

1 Publication practices during the COVID-19 2 pandemic: Biomedical preprints and peer- 3 reviewed literature

4
5 Yulia V. Sevryugina^{1*} and Andrew J. Dicks²

6
7 ¹ University of Michigan Library, Ann Arbor, Michigan, United States of America

8 ² School of Information, University of Michigan, Ann Arbor, Michigan, United States of America

9
10 * Corresponding author

11 E-mail: yulias@umich.edu (YS)

12 13 **Abstract**

14 The coronavirus pandemic introduced many changes to our society, and deeply affected
15 the established in biomedical sciences publication practices. In this article, we present a
16 comprehensive study of the changes in scholarly publication landscape for biomedical sciences
17 during the COVID-19 pandemic, with special emphasis on preprints posted on bioRxiv and
18 medRxiv servers. We observe the emergence of a new category of preprint authors working in the
19 fields of *immunology*, *microbiology*, *infectious diseases*, and *epidemiology*, who extensively used
20 preprint platforms during the pandemic for sharing their immediate findings. The majority of these
21 findings were works-in-progress unfitting for a prompt acceptance by refereed journals. The
22 COVID-19 preprints that became peer-reviewed journal articles were often submitted to journals

23 concurrently with the posting on a preprint server, and the entire publication cycle, from preprint
24 to the online journal article, took on average 63 days. This included an expedited peer-review
25 process of 43 days and journal's production stage of 15 days, however there was a wide variation
26 in publication delays between journals. Only one third of COVID-19 preprints posted during the
27 first nine months of the pandemic appeared as peer-reviewed journal articles. These journal articles
28 display high Altmetric Attention Scores further emphasizing a significance of COVID-19 research
29 during 2020. This article will be relevant to editors, publishers, open science enthusiasts, and
30 anyone interested in changes that the 2020 crisis transpired to publication practices and a culture
31 of preprints in life sciences.

32

33 **Introduction**

34 The lifecycle for any research starts and ends with a scholarly communication. Despite a
35 variety of avenues to communicate research findings, the foundation of the modern publication
36 practices is a publication in a peer-reviewed journal. The peer-review system is, at present, deeply
37 engraved in scientific minds as the golden standard for research quality. Certainly, the peer-review
38 process improves the drafted manuscript, but previous studies showed that its positive effect on
39 the overall quality of the final report is minor [1]. Besides, the traditional peer-review system is
40 notorious for reviewer bias, lack of agreement between reviewers, harsh criticism concealed by
41 anonymity, multiple cycles of reviews and rejections by different journals, and associated delays
42 and expenses [2].

43 Alternatively, or additionally, authors may choose to deposit their manuscripts to preprint
44 servers, institutionally or privately supported repositories of preprints. A preprint is a complete
45 manuscript shared publicly prior to officially undergoing the peer-review process. Notably, it has

46 likely undergone many rounds of internal review by authors and colleagues involved in its
47 preparation to ensure that ideas presented are well-versed and that the study can be easily validated
48 by the supporting data. A preprint submitted to one of the main preprint servers undergoes basic
49 screening for its scope, plagiarism, ethical issues and compliance, often performed as quickly as
50 within 24 hours, after which it is published online with a digital object identifier (DOI) that allows
51 it to be citable and trackable. Once posted on a preprint server, preprint can be read, commented
52 on-site or by email, and further shared on the Web and through social media (e.g. Twitter).
53 Additionally, any revision of a preprint's content or status, such as publication in a peer-reviewed
54 journal, is time-stamped.

55 Over the last decade, biomedical sciences have been slowly adapting a preprint culture.
56 The bioRxiv server owned by Cold Spring Harbor Laboratory was launched in November 2013
57 with the purpose of covering all aspects of life sciences research [3], and by 2018, it accumulated
58 37,648 preprints. By the end of September 2020, this number has doubled to 97,194 preprints. The
59 medRxiv preprint server, launched in June 2019 by BMJ, Yale University, and Cold Spring Harbor
60 Laboratory to cover all aspects of research in the medical, clinical, and related health sciences [4],
61 contained 11,329 preprints on September 30, 2020. The first and the most well-known preprint
62 server, arXiv.org was launched in 1991, and by the end of September 2020 had accommodated
63 1,769,336 preprints [5] in the fields of physics, mathematics, computer science, quantitative
64 biology, quantitative finance, statistics, electrical engineering and systems science, and economics.
65 The preprint platforms were widely explored as hubs for disseminating scientific findings [6],
66 especially for early-stage researchers [7], which revealed a number of social and technical issues
67 associated with their adoption [8]. The top four concerns with preprint servers included scooping
68 of scientific results [9], poor data quality [10], spread of misinformation [11], and varied and non-

69 transparent deposition policies [6]. Additionally, many authors refrained from depositing preprints
70 due to unclear journal policies on whether a preprint will be accepted for a journal publication after
71 being deposited on a server [12]. Nevertheless, research communities, including life sciences [13],
72 persisted in their adoption of preprints [14], gradually realizing that a preprint submission offers
73 many advantages:

- 74 1. Allows authors to establish the scientific priority by timestamping the first public record
75 of the research study [15];
- 76 2. Provides authors with a community feedback that can help them to further improve the
77 manuscript quality [16,17];
- 78 3. Expedites research sharing [15,16];
- 79 4. Increases research visibility [18,19];
- 80 5. Streamlines the journal submission process [20];
- 81 6. Allows sharing studies that are difficult to publish in traditional journals (works-in-
82 progress, negative results, replications, contradictions) [16];
- 83 7. Provides an open-access publication record [16].

84 Those benefits are what make scientific publishers and funders embrace preprints and
85 acknowledge them as works of scholarly communication. In June 2020, the National Institutes of
86 Health (NIH) launched the Preprint Pilot [21] to index preprints that come from NIH-funded
87 projects in PubMed, the largest database of biomedical scholarly works. This pilot builds on the
88 role of PubMed Central as a repository for NIH-supported [22] peer-reviewed articles and NIH's
89 guide notice NOT-OD-17-050 [23]. Library system vendors such as EBSCO, Proquest, Ex Libris,
90 and OCLC WorldCat have been investing substantially in open access discovery tools for a variety
91 of open sources, including preprints [24]. The practice of inviting journal submissions from

92 preprint servers is wide spread [25]. In June 2020, MIT Press and the Berkeley School of Public
93 Health launched a new COVID-19 journal, *Rapid Reviews: COVID-19*, which, after a thorough
94 peer-review, will publish preprint articles with good research and discredit those with bad [26].
95 Considering the reported influence of preprints on policy-making during the COVID-19 pandemic
96 [27] and ongoing article retractions [28], the concern about the quality of un-refereed preprints is
97 genuine [29]. The Sinai Immunology Review Project is an example of an institutional effort to
98 review and validate the COVID-19 related preprints posted to medRxiv and bioRxiv servers [30].
99 The Review Commons platform, launched in December 2019, is another initiative on that
100 direction, where preprint authors are offered an opportunity to request a journal-independent
101 portable peer review [31]. Lastly, it is worth mentioning new editorial policies from *eLife* that
102 resulted in the launch of Preprint Review [32] and made the preprint deposition mandatory prior
103 to a journal submission [33].

104 Since the first report on a pneumonia of unknown origin in Wuhan, China on Dec 30, 2019
105 [34], over 10,000 articles have been published in scholarly journals and on preprint servers, as well
106 as about 4,000 clinical trials registered. These figures attest an outstanding response of the
107 scientific community to what was in January 2020 simply known as the “novel coronavirus” [35].
108 As our health workers fought on the front lines against coronavirus and general population
109 followed rapidly changing policies, the scientific community immediately sought to contribute by
110 sharing and exchanging newly discovered information promptly and openly. Fortunately, the
111 necessary tools were already in-place and biomedical researchers quickly adopted preprint servers
112 as suitable platforms for timely, open, and transparent scientific communication. In this paper, we
113 set our goal to assess all aspects related to this adoption by analyzing the effect of COVID-19
114 pandemic on publication practices in biomedical sciences.

115

116 **Methods**

117 **Scope.** The scope of this study is COVID-19 related literature. For preprints, we focused on
118 medRxiv and bioRxiv preprint servers that compiled all their COVID-19 SARS-CoV-2 preprints
119 in the database that at the time of this study contained 1,964 bioRxiv and 7,258 medRxiv preprints.
120 For peer-reviewed scholarly works, we referred to journal and review articles indexed in PubMed,
121 of which about 1,400 had associated preprints.

122 **Timeline.** The first media statement on ‘viral pneumonia’ in Wuhan, People’s Republic of
123 China appeared on Dec 31, 2019 [36]. By the end of January, the first research reports appeared
124 as preprints, clinical trials, and journal articles [37]. In this manuscript, we focus on preprints
125 deposited between Jan 1 and Sept 30, 2020. The publication rates for these preprints were
126 evaluated on Sept 30 and on Dec 7, 2020. The deposition rate for preprints was evaluated during
127 Jan 1 – Nov 30, 2020. Altmetric analysis is based on articles published during Jan 1 - Nov 19,
128 2020, and includes articles with a publication date of Dec 1, 2020 (retrieved on Nov 19, 2020).
129 The latter date was included because we noticed that when date is unknown, Dimensions
130 automatically assigns it the first of the month. Similar enrichment of dates on the first of the year
131 and each month was observed in Crossref [38].

132 **Terminology.**

133 *Preprint* is defined according to the COPE (Committee of Publication Ethics) definition
134 [39]: “A preprint is a scholarly manuscript posted by the author(s) in an openly accessible platform,
135 usually before or in parallel with the peer review process.” As the main subjects of this paper are
136 COVID-19 related preprints posted on bioRxiv and medRxiv preprint servers, we will often

137 address to them as simply “preprints”. When discussing other preprints, preprint servers, or
138 publication topics, we will specify it in the text.

139 *Preprint server* – a repository of preprints. This study specifically focuses on bioRxiv and
140 medRxiv preprint servers.

141 *Preprint category* – a subject-filed category defined by a preprint server and selected by
142 authors during the deposition process.

143 *Journal category* – a category assigned by Scopus to the journal to define its scope.

144 *Published preprint* – a preprint of an article in a peer-reviewed journal.

145 *Elapsed time (T_{Σ})* – interval between the date when a preprint was deposited to the server
146 and publication date for its journal article analogue.

147 *Pre-submission time (t_{α})* – interval between the date when a preprint is deposited to the
148 server and the date when it is submitted to the journal.

149 *Review time (t_R)* – interval between the date when manuscript is submitted to the journal
150 and the date it is accepted for publication.

151 *Production stage time (t_{β})* – interval between the acceptance date for a manuscript and the
152 date the peer-reviewed journal article appears online.

153 **Data sources.** This paper examines data acquired from a number of sources, including the
154 database of COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv [40], Rxivist [41],
155 Crossref [42], E-utilities [43], Dimensions [44], CORD-19 [45], and CADRE[46].

156 Metadata for each individual COVID-19 preprint deposited to bioRxiv or medRxiv was
157 gathered by accessing the bioRxiv database of COVID-19 SARS-CoV-2 preprints from medRxiv
158 and bioRxiv, to which we will further refer as BioRxiv API [40]. Data were retrieved in JavaScript

159 Object Notation (JSON) format. Data analysis and visualization was done in Python (pandas,
160 numpy, requests, matplotlib, bokeh, and seaborn) using Jupyter Notebook.

161 To search PubMed, we used Entrez Programming Utilities (E-utilities) [43], an application
162 programming interface (API) that allows searching 38 databases from the National Center for
163 Biotechnology Information (NCBI). For E-Utilities, data were downloaded *via* CSV and converted
164 to Microsoft Excel for further analysis and visualization.

165 Rxivist [41] is a Python-based web crawler that parses the bioRxiv website, detects newly
166 posted preprints, and stores metadata about each item in a PostgreSQL database. The metadata we
167 extracted contained title, authors, submission date, category, DOI for preprint and, if published,
168 the new DOI and the journal of publication.

169 Crossref [42] is an official DOI registration agency of the International DOI Foundation
170 that establishes a cross-publisher citation linking system for academic that include journals,
171 conference proceedings, books, data sets, etc. It works with thousands of publishers to provide
172 authorized access to their metadata including DOI, publication date and other basic information.

173 CORD-19 or COVID-19 Open Research Dataset [47] is a free resource of over 200,000
174 scholarly articles about COVID-19, SARS-CoV-2, and related coronaviruses prepared by the
175 Allen Institute for AI (AI2) in collaboration with many partners and released on March 16, 2020.
176 We used its 2020.09.02 release downloaded on 2020.09.30 from CADRE [46] for metadata
177 associated with refereed journal articles.

178 Dimensions [48] is a comprehensive database that links scholarly outputs to a research
179 analytics suite to track the impact of research across its life cycle. Dimensions tracks many preprint
180 servers [49] but we only used it for bioRxiv and medRxiv preprints (see Data Flow Chart in SI for
181 the links between used datasets).

182 **Statistical analysis.** Descriptive analysis of the data, Student's t-test, and a one-way
183 ANOVA were conducted on the Statistical Package for Social Sciences version 27 (SPSS). Data
184 are presented as means (M) \pm standard deviations (SDs), accompanied by medians and modes,
185 when necessary.

186 **Altmetric data.** Altmetric Attention Scores were retrieved from Dimensions by querying for
187 articles published between Jan 1 and Nov 19, 2020, in selected journals. For COVID-19 related
188 publications in a journal, we used the recommended query for COVID-19 [50] supplemented with
189 the query for journal id. For journal publications unrelated to COVID-19, we gathered all articles
190 from the journal and removed those related to COVID-19 from the dataset. For COVID-19 articles
191 that were associated with preprints, we matched DOI's of articles found by Dimensions to DOI's
192 of articles we identified earlier as published preprints. To identify articles that were not associated
193 with preprints, we removed from the dataset of COVID-19 related publications all articles that had
194 associated bioRxiv and medRxiv preprints.

195 **Data challenges and study limitations.**

196 **Analysis of published preprints.** When a preprint is published in a peer-review journal, a
197 reference to the new DOI of the journal article appears next to its title, and DOIs of a preprint and
198 a published article are permanently linked in indexing platforms and tools, which pull from various
199 APIs. Rxivist [41] showed to be an excellent tool for extracting published DOIs for preprints
200 eventually appearing as peer-reviewed journal articles but only when bioRxiv records linked
201 preprints to their external publications. Rxivist also had a two weeks delay in updating its metrics,
202 and it might be of this delay that some peer-reviewed preprint analogues were missing from
203 Rxivist. Additionally, at the time of our study, Rxivist did not include medRxiv preprints in its
204 database, which changed after Nov 27, 2020. We found that the most reliable method of extracting
205 metadata about each individual preprint was by accessing the BioRxiv API [40]. Using the Python
206 library requests, we were able to extract information about each preprint based on DOI, which
207 gave us a column called ‘published.’ Within this column, if the preprint was also published in a
208 journal, the metadata provided the DOI that corresponded to the published version of the paper.
209 To ensure we found all published preprints, we also accessed data from Crossref, Dimensions, and
210 CORD-19 APIs. To establish the linkage between the preprints and corresponding peer-reviewed
211 journal articles we performed both, DOI and title matching. All channels were then combined and
212 duplicates were dropped. For detailed demonstration of data obtained by every data channel, see
213 Published Collections in SI.

214 To validate whether we found all peer-reviewed preprint versions based on a combination
215 of Rxivist, Crossref, CORD-19, Dimensions, and BioRxiv API, we randomly selected a sample of
216 100 preprints that our data returned as “unpublished” from both bioRxiv and medRxiv, and
217 searched Google Scholar by title. Our analysis of “unpublished” preprints returned 10% of bioRxiv
218 and 4% of medRxiv preprints as being published in refereed journals. All found journal

219 publications had slight modifications in article titles or authors' list, and the original "unpublished"
220 preprints were not linked on preprint servers to the corresponding published versions. In
221 comparison, this false-negative rate is lower than the 37.5%, reported by Blekhman *et al.* [51]. All
222 manually found journal article versions of "unpublished" preprints were manually added to data
223 discussed in this article.

224 **Double DOI.** When we looked for published preprints based on title matching, we encountered
225 a few instances when two published DOIs existed for a peer-reviewed preprint version. In one
226 case, it was erratum for the paper and in the other case it was a publication on another preprint
227 server. In both cases, we used only the DOI for the article in the peer-reviewed journal and
228 publication on another preprint server was removed from further analysis. We also encountered a
229 few cases when preprints with different DOIs were linked to the same DOI of the published
230 version. On inspection, preprints with different DOIs were somewhat similar in titles and authors'
231 list but not identical. For our analysis, we kept only one DOI for a preprint that was published
232 earlier.

233 **PubMed.** As mentioned in the Introduction, the NIH Preprint Pilot started in June 2020 and at
234 this stage, it primarily focuses on NIH-supported and COVID-19 related preprints from various
235 servers. By Sept 26, PubMed indexed 1,048 preprints from medRxiv, bioRxiv, ChemRxiv, arXiv,
236 Research Square, and SSRN, of which 1,043 were on COVID-19, and this constituted only 11.5%
237 of 9,072 medRxiv and bioRxiv COVID-19 related preprints from the BioRxiv API. For these
238 reasons, we did not use PubMed as a data source for preprints. We used PubMed (through E-
239 Utilities) to obtain metadata on peer-reviewed articles of “Journal Article” and “Review” article
240 types as the most traditional types of scholarly output. These two types constitute about 24% [52]
241 of all PubMed publications that include 187 different publication types [53].

