings to methods, results, and limitations sections.

Academic publishing needs to innovate and improve, but assessments of evidence quality remains a critical part of the scientific discovery and dissemination process.

Keywords: pre-prints; data dissemination; public health emergency; pandemic response; data sharing; COVID-19

(Received 20 April 2023; accepted 14 June 2023)

Background

Throughout the COVID-19 pandemic, dramatic press headlines about new pandemic "discoveries" related to the novel virus have driven practice carried out by clinicians, fear experienced by the public, and sometimes policy enacted by public health and governmental authorities. For the first time on a large scale, reports of scientific findings were commonly disseminated based on information gleaned from unpublished observational, laboratory, and modeling studies posted on pre-print servers without vetting or substantive quality evaluations.

Early in the pandemic, pre-print servers played an important role in speeding the rapid dissemination of evidence and allowed for widespread access to findings by the public and those without access to published articles behind a paywall. Recognizing the importance of emerging research to guide practice, a collaborative

Corresponding author: Westyn Branch-Elliman; Emails: wbranche@bidmc.harvard. edu; westyn.branch-elliman@va.gov

Cite this article: Doron S, Branch-Elliman W. Pandemic scientific data sharing recommendations: examining and re-imagining pre-print servers after the end of the world-wide emergency. *Antimicrob Steward Healthc Epidemiol* 2023. doi: 10.1017/ash.2023.410

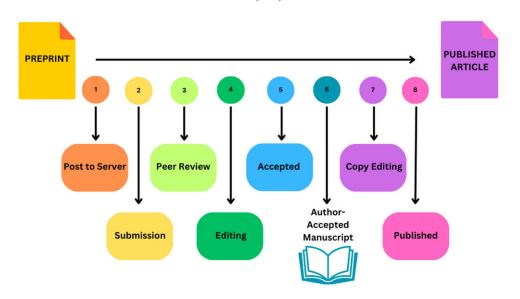
of major medical journals supported the use of pre-print servers to ensure equitable access to scientific advancements.¹

Now that the WHO has declared the end of the COVID-19 public health emergency,² should the practice of sharing and disseminating these unvetted studies continue? Will pre-prints be one of the pandemic innovations that stands the test of time? If so, how does the pre-print process need to evolve to safely support clinical advancements?

Pre-print servers: background and evolution

Preprint servers are sites where authors can post their scientific papers prior to submission, review, or acceptance in a traditional journal (Fig. 1). They existed prior to the pandemic, but their use was limited, and they were not vehicles for widespread dissemination of clinical or public health studies. The first preprint server, arXiv, was created in 1991 and covered non-medical sciences. Servers dedicated to medical and health sciences (such as medRxiv) and life sciences (such as bioRxiv) were established much later but did not rocket to popularity or serve as major information sources until the pandemic began. The most commonly used pre-print servers are not-for-profit and provide varying levels of basic screening prior to approving manuscripts

© VA, 2023. This is a work of the US Government and is not subject to copyright protection within the United States. Published by Cambridge University Press on behalf of The Society for Healthcare Epidemiology of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4. 0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



What is a preprint?

Figure 1. Steps in the academic publishing process.

(which must be original research, not case reports, editorials, or narrative reviews). Authors may submit revised versions of the same manuscript at any time, and these are tracked by the system, providing a historical record of the major changes. Some pre-print servers have affiliations with journals, particularly those that signed on to the collaborative mentioned above, facilitating direct transfer of files and saving authors time during the submission process.

The global pandemic: a catalyst for speeding and transforming data dissemination

As the year 2020 dawned, the established mechanisms by which research findings are made available were too slow to keep pace with the needs of the healthcare and scientific communities. Even for studies that go on to be accepted, the editorial process for publication in a peer reviewed journal, which includes multiple reviews and revisions, easily takes many months, and sometimes more than a year. Clearly, with a deadly and unknown disease ravaging the globe, and new discoveries being made every day, business as usual would have been inadequate and inappropriate. The pre-print servers offered a solution to these problems.

