Robustness of evidence reported in preprints during peer review

Lindsay Nelson, Honghan Ye, Anna Schwenn, Shinhyo Lee, Salsabil Arabi, B Ian Hutchins

Scientists have expressed concern that the risk of flawed decision making is increased through the use of preprint data that might change after undergoing peer review. This Health Policy paper assesses how COVID-19 evidence presented in preprints changes after review. We quantified attrition dynamics of more than 1000 epidemiological estimates first reported in 100 preprints matched to their subsequent peer-reviewed journal publication. Point estimate values changed an average of 6% during review; the correlation between estimate values before and after review was high (0·99) and there was no systematic trend. Expert peer-review scores of preprint quality were not related to eventual publication in a peer-reviewed journal. Uncertainty was reduced during peer review, with CIs reducing by 7% on average. These results support the use of preprints, a component of biomedical research literature, in decision making. These results can also help inform the use of preprints during the ongoing COVID-19 pandemic and future disease outbreaks.

Information Schoo

l (L Nelson BSc, A Schwenn BSc,
S Lee MSc, S Arabi BSc,
Prof B I Hutchins PhD) **and
Department of Statistics**
(H Ye PhD), **School of Computer,
Data and Information Sciences,
College of Letters and Science,
University of Wisconsin-
Madison, Madison, WI, USA**

Correspondence to:
Prof B Ian Hutchins, Information
School, School of Computer,
Data and Information Sciences,
University of Wisconsin-
Madison, Madison 53706, WI,
USA
bihutchins@wisc.edu

Introduction

As the use of preprint articles substantially increased during the COVID-19 pandemic,[1–5] researchers hypothesised that the use of unreviewed scientific articles could mislead personal and public health decision making. In bypassing the quality control of peer review, improvements to an article that would have been made during review are not incorporated and articles that are of too low quality to be published in a peer-reviewed journal might be communicated to the scientific community and the public.[6–9] Nevertheless, the US Government has encouraged policies that promote the use of preprints to increase scientific communication. When the National Institutes of Health (NIH)[10] announced their policy supporting the citation of preprints, whether biomedical researchers would use preprint articles as the US Government intended (ie, as "a complete and public draft of a scientific document")[10] or whether largely incomplete studies would be published was unclear. The National Library of Medicine also created a preprint policy to include some preprint articles in PubMed.[11]

Currently, little is known about how much the evidence base of scientific articles changes during peer review. A retrospective observational study published in 2020 suggested that data reported in preprints might systematically overestimate epidemiological data, affecting health-care policy decision making.[5] Scientists, journalists, and members of the public can therefore question how much the evidence presented in a preprint might change if it was peer reviewed. This question considers the robustness of evidence base in preprint versions of an article as it assesses the resilience of this evidence to the processes of peer review. In this Health Policy paper, we aimed to quantify how much data reported in preprints change from preprint versions to peer-reviewed, published versions.

Methods

Search strategy

We aimed to identify studies that reported original data on COVID-19 measurements, had a preprint deposited ahead of journal publication, and had a peer-reviewed publication linked to the preprint. The NIH iSearch COVID-19 Portfolio, part of the NIH iCite web service,[12–15] provides links between preprints deposited on the servers arXiv, bioRxiv, ChemRxiv, medRxiv, Preprints.org, Qeios, and Research Square and the version published in a peer-reviewed journal indexed in PubMed.

The NIH iSearch COVID-19 Portfolio uses artificial intelligence, machine learning approaches, and curation by biomedical research experts to comprehensively index preprint and peer-reviewed publications on COVID-19.[16,17] In addition, the COVID-19 Portfolio integrates natural language processing and detailed search engine functionality to allow users to identify articles that contain relevant epidemiological concepts. We searched the COVID-19 Portfolio for items published between Jan 1 and Sept 29, 2020, that contained the following keywords or their synonyms in the title or abstract: basic reproduction rate, incidence, case fatality rate, or infection fatality rate (appendix p 10). Articles that contained more than one keyword were analysed separately.

