

1 Preprinting a pandemic: the role of 2 preprints in the COVID-19 pandemic

3

4 Nicholas Fraser^{1,#}, Liam Brierley^{2,#}, Gautam Dey³, Jessica K Polka⁴, Máté Pálffy⁵ & Jonathon Alexis
5 Coates^{6,*}

6

7

8

9 ¹ Leibniz Information Centre for Economics, Düsternbrooker Weg 120, 24105 Kiel, Germany

10 ² Department of Health Data Science, University of Liverpool, Brownlow Street, Liverpool, L69 3GL,
11 UK

12 ³ MRC Lab for Molecular Cell Biology, UCL, Gower Street, London WC1E 6BT, UK

13 ⁴ ASAPbio, 600-16th St Ste N312E MC2200, San Francisco, CA 94143-2517 San Francisco, CA, USA

14 ⁵ The Company of Biologists, Bidder Building, Station Road, Histon, Cambridge CB24 9LF, UK

15 ⁶ Hughes Hall College, University of Cambridge, Wollaston Rd, Cambridge, CB1 2EW, UK

16

17

18

19 # These authors contributed equally to this work

20 * Correspondence: jc2216@cam.ac.uk

21

22 Short title: The role of preprints in the 2020 COVID-19 pandemic

23

24

25 Abstract

26 The world continues to face an ongoing viral pandemic that presents a serious threat to human
27 health. The virus underlying the COVID-19 disease, SARS-CoV-2, has caused over 3.2 million confirmed
28 cases and 220,000 deaths between January and April 2020. Although the last pandemic of respiratory
29 disease of viral origin swept the globe only a decade ago, the way science operates and responds to
30 current events has experienced a paradigm shift in the interim. The scientific community has
31 responded rapidly to the COVID-19 pandemic, releasing over 16,000 COVID-19 related scientific
32 articles within 4 months of the first confirmed case, of which at least 6,000 were hosted by preprint
33 servers. We focused our analysis on bioRxiv and medRxiv, two growing preprint servers for biomedical
34 research, investigating the attributes of COVID-19 preprints, their access and usage rates,
35 characteristics of their sharing on online platforms, and the relationship between preprints and their
36 published articles. Our data provides evidence for increased scientific and public engagement (COVID-
37 19 preprints are accessed and distributed at least 15 times more than non-COVID-19 preprints) and
38 changes in journalistic practice with reference to preprints. We also find evidence for changes in
39 preprinting and publishing behaviour: COVID-19 preprints are shorter, with fewer panels and tables,
40 and reviewed faster. Our results highlight the unprecedented role of preprints and preprint servers in
41 the dissemination of COVID-19 science, and the likely long-term impact of the pandemic on the
42 scientific publishing landscape.

43

44

45

46

47

48

49

50

51

52

53

54 Introduction

55 The first quarter of 2020 has been defined by the COVID-19 outbreak, which has escalated to
56 pandemic status, and caused over 3.2 million cases and 220,000 deaths within 4 months of the first
57 reported case [1,2]. The causative pathogen was rapidly identified as a novel virus within the family
58 *Coronaviridae* and was named severe acute respiratory syndrome coronavirus 2 (or ‘SARS-CoV-2’) [3].
59 Although multiple coronaviruses are ubiquitous among humans and cause only mild disease,
60 epidemics of newly emerging coronaviruses were previously observed in SARS coronavirus in 2002 [4]
61 and Middle East respiratory syndrome (MERS) coronavirus in 2012 [5]. The unprecedented extent and
62 rate of spread of COVID-19 has created a critical global health emergency and academic communities
63 have raced to actively respond through research developments.

64 Research developments have traditionally been communicated via published journal articles or
65 conference presentations. Traditional scientific publishing involves the submission of manuscripts to
66 an individual journal, which then organises peer review. Authors often conduct additional experiments
67 or analyses to address the reviewers’ concerns in one or more revisions. Even after this lengthy
68 process is concluded, almost half of submissions are rejected and require re-submission to a different
69 journal [6]. The median time between the date a preprint is posted and the date at which the first DOI
70 of a journal article is registered is 166 days [7]. Escalating demands made by reviewers and editors are
71 lengthening the publication process still further [8,9].

72 Further compounding the issues with traditional publishing, public funds are often used to conduct
73 research, pay direct publication costs and then pay once again for institutional subscriptions. Lack of
74 access to research articles due to these “paywalls” has a disproportionately negative effect on
75 scientific participation in developing countries [10]. Recent years have seen concerted efforts to
76 reduce paywalls as a barrier to scientific advances, the most prominent example being the Plan S
77 initiative (<https://www.coalition-s.org/>) which requires researchers supported by a large number of
78 national and international funding agencies to publish all of their work in open repositories or open
79 access journals by 2021. However, more than half of the newly-published global scientific literature
80 remains behind journal paywalls [11].

81 Preprints are publicly-accessible scientific manuscripts that have not yet been certified by peer review
82 [12]. While experiments with preprints date back to the 1960s [13], the physics and mathematics
83 communities have been sharing papers on arXiv, a preprint server launched in 1991 [14]. Initial efforts
84 to launch preprint servers in the life sciences were met with challenges, such as opposition from
85 traditional publishers and little interest from biologists [12,15,16]. However, in 2013 two new preprint
86 initiatives launched: PeerJ Preprints, from the publisher PeerJ, and bioRxiv, from Cold Spring Harbor

87 Labs. The latter established partnerships with journals that enabled simultaneous preprint posting at
88 the time of submission [17]. More recently, CSHL, in collaboration with Yale and BMJ, launched
89 medRxiv, a server for the medical sciences [15]. Preprint platforms have subsequently flourished, with
90 two-thirds of preprints eventually being published in peer-reviewed journals [7].

91 While funders and institutions explicitly encouraged pre-publication data sharing in the context of the
92 recent Zika and Ebola outbreaks [18], usage of preprints remained modest through these epidemics
93 [19]. The COVID-19 crisis represents the first time that preprints have been widely used to
94 communicate during an epidemic.

95 We assessed the role of preprints in the current COVID-19 pandemic between January 1st and April
96 30th, determining how preprint servers are being used, how preprints are being disseminated and how
97 they change in their published versions. We found that preprint servers hosted a large amount of
98 COVID-19 related science, that this was being accessed and downloaded in far greater volume than
99 other preprints on the same servers and that this was widely shared across multiple online platforms.
100 Moreover, we determined that COVID-19 preprints are shorter and are reviewed faster. Taken
101 together, our data demonstrates the importance of rapidly and openly sharing science in the context
102 of a global pandemic and the essential role of preprints in this endeavour.

103

104 Results

105 COVID-19 preprints were posted early in the pandemic

106 The COVID-19 pandemic has rapidly spread across the globe, from 3 patients in the city of Wuhan on
107 the 27th December 2019 to over 3.2 million confirmed cases worldwide by the end of April 2020 (Fig.
108 1A). Following the declaration of COVID-19 as a pandemic by the WHO on 11th March [20], the number
109 of cases grew exponentially in March, despite interventions by governments [21]. The scientific
110 community responded rapidly as COVID-19 emerged as a serious threat, with publications appearing
111 within weeks of the first reported cases (Fig. 1B, data from [22]). By the end of January 2020, 166
112 scientific articles related to COVID-19 had been published in either a peer-reviewed journal or on a
113 preprint server. When compared to other recent outbreaks of global significance caused by emerging
114 RNA viruses, the response to COVID-19 has been much more rapid. In the first 4 months of the COVID-
115 19 outbreak, 2,527 preprints were posted to bioRxiv and medRxiv alone; in comparison, only 78 Zika
116 virus, and 10 Western Africa Ebola virus preprints were posted to bioRxiv and medRxiv during the
117 respective time periods in which the outbreaks occurred. This surge in COVID-19 preprints is not
118 explained by general increases in preprint server usage as the proportion of epidemic-related

119 preprints was significantly greater for COVID-19 (Chi-square; $\chi^2 = 1641.6$, $df = 2$, $p < 0.001$)
120 (Supplemental Fig. 1A).