242 In analyzing PubMed dates, we found that articles with a missing day-of-publication were
243 coded as being published on January 1st; a similar issue was reported earlier for Crossref dates
244 [38]. Based on low number of preprints in January, we decided to avoid discussing January data
245 for PubMed (this month is omitted in Fig 2).

246 **Categories.** In general, we used a single category for a preprint as indicated in metadata from
247 the BioRxiv API. However, as of September 25, we found six out of total 1,956 of COVID-19
248 related bioRxiv preprints (0.3%) that displayed two categories. Since this contradicted the servers’
249 statement that “Only one subject area can be selected for an article”, we omitted the additional
250 category in our analysis. The journals’ scope categories were extracted from Crossref [54].

251 **Publication process delays.** Preprints deposition dates were extracted from Crossref. For
252 journal articles received, accepted, and published online dates, we used E-Utilities:
253 PubMedPubDate@PubStatus = “received”; PubMedPubDate@PubStatus = “accepted” and
254 ArticleDate@DateType="Electronic". When ArticleDate@DateType="Electronic" from PubMed
255 was not available, we substituted it with the “created-date” from Crossref.

256 Before deciding on which dates to use in our studies, we carefully analyzed those used in
257 previous studies and noted some inconsistency between different authors (Table 4). Unfortunately,
258 dates from different sources intended to represent the article publication date differ between each
259 other (see Article Publication Dates in SI) and are often incomplete. The amount of missing dates
260 in our dataset was 44.7% for “pub_date” from PubMed and over 90% for “published-online” or
261 “published-print” from Crossref. In this article, we referred to refereed journals to resolve
262 disputable publication dates. Our scrupulous analysis showed that
263 ArticleDate@DateType="Electronic" from PubMed is the date that exactly corresponds to the
264 online publication date of a journal article. When this date was not available (6.6% of the cases),
265 we used the “date-created” from Crossref that indicates the day when DOI for the article is created.
266 The “date-created” from Crossref differed from the online publication date of a journal article in
267 24% of the cases. This number was only 15.4% for “pub_date” from PubMed, but the latter had
268 high percentage of missing dates, while the “date-created” from Crossref had none. These

269 variations in measuring article publication dates may lead to some discrepancy between reports
270 and inconsistent statistical analysis. Based on our experience, we recommend extracting article
271 publication dates either directly from the journal articles or using
272 ArticleDate@DateType="Electronic" in PubMed and/or "date-created" from Crossref.

273 To assess the preprint *pre-submission time*, we subtracted the preprint deposition date from
274 the date the journal articles was "received". To assess the *review time*, we subtracted the date the
275 journal articles was "received" from the date it was "accepted". To assess the *production stage*
276 *time*, we subtracted the date the journal article was "accepted" from the date it was posted online
277 ("Electronic"). To evaluate the *elapsed time*, we subtracted the preprint deposition date from the
278 date the journal articles was posted online. One COVID-19 related journal article [55] was
279 excluded from our analysis of publication delays due to a peculiar date of submission, in May
280 2019, long before the emergence of the first report on COVID-19.

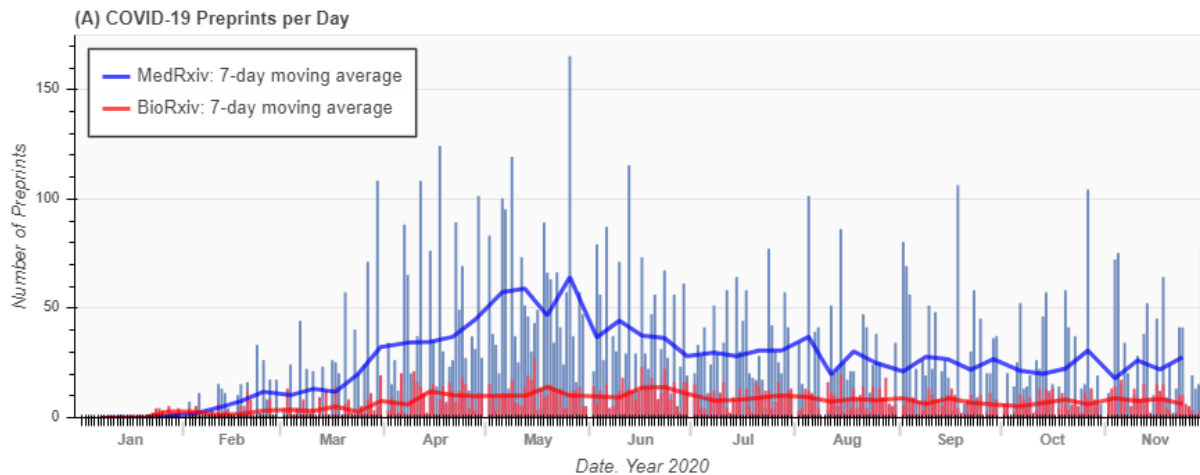
281

282 **Results**

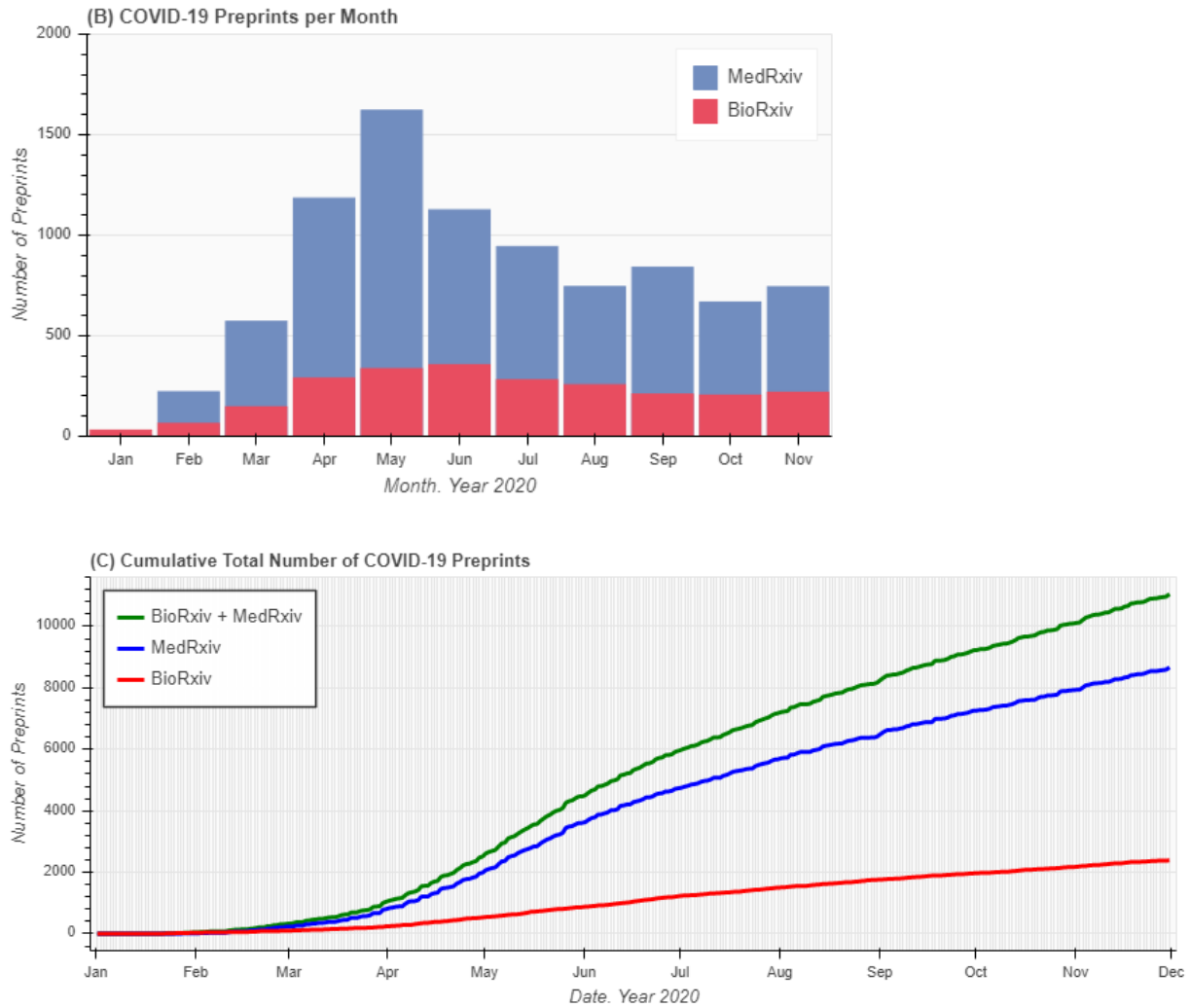
283 **1. Submission rates during the coronavirus pandemic**

284 With the COVID-19 outbreak, both bioRxiv and medRxiv experienced a surge of COVID-
285 19 preprint submissions. The first preprint on COVID-19 appeared on Jan 13 of 2020 [56], and the
286 number of submissions increased at a staggering rate afterwards (Fig 1). By Dec 4, 2020, the
287 number of COVID-19 preprints reached 8,750 on medRxiv and 2,416 on bioRxiv. In comparison,
288 we identified only 187 Zika related, and 17 Ebola related preprints deposited to bioRxiv during
289 the entire duration of each epidemic (for Zika: Nov 2015 - Aug 2017; for Ebola: May 2014 - Jan
290 2016).

291 Consistently throughout the pandemic, medRxiv experienced a significantly higher flux of
292 COVID-19 preprints as compared to bioRxiv (Table 1 and Scholarly Output in SI). On average,
293 medRxiv preprints on COVID-19 constituted 78% (SD = 2%) of combined bioRxiv and medRxiv
294 preprints on any single month, except January, when the number of COVID-19 related medRxiv
295 preprints was only 27% of COVID-19 related bioRxiv preprints. May was the most productive
296 month for authors of medRxiv preprints. In June, the number of medRxiv COVID-19 preprints
297 declined by 31%, while the number of bioRxiv preprints increased by 6%. After June, we noted a
298 slow decline in the number of COVID-19 preprints on both servers, with the total number of
299 bioRxiv and medRxiv COVID-19 preprints in November representing only 49% of those deposited
300 in May.



301



302

303

304 **Fig 1. Trends in submission of COVID-19 preprints.**

305 COVID-19 Preprints, bioRxiv (red), medRxiv (blue), and combined (green): (A) Daily; (B)
306 monthly; (C) cumulative during Jan 1 – Nov 30, 2020.

307

308 We then analyzed how the high flux of COVID-19 preprints compares to that of non-
309 COVID-19 preprints and biomedical journal articles published during the same period (Fig 2,
310 Table 1). Notably, during the pandemic, we observe the growth in preprints submissions not only
311 on COVID-19 but also on other topics unrelated to coronavirus. Consistently, bioRxiv received
312 significantly more of non-COVID-19 preprints than medRxiv. While the COVID-19 submissions

313 have been slowly declining since June, those unrelated to coronavirus are relatively stable for
 314 bioRxiv and have been increasing for medRxiv (see Scholarly Output in SI), indicating a growing
 315 popularity of medRxiv server.

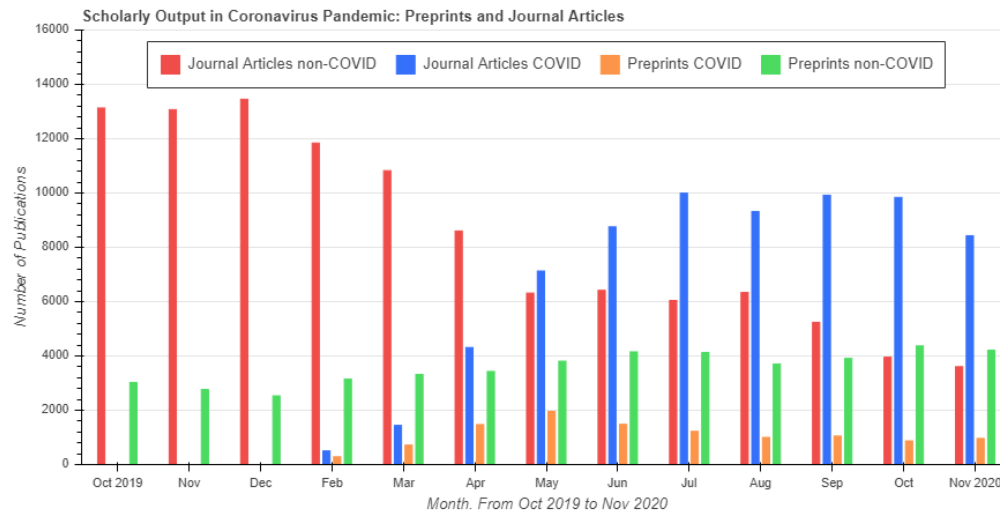
316

317 **Table 1. Descriptive statistics and independent samples t-test(s) for preprint submissions**
 318 **prior to and during the COVID-19 pandemic.^a**

Timeline	Prior to	During 2020				
Preprint platform	Combined bioRxiv & medRxiv	medRxiv			bioRxiv	
Topics	Others	COVID-19		Others		COVID-19
<i>M</i>	2772	3817	866	429	3387	236
<i>SD</i>	247	413	383	186	263	88
<i>N</i>	3	10	10	10	10	10
<i>t-test(s)</i>	$t(11) = 4.09, p = 0.002,$ <i>Cohen's d = 2.69</i>	$t(18) = 3.25, p = 0.004,$ <i>Cohen's d = 1.45</i>		$t(11) = -35.99, p < 0.001,$ <i>Cohen's d = -16.10</i>		
				$t(18) = -29.09, p < 0.001,$ <i>Cohen's d = -13.01</i>		

319 ^a Prior to pandemic is during Oct-Dec, 2019, and during the pandemic includes Feb-Nov, 2020.

320



321

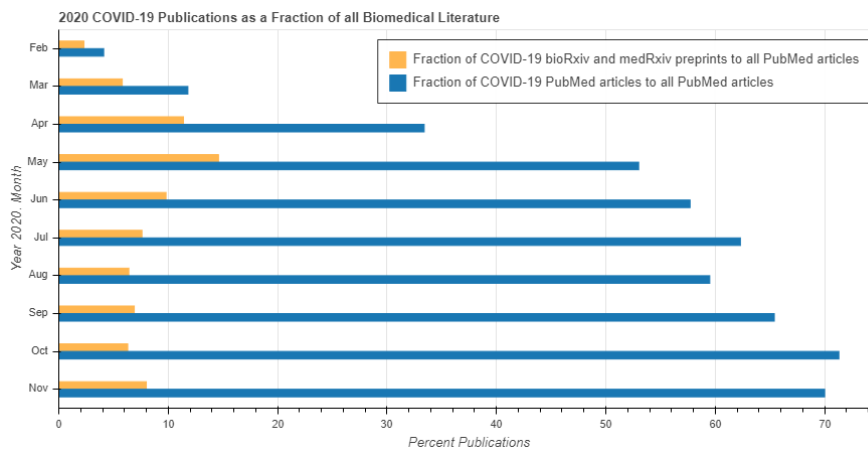
322 **Fig 2. Scholarly output during the coronavirus pandemic.**

323 Combined medRxiv and bioRxiv preprints related to COVID-19 (orange); combined medRxiv and
324 bioRxiv preprints unrelated to COVID-19 (green); PubMed journal articles related to COVID-19
325 (blue); PubMed journal articles unrelated to COVID-19 (red), prior to (Oct – Dec, 2019) and
326 during (Feb – Nov, 2020) the coronavirus pandemic.

327

328 The growth of preprints, however outstanding, is lagging behind the amount of peer-
329 reviewed journal articles (see Scholarly Output in SI). In 2019, preprints represented 2.6% of all
330 biomedical literature in PubMed [57]. This percentage was derived with respect to biomedical
331 preprints from multiple servers and all biomedical literature indexed in PubMed [58], while our
332 study is defined by bioRxiv and medRxiv preprints in relation to “Journal Article” and “Review”
333 article types in PubMed. Based on our analysis, in February, the amount of COVID-19 preprints
334 from medRxiv and bioRxiv constituted only 2% of biomedical articles on all topics but this fraction
335 increased to 15% in May (Fig 3). The number of peer-reviewed articles on COVID-19 has been
336 growing since the start of pandemic reaching a peak in July. In contrary, the number of unrelated
337 to coronavirus peer-reviewed literature has been slowly declining. As a result, the fraction of

338 COVID-19 journal articles with respect to all articles indexed in PubMed has been increasing since
339 the start of pandemic and reached 71% in October. At that time, the amount of COVID-19 bioRxiv
340 and medRxiv preprints was at 9% with respect to COVID-19 peer-reviewed literature in PubMed,
341 but this fraction was as high as 57% in February 2020. Thus, early in pandemic, there were over
342 half as many preprints as there were peer-reviewed articles about the newly emerged coronavirus.
343



344

345 **Fig 3. Percent ratio of COVID-19 publications to all journal articles in PubMed.**

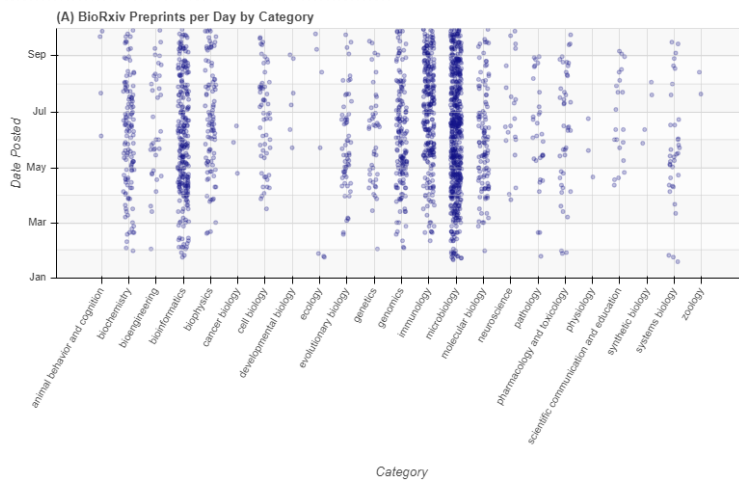
346 Combined COVID-19 medRxiv and bioRxiv preprints (orange) and COVID-19 journal and review
347 articles (blue).