For the first time, pre-print servers became important sources of medical information, with the papers therein amplified in traditional and non-traditional news outlets and even included in deliberations about important public health decisions, like school policy, mask mandates, and vaccine recommendations. Studies posted on pre-print servers impacted citation and alt-metric scores—two measures of research impact—and some had substantial impacts on clinical care decisions and policy without high-quality—or occasionally even verified—data to back up their claims.^{3,4}

Pre-prints: the good, the bad, the ugly

The good

There are a number of benefits associated with posting manuscripts to pre-print servers prior to publication, including (1) the findings are immediately widely available for viewing and citing and, unlike many journal sites, without a pay wall; (2) feedback received from the research community and (unlike what scientists might otherwise get when they present their work at conferences) the general public can be used to improve the quality of the manuscript prior to submission to journals and/or during the revision process; (3) leveraging the timestamp on the posting can avoid disputes about originality of research or the timing of findings compared to that of other researchers, and (4) because the sites include information about version numbers and revisions, they can serve as a mechanism for promoting transparency and accountability. As noted above, authors benefit from increased visibility and alt-metric scores, which are a measure of publication impact. Researchers can also benefit by using pre-print servers as a mechanism for sharing preliminary research findings used to support grant proposals.

The bad

Although dissemination of COVID-19-related research through pre-prints was widely accepted and adopted before major therapeutic and preventative milestones were reached, the approach has many limitations, which have important implications for public health and science communication. Quality checks are limited, and there is no standardized process for ensuring accuracy. Among clinical trials that are ultimately published, findings do not change substantially following peerreview, although presentation of results tends to be more complete and with less perceived spin.^{5,6} However, the same cannot be said for other study designs, or for clinical trial results that never go on to get published. A recent study found that one in five randomized controlled trials remained unpublished 12 months after posting, and those that were not published were less complete and more highly spun than those that had undergone the peer review process and were ultimately published in a journal.⁷ Findings are even more stark for other study designs. Although published literature on the topic is inconsistent, estimates are that only approximately two-thirds of pre-prints go on to get published⁸ of these, about 17% undergo major changes during the revision process,⁹ meaning that almost *half* of the papers posted to pre-print servers are either changed substantially or never make it through peer-review.

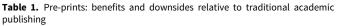
The ugly

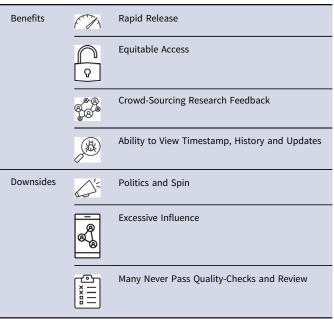
Anchoring bias is a well-described concept in psychology and refers to the strong tendency of humans to weight the first piece of information they receive more heavily than future information.¹⁰ Disclaimers are insufficient to overcome this basic human instinct. Many early reports are later debunked by better quality research, and promising interventions tested in laboratory or modeling settings often are found to be ineffective when tested in human populations. This tension is amplified because laboratory, modeling, and simulation studies are inherently faster to conduct. Human subjects are protected, drugs are regulated before they can be tested, and recruitment, outcomes assessments, and data analysis all take time.

Once data are available, findings and early presentations are widely shared on social and traditional media outlets, impacting perceptions and, at times, policy-making decisions. Pre-print servers do not create these information dissemination problems, but they do accelerate them. Thus, even updated policy about the use of pre-print servers is unlikely to be sufficient on its own to change the problems posed by factors inherent to the speed of the scientific discovery process or to how data are shared and discussed in public forums.