121 By the end of March, at least 5000 scientific articles were published relating to COVID-19; by the end
122 of April this number had tripled to more than 16,000. A large proportion of these articles (>6000) were
123 manuscripts hosted on preprint servers (Fig. 1B). Despite being one of the newest preprint servers,
124 medRxiv hosted the largest number of preprints (~2000), whilst other preprint servers (with the
125 exception of SSRN which hosts social sciences and humanities preprints) were each found to host
126 <1000 preprints (Fig. 1C). Eleven of the 31 preprint servers included in our dataset hosted over 100
127 COVID-19 related preprints each. It is important to note, however, that this preprint data is not
128 exhaustive, and several preprint servers that may be expected to also host large amounts of COVID-
129 19 research (e.g. RePEc, for economics research) are not included; the amount of research hosted by
130 preprint servers is likely an underestimate of the true amount [22].

131 Following a steep increase in the posting of COVID-19 research, traditional publishers adapted new
132 policies to support the ongoing public health emergency response efforts (Fig. 1D). Following multiple
133 public calls from scientists [23], over 30 publishers agreed to make all COVID-19 work freely accessible
134 by the 16th March [24,25]. Shortly after this, publishers (for example eLife [26]) began to alter peer-
135 review policies in an attempt at fast-tracking COVID-19 research. Towards the end of April, OASPA
136 issued an open letter of intent to maximise the efficacy of peer review [27]. The number of open-
137 access COVID-19 journal articles suggests that journals have largely been successful with these new
138 policies (Supplemental Fig. 1B).

139 [Attributes of COVID-19 preprints posted between January and April 2020](#)

140 Having observed that a large proportion of the scientific literature was hosted by multiple preprint
141 servers (Fig. 1B), we focused our following investigation on two of the most popular preprint servers
142 in the biomedical sciences: bioRxiv and medRxiv.

143 Between January and April 2020, 14,812 preprints were deposited between bioRxiv and medRxiv, of
144 which the majority (12,285, 82.9%) were non-COVID-19 preprints (Fig. 2A). The numbers of non-
145 COVID-19 related preprints deposited each week did not dramatically change over this period.
146 However, the number of COVID-19 preprints posted per week increased, peaking at over 250 in the
147 week beginning 6th April. The observed increase in COVID-19 preprints, did not seem to impact on the
148 number of non-COVID-19 related preprints being posted within any given week (Fig. 2A). When the

149 data was broken down by server, it was evident that whilst posting of preprints to bioRxiv had
150 remained relatively steady, preprints posted to medRxiv increased with time (Supplemental Fig. 2A).

151 This increase in posting poses challenges for the timely screening of preprints; we therefore analysed
152 the screening times of bioRxiv and medRxiv over this period. Only marginally faster screening was
153 detected for COVID-19 preprints than for non-COVID-19 preprints (Fig. 2B) when adjusting for
154 differences between servers (two-way ANOVA, interaction term; $F_{1,14808} = 69.13$, $p < 0.001$). Whilst
155 COVID-19 preprints were screened < 1 day quicker from mean differences observed within both
156 servers (Tukey HSD; both $p < 0.001$), larger differences were observed between servers (Supplemental
157 Fig. 2B), with bioRxiv screening preprints on approximately 2 days quicker than medRxiv for both
158 preprint types (both $p < 0.001$).

159 We next investigated the geographical distribution of preprint authors. Non-COVID-19 preprints most
160 commonly featured a corresponding author (which we assumed to be senior author) based in the
161 United States (US), with significant authorship also originating within the United Kingdom (GB) and
162 China. Considering COVID-19 preprints, China instead had the most corresponding authors (almost
163 20%), followed by the US and GB (Fig. 2C). We found that most countries posted their first COVID-19
164 preprint near to the time of their first confirmed COVID-19 case (Fig. 2D), with weak positive
165 correlation considering calendar days of both events (Spearman's rank; $\rho = 0.39$, $p = 0.001$). Countries
166 posting a COVID-19 preprint in advance of their first confirmed case were mostly higher-income
167 countries (e.g. US, GB, New Zealand, Switzerland). COVID-19 preprints were deposited from every
168 inhabited continent, revealing the global response to the pandemic.

169 The number of authors may give an indication as to the amount of work, resources used, and the
170 extent of collaboration in a paper. We therefore investigated the distribution of size of authorship
171 teams across preprints. While the average number of authors of COVID-19 and non-COVID-19
172 preprints did not differ, COVID-19 preprints showed slightly more variability in authorship team size
173 (median, 6 [IQR 8] vs 6 [IQR 5]). Single-author preprints were almost three times more common among
174 COVID-19 than non-COVID-19 preprints (Fig. 2E).

175 bioRxiv and medRxiv allow authors to select from a number of different Creative Commons
176 (<https://creativecommons.org/>) license types when depositing their work: CC0 (No Rights Reserved),
177 CC-BY (Attribution), CC BY-NC (Attribution, Non-Commercial), CC-BY-ND (Attribution, No-Derivatives),
178 CC-BY-NC-ND (Attribution, Non-Commercial, No-Derivatives). Authors may also select to post their
179 work without a license (i.e. All Rights Reserved). A previous analysis has found that bioRxiv authors
180 tend to post preprints under the more restrictive license types [28], although there appears to be

181 some confusion amongst authors as to the exact meaning of each license type [29]. We assessed
182 whether authors choose different license types when posting COVID-19 versus non-COVID-19
183 preprints (Fig. 2F). Authors of COVID-19 preprints were more likely to choose CC-BY-NC-ND or CC-BY-
184 ND than those of non-COVID-19 preprints, and less likely to choose CC-BY and CC (Fisher's exact, 1000
185 simulations; $p < 0.001$).

186 Preprint servers offer authors the opportunity to post new versions of a preprint, to improve upon or
187 correct mistakes in an earlier version. Predominantly, preprints existed as only a single version for
188 both COVID-19 or non-COVID-19 work with very few preprints existing beyond two versions (Fig. 2G).
189 COVID-19 preprints did not discernibly differ in number of versions compared with non-COVID-19
190 preprints (median, 1 [IQR 1] vs 1 [IQR 0]).

191 The speed with which COVID-19 preprints are being posted suggests that researchers have changed
192 the way in which they share results. To investigate this, we compared the word counts of COVID-19
193 preprints and non-COVID-19 preprints from bioRxiv. We found that COVID-19 preprints are indeed on
194 average shorter in length than non-COVID-19 preprints (median, 3432 [IQR 2597] vs 6143 [IQR 3363];
195 Mann-Whitney, $p < 0.001$) (Fig. 2H). This supports anecdotal observations that preprints are being
196 used to share more work-in-progress data than a complete story. We also found that COVID-19
197 preprints contain fewer references than non-COVID-19 preprints, reflecting the new, emerging COVID-
198 19 field (median, 30.5 [IQR 29] vs 51 [IQR 31]; $p < 0.001$) (Fig. 2I).

199

200 Extensive access of preprint servers for COVID-19 research

201 Throughout our time window, COVID-19 preprints received abstract views at a rate over 15 times that
202 of non-COVID-19 preprints (Fig. 3A) (time-adjusted negative binomial regression; odds ratio = 15.6, z
203 = 143.8, $p < 0.001$). There was minimal change in total abstract views over time for COVID-19 and non-
204 COVID-19 preprints, with each additional calendar week in posting date resulting in a 6.3% reduction
205 in odds of views (odds ratio = 0.937, $z = -44.56$, $p < 0.001$), suggesting that most preprints receive the
206 majority of views near the time of posting.

207 We found similar results when comparing the pdf downloads of COVID-19 and non-COVID-19
208 preprints, with COVID-19 preprints receiving almost 30 times more downloads (Fig. 3B) (odds ratio =
209 28.9, $z = 155.1$, $p < 0.001$). Again, there was negligible change in the rate of pdf downloads between
210 posting times for all examined preprints, with each additional calendar week in posting date resulting
211 in an 8.1% reduction in rate of downloads (odds ratio = 0.919, $z = -51.07$, $p < 0.001$). This further

212 suggested most preprints receive their heaviest usage near to time of posting, with the highest
213 observed usage for COVID-19 preprints occurring on the week commencing 20th January.

214 To confirm that usage of COVID-19 and non-COVID-19 preprints was not an artefact of differing
215 preprint server reliance during the pandemic, we compared usage to September 2019 – April 2020, as
216 a non-pandemic control period. We observed a slight decrease in abstract views (Supplemental Fig.
217 3A) and pdf downloads (Supplemental Fig. 3B) in March 2020, but otherwise, the usage data did not
218 differ from that prior to the pandemic.