Data analysis

The curating process followed guidelines to ensure consistency. Curators (LN, AS, SL) first verified that a version of each preprint was available in a peer-reviewed journal. Any articles that the NIH iSearch COVID-19 Portfolio identified had a subsequent version were excluded from analysis if the subsequent version was also a preprint and no journal-published version existed. If the preprint and published versions were suitable, the article version was verified as the initial version of the preprint, not a subsequent version that included changes due to peer review. Only the initial preprint submission was included in the analysis as subsequent updates to the preprint might have incorporated changes because of ongoing peer review, thereby underestimating the amount of change during peer review. All sections of articles, including supplementary materials included in either the preprint, published version, or both, were included in the analysis.
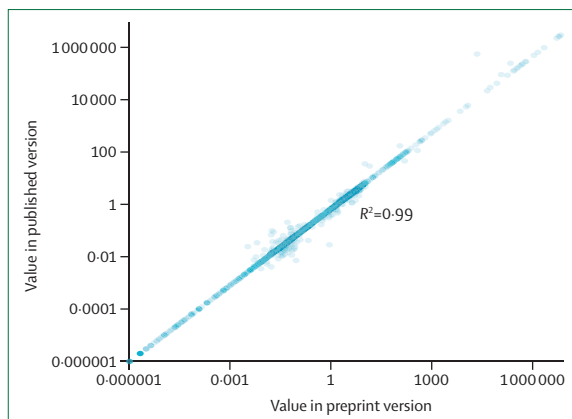
***Figure 1:* Robustness of preprint data during review**
Log scale comparison of epidemiological estimate values reported in preprints *vs* their matched values reported in peer-reviewed publications ($R^2 > 0.99$).

Curators examined identified preprint articles in random order. If there were data present on the search terms of interest (case fatality rate [CFR], incidence fatality rate [IFR], disease incidence, and basic reproduction number [R_0]), point estimates and their CIs were recorded. If a preprint mentioned one estimate (eg, CFR) in the title or abstract but also reported data for a different estimate that was not mentioned in the title or abstract (eg, R_0),[18,19] the data for the additional estimate were excluded as possibly ancillary. The same process was done when analysing the journal article. No data were collected if curators had to guess the value (eg, data were presented in a scatter plot but the estimate was not in the full text). Data points that were repeated (eg, mentioned in the abstract and in a table with accompanying data points) were deduplicated. Articles that reproduced data from another paper were excluded unless they also provided their own estimates.

Peer-review scores
High data alteration during peer review, either in terms of magnitude or censorship, might be expected in papers that were not published because of their low quality. This argument is untestable because the counterfactual peer-reviewed manuscripts do not exist; however, it makes a testable prediction. If article quality affects the outcome of a preprint being published in a peer-reviewed journal, then this outcome should be statistically related to independent peer-review assessments of preprint quality.

To test this hypothesis, we aggregated peer-review scores of preprint quality from the Rapid Reviews: COVID-19 database,[20] which solicits transparent reviews of preprints. These so-called overlay review scores are produced by experts in the research area and include members of the National Academy of Medicine, the University of California, Lawrence Berkeley National Laboratory, and the University of Washington. Overlay peer review is an acknowledged form of quality review: the National Library of Medicine recently began accepting

overlay journals into PubMed for indexing.[21] We averaged the review scores assigned to each preprint, similar to the aggregation of peer-review scores conducted by NIH.[15]

Data transformation and statistics
Percentage data were normalised onto a 0–1 scale for consistency. To test for systematic differences between the matched preprint and publication data points, their ratio was calculated and the log ratio values were tested with a two-sided exact Wilcoxon signed rank test. The same procedure was used to detect systematic differences in ranges of CIs before and after peer review (appendix p 2).

We used logistic models to examine the relationship between peer-review scores and publication probability (appendix p 8). Area of research was also included as an independent variable.

Results
We identified 100 matched preprint–journal-article pairs using the NIH iSearch COVID-19 Portfolio.[12–15,17] On average, matched articles reported 19 epidemiological point estimates per paper. We analysed a total of 1921 data points appearing in the preprint version, the publication version, or both. Of the 1606 data points reported in preprints, 173 (11%) were deleted during peer review. Consequently, 1433 estimates (89%) persisted after undergoing peer review. An additional 315 (18%) were added during peer review.