348

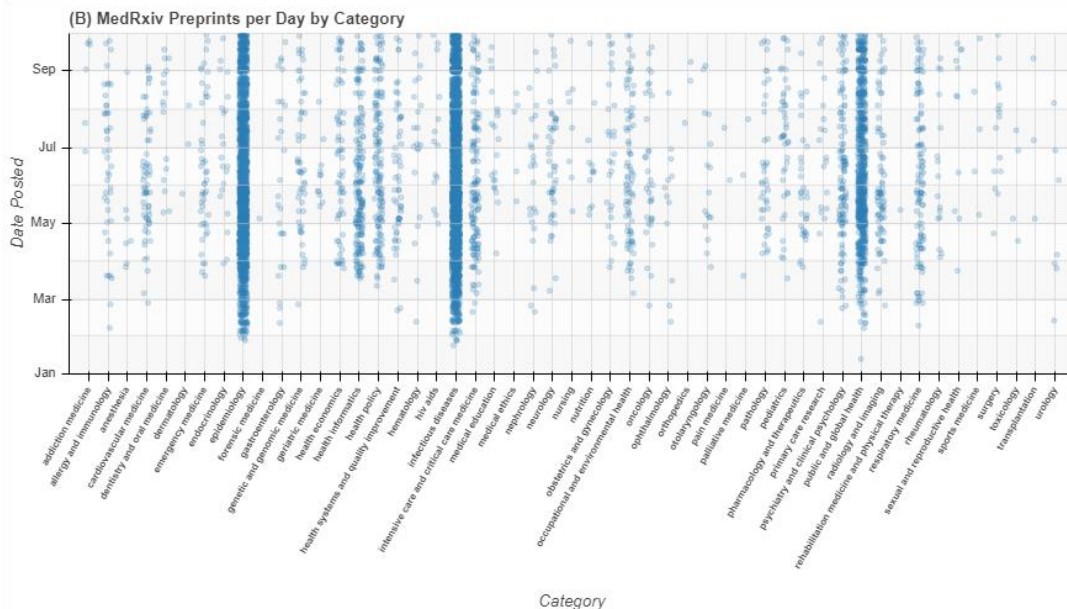
349 2. Preprints categories

350 Preprints are deposited in a certain category that defines their subject area. As expected,
351 *microbiology*, the field that studies microscopical organisms, such as viruses, is the top category
352 for bioRxiv preprints on COVID-19 (Fig 4A). Other leading bioRxiv categories for COVID-19
353 preprints are *bioinformatics* that analyzes and interprets large and complex biological data, and
354 *immunology* that studies the immune system. The top three leading categories for COVID-19

355 medRxiv preprints are *infectious diseases, epidemiology, and public and global health* (Fig 4B).
356 Clearly, the study of *infectious diseases* focuses on infectious agents, *epidemiology* deals with the
357 possible control of diseases, and *public and global health* focuses on disease prevention, all of
358 them closely related to our understanding of the nature of coronavirus, and the development of
359 corresponding diagnostic methods and treatment. The three leading categories for COVID-19
360 preprints in bioRxiv incorporate 1.3 times as many preprints as the remaining 20 subject
361 categories. The three leading categories for COVID-19 preprints in medRxiv incorporate 3.2 times
362 as many preprints as the remaining 48 subject categories (see Categories Analysis in SI). As
363 expected, the leading categories have the highest daily accumulation rates (Fig 5).



364

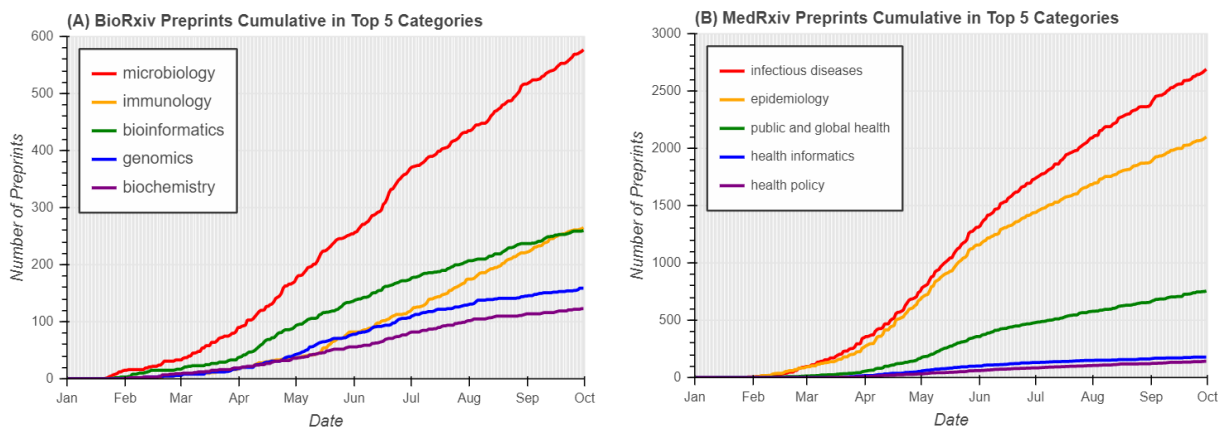


365

366 **Fig 4. Daily COVID-19 preprint submissions by preprint category.**

367 (A) BioRxiv and (B) medRxiv preprint submissions (represented by dots).

368



369

370 **Fig 5. Cumulative COVID-19 preprint submissions in the top 5 categories.**

371 (A) BioRxiv and (B) medRxiv.

372

373 We also analyzed categories for bioRxiv preprints unrelated to COVID-19 (see Categories

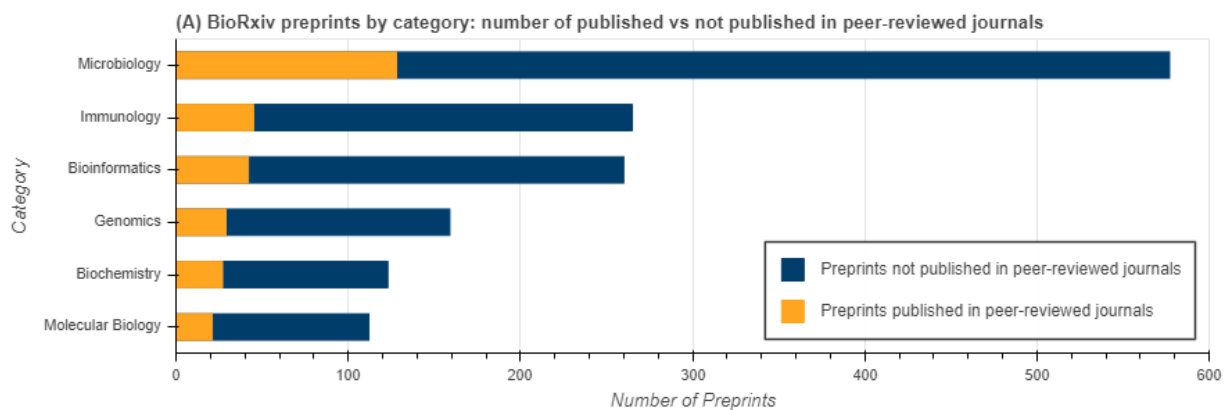
374 Analysis in SI) deposited into the server during Jan 1 – Sept 30, 2020. We found that the majority

375 of preprints unrelated to COVID-19 were deposited into categories of *neuroscience*, *microbiology*,
376 *and bioinformatics*. In fact, *neuroscience* was the leading category through Nov 2018 as was
377 demonstrated in Blekhman's study [51]. That same study named *bioinformatics*, *genomics*, and
378 *evolutionary biology* as other leading categories. Interestingly, the field of *bioinformatics* remains
379 among the leading categories for both COVID-19 related and unrelated preprints, prior to or during
380 the coronavirus pandemic, which demonstrates the growing importance of this field for all subject
381 areas, its interdisciplinary nature, and the familiarity of bioinformatics research community with
382 the preprint culture.

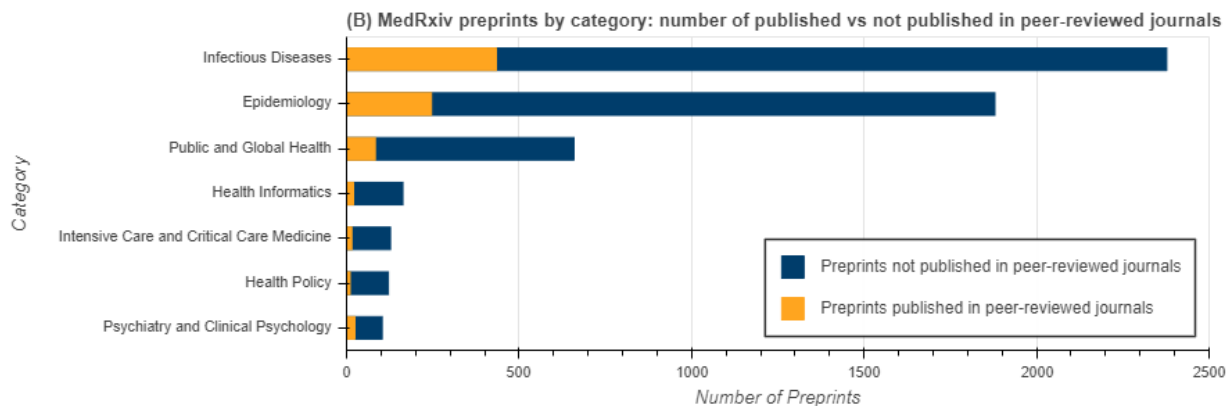
383 **3. Published preprints**

384 Preprints deposited on a preprint server can later undergo a peer-review process to be
385 published in refereed journals. Here, we will refer to a preprint that has a peer-reviewed journal
386 article analogue as a published preprint. Once a preprint is published in peer-reviewed journal,
387 preprint and article's DOIs are permanently linked in indexing databases and on the preprint server;
388 nevertheless, establishing this association can sometimes be challenging. For instance, the title of
389 the research study or the authors' list may change during the publication process, complicating
390 auto-matching the preprint and its published version. Computing the number of preprints published
391 as journal articles requires tedious analysis and the design of matching algorithms [59]. To ensure
392 we find all published preprints, we combined data from various APIs (BioRxiv API, Crossref,
393 Dimensions, Rxivist, and COVID-19) and matched publications based on both, DOIs and
394 publication titles (see Published Collections in SI). We found that during Jan 1 – Sept 30, on
395 average, 18% of COVID-19 preprints deposited to bioRxiv and medRxiv servers resurfaced as
396 peer-reviewed journal articles.

397 As expected, the majority of published preprints were those deposited into the leading
398 categories: *microbiology*, *immunology*, and *bioinformatics* represented 59% of all published
399 bioRxiv preprints, while *infectious diseases*, *epidemiology*, and *public and global health* contained
400 76% of all published medRxiv preprints (Fig 6). The publication rates vary across the preprint
401 categories (Fig 7). Thus, COVID-19 preprints in bioRxiv categories of *microbiology* and
402 *biochemistry* display the highest publication rates of 22%. The *infectious diseases* category in
403 medRxiv contained 18% of preprints that later appeared as journal articles and it alone provided
404 43% of all published medRxiv preprints on COVID-19. At the same time, other leading categories,
405 *epidemiology* and *public and global health*, both had publication rates of 13%, which is lower than
406 an average rate for medRxiv. The lowest publication rates were observed for COVID-19 preprints
407 deposited into the *neuroscience* category in bioRxiv and *health economics* in medRxiv. The
408 highest publication rate of 30% was observed for *pathology* in medRxiv, although it was only 10%
409 for bioRxiv.



410



411

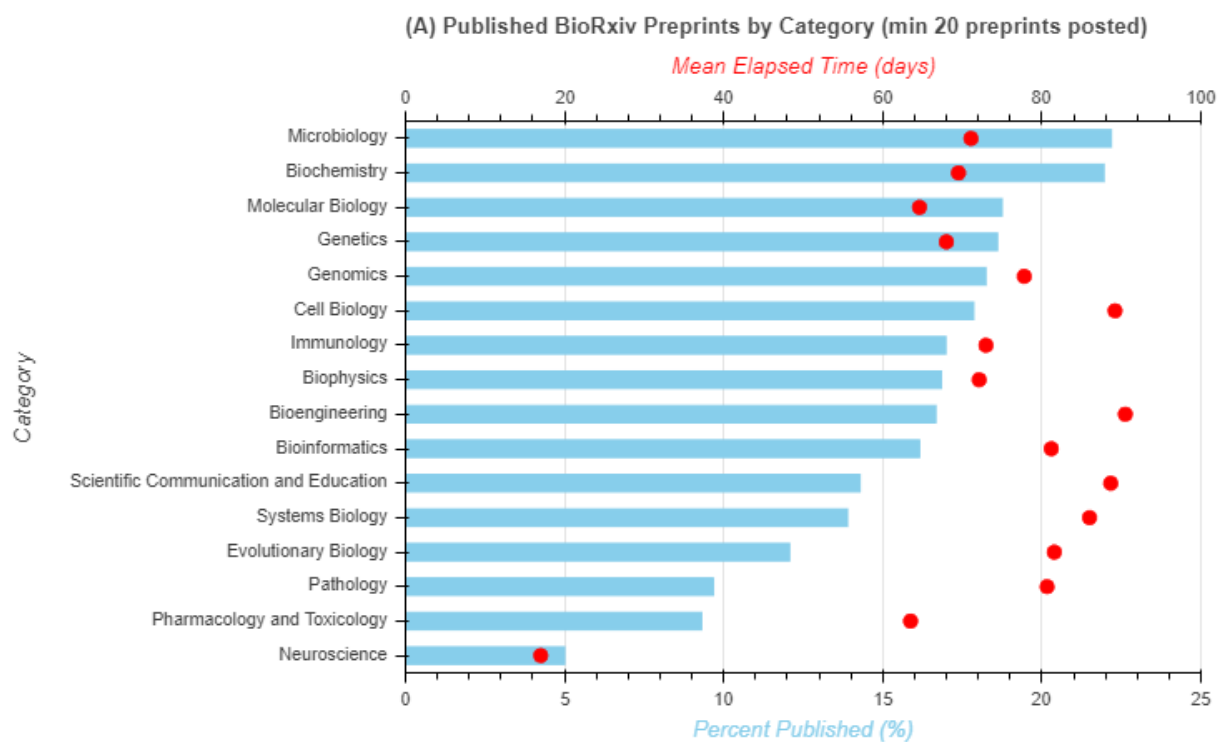
412 **Fig 6. Published vs. not published COVID-19 preprints by category.**

413 (A) BioRxiv and (B) medRxiv preprints that have journal article analogues (yellow) and those that

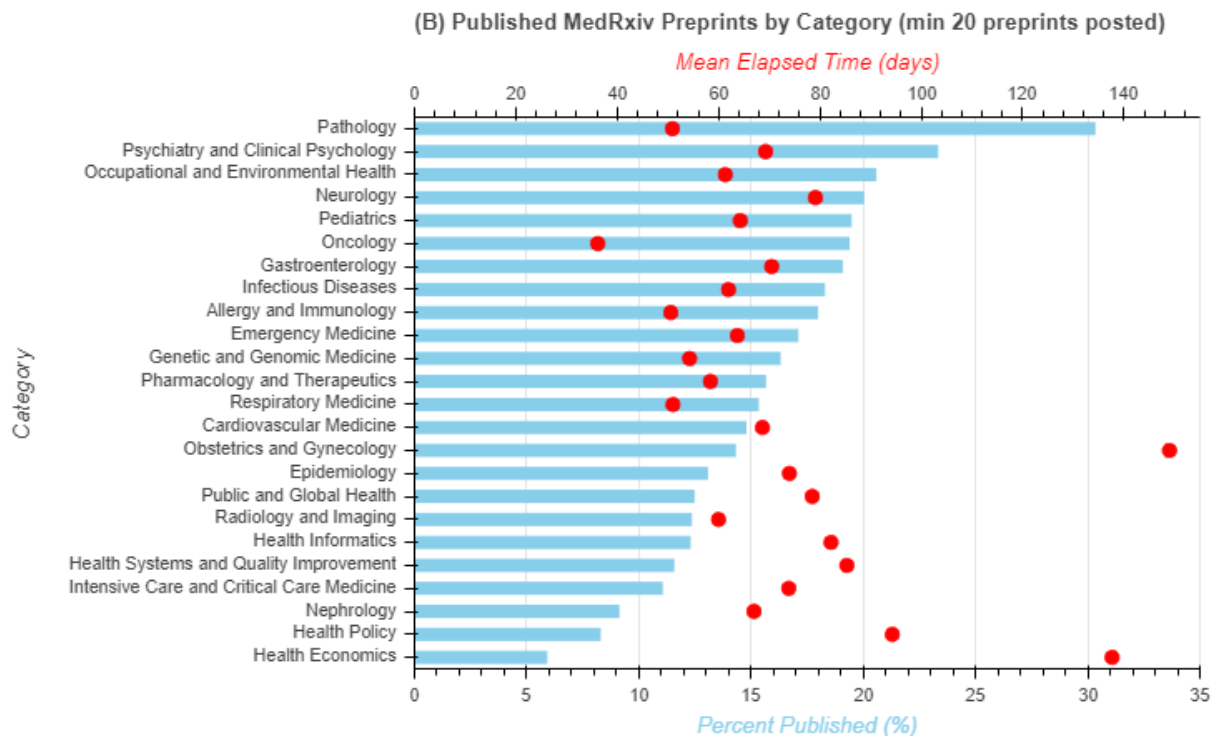
414 exist only as preprints (blue). The total number of preprints in the category is the sum of orange

415 and blue bars. Only categories with at least 100 preprints are displayed for clarity purposes.

