Publication rate and citation counts for preprints released during the COVID-19 pandemic: The good, the bad and the ugly

Diego Anazco [Equal first author, 1], Bryan Nicolalde [Equal first author, 1], Isabel Espinosa [1], Jose Camacho [2], Mariam Mushtaq [1], Jimena Gimenez [1], Enrique Teran [Corresp. 1]

[1] Colegio de Ciencias de la Salud, Universidad San Francisco de Quito, Quito, Ecuador

[2] Colegio de Ciencias e Ingenieria, Universidad San Francisco de Quito, Quito, Ecuador

Corresponding Author: Enrique Teran
Email address: eteran@usfq.edu.ec

Background. Preprints are preliminary reports that have not been peer-reviewed. In December 2019, a novel coronavirus appeared in China, and since then, scientific production, including preprints, has drastically increased. In this study, we intend to evaluate how often preprints about COVID-19 were published in scholarly journals and cited. **Methods.** We searched the iSearch COVID-19 portfolio to identify all preprints related to COVID-19 posted on bioRxiv, medRxiv, and Research Square from January 1, 2020, to May 31, 2020. We used a custom-designed program to obtain metadata using the Crossref public API. After that, we determined the publication rate and made comparisons based on citation counts using non-parametric methods. Also, we compared the publication rate, citation counts, and time interval from posting on a preprint server to publication in a scholarly journal among the three different preprint servers. **Results.** Our sample included 5061 preprints, out of which 288 were published in scholarly journals and 4773 remained unpublished (publication rate of 5.7%). We found that articles published in scholarly journals had a significantly higher total citation count than unpublished preprints within our sample (p<0.001), and that preprints that were eventually published had a higher citation count as preprints when compared to unpublished preprints (p<0.001). As well, we found that published preprints had a significantly higher citation count after publication in a scholarly journal compared to as a preprint (p<0.001). Our results also show that medRxiv had the highest publication rate, while bioRxiv had the highest citation count and shortest time interval from posting on a preprint server to publication in a scholarly journal. **Conclusions.** We found a remarkably low publication rate for preprints within our sample, despite accelerated time to publication by multiple scholarly journals. These findings could be partially attributed to the unprecedented surge in scientific production observed during the COVID-19 pandemic, which might saturate reviewing and

editing processes in scholarly journals. However, our findings show that preprints had a significantly lower scientific impact, which might suggest that some preprints have lower quality and will not be able to endure peer-reviewing processes to be published in a peer-reviewed journal.

1 **Publication rate and citation counts for preprints released during the COVID-19**
2 **pandemic: The good, the bad and the ugly**

4 Diego Anazco[1a], Bryan Nicolalde[1a], Isabel Espinosa[1], Jose Camacho[2], Mariam Mushtaq[1], Jimena
5 Gimenez[1], Enrique Teran[1]
6
7 [1]Colegio de Ciencias de la Salud, Universidad San Francisco de Quito, Quito, Ecuador
8 [2]Colegio de Ciencias e Ingenieria, Universidad San Francisco de Quito, Quito, Ecuador
9
10 [a] *Both authors contributed equally*
11
12 Corresponding Author:
13 Enrique Teran; MD, PhD
14 Colegio de Ciencias de la Salud
15 Universidad San Francisco de Quito
16 Quito, Ecuador
17 Email address: eteran@usfq.edu.ec

19
20
21
22

23   **Abstract**
24   **Background.** Preprints are preliminary reports that have not been peer-reviewed. In December
25   2019, a novel coronavirus appeared in China, and since then, scientific production, including
26   preprints, has drastically increased. In this study, we intend to evaluate how often preprints about
27   COVID-19 were published in scholarly journals and cited.
28   **Methods.** We searched the iSearch COVID-19 portfolio to identify all preprints related to
29   COVID-19 posted on bioRxiv, medRxiv, and Research Square from January 1, 2020, to May 31,
30   2020. We used a custom-designed program to obtain metadata using the Crossref public API.
31   After that, we determined the publication rate and made comparisons based on citation counts
32   using non-parametric methods. Also, we compared the publication rate, citation counts, and time
33   interval from posting on a preprint server to publication in a scholarly journal among the three
34   different preprint servers.
35   **Results.** Our sample included 5061 preprints, out of which 288 were published in scholarly
36   journals and 4773 remained unpublished (publication rate of 5.7%). We found that articles
37   published in scholarly journals had a significantly higher total citation count than unpublished
38   preprints within our sample ($p<0.001$), and that preprints that were eventually published had a
39   higher citation count as preprints when compared to unpublished preprints ($p<0.001$). As well,
40   we found that published preprints had a significantly higher citation count after publication in a
41   scholarly journal compared to as a preprint ($p<0.001$). Our results also show that medRxiv had
42   the highest publication rate, while bioRxiv had the highest citation count and shortest time
43   interval from posting on a preprint server to publication in a scholarly journal.
44   **Conclusions.** We found a remarkably low publication rate for preprints within our sample,
45   despite accelerated time to publication by multiple scholarly journals. These findings could be
46   partially attributed to the unprecedented surge in scientific production observed during the
47   COVID-19 pandemic, which might saturate reviewing and editing processes in scholarly
48   journals. However, our findings show that preprints had a significantly lower scientific impact,
49   which might suggest that some preprints have lower quality and will not be able to endure peer-
50   reviewing processes to be published in a peer-reviewed journal.
51
52