Point estimate values changed an average of 6% during review. We observed similar results at the article level (appendix pp 3–4). Furthermore, we did not observe any systematic increase or decrease in values after peer review, indicating that preprint values are not generally systematically overinflated (Wilcoxon signed rank test; p=0.74). The correlation between preprint and peer-reviewed estimate values was high, more than 0.99 (figure 1). These results held at all scales of the dataset (appendix pp 5–6). Accordingly, an assessment of agreement between preprint and publication point estimates using a non-parametric Bland–Altman analysis[22] showed high agreement between data from preprints and their published versions (appendix p 6).

For infectious disease research, time-varying estimates constitute a mechanism of change from preprint to peer-reviewed versions of articles that might contribute a unique form of variance in the underlying data compared with other areas of research. We analysed the correlation between time from posting a preprint online to publication versus the fold-change of point estimates in both preprint and publication versions and found no significant correlation ($R^2 < 0.01$; p=0.19; appendix p 3). Other areas of research might show less change than was measured here.

Outliers in this data analysis were most extreme in the incidence dataset as authors updated their estimates of the number of people infected, which sometimes

increased substantially. However, after subdividing the data into categories based on the estimate type (appendix p 6), none of these changes in estimate values were statistically significant (Wilcoxon signed rank test; CFR $p=0.81$; IFR $p=0.18$; incidence $p=0.13$; R_0 $p=0.23$). Thus, there were small changes to the 89% of data points that persisted through peer review that were not statistically or practically significant.

We identified 67 articles in the Rapid Reviews: COVID-19 database in the areas of biology, medicine, and public health with a timeframe that overlapped the papers from the NIH iSearch COVID-19 Portfolio. A positive relationship between overlay peer review scores and publication probability would support the hypothesis that the quality of published preprints differs meaningfully from that of those that remain unpublished. We did not observe such a direct relationship (figure 2, appendix p 7). However, public health research is more likely to be published than biology research ($p=0.02$). Preprint age ($p=0.58$) and peer-review scores ($p=0.88$) were not statistically significant independent variables, suggesting that time since posting a preprint online is not rate-limiting in this sample (an average of 417 days had elapsed from the original preprint posting). These analyses do not support the hypothesis that article quality scores are significantly associated with subsequent publication in a peer-reviewed journal (figure 2; appendix p 7).

Another possible effect of peer review is that it could improve evidence not by changing the central tendency, but by reducing uncertainty of estimates. This process could be done by modifying the data or experimental or analytical procedures in a way that reduces CIs. Therefore, we examined the change in point estimates' corresponding CIs (n=495). Similar to changes in point estimates, the amount of change in the CI ranges was small: 7·4% (geometric mean). Unlike point estimates, CIs showed a systematic tendency to decrease between preprint and published versions of articles, indicating that CI ranges reduced during peer review (Wilcoxon signed rank test; $p<0.001$; figure 2). Typically, this reduction was a result of increased sample size after review; authors added estimates (appendix p 3) and added data to their existing estimates during review. Based on these results, measurements of uncertainty, such as CI ranges, can be expected to decrease slightly during the peer-review process.