219 We investigated usage across additional preprint servers (data kindly provided by each of the servers).
220 We found that COVID-19 preprints were consistently downloaded more than non-COVID-19 preprints
221 during our timeframe, regardless of which preprint server hosted the science (Supplemental Fig. 3C),
222 though the gap in downloads varied between server (two-way ANOVA, interaction term; $F_{4,276544} =$
223 586.9 , $p < 0.001$). Server usage differences were more pronounced for COVID-19 preprints; multiple
224 post-hoc comparisons confirmed that bioRxiv and medRxiv received significantly higher usage per
225 COVID-19 preprint than all other servers for which data was available (Tukey HSD; all p values < 0.001).
226 However, for non COVID-19 preprints, the only observed pairwise differences between servers
227 indicated greater bioRxiv usage than SSRN or Research Square (Tukey HSD; all p values < 0.001). This
228 suggests specific attention has been given disproportionately to bioRxiv and medRxiv as repositories
229 for COVID-19 research.

230 COVID-19 preprints were shared more widely than non-COVID-19 preprints

231 Based on citation data from Dimensions, we found that COVID-19 preprints are being cited much more
232 often than non-COVID-19 preprints (time-adjusted negative binomial regression; odds ratio = 71.1, z
233 = 49.2, $p < 0.001$) (Fig. 4A), although it should be noted that only a minority of preprints received at
234 least one citation (30.6 % vs 5.5 %). We next investigated the ten highest cited COVID-19 preprints
235 (Table 1). The highest cited preprint had 127 citations, with the 10th most cited COVID-19 preprint
236 receiving 48 citations, with much of the highest cited preprints focus on the viral cell receptor,
237 angiotensin converting enzyme 2 (ACE2) or the epidemiology of COVID-19.

238

239 Utilising data from Altmeteric, we also investigated sharing of preprints on Twitter to assess the
240 exposure of wider public audiences to preprints. COVID-19 preprints were shared more often than
241 non-COVID-19 preprints (odds ratio = 14.8, $z = 91.55$, $p < 0.001$) (Fig. 4B). The most tweeted non-
242 COVID-19 preprint received 1,323 tweets, whereas 8 of the top 10 tweeted COVID-19 preprints were
243 tweeted over 10,000 times each (Table 2). Many of the top 10 tweeted COVID-19 preprints were
244 related to transmission, re-infection or seroprevalence and association with the BCG vaccine. The

245 most tweeted COVID-19 preprint (29,984 tweets) was a study investigating antibody seroprevalence
246 in California [30], whilst the second most tweeted COVID-19 preprint was a widely criticised (and later
247 withdrawn) study linking the SARS-CoV-2 spike protein to HIV-1 glycoproteins.

248

249 To better understand the main discussion topics associated with the top-10 most tweeted preprints,
250 we analysed the hashtags used in original tweets (i.e. excluding retweets) mentioning those preprints
251 (Supplemental Fig. 4A). After removing some highly inflated hashtags directly referring to the virus
252 (e.g. “#coronavirus”, “#COVID-19”), we found that the most dominant hashtag among tweets
253 referencing preprints was “#chloroquine”, a major controversial topic associated with two of the top
254 ten most tweeted preprints. Other prominent hashtags contained a mixture of direct, neutral
255 references to the disease outbreak such as “#coronavirusoutbreak” and “#Wuhan”, and some more
256 politicised terms, such as “#fakenews” and “#covidisalie”, associated with conspiracy theories.

257

258 As well as featuring heavily on social media, COVID-19 research has also saturated print and online
259 news media. We found that COVID-19 preprints had over two-hundred fold odds of being shared in
260 news articles than non-COVID-19 preprints (odds ratio = 220.4, $z = 39.27$, $p < 0.001$), although as with
261 citations, only a minority were mentioned in news articles at all (26.9% vs 6.7%) (Fig. 4C). The top non-
262 COVID-19 preprints were reported in less than 100 news articles in total, whereas the top COVID-19
263 preprints were reported in over 300 news articles (Table 3). Similarly, when we investigated the
264 sharing of preprints across blogs, we found that COVID-19 preprints were shared more than non-
265 COVID-19 preprints (odds ratio = 9.48, $z = 29.2$, $p < 0.001$) (Fig. 4D). We noted that several of the most
266 widely-disseminated non-COVID-19 preprints featured generalised topics still relevant to infectious
267 disease research, e.g. human respiratory physiology and personal protective equipment (Tables 2 and
268 3).

269

270 We next investigated if there was a correlation between these different usage indicators (citations,
271 tweets, news articles and blogs). In general, we observe much weaker correlation between all
272 indicators for non-COVID-19 preprints compared to COVID-19 preprints (Fig. 4E and 4F). For COVID-
273 19 preprints, we found weak correlation between the numbers of citations and Twitter shares
274 (Spearman’s $\rho = 0.37$, $p < 0.001$), and the numbers of citations and news articles ($\rho = 0.41$, $p < 0.001$)
275 (Fig. 4E), suggesting that the preprints cited mostly within the scientific literature differed to those
276 that were mostly shared by the wider public on other online platforms. There was a stronger
277 correlation between COVID-19 preprints that were most blogged and those receiving the most
278 attention in the news ($\rho = 0.58$, $p < 0.001$). Moreover, there was a strong correlation between COVID-

279 19 preprints that were most tweeted and those receiving the most attention in the news ($p = 0.52$, p
280 < 0.001), suggesting similarity between preprints shared on social media and in news media (Fig. 4E).
281 Indeed, of the top ten COVID-19 preprints that were tweeted or mentioned in news articles, five
282 appeared in both lists (Supplemental Fig. 4B).

283

284 As the sentiment of tweet text content associated with each of the 10 most tweeted COVID-19
285 preprints was scored to be generally positive (Supplemental Fig. 4C), we decided to examine topics
286 associated with the most shared COVID-19 preprints. We analysed the hashtags used on twitter for 3
287 of the preprints that were amongst the top ten most tweeted, top ten most mentioned in news articles
288 and top ten most blogged (Tables 1-4; Supplemental Fig. 4D-I). Diverse topics appeared in the
289 discussions following each individual preprint; the most tweeted preprint [30] was associated with
290 hashtags such as “#endthelockdown”, “#drfauci” and “#billgates” (Supplemental Fig. 4D & E), whilst
291 the fifth most tweeted article [31] was associated with hashtags related to prevention measures, for
292 example, “flattenthecurve”, “#washyourhands” and “#socialdistancing” (Supplemental Fig. 4F & G). A
293 preprint demonstrating a lack of efficacy of hydroxychloroquine [32] was dominated by the hashtag
294 “#fakenews” and “#hydroxychloroquine” (Supplemental Fig. 4H & I).

295

296 Our data reveals that COVID-19 preprints received an unprecedented amount of attention from
297 scientists, news organisations and the general public, representing a departure for how preprints are
298 normally shared (considering observed patterns for non-COVID-19 preprints).

299

300 **Table 1. Top 10 cited COVID-19 preprints**

301 **Table 2. Top 10 tweeted COVID-19 preprints**

302 **Table 3. Top 10 COVID-19 preprints covered by news organisations**

303

304 [Publishing and peer review of preprints during the pandemic](#)

305 We have demonstrated that preprint servers are seeing unprecedented use in response to the COVID-
306 19 pandemic (Figs. 1 & 2). Many traditional publishers adapted their policies in response to the
307 pandemic to better facilitate the communication and sharing of COVID-19 research (Fig. 1D). Within
308 our timeframe, 4% of COVID-19 preprints were published by April, a significant increase compared to
309 the 3% of non-COVID preprints that were published ($\chi^2 = 6.77$, $df = 1$, $p = 0.009$) (Fig. 5A). These
310 published COVID-19 preprints were split across many journals, with clinical or multidisciplinary

311 journals tending to publish the most papers that were previously preprints (Supplemental Fig. 5A). To
312 determine how publishers were prioritising COVID-19 research, we compared the time from preprint
313 posting to publication in a journal. Delay from posting to subsequent publication was significantly
314 accelerated for COVID-19 preprints by a mean difference of 25.7 days compared to non-COVID-19
315 preprints posted in the same time period (two-way ANOVA; $F_{1,289} = 69.8$, $p < 0.001$). This did not appear
316 driven by any temporal changes in publishing practices, as non-COVID preprints were similar to
317 expectation of our control timeframe of September - January (Fig. 5B). COVID-19 preprints also
318 appeared to have significantly accelerated publishing regardless of publisher (two-way ANOVA,
319 interaction term; $F_{6,283} = 0.41$, $p = 0.876$) (Supplemental Fig. 5B).