53 **Introduction**
54 A preprint is a preliminary report of scientific work that is shared publicly before it has been
55 peer-reviewed. Most preprints have a digital object identifier (DOI) and can be cited by other
56 research articles (Kaiser, 2019). There are multiple servers that host preprints related to
57 biological and medical sciences, such as bioRxiv (https://www.biorxiv.org/) and medRxiv
58 (https://www.medrxiv.org/), operated by Cold Spring Harbor Laboratory; and Research Square,
59 another widely used platform during the COVID-19 pandemic
60 (https://www.researchsquare.com/). The validity and usage of information from preprints are
61 being discussed. Advocates claim that preprints accelerate access to findings and improve the
62 quality of published papers by permitting faster feedback before publication. Also, the audience
63 for preprints might be larger because many scholarly journals do not have open access (Kaiser,
64 2017). Those who are against the use of preprints state that many of these investigations may be
65 flawed due to the lack of a peer-review process (Chalmers & Glasziou, 2016). Thus, readers
66 should be aware that preprints might contain errors and that information reported there has not
67 been peer-reviewed or endorsed by the scientific or medical community (Brainard, 2020; Kaiser,
68 2017, 2019).
69 Coronavirus disease 2019 (COVID-19) was first described in December 2019 in Wuhan, China,
70 and the World Health Organization (WHO) declared it as a pandemic on March 12, 2020 (WHO,
71 2020). Due to the drastic increase in scientific production about this novel disease, scholarly
72 journals have been overwhelmed, and in response, some journals have accelerated their usual
73 processes to shorten the time it takes to publish an article (Brainard, 2020). Nonetheless, due to
74 the imperative need for rapid information, authors have been urged to share relevant information
75 as soon as possible; thus, an increasing number of studies have been posted as preprints (Rubin
76 et al., 2020). The International Committee of Medical Journal Editors (ICMJE) stated that in the
77 case of public health emergencies, information with immediate implications for public health
78 should be rapidly disseminated, without concerns that this will prevent subsequent publication in
79 a scholarly journal (ICMJE, 2020). The WHO and many scholarly journals require that authors
80 share their work initially using a preprint server before it goes through formal peer-reviewing
81 and editing processes (Fidahic et al., 2020; Moorthy et al., 2020).
82 Thus, the objective of this study was to determine the publication rate in scholarly journals for
83 preprints about COVID-19 posted from January 1, 2020, to May 31, 2020, in medRxiv, bioRxiv,
84 and Research Square. Also, we evaluated how often preprints or published papers were cited,
85 and compared preprint servers in terms of publication rate, citation count, and time to
86 publication.
87
88 **Materials & Methods**
89 *Search strategy*
90 We used the iSearch COVID-19 portfolio (https://icite.od.nih.gov/covid19/search/), a
91 comprehensive platform developed by the National Institute of Health (NIH) that hosts
92 publications on scholarly journals and preprints related to SARS-CoV-2 and COVID-19, to