Discussion

The results of this study augment a small but growing literature on the reliability of data in preprints and changes to research articles caused by peer review, especially regarding COVID-19. Almost a quarter of COVID-19-related research is hosted by preprint servers which have been broadly disseminated by members of the scientific community and the public.[23] Metascience studies suggest that the discrepancy between preprints and peer-reviewed articles is small and the quality of
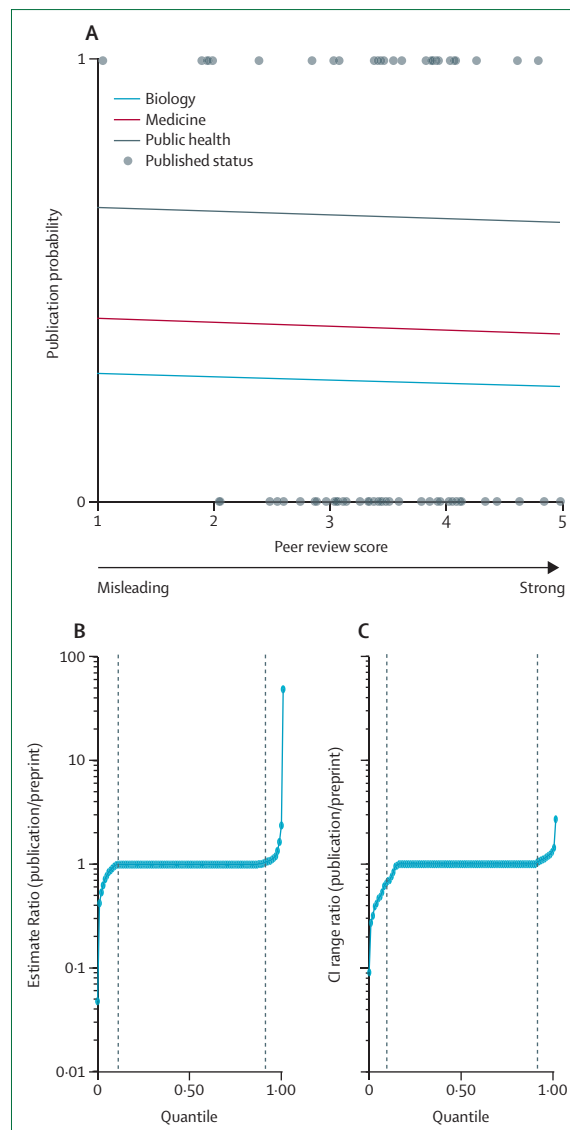
Figure 2: **Correlates of peer review**
(A) Rug plot and line plot of fitted logistic regression controlling for area of research. 10% jitter was added to the x-axis rug plot data points to facilitate visualisation of otherwise overlapping points. (B) Sorted ratios of the peer-reviewed point estimates to the matched preprint value. (C) Sorted ratios of the CI range in the published *vs* preprint versions of articles. Grey lines indicate the tenth and ninetieth quantiles.

reporting is within comparable range, supporting the validity of communicating research findings in preprints before review.[24,25] The results of our study and others[24–30] suggest that the reliability of data reported in preprints is generally high. Although there are measurable effects on research articles after peer review, such as the observed reduction in CIs, effect sizes are small. The amount of change to articles during peer review is small and expert opinions of article quality are not significantly different for preprints that are published versus preprints that are not published (figure 2). Overall, articles submitted to

preprint servers by researchers, especially on COVID-19, are largely complete versions of similar quality to published papers and can be expected to change little during peer review. An earlier version of this paper appeared as a preprint.[31]

This study is not an experimental trial of the marginal effects of peer review on primary data and should not be interpreted as such. Instead, we quantify the amount of expected change from the time preprints are posted online until the time published versions are available. Our findings support the use of preprints in decision making as a component of biomedical research literature, and could help inform the use of preprints during the ongoing COVID-19 pandemic and future disease outbreaks. Future research should test the generalisability of these findings to other areas of research and to other time periods.

Contributors
BIH and HY conceptualised this Health Policy, did the formal analysis, used the software, and reviewed and edited the final draft. Data curation was done by LN, AS, SL, and SA. BIH supervised the process, and LN and BIH did all administration. The original draft was written by LN, HY, AS, SL, SA, and BIH.

