Rapid Reviews COVID-19

Review 1: "Preprinting a pandemic: the role of preprints in the COVID-19 pandemic"

Noah Haber

Published on: Aug 11, 2020

DOI: 10.21428/2e3983f5.ee267763

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Broadly speaking, this is a fascinating article with a large number of useful, interesting, and surprising results and conclusions. The effort required to have produced this article, pulling from such a large number of sources so quickly, is impressive. I was surprised by several of the findings, even being a person who follows the meta literature quite closely. I am further impressed with the detailed and transparent reporting of the methods section, the descriptions of why decisions were made, and the sheer scope of sources and tools used.

I have found the major claims in this article to be relatively well-founded and justified by the methods and the data. Specifically, the I generally find that the claims about the properties of the COVID articles (more engagement, length, sheer volume of articles, etc) are relatively robust findings, and are applicable in the scientific meta practice. As a descriptive paper of the pre-print landscape of COVID-19, this is an excellent source of information.

However, the study also contains a number of errors and misinterpretations in its current form, particularly with regard to interpretation of statistical inference. One such error is in the abstract and conclusions, and repeated throughout the manuscript: the odds ratios are nearly universally interpreted as rate ratios. For example, the abstract incorrectly states that "COVID-36 19 preprints are accessed and distributed at least 15 times more than non-COVID-19 preprints," which is misinterpreted from an odds ratio. The two are strongly different measures and do not approximate one another in this context. I strongly recommend changing all odds ratio calculations to rate ratio calculations, since RRs are much more generally interpretable (and clearly what the authors prefer). If not, the authors should explicitly state that these are ratios of odds, not probabilities.

A second major statistical issue regards the sample size attribution of different literatures. While the properties of individual papers that are COVID-19 vs. other papers are individual units, the literatures as a whole should not be. For example, the paper compares the relative proportion of the literature that was COVID-19-related with the relative proportion of the literature that was Zika-related, and concludes that they are different with $p < 0.001$. However, because the comparison at the level of interest is a comparison of two binomial proportions, the effective sample size is 2, not the number of studies as claimed. This issue is repeated in a number of areas throughout the analysis. It is not an existential threat to the main conclusions, but it is misleading to claim that level of precision. Further, I am not sure it is meaningful to

make the Zika comparison at all. As platforms get larger, they also experience more rapid proportionate growth in emerging topics in general. Had the Zika outbreak happened in 2020, I imagine we would have seen larger proportionate (not just count of papers) growth in the number of papers (albeit still almost certainly less relative to COVID).

Broadly, this document may be suffering from doing too much, which leaves too little room to discuss the limitations of the measures or to pinpoint or discuss areas of improvement. I would strongly suggest paring down, and potentially moving some sections of the paper into appendices or a separate publication. A few areas I find that are weaker arguments are the semantic analysis, and the documenting changes between preprint and publication, and the review/transparency sections. These tests are too limited to be used conclusively, and there is little room for discussion of the weaknesses and limitations of these tests. These are potentially useful for a separate publication.

Overall, I found this pre-print to be useful, and would recommend it be edited and proceed to the full publication stage for further review and critique.