320 As a response to the pandemic, many labs have shifted their focus to COVID-19 research, with much
321 discussion over how appropriate this might be [33]. To quantify whether this was detectable within
322 the preprint literature, for each corresponding author associated with a COVID-19 preprint we traced
323 back their most recent previous preprint (COVID-19 or non-COVID-19) and compared the server-
324 deposited categories of both. Most senior authors were not drastically changing fields, with category
325 differences generally spanning reasonably related areas (for example, some authors previously
326 posting preprints in evolutionary biology have posted COVID-19 preprints in microbiology) (Fig. 5C).
327 This suggests that - at least within the life sciences - principal investigators are utilising their labs' skills
328 and resources in a responsible manner in their contributions to COVID-19 research.

329 Independent COVID-19 review projects have arisen to publicly review COVID-19 preprints [34]. To
330 determine the extent of non-journal-organised, public, peer-review we quantified the number of
331 comments for preprints posted between January and April. We found that non-COVID-19 preprints
332 were rarely commented upon, in comparison to COVID-19 preprints (time-adjusted negative binomial
333 regression; odds ratio = 27.9, $z = 32.0$, $p < 0.001$) (Fig. 5D); the most commented non-COVID-19
334 preprint received only 15 comments, whereas the most commented COVID-19 preprint had over 500
335 comments on the 30th April (Table 4). One preprint, which had 127 comments was retracted within 3
336 days of being posted following intense public scrutiny [35]. Comparing the sentiment score of the top
337 10 most commented COVID-19 preprints revealed a broadly positive sentiment within the comments
338 (Supplemental Fig. 5C). In contrast, an overwhelming majority of preprints that were subsequently
339 published were not associated with transparent reviews (Supplemental Fig. 5D) and many had similar
340 data availability to their preprint version (Supplemental Fig. 5E). Collectively these data suggest that
341 the most discussed or controversial COVID-19 preprints are being rapidly and publicly scrutinised, with
342 flawed preprints being either removed or updated.

343 Having established that public scrutiny was occurring for at least a portion of the COVID-19 preprints,
344 we assessed the extent to which published COVID-19 articles that were previously preprints had
345 changed during the publication process. We randomly sampled an equal number of published non-
346 COVID-19 articles that were previously preprints to act as a control sample and then qualitatively and
347 quantitatively scored preprint-paper pairs. Over 75% of preprints did not have any change in the
348 author list, with 15.8% of COVID-19 preprints having authors added for publication compared to 6.06%
349 of non-COVID preprints (Supplemental Fig. 5F). We assessed the difference between abstracts,
350 classifying whether the published abstract had no change, a softening or strengthening of the wording,
351 or a major change in the conclusions. We found that 61.3% of COVID-19 preprints did not have
352 significantly altered abstracts following publication (Fig. 5E). However, 26.7% of the COVID-19
353 abstracts did have altered wording or numbers that strengthened or softened the data and
354 conclusions, with 4.9% displaying major changes in the conclusions. Among non-COVID-19 abstracts,
355 77.7% did not have significantly altered abstracts, 20.2% had altered wording or numbers in the
356 abstract with 1.01% having major changes following peer review.

357 We next assessed the content of the preprint-paper pairs, focussing on the figures and tables. For
358 both COVID-19 and non-COVID-19 preprints, over 60% did not exhibit any additions, removals, or
359 rearrangements from the preprinted manuscript (Fig. 5F). Where we did observe a change, this was
360 often a re-arranging of the panels across figures or between the main paper and supplementary
361 sections. Importantly, we scored over 24.75% of COVID-19 preprints as having significant content
362 added or removed from figures, a similar score to non-COVID-19 preprints (21.21%). Surprisingly,
363 61.3% of COVID-19 preprints and 62.6% of non-COVID-19 preprints had no panel additions, removals,
364 or rearrangements at all (Supplemental Fig. 5G). Furthermore, we found that COVID-19 preprints and
365 papers contained significantly fewer total numbers of panels and tables than non-COVID-19 preprints
366 and papers (two-way mixed ANOVA; $F_{1,198} = 16.0$, $p < 0.001$, mean difference = 4.7); though there was
367 no difference between preprint and paper pairs ($F_{1,199} = 0.294$, $p = 0.588$) (Supplemental Fig. 5H).

368 Our data demonstrates that there is a public scrutiny of high-attention COVID-19 preprints and for
369 preprints published within our timeframe there was little change in the number or arrangement of
370 figure panels and tables of preprints compared to the published paper. Tracking with our earlier
371 observations of diminished word counts, COVID-19 preprints have markedly fewer figure panels and
372 tables than other preprints.

373

374 **Table 4. Top 10 most commented COVID-19 preprints**

375

376 Discussion

377 Our results show that preprints have been widely adopted and used for the communication of COVID-
378 19 research, and in turn, the pandemic has left what is likely to be a lasting imprint on the preprint
379 and science publishing landscape.

380 The evolution of the preprint response to COVID-19 has been in stark contrast to previous major
381 infectious diseases outbreaks: Johansson et al. [19] found only 174 preprints in a range of repositories
382 (including bioRxiv) were posted in response to the 2015-2016 Zika virus outbreak, and 74 preprints
383 were posted in response to the 2013-2016 Western African Ebola virus outbreak. The number of
384 preprints posted in response to these two outbreaks was dwarfed by the number of peer-reviewed
385 journal articles, where 1,641 and 2,187 PubMed-indexed journal articles related to Ebola and Zika,
386 respectively, were published in the same period [19]. In comparison, in just 4 months following the
387 first case of COVID-19, 2,527 preprints have been posted to bioRxiv and medRxiv alone, and >40% of
388 the total COVID-19 literature to date has been posted via preprints (Figure 1B).

389 The need to rapidly communicate findings prior to a lengthy review process might be driving more
390 authors to post preprints in response to COVID-19 (Fig.3). A recent study involving qualitative
391 interviews of multiple research stakeholders found “early and rapid dissemination” to be amongst the
392 most often cited benefits of preprints [36]. These findings were echoed in a survey of ~4200 bioRxiv
393 users [12], and are underscored by the 6 month median lag between posting of a preprint and
394 subsequent journal publication [7,37]. Such timelines for disseminating findings are clearly
395 incompatible with the lightning-quick progression of a pandemic. An analysis of publication timelines
396 for 14 medical journals has shown that some publishers have taken steps to accelerate their publishing
397 processes for COVID-19 research, reducing the time for the peer-review stage (submission to
398 acceptance) on average by 45 days, and the editing stage (acceptance to publication) by 14 days [38],
399 yet this still falls some way short of the ~1-3 day screening time for bioRxiv and medRxiv preprints (Fig.
400 2B).

401 A number of additional motivations driving the increase in preprints in response to COVID-19 may fall
402 on a spectrum from altruistic (e.g. to make findings openly available for everyone) to egotistic (e.g. to
403 stamp a priority claim on a finding to prevent being “scooped”), all of which may be amplified by the
404 unique circumstances of the COVID-19 outbreak. Further studies on this aspect, for example through
405 quantitative and qualitative author surveys may help funders and other stakeholders that support the
406 usage of preprints to address some of the social barriers for their uptake [39].

407 bioRxiv and medRxiv included a banner to explain that preprints should not be regarded as conclusive
408 and not reported on in the news media as established information [40]. Despite the warning message,
409 COVID-19 preprints have received unprecedented coverage on online media platforms (Fig. 4). Twitter
410 has been a particularly notable outlet for communication of preprints, although questions of exactly
411 *who* is tweeting about COVID-19 research, and what that means in terms of societal impact, remain
412 open. Twitter might not fully reflect public interest in research, as tweets are overrepresented by
413 academic users [41], and Twitter metrics may largely be dominated by mechanical retweeting rather
414 than reflecting original thought [42], although engagement levels may be generally higher for
415 biomedical research than for research from other fields [43]. This is underscored by the relatively weak
416 correlation found between citations and other indicators of online sharing (Fig 4E): of the articles in
417 the top-10 most shared on twitter, in news articles or on blogs, only one is ranked amongst the top-
418 10 most cited articles (Supplemental Fig. 4B). Hashtags associated with individual, highly tweeted
419 preprints reveal some emergent themes that suggest communication of certain preprints can,
420 however, extend well beyond scientific audiences (Supplemental Fig. 4). These range from good public
421 health practice (“#washyourhands”) to right-wing philosophies and conspiracy theories, (“#fakenews”
422 and “#endthelockdown”). This type of misinformation is common to new diseases [44] and social
423 media platforms have recently released a statement outlining their plans to combat this issue [45]. It
424 is also interesting to note that several preprints received negatively by the scientific community are
425 amongst the most tweeted: the preprint (“Uncanny similarity of unique inserts in the 2019-nCoV spike
426 protein to HIV-1 gp120 and Gag”; [46]), was withdrawn within 3 days by the authors following critical
427 comments. These findings make clear that indicators of social media sharing, at least within the
428 context of COVID-19, should be carefully interpreted and not used as direct indicators or proxies of
429 scientific quality.