References
1 Gianola S, Jesus TS, Bargeri S, Castellini G. Characteristics of academic publications, preprints, and registered clinical trials on the COVID-19 pandemic. *PLoS One* 2020; **15:** e0240123.
2 Guterman EL, Braunstein LZ. Preprints during the COVID-19 pandemic: public health emergencies and medical literature. *J Hosp Med* 2020; **15:** 634–36.
3 Kupferschmidt K. Preprints bring 'firehose' of outbreak data. *Science* 2020; **367:** 963–64.
4 Lachapelle F. COVID-19 preprints and their publishing rate: an improved method. *medRxiv* 2020; published online Oct 13. https://doi.org/10.1101/2020.09.04.20188771 (preprint).
5 Majumder MS, Mandl KD. Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. *Lancet Glob Health* 2020; **8:** e627–30.
6 Bagdasarian N, Cross GB, Fisher D. Rapid publications risk the integrity of science in the era of COVID-19. *BMC Med* 2020; **18:** 192.
7 Kharasch ED, Avram MJ, Clark JD, et al. Peer review matters: research quality and the public trust. *Anesthesiology* 2021; **134:** 1–6.
8 Sheldon T. Preprints could promote confusion and distortion. *Nature* 2018; **559:** 445.
9 van Schalkwyk MCI, Hird TR, Maani N, Petticrew M, Gilmore AB. The perils of preprints. *BMJ* 2020; **370:** m3111.
10 National Institutes of Health. Reporting preprints and other interim research products. 2017. https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html (accessed Dec 7, 2021).
11 National Library of Medicine. NIH preprint pilot. 2020. https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/ (accessed Dec 8, 2021).
12 Hutchins BI. A tipping point for open citation data. *Quant Sci Stud* 2021; **2:** 433–37.
13 Hutchins BI, Baker KL, Davis MT, et al. The NIH Open Citation Collection: a public access, broad coverage resource. *PLoS Biol* 2019; **17:** e3000385.
14 Hutchins BI, Davis MT, Meseroll RA, Santangelo GM. Predicting translational progress in biomedical research. *PLoS Biol* 2019; **17:** e3000416.
15 Hutchins BI, Yuan X, Anderson JM, Santangelo GM. Relative citation ratio (RCR): a new metric that uses citation rates to measure influence at the article level. *PLoS Biol* 2016; **14:** e1002541.
16 Hutson M. Artificial-intelligence tools aim to tame the coronavirus literature. 2020. https://doi.org/10.1038/d41586-020-01733-7 (accessed Jan 25, 2021).
17 Santangelo G. New NIH resource to analyze COVID-19 literature: the COVID-19 Portfolio Tool. 2020. https://nexus.od.nih.gov/all/2020/04/15/new-nih-resource-to-analyze-covid-19-literature-the-covid-19-portfolio-tool/ (accessed Nov 20, 2020).
18 Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R_0). *Emerg Infect Dis* 2019; **25:** 1–4.
19 Heesterbeek JAP, Dietz K. The concept of R_0 in epidemic theory. *Stat Neerl* 1996; **50:** 89–110.
20 Dhar V, Brand A. Coronavirus: time to re-imagine academic publishing. *Nature* 2020; **584:** 192.
21 JMIR Publications. JMIRx Med first overlay journal accepted for PubMed and PubMed Central. 2022. https://xmed.jmir.org/announcements/314&text=view%20this%20exciting%20announcement%20by%20@jmirpub!&quote=View%20this%20exciting%20announcement%20by%20@jmirpub! (accessed Sept 13, 2022).
22 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; **8:** 135–60.
23 Fraser N, Brierley L, Dey G, et al. Preprinting the COVID-19 pandemic. *PLoS Biol* 2021; published online Feb 5. https://doi.org/10.1101/2020.05.22.111294 (preprint).
24 Carneiro CFD, Queiroz VGS, Moulin TC, et al. Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Res Integr Peer Rev* 2020; **5:** 16.
25 Klein M, Broadwell P, Farb SE, Grappone T. Comparing published scientific journal articles to their pre-print versions. *Int J Digit Libr* 2018; **20:** 335–50.
26 Bero L, Lawrence R, Leslie L, et al. Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation. *BMJ Open* 2021; **11:** e051821.
27 Oikonomidi T, Boutron I, Pierre O, Cabanac G, Ravaud P. Changes in evidence for studies assessing interventions for COVID-19 reported in preprints: meta-research study. *BMC Med* 2020; **18:** 402.
28 Brierley L, Nanni F, Polka JK, et al. Tracking changes between preprint posting and journal publication during a pandemic. *PLoS Biol* 2022; **20:** e3001285.
29 Wang Y, Cao Z, Zeng DD, Zhang Q, Luo T. The collective wisdom in the COVID-19 research: comparison and synthesis of epidemiological parameter estimates in preprints and peer-reviewed articles. *Int J Infect Dis* 2021; **104:** 1–6.
30 Hoppe TA, Arabi S, Hutchins BI. Predicting causal citations without full text. *bioRxiv* 2022; published online July 7. https://doi.org/10.1101/2022.07.05.498860 (preprint).
31 Nelson L, Ye H, Schwenn A, Lee C, Arabi S, Hutchins BI. Robustness of evidence reported in preprints during peer review. *Research Square* 2022; published online Feb 22. https://doi.org/10.21203/rs.3.rs-1344293/v1 (preprint).