93   obtain our sample of preprints. We searched on August 10, 2020, to identify all the preprints that
94   had been posted on three different preprint servers (medRxiv, bioRxiv, and Research Square)
95   from January 1, 2020, to May 31, 2020 (figure 1). The search yielded 5106 preprints, and after
96   excluding 45 duplicates, we had a sample of 5061 preprints. For duplicated preprints, we decided
97   to use of the version that was first uploaded to a preprint server. We gathered metadata for these
98   preprints, including the title, authors list, Digital Object Identifier (DOI), preprint server, and
99   date of posting on the preprint server.
100   *Publication status and citation count*
101   If a preprint is subsequently published in a scholarly journal, a link to the published version is
102   made available on the preprint server. We used Crossref (https://www.crossref.org/) to determine
103   the publication status and citation count of this sample. To gather specific metadata from the
104   Crossref application programming interface (API), we developed a program using Python
105   (Python Software Foundation, https://www.python.org/). The codes for this program are
106   available online (https://github.com/josecxz/Crossref-Adquisition). Using this method, we
107   retrieved validated data for citation counts and the DOI of the published version of a preprint, if
108   available. We could not gather data for 13 preprints using the API in Crossref, so a manual
109   search was performed on the respective preprint server, Google Scholar, and PubMed to
110   determine the publication status. However, we decided not to include citation counts for this set
111   of preprints, as this data was not validated and methods to assess citation counts might differ.
112   Also, duplicated preprints were removed before this search, and this data was not collected nor
113   considered for subsequent analyses. After identifying preprints that were published in a scholarly
114   journal, we searched Crossref using our program to determine the citation count for the
115   published versions and the date in which these articles were published in a journal. Both searches
116   on the Crossref API were conducted on August 11, 2020.
117   *Statistical analysis:*
118   After gathering our data, we observed that the number of citations varied significantly. Through
119   graphical exploration, we identified that our data had multiple outliers. In our dataset, the median
120   represented the center of the distribution more accurately than the mean, which is more sensitive
121   to extreme values. As well, we conducted a Shapiro-Wilk test that showed that our data had a
122   non-normal distribution (p<0.001). Hence, we decided to use non-parametric tests for our
123   statistical analyses regarding citation counts. We conducted Mann-Whitney U tests for our
124   citation counts analyses for independent samples, and we conducted a Wilcoxon signed-ranked
125   test for a paired analysis of citation counts before and after publication in a scholarly journal, as
126   described in the Results section.
127   We used a chi-squared test to compare publication rates for preprints on three different preprint
128   servers (bioRxiv, medRxiv, Research Square). The non-parametric Kruskal-Wallis test, followed
129   by a posthoc Dunn´s test for pairwise comparisons, was used to compare citation counts and the
130   time intervals from posting on a preprint server to publication in a scholarly journal.
131   We used a two-sided alpha value of 0.05 for our statistical analyses, except for the chi-squared
132   tests (one-sided alpha value of 0.05). We used the Bonferroni correction method to adjust

133   probability (p-adj) values in our posthoc analyses. All analyses were conducted on IBM SPSS
134   Statistics for Windows, version 25 (IBM Corp., Armonk, N.Y., USA).
135
136   **Results**
137   We included 5061 preprints from three preprint servers. In our sample, most of the preprints
138   were obtained from medRxiv (n=3258), followed by Research Square (n=1011) and bioRxiv
139   (n=792). By August 11, 2020, only 288 preprints (5.7%) had been published in a scholarly
140   journal. The median time interval from the date of posting on a preprint server to the date of our
141   search was 110 days (min 72 – max 205 days).
142   *Citation counts*
143   As mentioned before, citation count data was not available for 13 preprints (8 unpublished
144   preprints, 5 published preprints), so this set of preprints was excluded for the following analyses.
145   Citation counts ranged from 0 to 168 citations for preprints, while it ranged from 0 to 3587
146   citations for published articles. Unpublished preprints (n=4765) had a mean citation count of
147   1.99 (median: 0; Q1-Q3: 0-1), while published articles (n=283) had a total mean citation count
148   of 72.12 (median:9; Q1-Q3: 2-38). After conducting a Mann-Whitney U test, we determined
149   that published articles had a significantly higher total number of citations than unpublished
150   preprints ($p<0.001$) (Table 1).
151   For the set of preprints that were published at some point (n=283), we compared the number of
152   citations that these articles had before and after being published in a scholarly journal. The mean
153   number of citations for these articles before being published (preprint status) was 6.29 (median:
154   1; Q1-Q3: 0-5). By using a Wilcoxon signed-ranked test, we identified that the citation count for
155   the same set of articles was significantly higher after being published in a scholarly journal
156   (mean: 65.8; median: 7; Q1-Q3: 1-30) ($p<0.001$).
157   As well, we compared the citation counts of preprints that eventually got published (n=283) to
158   preprints that remained unpublished (n=4765). After conducting a Mann-Whitney U test, we
159   found that published preprints had a higher number of citations as preprints (mean: 6.29; median:
160   1; Q1-Q3: 0-5) than preprints that remained unpublished (mean: 1.99; median: 0; Q1-Q3: 0-1) (
161    $p<0.001$). These results suggest that preprints that were published eventually had a higher impact
162   in scientific production, even before they were published in a scholarly journal, than preprints
163   that remained unpublished.
164   In our sample, 3140 preprints (62%) had no citations at all. Within the set of preprints that
165   remained unpublished, 3047 preprints (64%) had no citations; while 93 preprints (32.9%) that
166   were eventually published had no citations as preprints. We conducted the same statistical
167   analyses previously described after excluding from preprints that eventually got published
168   (n=283) those without citations before (n=93) plus those with no citations once published
169   (n=50). However, from the latest 34 were already included in the previous list, then only an
170   additional 16 were removed for a final sample of 174 published preprints with at least one cited.
171   The statistical significance was not altered (Table 2).
172