430 The fact that news outlets are reporting extensively on COVID-19 preprints (Fig. 4B and 4C) represents
431 a marked change in journalistic practice: pre-pandemic bioRxiv preprints received, in comparison to
432 journal articles, very little coverage [37]. This cultural shift provides an unprecedented opportunity to
433 bridge the scientific and media communities to create a consensus on the reporting of preprints [47].
434 In the near future, we aim to examine whether this change in practice extends beyond the media to
435 governments and policy-making bodies

436 It is not just preprints serving to inform the global pandemic response: the pandemic, in turn, is having
437 a major impact on peer-review and traditional scientific communication practices. Are these changes
438 in practice having a knock-on effect on peer review? To address this in our data, we compared
439 preprint-paper pairs across a range of metrics (Fig. 5). We were surprised to find that there was little

440 change between the number of figure panels and tables between preprint and subsequent published
441 manuscript – for both COVID-19 and non-COVID-19 preprints. Where we did observe addition or
442 removal of content, we rarely categorised this as significantly altering the conclusions stated in the
443 abstract, though there were more incidences of major abstract changes among COVID-19 preprints
444 than non-COVID-19 preprints (4.9% vs 1.0%). This supports other recent observations suggesting little
445 change between preprints and their published paper [48]. While comparing the preprint-paper pairs,
446 we noticed a high number of pairs for which data was harder to access after publication (often due to
447 broken links to supplemental material). It remains important to recognise, however, that our data
448 suffers from survivorship and selection bias, as those preprints published within our short timeframe
449 are potentially more likely to be of a higher standard than preprints which are not published or take
450 longer to reach publication. This is particularly relevant for the subset of COVID-19 preprints which
451 undergo more version changes than non-COVID-19 preprints.

452 Readers cannot use the journal in which papers have been published as a mechanism to judge their
453 reception among peers. As most COVID-19-preprints were not yet published, concerns regarding
454 quality will persist. Despite increased publicity for established preprint-review services (such as
455 PRereview [27,49], there has been limited use of these platforms [50]. However, independent
456 preprint-review projects have arisen whereby reviews are posted in the comments section of preprint
457 servers and hosted on independent websites [34]. These more formal projects partly account for the
458 increased commenting on COVID-19 preprints (Fig. 5). However, it is clear that the general public are
459 also using the commenting systems in addition to scientists. Moreover, prominent scientists are using
460 social media platforms such as Twitter to publicly share concerns with poor quality COVID-19 preprints
461 or to amplify high-quality preprints [51].

462 Our data demonstrates the indispensable role that preprints, and preprint servers, are playing during
463 a global pandemic. By communicating science through open-access preprints, we are sharing at a
464 faster rate than allowed by the current journal infrastructure, with limited impact on the quality of
465 preprints that are subsequently published.

466

467 [Methods](#)

468

469 [Preprint Metadata for bioRxiv and medRxiv](#)

470 We retrieved basic preprint metadata (DOIs, titles, abstracts, author names, corresponding author
471 name and institution, dates, versions, licenses, categories and published article links) for bioRxiv and

472 medRxiv preprints via the bioRxiv Application Programming Interface (API; <https://api.biorxiv.org>).
473 The API accepts a 'server' parameter to enable retrieval of records for both bioRxiv and medRxiv. We
474 initially collected metadata for all preprints posted from the time of the server's launch, corresponding
475 to November 2013 for bioRxiv and June 2019 for medRxiv, until the end of our analysis period on 30th
476 April 2020 (N = 84,524). All data were collected on 1st May 2020. Note that where multiple preprint
477 versions existed, we included only the earliest version and recorded the total number of following
478 revisions. Preprints were classified as "COVID-19 preprints" or "Non-COVID-19 preprints" on the basis
479 of the following terms contained within their titles or abstracts (case-insensitive): "coronavirus",
480 "covid-19", "sars-cov", "ncov-2019", "2019-ncov", "hcov-19", "sars-2". For comparison of preprint
481 behaviour between the COVID-19 outbreak and previous viral epidemics, namely Western Africa Ebola
482 virus and Zika virus (Supplemental Fig. 1), the same procedure was applied using the keywords "ebola"
483 or "zebov", and "zika" or "zikv", respectively.

484 For a subset of preprints posted between 1st September 2019 and 30th April 2020 (N = 25,883), we
485 enhanced the basic preprint metadata with data from a number of other sources, as outlined below.
486 Note that this time period was chosen to encapsulate our 4-month analysis period from 1st January
487 to 30th April 2020 (N = 14,812), as well as the preceding 4-month period from September 1st to
488 December 31st 2019 (N = 11,071), to use for comparison purposes. Of the preprints contained in the
489 later 4-month analysis period, 2,527 (17.1%) contained COVID-19 related keywords in their titles or
490 abstracts.

491 For all preprints contained in the subset, disambiguated author affiliation and country data for
492 corresponding authors were retrieved by querying raw affiliation strings against the Research
493 Organisation Registry (ROR) API (<https://github.com/ror-community/ror-api>). The API provides a
494 service for matching affiliation strings against institutions contained in the registry, on the basis of
495 multiple matching types (named "phrase", "common terms", "fuzzy", "heuristics", and "acronyms").
496 The service returns a list of potential matched institutions and their country, as well as the matching
497 type used, a confidence score with values between 0 and 1, and a binary "chosen" indicator relating
498 to the most confidently matched institution. A small number (~500) of raw affiliation strings returned
499 from the bioRxiv API were truncated at 160 characters; for these records we conducted web-scraping
500 using the rvest package for R [52] to retrieve the full affiliation strings of corresponding authors from
501 the bioRxiv public webpages, prior to matching. For the purposes of our study, we aimed for higher
502 precision than recall, and thus only included matched institutions where the API returned a confidence
503 score of 1. A manual check of a sample of returned results also suggested higher precision for results
504 returned using the "PHRASE" matching type, and thus we only retained results using this matching
505 type. In a final step, we applied manual corrections to the country information for a small subset of

506 records where false positives would be most likely to influence our results by a) iteratively examining
507 the chronologically first preprint associated with each country following affiliation matching and
508 applying manual rules to correct mismatched institutions until no further errors were detected (n = 8
509 institutions); and b) examining the top 50 most common raw affiliation strings and applying manual
510 rules to correct any mismatched or unmatched institutions (n = 2 institutions). In total, we matched
511 19,002 preprints to a country (73.2%); for COVID-19 preprints alone, 1716 preprints (67.9%) were
512 matched to a country. Note that a similar, albeit more sophisticated method of matching bioRxiv
513 affiliation information with the ROR API service was recently documented by Abdill et al. [53].

514 Word counts and reference counts for each preprint were also added to the basic preprint metadata
515 via scraping of the bioRxiv public webpages (medRxiv currently does not display full HTML texts, and
516 so calculating word and reference counts was limited to bioRxiv preprints). Web scraping was
517 conducted using the rvest package for R [52]. Word counts refer to words contained only in the main
518 body text, after removing the abstract, figure captions, table captions, acknowledgements and
519 references. In a small number of cases, word counts could not be retrieved because no full-text
520 existed; this occurs as we targeted only the first version of a preprint, but in cases where a second
521 version was uploaded very shortly (i.e. within a few days) after the first version, the full-text article
522 was generated only for the second version. Word and reference counts were retrieved for 21,975 of
523 22,156 bioRxiv preprints (99.1%); for COVID-19 preprints alone, word and reference counts were
524 retrieved for 553 of 564 preprints (98.0 %). Word counts ranged from 583 to 39,953 words, whilst
525 reference counts ranged from 1 to 487 references.