Pre-prints: empirical evidence of impact

These concerns are not just theoretical but supported by empirical evidence. Likely driven in part by anchoring bias, despite these well-documented limitations of different study designs and analyses, by the time contradictory or higher-quality evidence became available, the first report has already made its lasting mark. For example, a study by one of the authors (WBE) measured the speed and scope of COVID-19 prescribing practices in the Veterans Health System and found that practice patterns changed rapidly after data release via pre-print servers and press releases, with limited additional change after peer-reviewed publication.¹¹ Effects were strongest early in the pandemic, when there were few treatment options, and waned as time went on. Studies on the efficacy of the bivalent COVID-19 vaccine in human subjects have been presented at regulatory meetings before peer review, allowing for expedited authorization and guideline development, potentially benefiting society by reducing delays.^{12,13} On the other hand, even after a pre-print purportedly demonstrating efficacy was withdrawn from the server due to serious ethical infractions,¹⁴ ivermectin has continued to be prescribed despite a preponderance of evidence demonstrating lack of efficacy.^{15,16} Other examples of practice changes driven by early studies with major limitations include "gaiter-gate".^{17,18} Despite rapidly available contradictory evidence, the damage was done. Gaiters were widely banned in schools based on an early report that they were inferior to other cloth masks and subsequent dissemination via traditional and social media sources.^{19,20} In another example, a research group using publicly available CDC data published a pre-print in May 2022 that calculated that COVID-19 was the 4th highest cause of death in children under the age of 19.²¹ This analysis was shared in both the FDA's VRBPAC meeting and the CDC's ACIP meeting to inform discussion about COVID-19 vaccination policy for children. Findings were also widely shared in traditional and social media outlets. A revised version of the pre-print,²² however, listed COVID-19 as the 7th highest cause of death, and in the final publication, COVID-19 was calculated to be the 8th highest cause





of death.²³ During the revision process, the estimated crude mortality rate of COVID-19 for children changed quite dramatically, from 7.2 down to 1.0 per 100,000 from the first version of the pre-print to the final, peer-reviewed published version of the manuscript.

Evolving pre-prints to meet current needs and conditions

Recognizing the limitations of pre-prints, their benefits, and their real-world impacts raises important questions about how the use of pre-print servers for disseminating medical evidence should evolve as we move into the next phase of living with COVID-19 (Table 1).²⁴ First, given the aforementioned data demonstrating that one of the biggest impacts of peer-review is to change framing and reduce spin, a pre-print policy change that could be trialed is presentation of methods, results, and limitations sections only without background and discussion sections, which are fundamentally more prone to personal opinion and political viewpoints. Another possibility is for pre-print servers to be re-organized, such that limitations are a required element that are presented before the results section. This would ensure that limitations are highlighted so that the public can view and comment on them; reading limitations first might also change the reader's perceptions about the implications of the results, partially (if not completely) addressing anchoring bias tendencies. Additional requirements for quality checks and release of underlying methodology (e.g., by requiring code to be shared with the release, or publication of models so that others can review) would also be options for improving pre-print transparency and evaluations. Development and application of advanced technology to identify falsified images, calculation errors, and plagiarism are additional considerations that should be considered, supported, and evaluated to improve evidence quality.

Evolving existing academic publishing and data sharing

The influence of pre-prints has not occurred in a vacuum. They are one piece of the puzzle and changes at many levels of the system are

a release and sharing

Speed	 Clinical trials prioritized for rapid review Laboratory, observational, and modeling studies undergo traditional vetting processes
Quality	 Develop systems for assessing quality (e.g., automated systems for reviewing images and text to confirm authenticity) Automated solutions to review tables, etc. for accuracy and reliability Requirements for releasing underlying protocols or code used to conduct the research
Anchoring	 Evaluation of novelty of findings and potential public health impact Limitations specified upfront
Spin	 Prereviewed releases focus on methods, results and limitations Background sections and discussion sections not included Consider open peer-review processes
Equity	 Key study findings universally available free of charge Develop new models for traditional publications (e.g., publish methods, results, and limitations) even if full openaccess model is not supported
Impact	 Evidence standards for traditional media release Requirements for quality checks prior to sharing Guidelines about completeness of data and methods prior to publication and sharing

needed to improve data dissemination. Before pre-prints are widely shared by influencers, covered by the media, and communicated to the public as "truth," uncertainty about the study's eventual conclusions and implications of the work needs to be acknowledged by those who are responsible for sharing and spreading the information.