416



417



418
 419 **Fig 7. Published COVID-19 preprints (in blue) and mean elapsed time (T_E , in red) by**
 420 **category.**

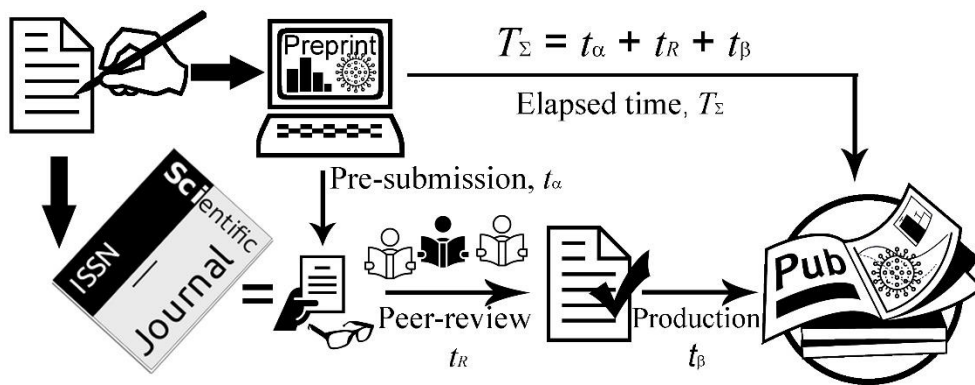
421 (A) BioRxiv and (B) medRxiv categories with at least 20 preprints published during Jan 1 – Sept
 422 30, 2020.

423
 424 The publication rates of 18% observed for COVID-19 related preprints on Sept 30, is low
 425 as compared to publication rate of 42% [51] or 70% [3] reported for pre-pandemic preprints or for
 426 previous health crises. For instance, for Ebola, 60% of the deposited preprints were published in
 427 PubMed-indexed journals, and for Zika, the corresponding amount was 48% [60]. One limitation
 428 to our study is that articles that are undergoing the peer-review or various editorial processes are
 429 invisible to us. The dependence of publication rates on the analysis timespan was noted by
 430 Blekman *et al.* [51], who showed that publication rate for preprints was 67% during the
 431 comprehensive study that covered the period of several years (2013-2016) but only 20% during

432 the last year of their study (2018) [51]. Similarly, our reported publication rates are higher than
433 rates reported by two independent studies of COVID-19 preprints deposited by the end of April
434 2020. One study reported that 6.1% of 2,102 COVID-19 related preprints from Dimensions
435 resurfaced as peer-review journal articles [61], while another study reported that only 4% of
436 COVID-19 related bioRxiv and medRxiv preprints were published [62]. We further evaluated the
437 possibility that publication rates for COVID-19 preprints deposited into medRxiv and bioRxiv
438 servers between Jan 1 and Sept 30 were underestimated, especially considering the length of the
439 publication process for COVID-19 related preprints (see next section). To that end, we reanalyzed
440 publication rates on Dec 7, 2020, for COVID-19 preprints deposited between Jan 1 and Sept 30,
441 2020, and we found them at 34% and 29%, for bioRxiv and medRxiv preprints, respectively.
442 Despite being higher than the publication rate of 18% derived from our data in October, reanalyzed
443 publication rates are still low.

444 4. Publication delays for published preprints

445



446

447 **Scheme 1. Preprint publication process** [63].

448

449 We analyzed various delays involved in the publication process using the model depicted
450 in Scheme 1. In our proposed model, the interval between preprint posting on a preprint server and
451 online publication of the corresponding peer-reviewed journal article constitutes the *elapsed time*
452 (T_{Σ}), which can be further expressed as:

$$453 \quad T_{\Sigma} = t_{\alpha} + t_R + t_{\beta} \quad (1),$$

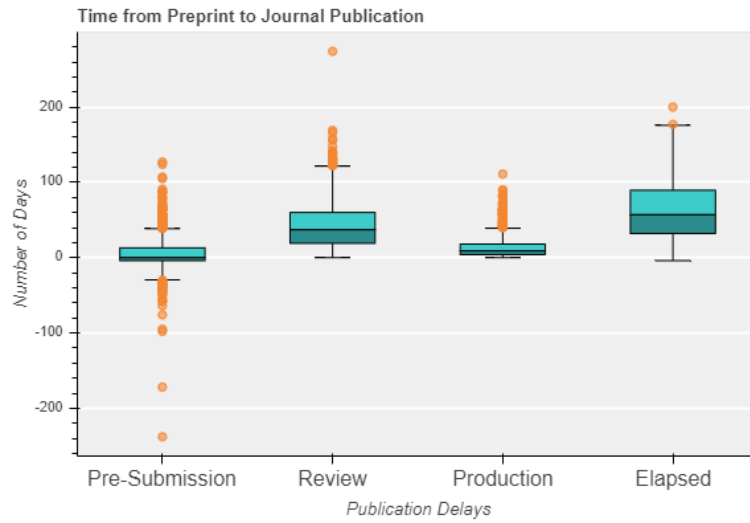
454 where *pre-submission time* (t_{α}) is the interval between preprint posting on a preprint server
455 and its submission to the peer-reviewed journal; *peer-review time* (t_R) is the duration of the peer-
456 review process; and *production stage time* (t_{β}) is the interval between article official acceptance
457 statement and its publication online.

458 The descriptive statistics for these publication delays are summarized in Table 2 and Fig 8
459 and will be discussed in detail below. It is worth noting that none of the publication delays display
460 a standard Gaussian distribution (Fig 9), thus we discuss both their medians and means.

461
462 **Table 2. Publication delays (in days) for combined medRxiv and bioRxiv COVID-19**
463 **preprints deposited during Jan 1 – Sept 30, 2020.**

Publication delays	Symbol	Mean	SD	Median	IQR	Mode	N
<i>Pre-submission</i>	t_{α}	5.6	23.9	0	17	-1	973
<i>Review</i>	t_R	43.4	32.4	37	41	15	950
<i>Production</i>	t_{β}	14.6	15.7	9	14	4	916
<i>Elapsed time</i>	T_{Σ}	63.4	39.5	57	58	42	1099

464

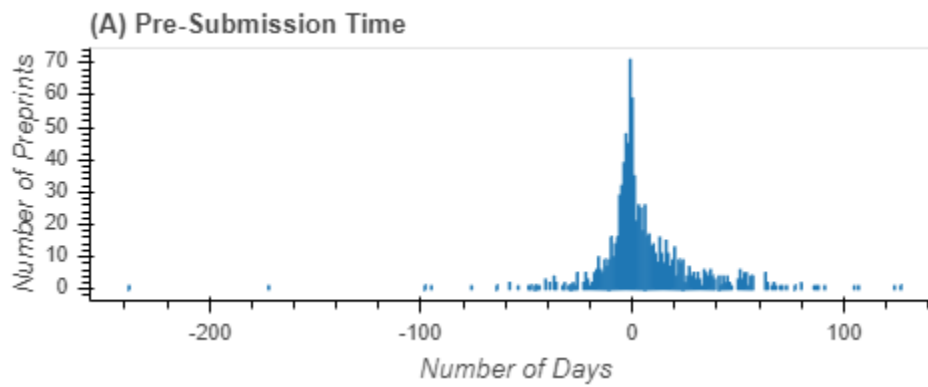


465

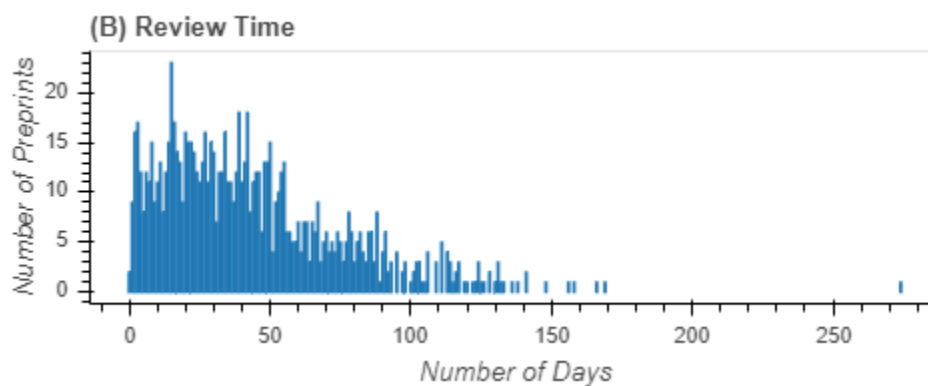
466 **Fig 8. Time from preprint to journal publication.**

467 Box plot displaying publication delays for combined medRxiv and bioRxiv COVID-19 preprints
468 deposited during Jan 1 – Sept 30, 2020.

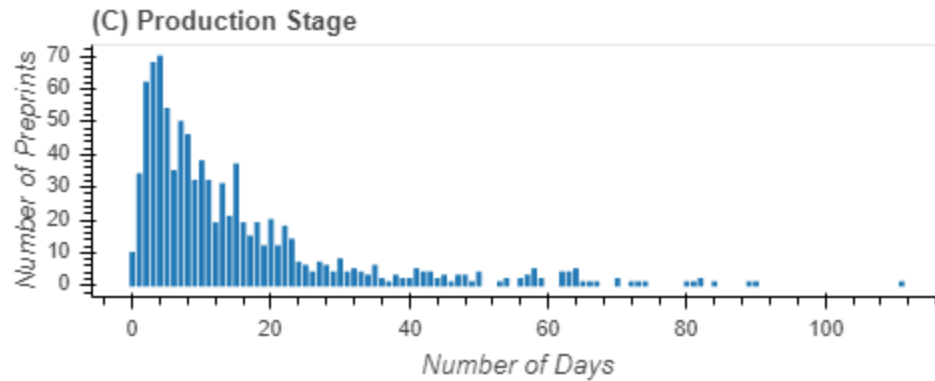
469



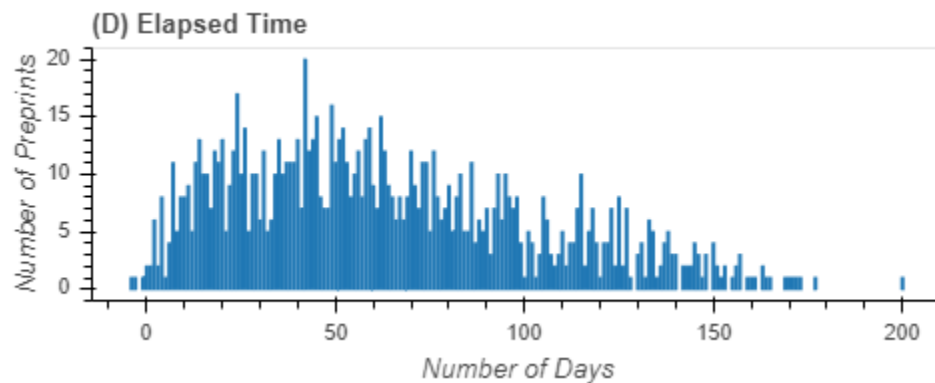
470



471



472



473

474 **Fig 9. Publication delays from preprint to journal publication.**

475 Number of COVID-19 combined medRxiv and bioRxiv preprints with given number of days that
476 correspond to various publication delays: (A) *Pre-submission time*; (B) *Review time*; (C)
477 *Production stage time*; and (D) *Elapsed Time*.

478

479 **4.1. Elapsed Time (T_{Σ}) - Time from a preprint to its journal article version**

480 For bioRxiv COVID-19 preprints, the evaluated *elapsed time* (T_{Σ}) varies anywhere
481 between 0 and 177 days, with a mean of 68.5 days (Table 3). For medRxiv COVID-19 preprints,
482 T_{Σ} varies anywhere between -4 to 200 days with the mean of 61.4 days. The difference in elapsed
483 time between medRxiv and bioRxiv preprints is statistically significant (Table 3). The mean T_{Σ} for
484 combined bioRxiv and medRxiv COVID-19 preprints deposited between Jan 1 and Sept 30, 2020

485 is 63.4 days (Table 2), which is significantly faster than 166 [51] or 155 days [64] reported for
 486 bioRxiv preprints during 2013-2018; and faster than 150 days reported for Zika or Ebola preprints
 487 [60] (Table 4). Curiously, our mean T_{Σ} is significantly longer than 22.5 days reported by Coates *et*
 488 *al.* for bioRxiv and medRxiv preprints published by the end of April 2020 [62], although their data
 489 may be unintentionally biased towards low mean T_{Σ} values considering only 4% of preprints were
 490 published at that time.

491
 492 **Table 3. Descriptive statistics for publication delays (in days) for bioRxiv and medRxiv, as**
 493 **well as Student’s t-test to evaluate their discrepancy.**

Delays	bioRxiv			medRxiv			bioRxiv vs. medRxiv <i>t</i> -test			
	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>Cohen’s d</i>
t_{α}	283	8.2	25.7	690	4.6	23.1	971	-2.16	0.03	-0.15
t_R	273	45.8	33.8	677	42.5	31.8	948	-1.43	0.16	
t_{β}	267	14.7	15.1	649	14.6	16.0	914	-0.09	0.93	
T_{Σ}	301	68.5	39.1	798	61.4	39.6	1097	-2.66	0.008	-0.18

494
 495
 496 **Table 4. Elapsed time (T_{Σ} , in days) for COVID-19 preprints as compared to previous studies**
 497 **(*df* = 1098).**

T_{Σ}	Retrieval of article publication dates	<i>Student’s t</i> -test		
		<i>t</i>	<i>p</i>	<i>Cohen’s d</i>
[Ref]				

166	BioRxiv preprints (2013-2018). Date corresponds to	-86.06	< 0.001	-2.60
[51]	Crossref's "published-online" date or, when not available, "published-print" date.			
155	BioRxiv preprints (2013-2017). Date corresponds to	-76.84	< 0.001	-2.32
[64]	Crossref's "created-date".			
150	Zika (2015-2017) and Ebola (2014-2016) related	-72.65	< 0.001	-2.19
[60]	preprints. Date corresponds to "pub_date" in PubMed.			
22.5	COVID-19 medRxiv and bioRxiv preprints (Jan-	34.24	< 0.001	1.03
[62]	April, 2020). Dates retrieval method is not specified.			