526 Our basic preprint metadata retrieved from the bioRxiv API also contained DOI links to published
527 versions (i.e. a peer-reviewed journal article) of preprints, where available. In total, 2710 records in
528 our preprint subset (10.5%) contained links to published articles, although of COVID-19 preprints only
529 101 preprints contained such links (4.0%). It should be noted that COVID-19 articles are heavily
530 weighted towards the most recent months of the dataset and have thus had less time to progress
531 through the journal publication process. Links to published articles are likely an underestimate of the
532 total proportion of articles that have been subsequently published in journals – both as a result of the
533 delay between articles being published in a journal and being detected by bioRxiv, and bioRxiv missing
534 some links to published articles when e.g. titles change significantly between the preprint and
535 published version [37]. Published article metadata (titles, abstracts, publication dates, journal and
536 publisher name) were retrieved by querying each DOI against the Crossref API
537 (<https://api.crossref.org>), using the rcrossref package for R [54]. We also retrieved data regarding the
538 open access status of each article by querying each DOI against the Unpaywall API, via the roadoi
539 package for R [55].

540

541 Usage, Altmetrics and Citation Data

542 For investigating the rates at which preprints are used, shared and cited, we collected detailed usage,
543 altmetrics and citation data for all bioRxiv and medRxiv preprints posted between 1st September 2019
544 to 30th April 2020 (i.e. for every preprint where we collected detailed metadata, as described in the
545 previous section). Collection of all usage, altmetrics and citation data were conducted on 1st May
546 2020.

547 Usage data (abstract views and pdf downloads) were scraped from the bioRxiv and medRxiv public
548 webpages, using the rvest package for R (Wickham, 2019). bioRxiv and medRxiv webpages display
549 abstract views and pdf downloads on a calendar month basis; for subsequent analysis (e.g Figure 4),
550 these were summed to generate total abstract views and downloads since the time of preprint
551 posting. In total, usage data were recorded for 25,865 preprints (99.9%) – a small number were not
552 recorded, possibly due to server issues during the web scraping process. Note that bioRxiv webpages
553 also display counts of full-text views, although we did not include these data in our final analysis. This
554 was partially to ensure consistency with medRxiv, which currently does not provide display full HTML
555 texts, and partially due to ambiguities in the timeline of full-text publishing – the full text of a preprint
556 is added several days after the preprint is first available, but the exact delay appears to vary from
557 preprint to preprint. We also compared rates of PDF downloads for bioRxiv and medRxiv preprints
558 with a number of other preprint servers (Preprints.org, SSRN, and Research Square) (Supplemental
559 Fig. 3C) - these data were provided directly by representatives of each of the respective preprint
560 servers.

561 Counts of multiple altmetric indicators (mentions in tweets, blogs, and news articles) were retrieved
562 via Altmetric (<https://www.altmetric.com>), a service that monitors and aggregates mentions to
563 scientific articles on various online platforms. Altmetric provide a free API (<https://api.altmetric.com>)
564 against which we queried each preprint DOI in our analysis set. Importantly, Altmetric only contains
565 records where an article has been mentioned in at least one of the sources tracked, thus, if our query
566 returned an invalid response we recorded counts for all indicators as zero. Coverage of each indicator
567 (i.e. the proportion of preprints receiving at least a single mention in a particular source) for preprints
568 were 99.1%, 9.6%, and 3.5% for mentions in tweets, blogs and news articles respectively. The high
569 coverage on Twitter is likely driven, at least in part, by automated tweeting of preprints by the official
570 bioRxiv and medRxiv twitter accounts. For COVID-19 preprints, coverage was found to be 100.0%,
571 16.6% and 26.9% for mentions in tweets, blogs and news articles respectively.

572 COVID-19 preprints may receive large volumes of usage and attention as a result of their perceived
573 quality being either high or low. To quantitatively capture whether high-usage preprints were well-
574 received by both public audiences, we firstly retrieved all tweets linking to the top ten most-tweeted
575 preprints. Tweet IDs were retrieved via the Altmetric API service, and then queried against the Twitter
576 API using the rtweet package [56] for R, to retrieve full tweet content (e.g. tweet text, hashtags). We
577 examined the positivity or negativity of each tweet text by calculating the average sentiment polarity
578 scores over all sentences using the sentimentr package for R [57]. Polarity of terms was determined
579 using an adjusted Semantic Orientation CALculator (SO-CAL) lexicon [58], neutralising scores for
580 various common scientific or infectious disease related terms, e.g. “respiratory”, “cellular”, “abstract”.
581 Polarity was adjusted for valence shifters, i.e. words or phrases that contextually alter sentiment, for
582 example the term “not” in the statement “this preprint is not interesting” would negate the otherwise-
583 positive term “interesting”. To avoid any potential bias, preprint title strings were excluded from tweet
584 texts before sentiment was calculated.

585 Citations counts for each preprint were retrieved from the scholarly indexing database Dimensions
586 (<https://dimensions.ai>). An advantage of using Dimensions in comparison to more traditional citation
587 databases (e.g. Scopus, Web of Science) is that Dimensions also includes preprints from several
588 sources within their database (including from bioRxiv and medRxiv), as well as their respective citation
589 counts. When a preprint was not found, we recorded its citation counts as zero. Of all preprints, 3707
590 (14.3%) recorded at least a single citation in Dimensions. For COVID-19 preprints, 774 preprints
591 (30.6%) recorded at least a single citation.

592

593 [Comments](#)

594 BioRxiv and medRxiv html pages feature a Disqus (<https://disqus.com>) comment platform to allow
595 readers to post text comments. Comment counts for each bioRxiv and medRxiv preprint were
596 retrieved via the Disqus API service (<https://disqus.com/api/docs/>). Where multiple preprint versions
597 existed, comments were aggregated over all versions. As with preprint perceptions among public
598 audiences on Twitter, we then examined perceptions among academic audiences by examining
599 comment sentiment. Text content of comments for COVID-19 preprints were provided directly by the
600 bioRxiv development team. Sentiment polarity scores were calculated for each comment on the top
601 ten most-commented preprints using the lexicon and protocol previously described for the analysis of
602 tweet sentiment.

603 Screening time for bioRxiv and medRxiv

604 To calculate screening time, we followed the method outlined by Steve Royle [59]. In short, we
605 calculate the screening time as the difference in days between the preprint posting date, and the date
606 stamp of submission approval contained within bioRxiv and medRxiv DOIs (only available for preprints
607 posted after December 11th 2019). bioRxiv and medRxiv preprints were filtered to preprints posted
608 between January 1st – April 30th 2020, accounting for the first version of a posted preprint.

609

610 Comparisons between preprints and their published articles

611 We identified all bioRxiv and medRxiv preprints from our bioRxiv and medRxiv preprints that have
612 been published in peer-reviewed journals (using journal DOIs extracted from the preprint metadata),
613 resulting in a set of 101 preprint-paper pairs. We generated a control set of 101 non-COVID-19
614 preprint-paper pairs by drawing a random subset of all bioRxiv and medRxiv preprints published in
615 peer reviewed journals within the extended analysis period (1st September 2019 and 30th April 2020;
616 see “Preprint Metadata for bioRxiv and medRxiv” for additional details), preserving the same ratio of
617 bioRxiv:medRxiv preprints as in the COVID-19 set. Each preprint-paper pair was then scored
618 independently by two referees using a variety of quantitative and qualitative metrics reporting on
619 changes in data presentation and organisation, the quantity of data, and the communication of
620 quantitative and qualitative outcomes between paper and preprint (using the reporting questionnaire
621 provided as supplemental material). Of particular note: individual figure panels were counted as such
622 when labelled with a letter, and for pooled analyses a full table was treated as a single-panel figure.
623 The number of figures and figure panels was capped at 10 each (Any additional figures/panels were
624 pooled), and the number of supplementary items (files/figures/documents) was capped at 5. In the
625 case of preprints with multiple versions, the comparison was always restricted to version 1 of the
626 preprint. Any conflicting assessments were resolved by a third independent referee, resulting in a final
627 consensus report for 99 non-COVID-19 and 101 COVID-19 preprint-paper pairs (excluding 10 pairs not
628 meeting the initial selection criteria or those still awaiting post-publication reviews). This final dataset
629 was used to generate the graphs in Fig. 5E, 5F and Supplementary Fig. 5D-G.