173 *Comparing preprint servers*
174 Preprints in medRxiv had the highest number of published preprints and the highest publication
175 rate (n=225; 6.9%), compared to preprints in bioRxiv (n=45; 5.7%) and Research Square (n=18;
176 1.8%). We performed a chi-squared test that confirmed a significant difference regarding
177 publication status on these databases (df=2; χ^2=37.77; p<0.001). Post-hoc pairwise comparisons
178 revealed significant differences when comparing bioRxiv to Research Square (df=1, χ^2=20.05,
179 p<0.001), and medRxiv to Research Square (df=1; χ^2=37.36; p<0.001), however, we did not find
180 a significant difference when comparing bioRxiv to medRxiv (df=1; χ^2=1.53; p=0.215).
181 In our sample, the median time interval from posting on a preprint server to publication on a
182 scholarly journal was 24 days (minimum: 0; maximum 117). BioRxiv had the mean shortest time
183 interval (26 days; median: 20; Q1-Q3: 13-38.8), followed by medRxiv (mean: 27.7 days;
184 median: 23; Q1-Q3: 11-42.5) and Research Square (mean: 65.7 days; median: 55.5; Q1-Q3: 74-
185 85.8; p<0.001) (figure 2). In a post-hoc Dunn's test we identified a significant difference
186 between bioRxiv and Research Square (z=-113.77; p-adj <0.001), and, between medRxiv and
187 Research Square (z=-111.11; p-adj <0.001), however, we did not find a significant difference
188 between bioRxiv and medRxiv (z=-2.66; p-adj=1.000).
189 BioRxiv preprints had the highest mean number of citations (5.50; median: 2; Q1-Q3: 0-6),
190 followed by medRxiv preprints (mean: 2.06; median: 0; Q1-Q3: 0-1) and Research Square
191 (mean: 0.23; median: 0; Q1-Q3: 0-0; p<0.001) (Table 3). In posthoc pairwise comparisons, we
192 identified significant differences between all pairs of preprint servers in terms of the number of
193 citations (all p-adj<0.001).
194
195 **Discussion**
196 Preprints intend to accelerate the access to preliminary data for the scientific community, mainly
197 to rapidly inform about time-sensitive issues and receive rapid feedback before entering a peer-
198 review process, which is a requirement for publication in the majority of indexed journals
199 (Mudrak, 2020). medRxiv and bioRxiv, two of the most widely used preprint servers, have a
200 disclaimer on their homepage that states: *"these are preliminary reports that have not been peer-*
201 *reviewed. They should not be regarded as conclusive, guide clinical practice/health-related*
202 *behavior, or be reported in news media as established information"* (medRxiv, 2020).
203 Scientific production has drastically increased due to the COVID-19 pandemic, and
204 approximately 900 articles, including published papers and preprints, were available before
205 March 12, 2020 (Callaway et al., 2020). During the pandemic, multiple scholarly journals
206 required authors to share their papers simultaneously using a preprint server (Fidahic et al.,
207 2020).
208 We found that the publication rate for preprints was very low (5.7%), even after a considerably
209 long follow-up to determine the publication status. Palayew et al described that on average, 367
210 articles related to COVID-19 were published weekly in scholarly journals and that the median
211 time from submission to acceptance was just 6 days, which is remarkably shorter than the usual
212 time to accept that scholarly journals had before the pandemic (Palayew et al., 2020). Thus, we