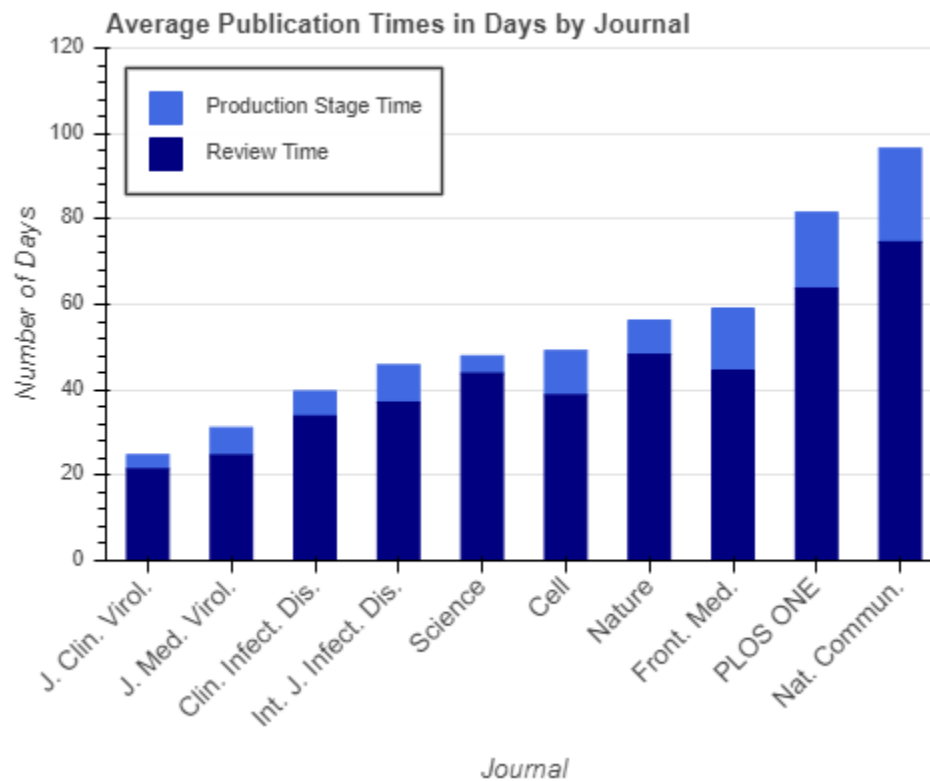
498

499 We also explored whether T_{Σ} can explain the different publication rates for preprint
500 categories. For example, medRxiv preprints in *health policy* ($M = 94.3$, $SD = 41.1$, $N = 10$) and
501 *health informatics* ($M = 82.2$, $SD = 41.8$, $N = 20$) have long mean T_{Σ} and display low publication
502 rates, while preprints in *pathology* display the shortest T_{Σ} ($M = 50.9$, $SD = 33.0$, $N = 10$) and, in
503 fact, have high publication rate (Fig 7 and Categories Analysis in SI). On the contrary, bioRxiv
504 preprints in *cell biology* display the longest T_{Σ} ($M = 89.2$, $SD = 46.4$, $N = 10$) but not the lowest
505 publication rate. Pearson's analysis showed that *elapsed time* and publication rate do not correlate
506 for bioRxiv preprints and only slightly correlate for medRxiv preprints (for medRxiv: $r(41) =$
507 -0.39 , $p = 0.011$; for bioRxiv: $r(19) = -0.13$, $p = 0.581$). As publication rates settle, we plan to re-
508 analyze whether the disciplinary publication trends are affected by the duration of the publication
509 process.

510 4.2. Review time (t_R) and production stage time (t_β)

511 While the *elapsed time* describes how quickly it takes for a preprint to resurface as a journal
512 article, it is unable to distinguish whether the successful expediting of a publication process is a
513 result of the new policies implemented by journal publishers or the old practices of posting a
514 preprint on a preprint server shortly after or even prior to its submission to the peer-reviewed
515 journal. In our quest to explain the expedited publication times, we analyzed *review time* (t_R) and
516 the *production stage period* (t_β) (Fig 10).

517



518

519 **Fig 10. Mean publication delays.** Review time (t_R , dark blue bar) and production stage time (t_β ,
520 light blue bar) are displayed in days for the top 10 journals publishing COVID-19 medRxiv and
521 bioRxiv preprints. The entire two-colored bar represents the total publication time.

522

523 We found that a mean *review time* (t_R) for COVID-19 related bioRxiv and medRxiv
524 preprints is 43.4 days (Table 2), which is significantly shorter ($t(949) = -53.85, p < 0.001, Cohen's$
525 $d = -1.75$) than a standard *review time* of 100 days [65]. The difference in the mean t_R for COVID-
526 19 related medRxiv and bioRxiv preprints is not statistically significant (Table 3). Our data
527 showed that t_R for COVID-19 articles has significantly decreased ($t(949) = -92.86, p < 0.001,$
528 $Cohen's d = -3.01$), by about 70%, during the pandemic, as compared to the earlier analysis by
529 Björk and Solomon, who in 2013 reported that biomedicine journals exercised a mean t_R of 141
530 days [66]. We found that t_R for COVID-19 articles associated with preprints was significantly
531 shorter than 51.0 days ($t(949) = -7.23, p < 0.001, Cohen's d = -0.24$) reported for all coronavirus
532 articles [61]; and significantly shorter than 84 days ($t(949) = -38.63, p < 0.001, Cohen's d = -1.25$)
533 reported for articles published on topics other than COVID-19 during 2020 [67]. Although both
534 studies describe the period of early pandemic, since January and until the end of April, the former
535 study reports a mean review time of 51.0 days [61], while the latter a median *review time* of 6 days
536 [67]. This discrepancy in early data is likely due to a severe skew in frequency distribution for t_R .

537 The major advance in speeding up the peer-review process was observed for *PLOS ONE*.
538 For COVID-19 related articles published in *PLOS ONE* between Jan 1 and Oct 23, 2020, the peer-
539 review took on average 63.7 days. This is significantly shorter than 125 days in 2016 [68] ($t(59) =$
540 $-15.13, p < 0.001, Cohen's d = -1.95$) and 126.6 days in April 2020 [61] ($t(59) = -15.52, p <$
541 $0.001, Cohen's d = -2.00$). Of the top ten journals that published medRxiv and bioRxiv COVID-
542 19 preprints, *Journal of Clinical Virology* had the shortest mean review time of 21.4 days and
543 *Nature Communications* had the longest mean review time of 74.6 days (Table 5, Fig 10).

544

545 **Table 5. Descriptive statistics for publication delays (in days) for the top 10 journals that**
 546 **published medRxiv and bioRxiv COVID-19 preprints.**

<i>Journal</i>	N	Review		Production		Publication		Elapsed Time		
		(t_R)		(t_β)		$(t_R + t_\beta)$		(T_Σ)		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	N
<i>J. Clin. Virol.</i>	15	21.4	41.7	3.3	2.3	24.7	41.2	26.2	29.7	18
<i>J. Med. Virol.</i>	26	24.7	21.7	6.5	4.5	31.1	22.3	38.3	28.8	27
<i>Clin. Infect. Dis.</i>	18	33.8	13.9	6.8	13.0	39.7*	20.3*	46.8	33.5	32
<i>Int. J. Infect. Dis.</i>	23	37.0	20.3	8.7	9.1	45.8	19.7	53.7	31.1	24
<i>Science</i>	21	43.8	37.4	4.1	2.3	47.9	37.8	51.2	41.8	25
<i>Cell</i>	15	38.7	21.4	10.4	7.8	49.1	16.4	43.5	16.4	15
<i>Nature</i>	13	48.2	27.0	8.0	2.0	56.3	28.3	52.3	28.4	14
<i>Front. Med.</i>	15	44.5	20.5	14.5	4.7	59.1	21.1	71.8	39.0	18
<i>PLOS ONE</i>	60	63.7	31.4	17.8	7.3	81.5	32.5	89.4	36.2	83
<i>Nat. Commun.</i>	19	74.6	58.9	21.9	9.8	96.5	66.4	102.7	30.9	23

547 * - based on 26 publications.

548

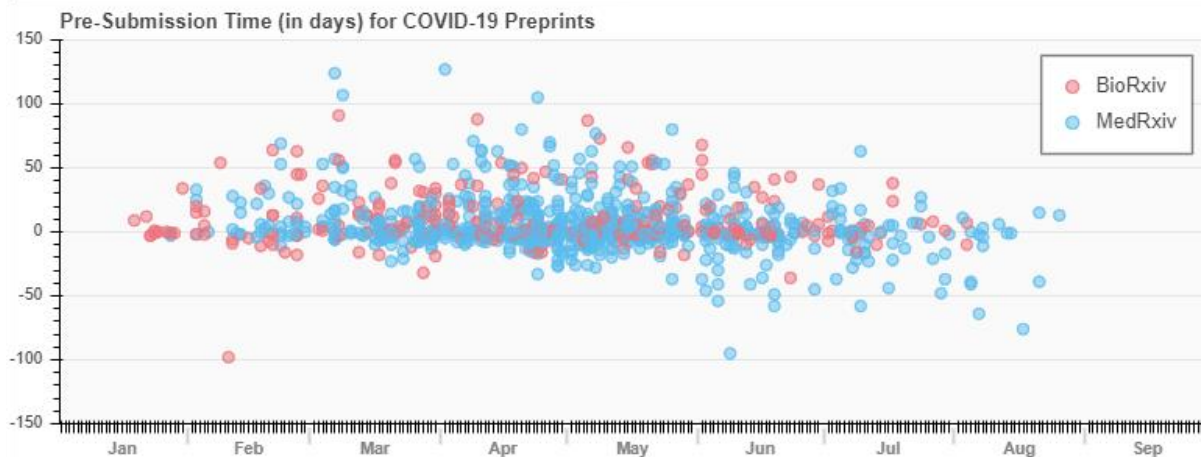
549 The mean *production stage time* (t_β) for COVID-19 related bioRxiv and medRxiv preprints
 550 is 14.6 days, about one third of the average t_R found above for the same set of articles (Table 2).
 551 The difference in t_β for medRxiv and bioRxiv preprints on COVID-19 is not significant (Table 3).
 552 As compared to the t_β of 147 days reported by Björk and Solomon in 2013 [66], t_β reduced by 90%
 553 for COVID-19 articles associated with preprints; this difference being statistically significant
 554 ($t(915) = -254.89, p < 0.001, \text{Cohen's } d = -8.42$). However, a study early in the pandemic reported

555 a significantly shorter $t_{\beta} = 9.3$ days [61] ($t(915) = 10.17, p < 0.001, \text{Cohen's } d = 0.34$), although
556 this value was obtained for all published coronavirus articles, not just those associated with
557 preprints. Of the top ten journals that published COVID-19 bioRxiv and medRxiv preprints,
558 *Journal of Clinical Virology* has the shortest t_{β} of 3.3 days and *Nature Communications* has the
559 longest t_{β} of 21.9 days (Table 5, Fig 10).

560

561 **4.3. Pre-submission time (t_{α})**

562 We found that an average *pre-submission time* (t_{α}) for COVID-19 related preprints is 5.6
563 days (Table 2), a positive value implying that, on average, authors posted their manuscript to the
564 preprint server before advancing their preprints to journal publishers (Fig 11). Authors of bioRxiv
565 COVID-19 preprints waited longer than authors of medRxiv COVID-19 preprints; this difference
566 being statistically significant (Table 3). The distribution of t_{α} frequencies indicates a median at 0
567 days (Table 2, Fig 9). A more detailed analysis showed that 44% of the COVID-19 preprints were
568 deposited to bioRxiv or medRxiv servers after being submitted to the journal (negative t_{α}) and
569 only 28% of preprints were posted more than 10 days before they were submitted to the journal
570 where they were published. Our results mirror earlier findings by Anderson [69], who reported
571 those values as 57% and 29%, respectively, for papers that had preprint analogues and were
572 published in *Nature* journals between 2013 and 2018. This authors' behavior was also observed
573 for arXiv preprints [70] and explained by a widespread fear for "getting scooped" when making
574 details of research publicly available. We thus confirm that the previously reported trends are
575 restated for COVID-19 preprints deposited between Jan 1 and Sept 30, 2020. We believe that
576 during the pandemic, bioRxiv and medRxiv preprint servers were not used to gather a community
577 feedback because authors primarily pursued rapid and open dissemination of critical research.



578

579 **Fig 11. Pre-submission time for COVID-19 published preprints.**

580 The t_{α} , in days, is plotted for bioRxiv (red) and medRxiv (blue) preprints deposited during Jan 1 –
581 Sept 30, 2020. The 0 date is the date the preprint was submitted to the peer-reviewed journal and
582 a positive t_{α} indicates that the preprint was deposited before being submitted to the journal.

583

584 **5. Journals**

585 **5.1. Top journals by number of submissions**

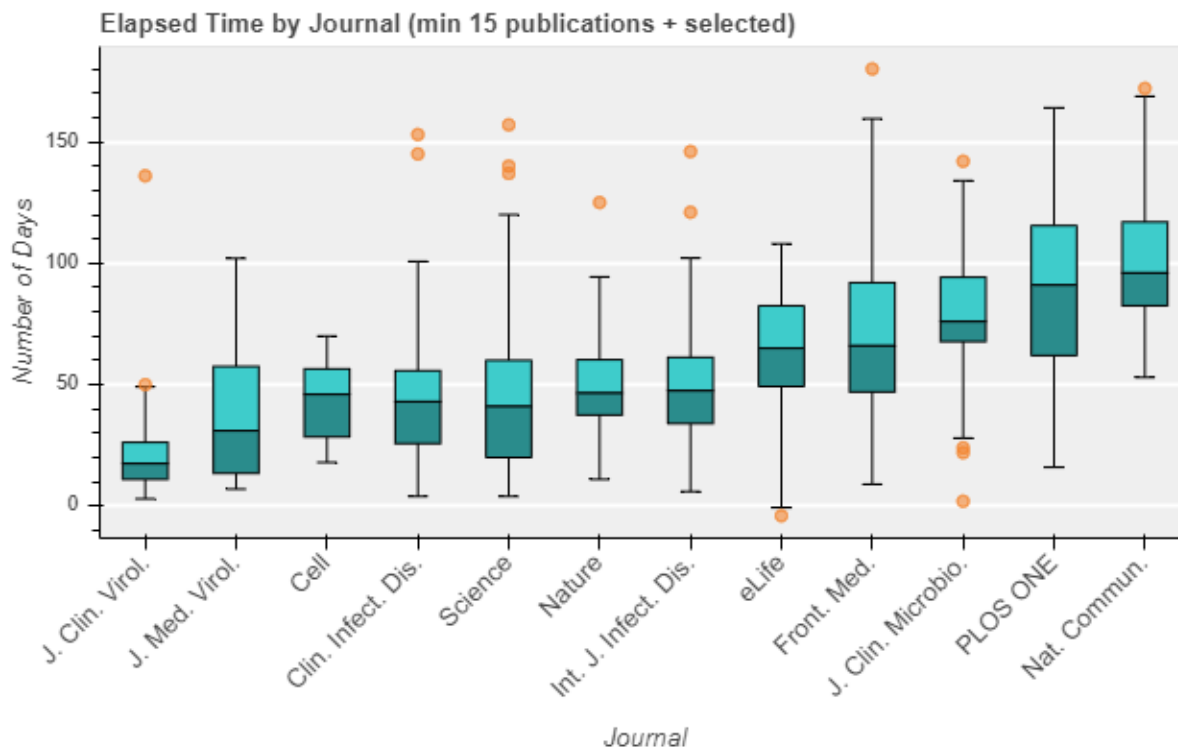
586 By Oct 23, 2020, 1,359 preprints from both, medRxiv and biorxiv appeared as peer-
587 reviewed journal articles in 525 different academic journals, with only 25 journals (4.8%)
588 publishing at least ten preprints (see Leading Journals in SI). The top three journals publishing the
589 majority of bioRxiv preprints were *Science*, *Nature*, and *Nature Communications*. The top three
590 journals publishing the majority of medRxiv preprints and, in fact, the majority of combined
591 medRxiv and bioRxiv COVID-19 related preprints were *PLOS ONE*, *Clinical Infectious Diseases*,
592 and *Journal of Medical Virology*.

593 Previous analysis of bioRxiv preprints showed that *Scientific Reports*, *eLife*, and *PLOS*
594 *ONE* published the majority of server's preprints by the end of 2018 [51]. While *PLOS ONE*

595 remains a leader in publishing COVID-19 related preprints, the majority of them (91.6%) are from
596 medRxiv. *Scientific Reports* during this period published only 0.4% of COVID-19 articles, of
597 which 20.7% were originally preprints and 75% of which were specifically from bioRxiv. In
598 comparison, two studies early in the pandemic reported that the majority of preprints were
599 published by *Nature*, *Cell*, and *Science* (for COVID-19 preprints indexed in Dimensions) [61] and
600 by *Journal of Medical Virology* (for COVID-19 bioRxiv and medRxiv preprints) [37].

601 **5.2. Elapsed time (T_{Σ}) by journals**

602 We reasoned that variations in the set of journals publishing the majority of preprints could
603 be explained with difference in elapsed times for each journal (Fig 12). Indeed, in *post hoc*
604 Bonferroni-corrected pairwise comparison, we found that the *Journal of Medical Virology*, *Nature*,
605 *Cell*, and *Science* had significantly shorter elapsed time (T_{Σ}) as compared to *PLOS ONE* ($p < .001$
606 – 0.03). A one-way ANOVA test showed that there was a significant main effect ($F(5, 175) =$
607 19.33, $p < .001$, $\eta_p^2 = 0.36$). The other four journals did not differ between each other. This
608 explains why early in the pandemic, *Journal of Medical Virology*, *Nature*, *Cell*, and *Science*
609 incorporated more published preprints than *PLOS ONE*. As can be seen from Fig 12, it took longest
610 for COVID-19 preprints to be published in *Nature Communications* or *PLOS ONE*, and quickest
611 - in *Journal of Clinical Virology* (Table 5).



612

613 **Fig 12. Elapsed time by journal.**

614 Box plot depicts quartiles (0.0, 0.25, 0.5, 0.75, 1.0) and outlier data for T_{Σ} of combined COVID-
615 19 related medRxiv and bioRxiv preprints published as journal articles in selected journals.