630

631 Statistical analyses

632 Preprint counts were compared across categories (e.g., COVID-19 or non-COVID-19) using Chi-square
633 tests or, in cases where any expected values were < 5, with Fisher’s exact tests using Monte Carlo
634 simulation. Quantitative preprint metrics (e.g. word count, comment count) were compared across
635 categories using Mann-Whitney tests and correlated with other quantitative metrics using Spearman’s
636 rank tests for univariate comparisons.

637 For time-variant metrics (e.g. views, downloads, which may be expected to vary with length of preprint
638 availability), we analysed the difference between COVID-19 and non-COVID-19 preprints using
639 generalised linear regression models with calendar days since Jan 1st 2020 as an additional covariate
640 and negative binomially-distributed errors. This allowed estimates of time-adjusted odds ratios
641 comparing COVID-19 and non-COVID-19 preprint metrics. Negative binomial regressions were
642 constructed using the function 'glm.nb' in R package MASS [60]. For multivariate categorical
643 comparisons of preprint metrics (e.g. screening time between preprint type and preprint server), we
644 constructed two-way factorial ANOVAs, testing for interactions between both category variables in all
645 cases. Pairwise post-hoc comparisons of interest were tested using Tukey's honest significant
646 difference (HSD) while correcting for multiple testing, using function 'glht' in R package multcomp
647 [61].

648

649 [Parameters and limitations of this study](#)

650 We acknowledge a number of limitations in our study. Firstly, to assign a preprint as COVID-19 or not,
651 we used keyword matching to titles/abstracts on the preprint version at the time of our data
652 extraction. This means we may have captured some early preprints, posted before the pandemic, that
653 had been subtly revised to include a keyword relating to COVID-19. Our data collection period was a
654 tightly defined window (January-April 2020) which may impact upon the altmetric and usage data we
655 collect as those preprints posted at the end of April would have had less time to accrue these metrics.
656 In addition, our data discussing preprint-paper differences suffers from survivorship and selection bias
657 in that we could only examine preprints that have been published and our findings may not be
658 generalisable to all preprints. A larger, more comprehensive sample would be necessary for more
659 conclusive statements to be made.

660

661 [Acknowledgements](#)

662 The authors would like to thank Ted Roeder, John Inglis and Richard Sever from bioRxiv and medRxiv
663 for providing information relating to comments on COVID-19 preprints. We would also like to thank
664 Martyn Rittman (preprints.org), Shirley Decker-Lucke (SSRN) and Michele Avissar-Whiting (Research
665 Square) for kindly providing usage data. Further thanks to Helena Brown and Sarah Bunn for
666 conversations regarding media usage and government policy.

667

668 Author contributions

669 Conceptualisation, N.F., L.B., G.D., J.K.P., M.P., J.A.C.; Methodology, N.F., L.B., G.D., J.A.C.; Software,
670 N.F., L.B.; Validation, N.F., L.B., J.A.C.; Formal analysis, N.F., L.B., G.D., J.A.C.; Investigation, N.F., L.B.,
671 G.D., J.K.P., M.P., J.A.C.; Resources, J.P. and J.A.C.; Data curation, N.F., L.B., G.D., J.A.C.; Writing –
672 original draft, N.F., L.B., G.D., J.K.P., M.P., J.A.C.; Writing – Review & editing, N.F., L.B., G.D., J.K.P.,
673 M.P., J.A.C.; Visualisation, N.F., L.B., G.D., J.A.C.; Supervision, J.A.C.; Project administration, J.A.C.
674

675 Data availability

676 All data and code used in this study are available on github ([https://github.com/preprinting-a-](https://github.com/preprinting-a-pandemic/pandemic_preprints)
677 [pandemic/pandemic_preprints](https://github.com/preprinting-a-pandemic/pandemic_preprints)), with the exception of data provided by preprint servers, publishers
678 and raw tweet data. Data not publicly available may be shared following permission from the
679 relevant provider and upon request to the corresponding author.

680

681 Declaration of interests

682 JP is the executive director of ASAPbio, a non-profit organization promoting the productive use of
683 preprints in the life sciences. GD is a bioRxiv Affiliate, part of a volunteer group of scientists that screen
684 preprints deposited on the bioRxiv server. MP is the community manager for preLights, a non-profit
685 preprint highlighting service. GD and JAC are contributors to preLights. The authors declare no other
686 competing interests.

687

688 References

- 689 1. WHO. COVID-19 situation report 19. 8 Feb 2020 [cited 13 May 2020]. Available:
690 [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200501-covid-19-](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200501-covid-19-sitrep.pdf)
691 [sitrep.pdf](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200501-covid-19-sitrep.pdf)
- 692 2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with
693 Pneumonia in China, 2019. *N Engl J Med.* 2020;382: 727–733. doi:10.1056/NEJMoa2001017
- 694 3. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The
695 species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and
696 naming it SARS-CoV-2. *Nat Microbiol.* 2020;5: 536–544. doi:10.1038/s41564-020-0695-z
- 697 4. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A Novel Coronavirus
698 Associated with Severe Acute Respiratory Syndrome. In:
699 <http://dx.doi.org/10.1056/NEJMoa030781> [Internet]. Massachusetts Medical Society; 7 Oct
700 2009 [cited 13 May 2020]. doi:10.1056/NEJMoa030781

- 701 5. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a
702 Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367: 1814–
703 1820. doi:10.1056/NEJMoa1211721
- 704 6. Wallach JD, Egilman AC, Gopal AD, Swami N, Krumholz HM, Ross JS. Biomedical journal speed
705 and efficiency: a cross-sectional pilot survey of author experiences. *Res Integr Peer Rev*.
706 2018;3: 1. doi:10.1186/s41073-017-0045-8
- 707 7. Abdill RJ, Blekhman R. Tracking the popularity and outcomes of all bioRxiv preprints. Pewsey E,
708 Rodgers P, Greene CS, editors. *eLife*. 2019;8: e45133. doi:10.7554/eLife.45133
- 709 8. Powell K. Does it take too long to publish research? *Nat News*. 2016;530: 148.
710 doi:10.1038/530148a
- 711 9. Vale RD. Accelerating scientific publication in biology. *Proc Natl Acad Sci*. 2015;112: 13439–
712 13446. doi:10.1073/pnas.1511912112
- 713 10. Evans JA, Reimer J. Open Access and Global Participation in Science. *Science*. 2009;323: 1025–
714 1025. doi:10.1126/science.1154562
- 715 11. Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, et al. The state of OA: a
716 large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*. 2018;6: e4375.
717 doi:10.7717/peerj.4375
- 718 12. Sever R, Roeder T, Hindle S, Sussman L, Black K-J, Argentine J, et al. bioRxiv: the preprint server
719 for biology. *bioRxiv*. 2019; 833400. doi:10.1101/833400
- 720 13. Cobb M. The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLOS*
721 *Biol*. 2017;15: e2003995. doi:10.1371/journal.pbio.2003995
- 722 14. Ginsparg P. It was twenty years ago today ... ArXiv11082700 Astro-Ph Physicscond-Mat
723 Physicsgr-Qc Physicshep-Ph Physicshep-Th Physicsphysics Physicsquant-Ph. 2011 [cited 13 May
724 2020]. Available: <http://arxiv.org/abs/1108.2700>
- 725 15. Rawlinson C, Bloom T. New preprint server for medical research. *BMJ*. 2019;365.
726 doi:10.1136/bmj.l2301
- 727 16. Tabor E. Prepublication culture in clinical research. *The Lancet*. 2016;387: 750.
728 doi:10.1016/S0140-6736(16)00330-5
- 729 17. Kaiser J, 2014, Am 12:00. BioRxiv at 1 year: A promising start. In: *Science | AAAS [Internet]*. 11
730 Nov 2014 [cited 13 May 2020]. Available: [https://www.sciencemag.org/news/2014/11/biorxiv-](https://www.sciencemag.org/news/2014/11/biorxiv-1-year-promising-start)
731 [1-year-promising-start](https://www.sciencemag.org/news/2014/11/biorxiv-1-year-promising-start)
- 732 18. Wellcome Trust. Sharing data during Zika and other global health emergencies | Wellcome. In:
733 *Wellcome.ac.uk [Internet]*. 10 Feb 2016 [cited 13 May 2020]. Available:
734 <https://wellcome.ac.uk/news/sharing-data-during-zika-and-other-global-health-emergencies>
- 735 19. Johansson MA, Reich NG, Meyers LA, Lipsitch M. Preprints: An underutilized mechanism to
736 accelerate outbreak science. *PLOS Med*. 2018;15: e1002549.
737 doi:10.1371/journal.pmed.1002549