213 consider that enough follow-up time was provided to identify most of the preprints in our sample
214 that will be published in a journal.
215 Our findings reveal a drastically lower publication rate for preprints related to COVID-19 when
216 compared to the publication rates of preprints related to other public health emergencies of
217 international concern (PHEIC). For example, during the Zika outbreak (Nov 2015 to Aug 2017),
218 174 preprints were posted, with a publication rate of 60%; while 75 preprints were posted during
219 the Ebola outbreak (May 2014 to Jan 2016), with a publication rate of 48% (Johansson et al.,
220 2018). During the COVID-19 pandemic, the number of preprints was unprecedented, and a
221 potential saturation of scholarly journals might contribute to this lower publication rate.
222 However, on a rapid search on the iSearch COVID-19 portfolio, we identified that 17564 articles
223 had been published in peer-reviewed journals indexed in PubMed during the same interval of
224 time (January 31, 2020, to May 31, 2020). These findings suggest that despite being swamped
225 with papers by the scientific community from all around the globe, scholarly journals have been
226 able to maintain a substantial output of peer-reviewed publications, and we could assume that
227 other factors, such as overall lower quality, might contribute to the low publication rate for
228 preprints in our sample.
229 Never before in history has the scientific community produced so much non-peer-reviewed data
230 (Dinis, 2020). However, by attempting to drastically increase the output of scientific data during
231 times of desperate need for information about a novel disease, we might overlook essential
232 processes needed to promote the quality and validity of this scientific production. In this report,
233 we found that articles published in a scholarly journal had a significantly higher citation count
234 than preprints, and even though we did not directly evaluate the quality of the preprints, the
235 citation count is an indicator of the scientific impact, which is one of the components within the
236 concept of scientific quality (Aksnes et al., 2019). Importantly, a significant proportion of
237 preprints in our sample were never cited by other scientific papers, which highlights the low
238 scientific impact that some preprints had, and probably did not merit a fast, non-peer-reviewed
239 publication. As well, we could infer that articles published in scholarly journals are used more
240 often by the scientific community to develop new papers, perhaps due to higher quality and the
241 confidence that a peer-reviewing process might provide. These findings are supported by the fact
242 that amongst preprints that were eventually published, the citation counts once published in a
243 scholarly journal were significantly higher than for the preprint version However, these might
244 also reflect longer times in article form compared to preprint form or growing awareness of the
245 research over time.
246 In a study conducted before the pandemic, there were small differences in quality between
247 preprint versions and articles published in scholarly journals (Klein et al., 2019). However, these
248 findings are not necessarily applicable to the current situation, as preprint production has
249 increased drastically due to the COVID-19 pandemic, and our results support the impression of a
250 lower scientific impact of preprints when compared to scholarly journals.
251 Our study also revealed differences among the three preprint servers analyzed. There, medRxiv
252 had a higher publication rate, while bioRxiv had the shortest interval from posting to publication

253 and had the highest citation counts. These differences could be related to the scope that preprint
254 servers have and how visible these preprint servers are for the scientific community, which could
255 determine the scientific impact that a certain publication will have.
256 Some preprints might contain essential and time-sensitive information, for example, the basic
257 reproduction number [R0], calculated using data on preprints was shown not to be different from
258 the one estimated in peer-reviewed articles (Majumder & Mandl, 2020). Similarly, preprints on
259 the viral sequence and structure have allowed for the early investigation of potential therapeutic
260 options and vaccines (Brainard, 2020; Kwon, 2020). While there is widespread agreement that
261 preprints could be useful in the current context, there are significant risks associated with the
262 potential spread of faulty data without appropriate third-party screening (Dinis, 2020).
263 The basic screening process employed by preprint servers may not be enough to avoid the
264 dissemination of flawed information (Rawlinson & Bloom, 2019). For example, a preprint that
265 was posted on bioRxiv suggested significant molecular similarities between SARS-CoV-2 and
266 HIV (Kwon, 2020). This preprint was later withdrawn from the server, however, by the time that
267 happened, it had already sparked controversy and conspiracy theories. To our concern, we found
268 that during the COVID-19 pandemic multiple preprints have been used in the development of
269 clinical guidelines and public health policies (Bhimraj et al., 2020; Nicolalde et al., 2020)
270 Although peer-review aims to be an exhaustive and in-depth process that improves the quality of
271 a manuscript, articles published in a peer-reviewed journal should not be taken as non-refutable
272 knowledge. To illustrate this, a couple of high-impact peer-reviewed articles were withdrawn
273 from two prestigious journals due to significant concerns on primary data validity (M. Mehra et
274 al., 2020; M. R. Mehra et al., 2020). Currently, 33 papers related to SARS-CoV-2 and COVID-
275 19 (13 preprints and 20 articles published in scholarly journals) have been retracted (Retraction
276 Watch, 2020). It is possible that due to the fast-track processes that many scientific journals
277 currently provide, the quality of peer-reviewed articles might also be compromised (Palayew et
278 al., 2020).
279 We acknowledge certain limitations to our study. Even though we had an extensive sample size,
280 and included the most used preprint servers, several other platforms host preprints. As well,
281 some preprints included in our sample might end up published in a scholarly journal after some
282 time, however, considering the shortened editorial processes due to the COVID-19 pandemic
283 (Palayew et al., 2020), we consider that we provided a sufficient follow-up period to identify
284 most of the preprints that will be eventually published within our sample. Additionally, our
285 analysis approach for citation counts did not permit adjusting for other variables, which could
286 contribute to confounding bias. We consider that a potential confounding factor is the time
287 interval from posting on a preprint server, as preprints that were available for a longer time
288 online might have a higher number of citations than more recent preprints. We consider that by
289 using comprehensive platforms, such as the iSearch COVID-19 portfolio, and resources that
290 include verified data, such as Crossref, our results are reliable.
291
292