616

617 **5.3. Alternative ranking for top journals**

618 In the previous section, we discussed journals that published the majority of bioRxiv and
619 medRxiv preprints based on the number of preprints. By this metrics, mega-journals that publish
620 a large number of articles will systematically appear as leaders in publishing preprints. An
621 alternative way to allocate leading journals is to evaluate what fraction of journal's publications
622 come from preprints. For this, we analyzed the percentage of published COVID-19 related
623 preprints with respect to all journal and review articles in a particular journal, as indexed in
624 PubMed between Jan 1 and Sept 30, 2020. We found that COVID-19 preprints published in the

625 mega-journal *PLOS ONE* constituted only 0.6% of its total publication volume (see Leading
626 Journals in SI). However, if we only compare to the research and review articles on COVID-19,
627 we find that 20.8% of *PLOS ONE* COVID-19 articles are former preprints. A similar scenario is
628 observed for other multidisciplinary journals, like *Science* or *Nature*. For example, 34.8% of
629 COVID-19 articles in *Nature Communications* come from preprints but they constitute only 0.5%
630 of the total publication volume for this journal. The COVID-19 portfolio of *BMC Medicine* and
631 *Journal of Clinical Microbiology* incorporated 36.4% and 33.3%, respectively, of former preprints.
632 These percentages become 5.3% and 5.7% when we consider what fraction of all articles in *BMC*
633 *Medicine* and *Journal of Clinical Microbiology*, respectively, are COVID-19 former preprints. As
634 compared to *Nature Communications* and *PLOS ONE*, these percentages are higher, indicating
635 that *BMC Medicine* and *Journal of Clinical Microbiology* have disciplinary scope that is adequate
636 for coronavirus research and is more specialized than that of *Nature Communications* or *PLOS*
637 *ONE*. Our analysis of COVID-19 papers associated with preprints shows that *Emerging Microbes*
638 *& Infections* published 68.8% of COVID-19 related preprints, which constituted 17.7% of this
639 journal's COVID-19 articles. *Emerging Microbes & Infections* is an open access peer reviewed
640 journal from Taylor & Francis, which scope, as indexed by Scopus [54], is *parasitology, infectious*
641 *diseases, microbiology, virology, immunology, drug discovery, and epidemiology*, categories that
642 are all very relevant to coronavirus research.

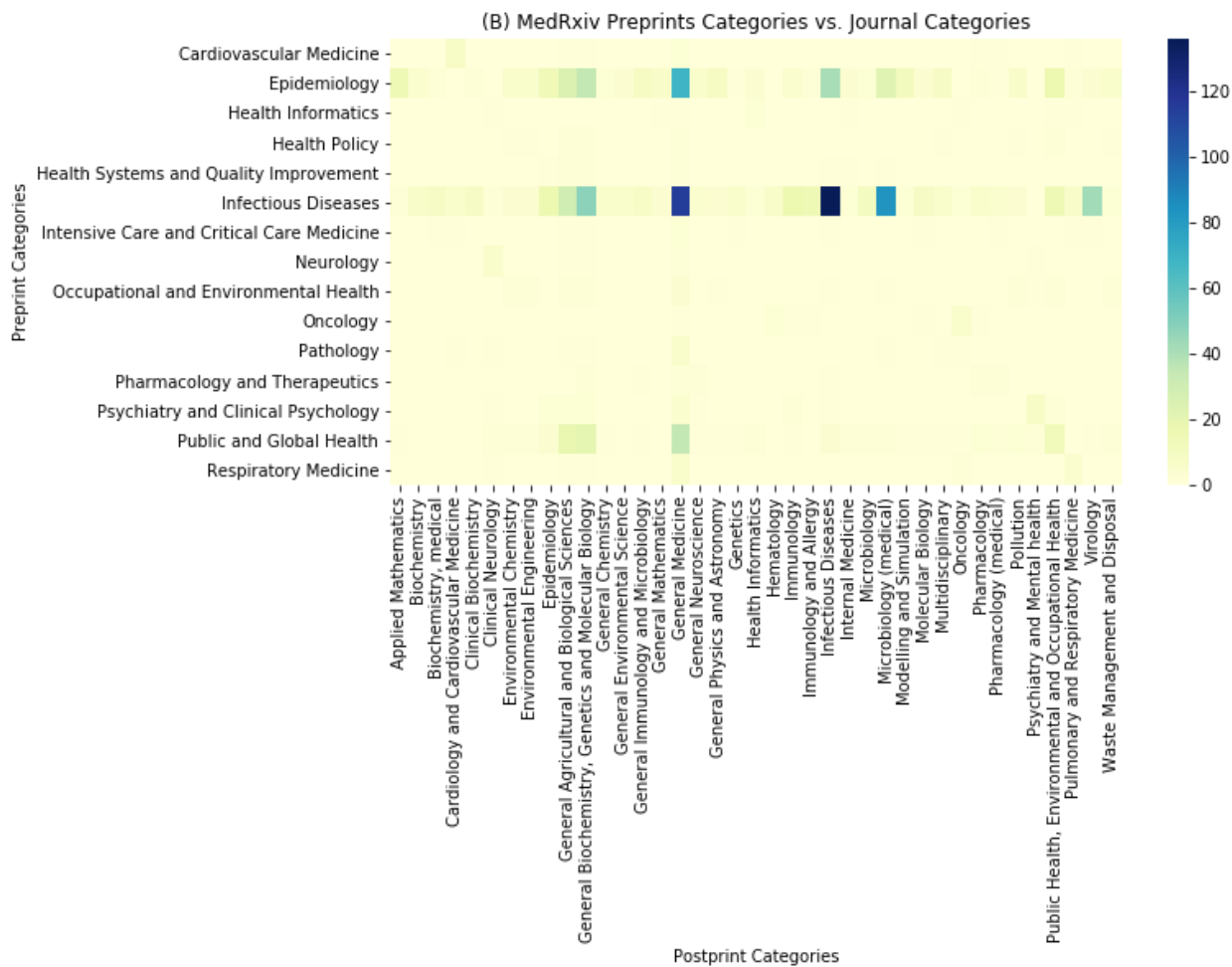
643 **5.4. Categories of preprints and corresponding journals**

644 To analyze the scope of journals that published COVID-19 preprints, we plotted preprint
645 subject categories vs. journal's scope categories, as defined by Scopus [54], for all published
646 preprints and their article analogues (Fig 13). We found that the majority of COVID-19 preprints
647 in both medRxiv and bioRxiv were published in journals whose scope is *general biochemistry*,

648 *genetics*, and *molecular biology*. Additionally, *microbiology* preprints from bioRxiv were
649 published in journals specialized in *microbiology*, *infectious diseases*, and *virology*. The latter
650 category is currently absent in either bioRxiv or medRxiv platforms but is listed among Scopus
651 categories. The majority of medRxiv preprints were published in journals whose scope is *general*
652 *medicine*. Preprints in *infectious diseases* and *epidemiology* were published in journals whose
653 scope is *infectious diseases* and *microbiology (medical)*. For both, medRxiv and bioRxiv, the
654 COVID-19 preprints in leading categories also resurfaced as journal articles in *multidisciplinary*
655 journals, of which *PLOS ONE*, *Science*, and *Nature Communication* published the majority of
656 them (see Leading Journals in SI).

657





659

660 **Fig 13. Preprint vs. journal categories for published COVID-19 preprints.**

661 (A) BioRxiv and (B) medRxiv preprint categories vs. journal categories of corresponding journal
662 articles. A preprint category may be associated with many journal categories as journal scope is
663 usually represented by multiple categories. Categories that appear less than 10 times are removed
664 for clarity purposes.

665

666 5.5. Publishers

667 The top two publishers for COVID-19 bioRxiv and medRxiv preprints were Elsevier and
668 Springer. By Sept 30, they published 40% of all COVID-19 preprints deposited into bioRxiv and

669 medRxiv during Jan 1 – Sept 30, although medRxiv preprints constituted the majority of their
670 publications (73% in Elsevier and 65% in Springer). Elsevier journals that published the majority
671 of COVID-19 preprints were *International Journal of Infectious Diseases*, *Journal of Clinical*
672 *Virology*, and *Cell*. Springer journals that published the majority of COVID-19 preprints were
673 *Nature Communications*, *Nature*, and *Scientific Reports*. The third major publisher for COVID-19
674 preprints was Oxford University Press, which journals like *Clinical Infectious Diseases*, *The*
675 *Journal of Infectious Diseases*, *Journal of Travel Medicine*, and *American Journal of Clinical*
676 *Pathology* published 7% of bioRxiv and medRxiv preprints on COVID-19 by Sept 30, 2020. The
677 third major publisher for COVID-19 bioRxiv preprints was the *American Society for Microbiology*
678 that alone published 7% of all COVID-19 preprints deposited into bioRxiv by Sept 30, 2020.

679 **6. Impact of preprints**

680 Previous studies of bioRxiv preprints deposited during 2013-2017 [64], 2014-2016 [71],
681 and 2015-2018 [18], all ascertain an altmetric advantage of articles that were associated with
682 preprints over articles that did not come from preprints. A notion of altmetric advantage was
683 derived from higher Altmetric Attention Scores, which summarize various mentions of the article
684 in the public policy documents, Wikipedia articles, mainstream news, blogs, and various social
685 platforms [72]. For example, both bioRxiv and medRxiv preprints are widely shared on Twitter
686 and even prior to pandemic, a single preprint was reported to generate on average 13-18 tweets
687 [69].

688 To assess the visibility of COVID-19 preprints, we compared the Altmetric Attention
689 Scores of COVID-19 related articles that had associated preprints to those that did not, and to
690 articles unrelated to COVID-19 that were published between Jan 1 and Nov 19, 2020 (Fig 14,
691 Table 6). We also stratified our results by journal to eliminate a potential effect of a journal's

692 impact factor (IF) or other journal-specific variables. For the top ten journals that published the
 693 majority of COVID-19 preprints, we found that Altmetric Attention Scores for articles that had
 694 associated preprints were slightly higher on average but not significantly different from articles
 695 that did not have associated preprints (Table 6). As expected, all COVID-19 related articles had
 696 an altmetric advantage over non-COVID-19 related publications in the same journals. For
 697 example, in *Nature*, COVID-19 related articles published between Jan 1 and Nov 19, 2020 had on
 698 average an altmetric score of 1077.8 ($SD = 1467.8, N = 253$). This is significantly higher ($t(263.3)$
 699 $= -7.76, p < 0.001, Cohen's d = -0.89$) than an average altmetric score of 354.3 ($SD = 722.3, N =$
 700 2766) for all other articles in the same journal published during the same period. Similar analysis
 701 of all ten journals resulted in average altmetric score of 332.8 for COVID-19 related works vs.
 702 91.0 for non-COVID-19 publications, this difference being statistically significant (Table 6). As
 703 seen from Fig 14, altmetric scores vary for different journals. In fact, we found a strong Pearson's
 704 correlation between the Altmetric Attention Scores for journal articles and journals' impact factors
 705 (for COVID-19 articles that were deposited as preprints, $r(8) = 0.99, p < 0.001$).

706

707 **Table 6. Descriptive statistics for Altmetric Attention Scores for COVID-19 related and**
 708 **unrelated publications, as well as Student's t-test to evaluate discrepancies.**

Altmetric Attention Scores	COVID-19 articles with preprints	COVID-19 articles without preprints	COVID-19 articles	Non-COVID- 19 articles
<i>N</i>	10	10	20	10
<i>M</i>	377.9	287.7	332.8	91.0
<i>SD</i>	540.9	393.3	462.6	114.2

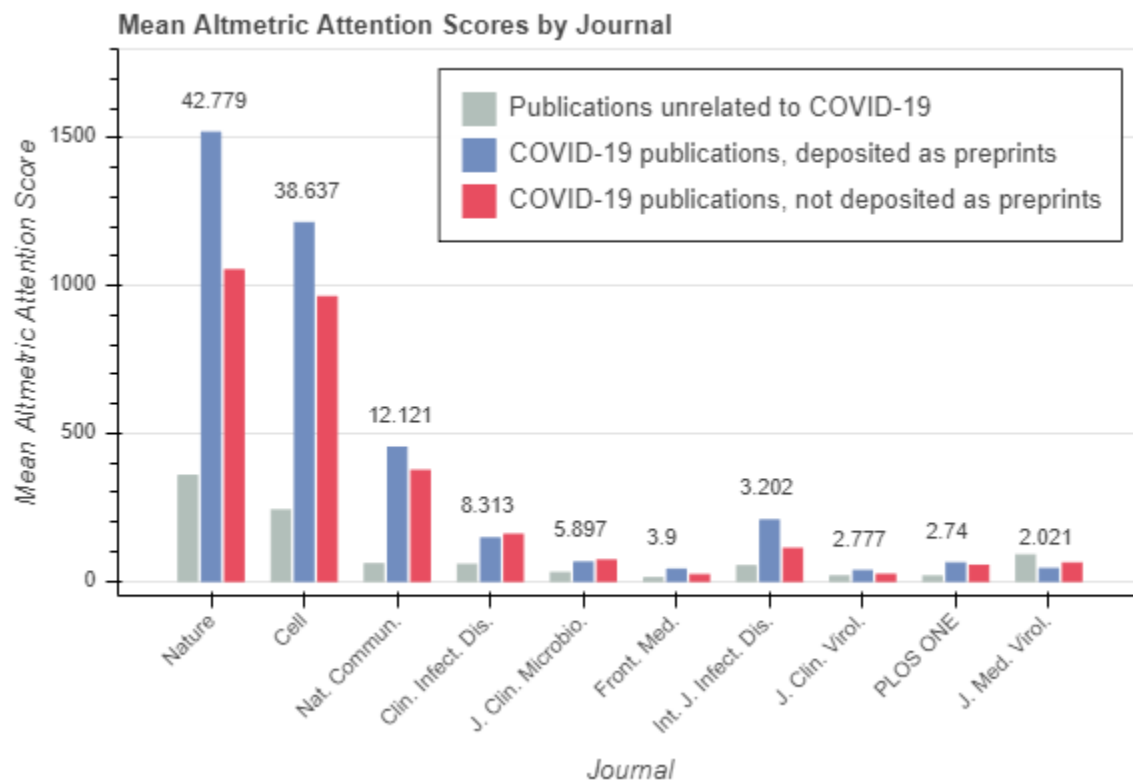
t-test(*s*)

t(18) = 0.43, *p* = 0.68

t(23.2) = 2.21, *p* = 0.04,

Cohen's d = 0.63

709



710

711 **Fig 14. Mean Altmetric Attention Scores by journal.**

712 For COVID-19 related articles that were deposited as preprints (blue), COVID-19 related articles
713 that were not deposited as preprints (red), and all non-COVID-19 articles published during the
714 same period (light blue). Impact factors are displayed above the bars and are from the Journal
715 Citation Reports [71].

716

717 Discussion

718 In this paper, we explored how publication practices in biomedical sciences reacted to an
719 emergency, such as COVID-19 pandemic. Our first focus was analyzing the usage of two major

720 biomedical preprint servers, bioRxiv and medRxiv. Following the deposition of the first preprint
721 on a “novel coronavirus” in mid-January 2020 [56], preprint submissions to these two platforms
722 increased rapidly. Submissions of new coronavirus related preprints reached 10 to 20 per day by
723 February and increased to about 150 per day by May (Fig 1). In addition to this incredible flow of
724 COVID-19 preprints, we observed about a 35% increase in deposition of preprints on all other
725 topics. Towards the end of 2020, the amount of posted COVID-19 related preprints has been
726 declining each month (Fig 1) but preprints unrelated to coronavirus continue to be submitted at a
727 relatively constant rate (Fig 2). Preprints traditionally represented only a small fraction of scholarly
728 literature; however, during the pandemic the fraction of combined medRxiv and bioRxiv COVID-
729 19 preprints over all journal and review articles indexed in PubMed in the given month increased
730 from 2% in February to 15% in May (Fig 3). A higher flux of biomedical preprints was also
731 observed for Zika and Ebola outbreaks [60], which was suggested to result from an increase in
732 research activities, and the statement of prestigious journal publishers and funders on the
733 importance of preprints and data sharing in public health emergencies [73]. In contrast, COVID-
734 19 pandemic was accompanied by a severe reduction in research activities due to social distancing
735 policies and stay-at-home orders issued around the world. We hence raised the question regarding
736 the origin of the amplified usage of preprint servers during the COVID-19 pandemic?

737 The answer to this question lies within the trends in the most active fields in each preprint
738 server. Our analysis revealed that the majority of COVID-19 related preprints in bioRxiv were
739 deposited in those fields that are most relevant to coronavirus research, such as *microbiology*,
740 *bioinformatics*, and *immunology* (Fig 4A); while the leading categories in all other topics during
741 the same period were *neuroscience*, *microbiology*, and *bioinformatics*. Note that prior to the
742 pandemic, the three leading categories in bioRxiv were *neuroscience*, *bioinformatics*, and

743 *genomics* [51]. Thus, while *neuroscience* researchers remain as the main users of the bioRxiv
744 platform, we see an emergence of a new community of preprint authors working in the fields of
745 *immunology* and *microbiology*. The latter can perhaps be further refined as *virology*, a category
746 currently absent in bioRxiv (and medRxiv), but listed among journal scope categories, where the
747 majority of *microbiology* preprints were later published. With regard to medRxiv server, the
748 leading categories for COVID-19 related preprints are *infectious diseases*, *epidemiology*, and
749 *public and global health* (Fig 4B); however, it is not possible to compare these figures with the
750 pre-COVID-19 data due to the relatively recent launch of this preprint server. To explain, medRxiv
751 was founded only six months prior to the pandemic and increased its volume by almost 16-fold
752 during 2020, with 66% of this increase attributed to preprints on COVID-19. Overall, we noted
753 that during the coronavirus pandemic preprint servers attracted a new pool of researchers from
754 previously underrepresented fields. This agrees well with an earlier report of Coates *et al.* who
755 found that 83% of COVID-19 preprint authors were posting a preprint for the first time [62]. It is
756 worth noting that the same study also found that most corresponding authors were not switching
757 to COVID-19 research from other fields, therefore, we propose that the most plausible factor
758 contributing to the observed surge of preprints is the extensive use of preprint repositories by new
759 COVID-19 researchers. This is remarkable considering preprints are rarely taken into
760 consideration for tenure or promotion decisions in biomedical sciences. Further studies monitoring
761 these new clients of preprint servers will be valuable in determining whether this is just a transient
762 behavior triggered by COVID-19 or whether these practices will persist beyond pandemic
763 duration.