Although pre-prints in their current iteration may not be the ultimate solution, innovation in medical publishing and dissemination of research findings is badly needed. Academic publishing is strongly biased toward publishing papers-and questioning findings less-if authored by those who are already "famous."25 Peer review is slow, biased, and often does not lead to major changes even when they are needed,²⁶ as occurred with the Surgisphere debacle, in which the underlying data supporting the harms of hydroxychloroquine for COVID-19 treatment could not be verified.²⁷ For topics of critical public health importance, rapid review pathways should be adopted to ensure that data from highquality randomized controlled trials are quickly available to inform clinical care decisions. Peer-reviewed manuscripts in traditional journals should be readily available and accessible without a subscription, and access needs to be expanded; Plan S in Europe²⁸ and the White House Office of Science and Technology policy on free, immediate, and equitable access are important policy changes aimed to address these barriers.²⁹ Open access should be the standard, and The Journal of the American Medical Association took big steps in this direction in December 2022.³⁰ The National Library of Medicine could require indexed journals to, at a minimum, make methods, results, and limitations openly accessible to the public and set quality standards to encourage expansion of open-access models.

The financial incentives of academic publishing also inherently create challenges. Academic journals rely on the donated time of unpaid reviewers, which may limit review quality. Journals requesting peer-review are asking for time and expertise—things that outside of academia are always compensated. The practice of donated reviews may be defensible for society journals that do not have publication fees and therefore providing an expert review free-of-charge is a *quid pro quo*. However, given the rise of for-profit journals and major publishing enterprises, some of which boast profit margins higher than those of Apple or Google,³¹ due in part to the free labor and expertise provided by reviewers—the compensation structure, or lack thereof, needs reevaluation. However, despite all of these caveats and limitations, recent data suggest that the speed and quality of peer review may actually have increased during the pandemic.^{32,33}

Next steps

What is the best path forward? Evidence of quality and our willingness to act upon findings need to be balanced (Table 2). Data from pivotal clinical trials should be reviewed and released rapidly, so that clinicians and policy makers can translate evidence into care immediately; recently released data on the harms associated with corticosteroids for inpatients without severe COVID-19 is an example that should immediately change practice.³⁴ For other studies, speed and content quality could be balanced by rapid posting of accepted, but not publication proofed, articles.

Mainstream and social media outlets should wait for observational and laboratory-based research to undergo a vetting process. Denise-Marie Ordway wrote in The Journalist's Resource in April 2020 that there are 6 things journalists must know before covering biomedical research pre-prints about COVID-19:³⁵ (1) pre-prints can be dangerous if doctors change their practice based on the results; (2) pre-prints are not peer reviewed, leaving them without opportunity for other experts to catch errors, argue with the authors' interpretation of findings, or request additional data or analyses; (3) best practice dictates that reporters should be explicit that findings reported in pre-prints are preliminary and, ideally, the opinion of an expert should be included to counter inappropriate conclusions; (4) pre-prints are best covered by experienced science journalists; (5) journalists should check with experts to determine whether findings from a pre-print are even worthy of coverage; (6) pre-prints are sometimes withdrawn, particularly after feedback from the scientific community alerts authors to flaws in their methodology or conclusions.³⁴

Conclusions

Policies recommending immediate posting of COVID-19-related research should be re-considered—case reports, observational, and modeling studies should be made publicly and freely available after an evaluation process and with quality standards. Academic publishing needs to innovate and improve—but evidence of quality assessments remain a critical part of the scientific discovery process. We need to take a step back and digest the data before it is disseminated and implemented. To achieve real improvement, we also need to re-think how and when information is shared.

Acknowledgments. We would like to thank Ms. Hannah Shimer for her assistance with referencing and figures.

Financial support. This study is unfunded. All authors declare no financial conflicts of interest.