Conclusions

Preprints have been used widely as a means to share information rapidly during the COVID-19 pandemic, and multiple organizations and scholarly journals have urged authors to share their research on preprint servers. However, the publication rate for preprints in our sample was remarkably low, despite accelerated editing processes by multiple scholarly journals. There are some important caveats related to the use of preprints. Notably, it should be highlighted that preprints include non-peer-reviewed data, and thus, should be considered as preliminary reports. We found that preprints that had sufficient quality to be published after going through peer-reviewing scrutiny were more cited than preprints that remained unpublished. In this sense, preprint servers might provide an opportunity for higher visibility and potentially improve the scientific impact of research articles. Peer-reviewing and editing processes play an essential role during the current infodemic, and researchers that initially post their preliminary reports on preprint servers should strive to publish their works in peer-reviewed journals. Also, researchers should be careful when using and citing preprints in their publications.

Acknowledgments

N/A

References

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*, *9*(1), 2158244019829575. https://doi.org/10.1177/2158244019829575

Bhimraj, A., Morgan, R. L., Shumaker, A. H., Lavergne, V., Baden, L., Cheng, V. C.-C., Edwards, K. M., Gandhi, R., Muller, W. J., O'Horo, J. C., Shoham, S., Murad, M. H., Mustafa, R. A., Sultan, S., & Falck-Ytter, Y. (2020). Infectious Diseases Society of America Guidelines on the Treatment and Management of Patients with COVID-19. *Clinical Infectious Diseases*, ciaa478. https://doi.org/10.1093/cid/ciaa478

Brainard, J. (2020, May 13). *Scientists are drowning in COVID-19 papers. Can new tools keep them afloat?* Science | AAAS. https://www.sciencemag.org/news/2020/05/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat

Callaway, E., Cyranoski, D., Mallapaty, S., Stoye, E., & Tollefson, J. (2020). The coronavirus pandemic in five powerful charts. *Nature*, *579*(7800), 482–483. https://doi.org/10.1038/d41586-020-00758-2

Chalmers, I., & Glasziou, P. (2016). Should there be greater use of preprint servers for publishing reports of biomedical science? *F1000Research*, *5*. https://doi.org/10.12688/f1000research.8229.1

Dinis, R. (2020, April 18). *COVID-19 research: Pandemic versus "paperdemic", integrity, values and risks of the "speed science."* https:// doi.org/10.1080/20961790.2020.1767754

Fidahic, M., Nujic, D., Runjic, R., Civljak, M., Markotic, F., Lovric Makaric, Z., & Puljak, L. (2020). Research methodology and characteristics of journal articles with original data, preprint

articles and registered clinical trial protocols about COVID-19. *BMC Medical Research Methodology*, *20*(1), 161. https://doi.org/10.1186/s12874-020-01047-2

ICMJE. (2020). http://icmje.org/recommendations/browse/publishing-and-editorial-issues/overlapping-publications.html

Johansson, M. A., Reich, N. G., Meyers, L. A., & Lipsitch, M. (2018). Preprints: An underutilized mechanism to accelerate outbreak science. *PLOS Medicine*, *15*(4), e1002549. https://doi.org/10.1371/journal.pmed.1002549

Kaiser, J. (2017). The preprint dilemma. *Science*, *357*(6358), 1344–1349. https://doi.org/10.1126/science.357.6358.1344

Kaiser, J. (2019, June 5). *Medical preprint server debuts*. Science | AAAS. https://www.sciencemag.org/news/2019/06/medical-preprint-server-debuts

Klein, M., Broadwell, P., Farb, S. E., & Grappone, T. (2019). Comparing published scientific journal articles to their pre-print versions. *International Journal on Digital Libraries*, *20*(4), 335–350. https://doi.org/10.1007/s00799-018-0234-1

Kwon, D. (2020). How swamped preprint servers are blocking bad coronavirus research. *Nature*, *581*(7807), 130–131. https://doi.org/10.1038/d41586-020-01394-6

Majumder, M. S., & Mandl, K. D. (2020). Early in the epidemic: Impact of preprints on global discourse about COVID-19 transmissibility. *The Lancet Global Health*, *8*(5), e627–e630. https://doi.org/10.1016/S2214-109X(20)30113-3

medRxiv. (2020). *About medRxiv*. https://www.medrxiv.org/content/about-medrxiv

Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D., & Patel, A. N. (2020). Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. DOI: 10.1056/NEJMoa2007621. *New England Journal of Medicine*. https://doi.org/10.1056/NEJMc2021225

Mehra, M., Ruschitzka, F., & Patel, A. (2020). Retraction: "Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis." *Lancet (London, England)*. https://doi.org/10.1016/ S0140-6736(20)31324-6