764 Once we established that increased volume arises from new users, we further inquired
765 whether, besides rapid dissemination of research results, they were attracted by any specific of the

766 known benefits for preprint platforms. Among many, we focused on: (i) opportunity to share
767 studies that are difficult to publish in traditional journals or those that would be rather works-in-
768 progress than complete reports, and/or (ii) access to public comments that could help improving
769 the manuscript prior to journal submission.

770 We first explored whether COVID-19 preprints were in-progress or complete works by
771 analyzing how many of the COVID-19 preprints later appeared in refereed journals, and used that
772 metric as an indicator for their suitability for publication. Our analysis showed that by Sept 30,
773 2020, only 18% of the bioRxiv and medRxiv preprints related to COVID-19 appeared as peer-
774 reviewed journal publications. Anticipating a dependence of publication rate on the analysis
775 timespan [51], we assessed the time it takes for a preprint to resurface as a journal article and found
776 that our study on Sept 30, 2020, undervalued publication rate for preprints deposited any time after
777 the end of July (mean $T_{\Sigma} = 63.4$ days, Table 2). To derive a meaningful publication rate, we
778 reanalyzed the same pool of COVID-19 related preprints (deposited to bioRxiv and medRxiv
779 servers by Sept 30) on December 7, 2020, which is 68 days past Sept 30 (mean bioRxiv $T_{\Sigma} = 68.5$
780 days, Table 2). As expected, we found higher publication rates for COVID-19 preprints, 33.8%
781 and 28.8% for bioRxiv and medRxiv servers, respectively; although these values remained lower
782 than those reported for pre-pandemic preprints (42% [51] or 70% [3]) and for previous health
783 crises, such as Ebola and Zika (60% and 48%, respectively) [60]. We therefore concluded that
784 COVID-19 preprints are mostly works-in-progress, in line with earlier suggestions [62]. In the
785 future, it will be important to reevaluate publication rates for COVID-19 related preprints. It is
786 likely that some COVID-19 preprints keep circulating through repeated cycles of journal
787 submissions and rejections, and thus remain invisible to our analysis.

788 We then assessed whether coronavirus researchers used bioRxiv and medRxiv preprint
789 servers to gather public feedback by examining the *pre-submission time* (t_α) for COVID-19
790 preprints. Our analysis of publication delays yielded a median t_α of 0 days for published COVID-
791 19 preprints, implying that preprints were submitted to preprint servers and to journals
792 simultaneously (Fig 9 and Fig 11). A more detailed analysis showed that only 28% of preprints
793 were deposited into servers for over 10 days prior to journal submission. This trend of shortening
794 the waiting period between preprint deposition to a server and its submission to a journal had been
795 already reported in studies involving bioRxiv [69] and arXiv preprints [70]. Note that *pre-*
796 *submission time* may include submissions to multiple journals, thus a low t_α value also means that
797 the majority of the so far published COVID-19 preprints were accepted to journals at their first
798 submission. Thus, we see no change in authors' behavior due to pandemic: COVID-19 manuscripts
799 deposition as preprints occurred concurrently to their submission to journals; this implies that
800 authors of COVID-19 preprints did not specifically pursue the pre-submission feedback when
801 posting their research study on a preprint server.

802 In line with previous reports [52,61,67], our analysis of publication delays demonstrated
803 another important change in scientific publishing in response to the COVID-19 pandemic, namely,
804 expedited publication process for COVID-19 related articles. For instance, COVID-19 article
805 versions of preprints are currently reviewed in about half the time (43 days) as compared with the
806 30-year average *peer-review time* of 100 days [65]. Further, COVID-19 articles associated with
807 preprints are being reviewed 70% faster and prepared for the final version 90% faster than
808 biomedical articles were prior to the pandemic ($t_R = 141$ days and $t_\beta = 147$ days, in 2013 [66]).
809 This acceleration is also evident when comparing articles associated with bioRxiv preprints:
810 COVID-19 related preprints transformed into peer-reviewed journal publications 2.5 times faster

811 during the study period, Jan 1 – Oct 23 2020, as compared to bioRxiv preprints before the
812 pandemic ($T_{\Sigma} = 63.4$ days in 2020 vs 155-166 days by 2018 [51,64]; Table 3). This is
813 unprecedented and was not observed for Zika or Ebola related preprints, where the median *elapsed*
814 *time* for both outbreaks (150 days [60]) was reported to be no different than the normal publication
815 timeline.

816 This really speaks to a success of modified journal editorial policies in expediting
817 publication process for coronavirus researchers. These initiatives [33,74] quickly led to a surge of
818 COVID-19 journal articles that in July represented about 62% of all journal and review articles
819 indexed in PubMed for that month. By the end of October 2020, 1,357 of COVID-19 preprints
820 from medRxiv and bioRxiv servers appeared in 525 different academic journals, with
821 multidisciplinary journals, such as *PLOS ONE*, *Science*, *Nature*, and *Nature Communications*,
822 accruing the highest number of former preprints (see Leading Journals in SI). Among the
823 specialized journals, *Clinical Infectious Diseases*, and *Journal of Medical Virology* published 6%
824 of all medRxiv preprints. The COVID-19 portfolio of *BMC Medicine*, *Journal of Clinical*
825 *Microbiology*, and *Nature Communications* was on one third composed of former preprints.
826 However, the highest fraction of preprints in all articles published by the journal was observed for
827 *Emerging Microbes & Infections*, where 69% of COVID-19 articles originated from preprints.
828 Journal champions varied at various moments throughout the pandemic, which was found to be
829 related to variations in publication times among the journals (Fig 12). For example, it took the
830 fastest for bioRxiv or medRxiv preprints to appear in *Journal of Clinical Virology* or *Journal of*
831 *Medical Virology* ($T_{\Sigma} = 26.2$ and 38.3 days, respectively), but longest in *PLOS ONE* and *Nature*
832 *Communications* ($T_{\Sigma} = 89.4$ and 102.7 days, respectively). On average, the publication process for
833 COVID-19 preprints posted to bioRxiv or medRxiv servers took about two months. This reflects

834 an earlier observation that a publication peak for COVID-19 preprints in May transfers to the
835 summit in July for journal article publications (Fig 2).

836 Complementary efforts of preprint servers and scholarly journals to disseminate knowledge
837 promptly, while differentiating reliable and important findings from those that may be misleading
838 attest to the utmost relevance of COVID-19 topic during 2020, as evident from Altmetric
839 Attention Scores for COVID-19 research articles (Fig 14). For example, COVID-19 related articles
840 published in *Nature* journal between Jan 1 and Nov 19, 2020 had a mean altmetric score of 1,077.8
841 as compared to 354.3 for articles on all other topics. The highest Altmetric Attention Score of
842 1,518.2 was observed for *Nature* journal articles associated with preprints. Despite the high
843 visibility of COVID-19 articles originated from preprints, our analysis showed no significant
844 altmetric advantage over COVID-19 articles that did not have a preprint version. This is in contrast
845 with previous findings of clear altmetric differential between articles deposited as bioRxiv
846 preprints and those that were not [18,64,71]. Considering the acute public interest in novel
847 coronavirus research, we believe that any scholarly work, whether it was an article, associated or
848 not associated with a preprint, received equivalent amount of social media attention during the
849 pandemic. The strong correlation we observed between the altmetric scores and journals' impact
850 factors for COVID-19 articles tells us how popularized was the professional arena in biomedical
851 sciences during the pandemic.

852 In summary, our analysis showed that early in pandemic, preprints were prevailing in
853 disseminating findings on the topic of the public health emergency. Preprint authors deposited
854 them into fields previously underrepresented on bioRxiv or medRxiv servers but those that were
855 directly related to our understanding of the newly emerged coronavirus and ways to prevent the
856 spread of the disease. This new category of authors mainly pursued rapid and transparent scientific

857 communication but was not specifically interested in pre-submission feedback as observed from
858 preprints submissions onto servers occurring concurrently with the journal submissions. We
859 believe the majority of COVID-19 preprints were immediate in-progress findings, not suitable for
860 the direct transfer to refereed journals; this conclusion was based on the low publication rate for
861 COVID-19 preprints. The originally estimated publication rates were reaffirmed two months later,
862 which is how long it took on average for a COVID-19 preprint to go through the peer-review and
863 production stages. Both stages of the publication process were significantly expedited for COVID-
864 19 publications; but they varied widely among different journals. The COVID-19 preprints that
865 resurfaced as journal articles display exceptionally high Altmetric Attention scores echoing a high
866 social media engagement in coronavirus research. The concerted efforts of journal and preprint
867 publishers, institutions, funders, and individuals to bring us information and to facilitate its sorting
868 and evaluation, deserve special applauds. It is our hope that as pandemic recedes, biomedical
869 sciences will keep on building their relationships with preprint servers, while journal publishers
870 will retain policies that were most helpful in expediting, revising, and sharing critical research
871 output during the pandemic.

872

873 **Acknowledgements**

874 We thank the University of Michigan Library for the financial support of this project and
875 CADRE Fellowship from the Research Cohort for the Study of Coronaviruses Program for
876 providing access to the COVID-19 dataset. We thank Dr. Craig Smith for statistics related
877 consultations. We also thank Dr. Oscar Tutusaus for his assistance with manuscript editing.

878 **References:**

- 879 1. Carneiro CFD, Queiroz VGS, Moulin TC, Carvalho CAM, Haas CB, Rayê D, Henshall DE,
880 De-Souza EA, Amorim FE, Boos FZ, Guercio GD, Costa IR, Hajdu KL, van Egmond L,
881 Modrák M, Tan PB, Abdill RJ, Burgess SJ, Guerra SFS, Bortoluzzi VT, Amaral OB.
882 Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical
883 literature. BioRxiv 581892 [Preprint]. 2020 [cited 2020 Oct 6]. Available from:
884 <https://www.biorxiv.org/content/10.1101/581892v4>.
- 885 2. Tennant JP, Ross-Hellauer T. The limitations to our understanding of peer review. Res Integr
886 Peer Rev. 2020 Apr 30;5:6. doi: 10.1186/s41073-020-00092-1. PMID: 32368354.
- 887 3. Sever, R, Roeder T, Hindle S, Sussman L, Black KJ, Argentine J, Manos W, Inglis JR.
888 bioRxiv: the preprint server for biology. BioRxiv 833400 [Preprint]. 2019 [cited 2020 Oct 20].
889 Available from: <https://www.biorxiv.org/content/10.1101/833400v1>.
- 890 4. Iacobucci G. New preprint server allows earlier sharing of research methods and findings.
891 BMJ. 2019 Jun 6;365:14110. PubMed PMID: 31171560.
- 892 5. Arxiv.org/help/stats [Internet]. ArXiv usage statistics [cited 2020 Oct 6]. Available from:
893 <https://arxiv.org/help/stats>.
- 894 6. Kirkham JJ, Penfold NC, Murphy F, Boutron I, Ioannidis JPA, Polka JK, Moher D. A
895 systematic examination of preprint platforms for use in the medical and biomedical sciences
896 setting. BioRxiv 063578v1 [Preprint]. 2020 [cited 2020 Dec 15]. Available from:
897 <https://www.biorxiv.org/content/10.1101/2020.04.27.063578v1>.
- 898 7. Rittman M. Preprints as a Hub for Early-Stage Research Outputs. Preprints 2018060243
899 [Preprint]. 2018 [cited 2020 Oct 6]. Available from:
900 <https://www.preprints.org/manuscript/201806.0243/v1>.

- 901 8. Penfold NC, Polka JK. Technical and social issues influencing the adoption of preprints in the
902 life sciences. *PLOS Genet.* 2020 Apr 20;16(4):e1008565. PubMed PMID: 32310942.
- 903 9. (a) ASAPbio [Internet]. Preprint FAQ on Scooping [cited 2020 Oct 6]. Available from:
904 <https://asapbio.org/preprint-info/preprint-faq#qe-faq-923>; (b) Tennant JP, Crane H, Crick T,
905 Davila J, Enkhbayar A, Havemann J, et al. Ten Hot Topics around Scholarly Publishing.
906 *Publications.* 2019;7(2):34.
- 907 10. Kling R. The Internet and unrefereed scholarly publishing. *Ann. Rev. Inf. Sci. Tech.*
908 2005;38(1):591–631.
- 909 11. Sheldon T. Preprints could promote confusion and distortion. *Nature.* 2018 Jul;559(7715):445.
910 PubMed PMID: 30042547.
- 911 12. Klebel T, Reichmann S, Polka J, McDowell G, Penfold N, Hindle S, Ross-Hellauer T. Peer
912 review and preprint policies are unclear at most major journals. *PLOS ONE.* 2020 Oct
913 21;15(10):e0239518. doi: 10.1371/journal.pone.0239518. PMID: 33085678.

- 914 13. Berg JM, Bhalla N, Bourne PE, Chalfie M, Drubin DG, Fraser JS, Greider CW, Hendricks M,
915 Jones C, Kiley R, King S, Kirschner MW, Krumholz HM, Lehmann R, Leptin M, Pulverer B,
916 Rosenzweig B, Spiro JE, Stebbins M, Strasser C, Swaminathan S, Turner P, Vale RD,
917 VijayRaghavan K, Wolberger C. Scientific Community. Preprints for the life sciences.
918 Science. 2016 May 20;352(6288):899-901. PubMed PMID: 27199406.
- 919 14. Kaiser J. The Preprint Dilemma. Science. 2017;357(6358):1344-1349. PubMed PMID:
920 28963238.
- 921 15. Bourne PE, Polka JK, Vale RD, Kiley R. Ten simple rules to consider regarding preprint
922 submission. PLOS Comput Biol. 2017 May 4;13(5):e1005473. PubMed PMID: 28472041.
- 923 16. Sarabipour S, Debat HJ, Emmott E, Burgess SJ, Schwessinger B, Hensel Z. On the value of
924 preprints: An early career researcher perspective. PLOS Biol. 2019;17(2):e3000151. PubMed
925 PMID: 30789895.
- 926 17. Slavov Lab [Internet]. Slavov N. 2017. Why I love preprints [cited 2020 Oct 6]. Available
927 from: <https://web.northeastern.edu/slavovlab/blog/2017/09/28/biomedical-preprints-benefits/>.
- 928 18. Davis PM, Fromerth MJ. Does the arXiv lead to higher citations and reduced publisher
929 downloads for mathematics articles? Scientometrics. 2007;71(2):203–215.
- 930 19. Fu DY, Hughey JJ. Releasing a Preprint is Associated with More Attention and Citations for
931 the Peer-Reviewed Article. Elife. 2019 Dec 6;8:e52646. PubMed PMID: 31808742.
- 932 20. Hoy MB. Rise of the Rxivs: How Preprint Servers are Changing the Publishing Process. Med.
933 Ref. Serv. Q. 2020;39(1):84-89. PubMed PMID: 32069196.
- 934 21. NCBI [Internet]. NIH Preprint Pilot [cited 2020 Oct 6]. Available from:
935 <https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/>.

- 936 22. NIH Public Access Policy [Internet]. When and how to comply [cited 2020 Oct 6]. Available
937 from: <https://publicaccess.nih.gov/>.
- 938 23. NIH [Internet]. Reporting Preprints and Other Interim Research Products (NOT-OD-17-
939 050) [cited 2020 Oct 6]. Available from: [https://grants.nih.gov/grants/guide/notice-files/NOT-
940 OD-17-050.html](https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html).
- 941 24. Jisc scholarly communications [Internet]. Perspectives on the open access discovery landscape
942 [cited 2020 Oct 6]. Available from:
943 [https://scholarlycommunications.jiscinvolve.org/wp/2019/04/24/perspectives-on-the-open-
944 access-discovery-landscape/](https://scholarlycommunications.jiscinvolve.org/wp/2019/04/24/perspectives-on-the-open-access-discovery-landscape/).
- 945 25. (a) Barsh GS, Bergman CM, Brown CD, Singh ND, Copenhaver GP. Bringing PLOS Genetics
946 Editors to Preprint Servers. PLOS Genet. 2016 Dec 1;12(12):e1006448. PubMed PMID:
947 27906975; (b) Barrett SCH. Proceedings B 2017: the year in review. Proc Biol Sci. 2018 Jan
948 10;285(1870):20172553. PubMed PMID: 29298940.
- 949 26. Inside Higher Ed [Internet]. McKenzie L. 2020 Jun 29. Debunking Bad COVID-19 Research
950 [cited 2020 Oct 6]. Available from: [https://www.insidehighered.com/news/2020/06/29/new-
951 mit-press-journal-debunk-bad-covid-19-research#.XvnIUdsSqNA.email](https://www.insidehighered.com/news/2020/06/29/new-mit-press-journal-debunk-bad-covid-19-research#.XvnIUdsSqNA.email).
- 952 27. Vlasschaert C, Topf JM, Hiremath S. Proliferation of Papers and Preprints During the
953 Coronavirus Disease 2019 Pandemic: Progress or Problems With Peer Review? Adv Chronic
954 Kidney Dis. 2020 Sep;27(5):418–426.
- 955 28. RetractionWatch, (2020b) [Internet]. Retracted coronavirus (COVID-19) papers.
956 RetractionWatch [cited 11 Nov 2020]. Available from: [https://retractionwatch.com/retracted-
957 coronavirus-covid-19-papers/](https://retractionwatch.com/retracted-coronavirus-covid-19-papers/).