Disclaimer. The views expressed are those of the authors, and do not necessarily reflect those of the VA healthcare system or the US Federal Government.

References cited

- "Coronavirus (COVID-19): sharing research data." Wellcome 2020. https:// doi.org/10.1007/s11192-021-03971-6.
- "Statement on the Fifteenth Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Coronavirus Disease (COVID-19) pandemic." World Health Organization 2023. https://www.who.int/news/ item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-internationalhealth-regulations-(2005)-emergency-committee-regarding-the-coronavirusdisease-(covid-19)-pandemic. Accessed May 19, 2023.
- Fraser N., Momeni F, Mayr P, Peters I. The Effect of BioRxiv Pre-prints on Citations and Altmetrics. Preprint, Scientific Communication and Education, 2019. https://doi.org/10.1162/qss_a_00043.
- Conroy G. Pre-prints boost article citations and mentions. Nature Index 2019. https://www.nature.com/nature-index/news/preprints-boost-article-citationsand-mentions
- Zeraatkar D, Pitre T, Leung G, *et al.* Consistency of Covid-19 trial pre-prints with published reports and impact for decision making: retrospective review. *BMJ Med* 2022;1. https://doi.org/10.1136/bmjmed-2022-000309. PMID: 36936583.
- Spungen H, Burton J, Schenkel S, Schriger DL. Completeness and Spin of medRxiv preprint and associated published abstracts of covid-19 randomized clinical trials. *JAMA*, 329(15), 1310–1312.
- Preprint and Associated Published Abstracts of COVID-19 Randomized Clinical Trials. JAMA 2023;329:1310–1312. https://doi.org/10.1001/jama. 2023.1784. PMID: 37071105.
- Llor C, Moragas A, Maier M. Evaluation of publication of COVID-19–related articles initially presented as pre-prints. *JAMA Netw Open* 2022;5:e2245745. https://doi.org/10.1001/jamanetworkopen.2022.45745. PMID: 36480205.
- Abdill RJ and Blekhman R. Tracking the Popularity and Outcomes of All BioRxiv Pre-prints: Preprint, Scientific Communication and Education, 2019. https://doi.org/10.7554/eLife.45133. PMID: 31017570.
- 10. Miller N. How different are pre-prints from their published versions? 2 studies explore. *The Journalist's Resource*, 2022.
- "Anchoring Bias." The Decision Lab. https://thedecisionlab.com/biases/ anchoring-bias. Accessed 5 June 2023.
- La J, Fillmore NR, Do NV, et al. Factors associated with the speed and scope of diffusion of COVID-19 therapeutics in a nationwide healthcare setting: a mixed-methods investigation. *Health Res Policy Sys* 2022;20:134. https:// doi.org/10.1186/s12961-022-00935-x. PMID: 36517793.
- Chalkias S., Harper C, Vrbicky K, et al. A bivalent omicron-containing booster vaccine against Covid-19. medRxiv 2022. https://doi.org/10.1101/ 2022.06.24.22276703.
- Wang Q, Hueda-Zavaleta M, Cáceres-DelAguila JA, Muro-Rojo C, De La Cruz-Escurra N, Benítes-Zapata VA. Antibody responses to omicron BA.4/ BA.5 bivalent mRNA vaccine booster Shot. Preprint, *Microbiology*, 2022. https://doi.org/10.1101/2022.10.22.513349.
- Reardon S. Flawed Ivermectin pre-print highlights challenges of COVID drug studies. *Nature* 2021;596:173–174. www.nature.com, https://doi.org/ 10.1038/d41586-021-02081-w.
- Lind JN, Lovegrove MC, Geller AI, Uyeki TM, Datta SD, Budnitz DS. Increase in outpatient Ivermectin dispensing in the US during the COVID-19 pandemic: a cross-sectional analysis. J General Intern Med 2021;36:2909– 2911. Springer Link. https://doi.org/10.1007/s11606-021-06948-6.
- Becker NV, Seelye S, Chua K-P, Echevarria K, Conti RM, Prescott HC. Dispensing of Ivermectin from veterans administration pharmacies during

the COVID-19 pandemic. JAMA Netw Open 2023;6:e2254859. https://doi.org/10.1001/jamanetworkopen.2022.54859. PMID: 36723943