Moorthy, V., Henao Restrepo, A. M., Preziosi, M.-P., & Swaminathan, S. (2020). Data sharing for novel coronavirus (COVID-19). *Bulletin of the World Health Organization*, *98*(3), 150–150. https://doi.org/10.2471/BLT.20.251561

Mudrak, B. (2020). *What are Preprints, and How Do They Benefit Authors? | AJE*. https://www.aje.com/arc/benefits-of-preprints-for-researchers/

Nicolalde, B, Añazco, D, Mushtaq, M, Aguilar, A, & Terán, E. (2020). *Current Pharmacological Therapy against COVID-19: A Latin American Perspective*. 59–68. https://doi.org/10.34141/LJCS2640133

Palayew, A., Norgaard, O., Safreed-Harmon, K., Andersen, T. H., Rasmussen, L. N., & Lazarus, J. V. (2020). Pandemic publishing poses a new COVID-19 challenge. *Nature Human Behaviour*, *4*(7), 666–669. https://doi.org/10.1038/s41562-020-0911-0

Rawlinson, C., & Bloom, T. (2019). New preprint server for medical research. *BMJ*, *365*. https://doi.org/10.1136/bmj.l2301

372   Retraction Watch. (2020, April 29). Retracted coronavirus (COVID-19) papers. *Retraction*

373   *Watch*. https://retractionwatch.com/retracted-coronavirus-covid-19-papers/

374   Rubin, E. J., Baden, L. R., Morrissey, S., & Campion, E. W. (2020). Medical Journals and the

375   2019-nCoV Outbreak. *New England Journal of Medicine*, *382*(9), 866–866.

376   https://doi.org/10.1056/NEJMe2001329

377   WHO. (2020, March 12). *WHO announces COVID-19 outbreak a pandemic*. World Health

378   Organization. http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-

379   19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic

380

Characteristics and analyses of citation counts data made for preprints posted on medRxiv, bioRxiv, and Research Square from January 1 to May 31, 2020.

Q1: quartile 1; Q3: quartile 3

1 **Table 1.** Characteristics and analyses of citation counts data made for preprints posted on
2 medRxiv, bioRxiv, and Research Square from January 1 to May 31, 2020. Q1: quartile 1; Q3:
3 quartile 3

	Unpublished preprints (n=4765)	Published preprints (n=283) (total citation count)	p-value
Mean	1.99	72.12	p<0.001
Range	168	3757	
Minimum	0	0	
Maximum	168	3757	
Median (Q1-Q3)	0 (0-1)	9 (2-38)	
	Published preprints (n=283) (citation count of preprint version)	Published preprints (n=283) (citation count of published version)	p-value
Mean	6.29	65.83	p<0.001
Range	170	3587	
Minimum	0	0	
Maximum	170	3587	
Median (Q1-Q3)	1 (0-5)	7 (1-30)	
	Unpublished preprints (n=4765)	Published preprints (n=283) (citation count of preprint version)	p-value
Mean	1.99	6.29	p<0.001
Range	168	170	
Minimum	0	0	
Maximum	168	170	
Median (Q1-Q3)	0 (0-1)	1 (0-5)	

5
6
7
8
9
10
11
12
13
14
15
16
17

Characteristics and analyses of citation counts data made for preprints posted on medRxiv, bioRxiv, and Research Square from January 1 to May 31, 2020 after excluding preprints without citations.

Q1: quartile 1; Q3: quartile 3

1 **Table 2.** Characteristics and analyses of citation counts data made for preprints posted on
2 medRxiv, bioRxiv, and Research Square from January 1 to May 31, 2020 after excluding
3 preprints without citations. Q1: quartile 1; Q3: quartile 3

	Unpublished preprints (n=1718)	Published preprints (n=174) (citation count of published version)	p-value
Mean	5.52	104.58	p<0.001
Range	167	3586	
Minimum	1	1	
Maximum	168	3587	
Median (Q1-Q3)	2 (1-5.75)	20.50 (6-72.50)	
	Published preprints (n=174) (citation count of preprint version)	**Published preprints (n=174) (citation count of published version)**	**p-value**
Mean	10.09	104.58	p<0.001
Range	169	3586	
Minimum	1	1	
Maximum	170	3587	
Median (Q1-Q3)	4 (2-9)	20.50 (6-70.25)	
	Unpublished preprints (n=1718)	**Published preprints (n=174) (citation count of preprint version)**	**p-value**
Mean	5.52	10.09	p<0.001
Range	167	169	
Minimum	1	1	
Maximum	168	170	
Median (Q1-Q3)	2 (1-5.75)	4 (2-9)	

5
6
7
8
9
10
11
12
13
14
15
16
17

Citation counts for preprints posted in medRxiv, bioRxiv and Research Square.