- 958 29. Ioannidis JPA. Coronavirus disease 2019: the harms of exaggerated information and non-
959 evidence-based measures. *Eur J Clin Invest*. 2020 Mar 23:e13223. doi: 10.1111/eci.13223.
960 PubMed PMID: 32202659.
- 961 30. Vabret N, Samstein R, Fernandez N, Merad M. Advancing scientific knowledge in times of
962 pandemics. *Nat Rev Immunol*. 2020 Jun;20(6):338. PubMed PMID: 32327718.
- 963 31. Review Commons [Internet]. About [cited 2020 Oct 31]. Available from:
964 <https://www.reviewcommons.org/about/>.
- 965 32. eLife [Internet]. New from eLife: Invitation to submit to Preprint Review [cited 2021 Jan 15].
966 Available from [https://elifesciences.org/inside-elifesciences/d0c5d114/new-from-elifesciences-invitation-to-](https://elifesciences.org/inside-elifesciences/d0c5d114/new-from-elifesciences-invitation-to-submit-to-preprint-review?gclid=Cj0KCQiA9P_BRC0ARIsAEZ6irgrw7tbLruYEqo0d_hvuQ01BqsyxjYZKq68NkZAUppBUJhzwmkbuzgaAi_ZEALw_wcB)
967 [submit-to-preprint-](https://elifesciences.org/inside-elifesciences/d0c5d114/new-from-elifesciences-invitation-to-submit-to-preprint-review?gclid=Cj0KCQiA9P_BRC0ARIsAEZ6irgrw7tbLruYEqo0d_hvuQ01BqsyxjYZKq68NkZAUppBUJhzwmkbuzgaAi_ZEALw_wcB)
968 [review?gclid=Cj0KCQiA9P_BRC0ARIsAEZ6irgrw7tbLruYEqo0d_hvuQ01BqsyxjYZKq6](https://elifesciences.org/inside-elifesciences/d0c5d114/new-from-elifesciences-invitation-to-submit-to-preprint-review?gclid=Cj0KCQiA9P_BRC0ARIsAEZ6irgrw7tbLruYEqo0d_hvuQ01BqsyxjYZKq68NkZAUppBUJhzwmkbuzgaAi_ZEALw_wcB)
969 [8NkZAUppBUJhzwmkbuzgaAi_ZEALw_wcB](https://elifesciences.org/inside-elifesciences/d0c5d114/new-from-elifesciences-invitation-to-submit-to-preprint-review?gclid=Cj0KCQiA9P_BRC0ARIsAEZ6irgrw7tbLruYEqo0d_hvuQ01BqsyxjYZKq68NkZAUppBUJhzwmkbuzgaAi_ZEALw_wcB).
- 970 33. Eisen MB, Akhmanova A, Behrens TE, Weigel D. Publishing in the time of COVID-19. *eLife*.
971 2020 Mar 25;9:e57162. PubMed PMID: 32209226.
- 972 34. ProMED International Society for Infectious Diseases [Internet]. Undiagnosed Pneumonia –
973 China (Hubei): RFI (2019 Dec 30) [cited 2020 Nov 23]. Available from:
974 <https://promedmail.org/promed-post/?id=6864153%20#COVID19>.
- 975 35. (a) Kupferschmidt K. Preprints bring 'firehose' of outbreak data. *Science*. 2020 Feb
976 8;367(6481):963-964. PubMed PMID: 32108094; (b) Callaway E. The COVID-19 crisis could
977 permanently change scientific publishing. *Nature*. 2020 June11;582:167-168; (c) STAT
978 [Internet]. Krumholz HM, Bloom T, Ross JS. 2020 Jan 31. Preprints can fill a void in times of
979 rapidly changing science [cited 2020 Oct 7]. Available from:
980 <https://www.statnews.com/2020/01/31/preprints-fill-void-rapidly-changing-science/>.

- 981 36. WHO [Internet]. Timeline of WHO's response to COVID-19 [cited 2020 Oct 7]. Available
982 from: <https://www.who.int/news-room/detail/29-06-2020-covidtimeline>.
- 983 37. Fidahic M, Nujic D, Runjic R, Civljak M, Markotic F, Lovric Makaric Z, Puljak L. Research
984 methodology and characteristics of journal articles with original data, preprint articles and
985 registered clinical trial protocols about COVID-19. BMC Med Res Methodol. 2020 Jun
986 22;20(1):161. PubMed PMID: 32571302.
- 987 38. Biooverlay.org [Internet]. Hensel Z. Tracking the popularity and outcomes of all bioRxiv
988 preprints [cited 2020 Dec 15]. Available at: [https://www.biooverlay.org/post/2019-03-tracking-
989 the-popularity-and-outcomes-of-all-biorxiv-preprints/](https://www.biooverlay.org/post/2019-03-tracking-the-popularity-and-outcomes-of-all-biorxiv-preprints/).
- 990 39. Committee on Publication Ethics [Internet]. Discussion document on preprints [cited 2020 Oct
991 6]. Available at: https://publicationethics.org/files/u7140/COPE_Preprints_Mar18.pdf.
- 992 40. Database of COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv [Internet].
993 BioRxiv.org [cited 2020 Oct 7]. Available from:
994 <https://connect.biorxiv.org/relate/content/181>.
- 995 41. Abdill RJ, Blekhman R. Rxivist.org: Sorting biology preprints using social media and
996 readership metrics. PLOS Biol. 2019 May 21;17(5):e3000269. PubMed PMID: 31112533.
- 997 42. Crossref [Internet]. Crossref REST API [cited 2020 Oct 7]. Available from:
998 <https://www.crossref.org/education/retrieve-metadata/rest-api/>.
- 999 43. Entrez Programming Utilities Help [Internet]. Bethesda (MD): National Center for
1000 Biotechnology Information (US); 2010 [cited 2020 Oct 7]. Available from:
1001 <https://www.ncbi.nlm.nih.gov/books/NBK25501/>.
- 1002 44. Dimensions [Internet]. Dimensions API [cited 2020 Oct 7]. Available from:
1003 <https://docs.dimensions.ai/dsl/>.

- 1004 45. Semantic Scholar [Internet]. CORD-19 - COVID-19 Open Research Dataset [cited 2020 Oct
1005 7]. Available from: <https://www.semanticscholar.org/cord19>.
- 1006 46. Mabry PL, Yan X, Pentchev V, Van Rennes R, McGavin SH, Wittenberg JV. CADRE: A
1007 Collaborative, Cloud-Based Solution for Big Bibliographic Data Research in Academic
1008 Libraries. *Frontiers in Big Data*. 2020 Nov 20;3:42.
- 1009 47. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, Funk K, Kinney R, Liu Z,
1010 Merrill W, Mooney P, Murdick D, Rishi D, Sheehan J, Shen Z, Stilson B, Wade AD, Wang K,
1011 Wilhelm C, Xie B, Raymond D, Weld DS, Etzioni O, Kohlmeier S. CORD-19: The Covid-19
1012 Open Research Dataset. *ArXiv [Preprint]*. 2020 Apr 22;arXiv:2004.10706v2. PubMed PMID:
1013 32510522.
- 1014 48. Mouratidis RW. Dimensions. *Journal of the Medical Library Association: JMLA*. 2019;
1015 107(3): 459–461. Available from: <https://doi.org/10.5195/jmla.2019.695>.
- 1016 49. Repositories and preprint servers tracked by Altmetric [Online]. Available from:
1017 [https://help.altmetric.com/support/solutions/articles/6000242541-repositories-and-preprint-](https://help.altmetric.com/support/solutions/articles/6000242541-repositories-and-preprint-servers-tracked-by-altmetric)
1018 [servers-tracked-by-altmetric](https://help.altmetric.com/support/solutions/articles/6000242541-repositories-and-preprint-servers-tracked-by-altmetric) [cited 2020 Oct 20].
- 1019 50. Real-world query used to extract publications related to COVID-19 [cited 29 Nov 2020]. In
1020 Dimensions API Lab [Internet]. Available from [https://api-lab.dimensions.ai/cookbooks/1-](https://api-lab.dimensions.ai/cookbooks/1-getting-started/5-Deep-dive-DSL-language.html)
1021 [getting-started/5-Deep-dive-DSL-language.html](https://api-lab.dimensions.ai/cookbooks/1-getting-started/5-Deep-dive-DSL-language.html).
- 1022 51. Abdill RJ, Blekhman R. Tracking the popularity and outcomes of all bioRxiv preprints. *Elife*.
1023 2019 Apr 24;8:e45133. PubMed PMID: 31017570.
- 1024 52. Kun A. Time to Acceptance of 3 Days for Papers about COVID-19. *Publications*. 2020;8:30.
- 1025 53. Publication Types [cited Nov 29 2020]. In NLM [Internet]. Available from:
1026 <https://www.nlm.nih.gov/mesh/pubtypes.html>.

- 1027 54. These subject categories only appear in the REST API of Crossref. They are imported from
1028 Scopus' ASJC codes and are applied at the journal-level, by association with the journal's
1029 ISSNs. The list of Scopus codes is available from: <https://github.com/plreyes/Scopus>.
- 1030 55. Carnahan RH, Crowe JE Jr et al. Potently neutralizing and protective human antibodies against
1031 SARS-CoV-2. *Nature*. 2020 Aug 20;584:443-461.
- 1032 56. You J, Expert P, Costelloe C. Using text mining to track outbreak trends in global surveillance
1033 of emerging diseases: ProMED-mail. *MedRxiv* 20017145 [Preprint]. 2020 Jan 13 [cited 2020
1034 Nov 13]. Available from: <https://medrxiv.org/cgi/content/short/2020.01.10.20017145>.
- 1035 57. Biology preprints over time [cited 2020 Sept 27]. In ASAPbio [Internet]. Available from:
1036 <https://asapbio.org/preprint-info/biology-preprints-over-time>.
- 1037 58. Publication Types [cited Nov 29 2020]. In NLM [Internet]. Available from:
1038 <https://www.nlm.nih.gov/mesh/pubtypes.html>.
- 1039 59. Jialiang L, Yao Y, Yu Z, Zhiyang Z, Xiaodong S. How many preprints have actually been
1040 printed and why: a case study of computer science preprints on arXiv. *Scientometrics*.
1041 2020;124:555-574.
- 1042 60. Johansson MA, Reich NG, Meyers LA, Lipsitch M. Preprints: An underutilized mechanism to
1043 accelerate outbreak science. *PLOS Med*. 2018 Apr 3;15(4):e1002549. PubMed PMID:
1044 29614073.
- 1045 61. Horbach SPGM. Pandemic Publishing: Medical journals drastically speed up their publication
1046 process for COVID-19. *Quantitative Science Studies*. 2020;1(3):1056–1067.
- 1047 62. Fraser N, Brierley L, Dey G, Polka JK, Pálffy M, Nanni F, Coates JA. Preprinting the COVID-
1048 19 pandemic. *BioRxiv* 111294 [Preprint]. 2020 May 22 [cited 2020 Nov 16]. Available from:
1049 <https://www.biorxiv.org/content/10.1101/2020.05.22.111294v2>.

- 1050 63. Sources of icons: (a) Borewicz S. Scientific journal icon, CC-BY-SA (2014) [Internet].
1051 Available from: Wikimedia Commons; (b) Designed by Freepik from Flaticon. Coronavirus,
1052 CC-BY [Internet]. Available from: Flaticon.com; (c) Adioma. Document Review, CC-BY
1053 [Internet]. Available from: Adioma.com; (d) Kyle Scott. Glasses, CC-BY [Internet]. Available
1054 from: Thenounproject.com; (e) Publication, CC-NC [Internet]. Available from HiClipArt.com;
1055 (f) Designed by srip from Flaticon. Read, CC-BY [Internet]. Available from: Flaticon.com.
- 1056 64. Fraser N, Momeni F, Mayr P, Peters I. The relationship between bioRxiv preprints, citations
1057 and altmetrics. *Quantitative Science Studies*. 2020;1(2):618–638.
- 1058 65. Powell K. Does it take too long to publish research? *Nature* 2016;530:148–151.
- 1059 66. Björk BC, Solomon D. The publishing delay in scholarly peer-reviewed journals. *Journal of*
1060 *Informetrics*. 2013;7(4):914-923.
- 1061 67. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV.
1062 Pandemic publishing poses a new COVID-19 challenge. *Nat Hum Behav*. 2020 Jul;4(7):666-
1063 669. doi: 10.1038/s41562-020-0911-0. PubMed PMID: 32576981.
- 1064 68. Himmelstein D. S. & Powell K. Zenodo Repository (2016) [Internet]. Analysis for "the history
1065 of publishing delays" blog post v1.0 [cited 2020 Nov 3]. Available from:
1066 <https://zenodo.org/record/45516#.X6Gt9YhKhnI>.
- 1067 69. Anderson K.R. bioRxiv: Trends and analysis of five years of preprints. *Learned Publishing*.
1068 2019 Nov 2;33:104-109.
- 1069 70. Larivière V, Sugimoto CR, Macaluso B, Milojevic S, Cronin B, Thelwall M. arXiv E-Prints
1070 and the Journal of Record: An Analysis of Roles and Relationships. *Journal of the Association*
1071 *for Information Science and Technology*. 2014;65(6):1157–1169.

- 1072 71. Serghiou S, Ioannidis JPA. Altmetric Scores, Citations, and Publication of Studies Posted as
1073 Preprints. JAMA. 2018 Jan 23;319(4):402-404. doi: 10.1001/jama.2017.21168. PubMed
1074 PMID: 29362788.
- 1075 72. Altmetric [Internet]. Sources of Attention [cited 2020 Nov 16]. Available from:
1076 <https://www.altmetric.com/about-our-data/our-sources/>.
- 1077 73. Sharing data during Zika and other global health emergencies [cited Dec 2 2020]. In Wellcome
1078 Trust [Internet]. Available from: [https://wellcome.org/news/sharing-data-during-zika-and-](https://wellcome.org/news/sharing-data-during-zika-and-other-global-health-emergencies)
1079 [other-global-health-emergencies](https://wellcome.org/news/sharing-data-during-zika-and-other-global-health-emergencies).
- 1080 74. Bauchner H, Fontanarosa PB, Golub RM. Editorial Evaluation and Peer Review During a
1081 Pandemic: How Journals Maintain Standards. JAMA. 2020;324(5):453–454.
1082 doi:10.1001/jama.2020.11764.

1083

1084 **Supporting Information**

1085 **Supplementary files:**

- 1086 1. Supporting information that contains:
- 1087 a. Data Flow Chart – shows the usage of various data platforms in this study;
- 1088 b. Published Collections – shows how various data were combined to account for all
1089 published preprints;
- 1090 c. Leading Journals – shows the number and relative fraction of COVID-19 preprints
1091 published in selected journals and highlights leading journals in publishing
1092 COVID-19 bioRxiv and medRxiv preprints;
- 1093 d. Publication Rates – shows publication rates for COVID-19 preprints.
- 1094 2. CSV files with raw data for:

- 1095 a. Altmetric Attention Scores – shows average altmetric scores and impact factors for
1096 selected journals;
- 1097 b. Analysis of Journal Categories – shows how journal categories relate to preprint
1098 categories for published bioRxiv and medRxiv COVID-19 preprints;
- 1099 c. Analysis of Preprint Categories – shows data on categories of bioRxiv and
1100 medRxiv preprints prior to and during the pandemic, both for COVID-19 related
1101 and unrelated works;
- 1102 d. ANOVA – shows the results of ANOVA test to evaluate whether one parameters
1103 for one journal were significantly different from parameters for a group of journals;
- 1104 e. Articles' Publication Pates – shows the variation in publication dates retrieved from
1105 different data sources;
- 1106 f. Publication Delays – shows various publication delays and their distribution for
1107 COVID-19 preprints published during 2020;
- 1108 g. Scholarly Output – shows the monthly accumulation rates for bioRxiv and
1109 medRxiv preprints, as well as for journal and review articles related and unrelated
1110 to COVID-19, as indexed in PubMed;
- 1111 3. Python code required to pull the data and visualize selected figures.

1112 **Data availability.** Source data for all figures have been provided in supporting files that
1113 were deposited in a Zenodo repository with DOI 10.5281/zenodo.4329576.

1114