- 18. Chiu A. Wearing a neck Gaiter may be worse than no mask at all, researchers find. Washington Post, 2020.
- Fischer EP, Fischer MC, Grass D, Henrion I, Warren WS, Westman E. Lowcost measurement of face mask efficacy for filtering expelled droplets during speech. *Sci Adv* 2020;6:eabd3083. https://doi.org/10.1126/sciadv.abd3083. PMID: 32917603
- Kielar M. Baldwinsville schools will not allow Neck Gaiters or Bandannas as masks. WSTM, 2020.
- 21. Parker-Pope T. Save the Gaiters! The New York Times, 2020. NYTimes.com.
- Flaxman S, Whittaker C, Semenova E, *et al.* Covid-19 Is a leading cause of death in children and young people ages 0–19 years in the United States. Preprint, *Infect Dis* (except HIV/AIDS), 2022. https://doi.org/10.1101/2022. 05.23.22275458.
- Covid-19 Is a leading cause of death in children and young people ages 0–19 years in the United States. Preprint, *Infect Dis.* (except HIV/AIDS), 2022. https://doi.org/10.1101/2022.05.23.22275458.
- 24. Flaxman, S, Whittaker C, Semenova E, et al. Assessment of COVID-19 as the underlying cause of death among children and young people aged 0 to 19 years in the US. JAMA Netw Open 2023; 6:e2253590. https://doi.org/10. 1001/jamanetworkopen.2022.53590.
- Affairs (ASPA), Assistant Secretary for Public. Fact sheet: COVID-19 public health emergency transition roadmap. *HHS.Gov*, 2023.
- Brainard J. Reviewers award higher marks when a paper's author is famous. Science 2022. https://doi.org/10.1126/science.ade8714. PMID: 36108016
- Janda G, Khetpal V, Shi X, Ross JS, Wallach JD. Comparison of clinical study results reported in MedRxiv pre-prints vs Peer-reviewed journal articles. JAMA Netw Open 2022;5:e2245847. https://doi.org/10.1001/ jamanetworkopen.2022.45847. PMID: 36484989
- Offord C. Lancet, NEJM retract Surgisphere studies on COVID-19 patients. *The Scientist Magazine*[®], TheScientist, 2020.
- 29. Schliltz M. Why Plan S | Plan S. European Science Foundation, 2018.
- 30. Memorandum for the Heads of Executive Departments and Agencies. *Federal Register*, 2021.
- Bibbins-Domingo K, Shields B, Ayanian JZ, et al. Public access to scientific research findings and principles of biomedical research—a new policy for the JAMA network. JAMA Oncol 2023;9:172. https://doi.org/10.1001/ jamainternmed.2022.6493. PMID: 36516051
- 32. Buranyi S. Is the staggeringly profitable business of scientific publishing bad for science? The Guardian, 2017.
- Perlis RH, Kendall-Taylor J, Hart K, *et al.* Peer review in a general medical research journal before and during the COVID-19 pandemic. *JAMA Netw Open*, 2023;6:e2253296. https://doi.org/10.1001/jamanetworkopen.2022. 53296. PMID: 36705922.
- 34. RECOVERY Collaborative Group, *et al.* Higher dose corticosteroids in hospitalised COVID-19 patients with hypoxia but not requiring ventilatory support (RECOVERY): a randomised, controlled, open-label, platform trial. Preprint, *Infect Dis*, (except HIV/AIDS). 2022. https://doi.org/10.1016/ s0140-6736(23)00510-x. PMID: 37060915
- Covering research pre-prints amid the Coronavirus: 6 things to know. The Journalist's Resource, 2020. https://journalistsresource.org/health/medicalresearch-pre-prints-coronavirus/.