Q1: quartile 1; Q3: quartile 3

2 **Table 3.** Citation counts for preprints posted on medRxiv, bioRxiv and Research Square. Q1:

3 quartile 1; Q3: quartile 3

	medRxiv (n=225)	BioRxiv (n=45)	Research Square (n=18)	p-value
Mean	2.06	5.50	0.23	p<0.001
Range	170	168	16	
Minimum	0	0	0	
Maximum	170	168	16	
Median (Q1-Q3)	0 (0-1)	2 (0-6)	0 (0-0)	

Figure 1

Timeline of data gathering for the study of preprints on COVID-19.
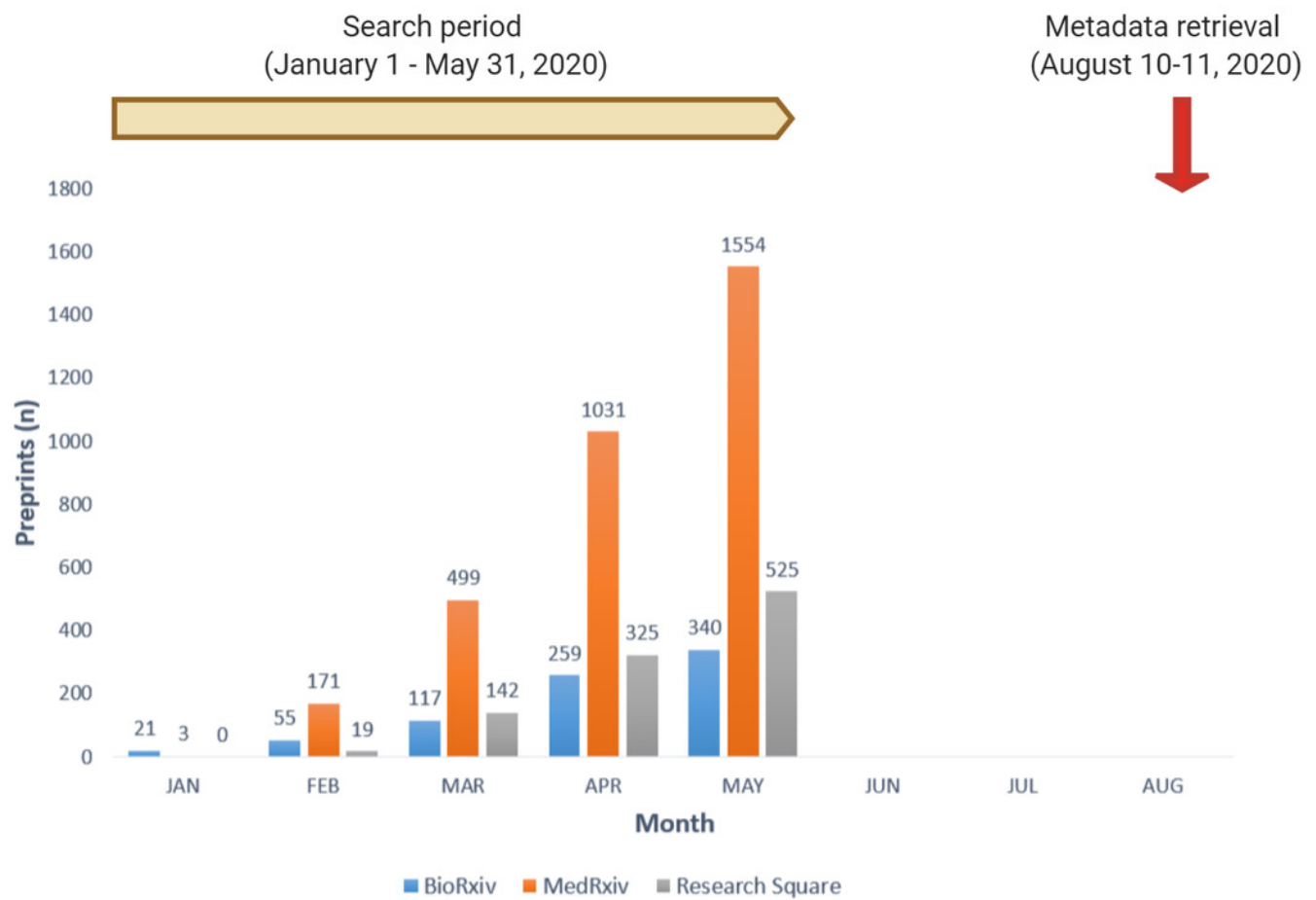
Figure 2

Time interval from posting on preprint server to publication in a scholarly journal

Time interval from posting on preprint server to publication on a scholarly journal

Kruskal-Wallis test p<0.001