- 738 20. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March
739 2020. 11 Mar 2020 [cited 13 May 2020]. Available:
740 [https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)
741 [media-briefing-on-covid-19---11-march-2020](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)
- 742 21. Cheng C, Barcelo J, Hartnett A, Kubinec R, Messerschmidt L. CoronaNet: A Dyadic Dataset of
743 Government Responses to the COVID-19 Pandemic. 2020 [cited 27 Apr 2020].
744 doi:10.31235/osf.io/dkvxy
- 745 22. Fraser N, Kramer B. covid19_preprints. 2020. doi:10.6084/m9.figshare.12033672.v14
- 746 23. Wellcome Trust. Coronavirus (COVID-19): sharing research data | Wellcome. 31 Jan 2020 [cited
747 21 May 2020]. Available: <https://wellcome.ac.uk/coronavirus-covid-19/open-data>
- 748 24. Nature Editorial. Calling all coronavirus researchers: keep sharing, stay open. *Nature*. 2020;578:
749 7–7. doi:10.1038/d41586-020-00307-x
- 750 25. Wellcome Trust. Publishers make coronavirus (COVID-19) content freely available and reusable
751 | Wellcome. 16 Mar 2020 [cited 21 May 2020]. Available: [https://wellcome.ac.uk/press-](https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable)
752 [release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable](https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable)
- 753 26. Eisen MB, Akhmanova A, Behrens TE, Weigel D. Publishing in the time of COVID-19. *eLife*.
754 2020;9: e57162. doi:10.7554/eLife.57162
- 755 27. OASPA. COVID-19 Publishers Open Letter of Intent - Rapid Review. In: OASPA [Internet]. 27
756 May 2020 [cited 13 May 2020]. Available: [https://oaspa.org/covid-19-publishers-open-letter-](https://oaspa.org/covid-19-publishers-open-letter-of-intent-rapid-review/)
757 [of-intent-rapid-review/](https://oaspa.org/covid-19-publishers-open-letter-of-intent-rapid-review/)
- 758 28. Himmelstein D. The licensing of bioRxiv preprints. Satoshi Village. 2016 [cited 19 May 2020].
759 Available: <https://blog.dhimmel.com/biorxiv-licenses/>
- 760 29. ASAPbio. [asapbio/licensing](https://github.com/asapbio/licensing). ASAPbio; 2018. Available: <https://github.com/asapbio/licensing>
- 761 30. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19 Antibody
762 Seroprevalence in Santa Clara County, California. *medRxiv*. 2020; 2020.04.14.20062463.
763 doi:10.1101/2020.04.14.20062463
- 764 31. Doremalen N van, Bushmaker T, Morris D, Holbrook M, Gamble A, Williamson B, et al. Aerosol
765 and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1. *medRxiv*. 2020;
766 2020.03.09.20033217. doi:10.1101/2020.03.09.20033217
- 767 32. Magagnoli J, Narendran S, Pereira F, Cummings T, Hardin JW, Sutton SS, et al. Outcomes of
768 hydroxychloroquine usage in United States veterans hospitalized with Covid-19. *medRxiv*.
769 2020; 2020.04.16.20065920. doi:10.1101/2020.04.16.20065920
- 770 33. Gog JR. How you can help with COVID-19 modelling. *Nat Rev Phys*. 2020; 1–2.
771 doi:10.1038/s42254-020-0175-7
- 772 34. Vabret N, Samstein R, Fernandez N, Merad M. Advancing scientific knowledge in times of
773 pandemics. *Nat Rev Immunol*. 2020; 1–1. doi:10.1038/s41577-020-0319-0
- 774 35. Oransky I, Markus A. Quick retraction of coronavirus paper was good moment for science. In:
775 *STAT* [Internet]. 3 Feb 2020 [cited 18 May 2020]. Available:

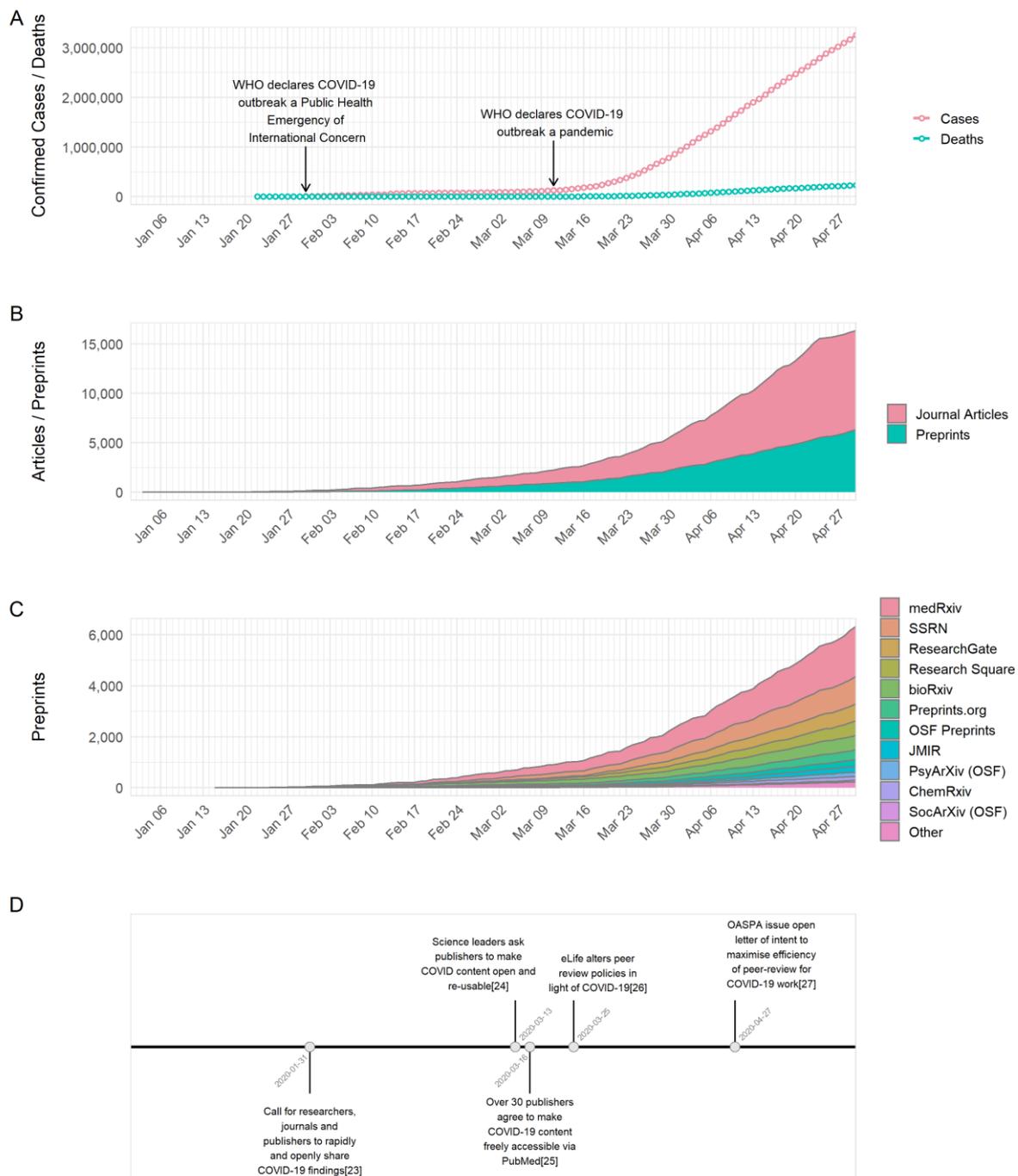
- 776 [https://www.statnews.com/2020/02/03/retraction-faulty-coronavirus-paper-good-moment-](https://www.statnews.com/2020/02/03/retraction-faulty-coronavirus-paper-good-moment-for-science/)
777 [for-science/](https://www.statnews.com/2020/02/03/retraction-faulty-coronavirus-paper-good-moment-for-science/)
- 778 36. Chiarelli A, Johnson R, Pinfield S, Richens E. Preprints and Scholarly Communication: An
779 Exploratory Qualitative Study of Adoption, Practices, Drivers and Barriers. *F1000Research*.
780 2019;8: 971. doi:10.12688/f1000research.19619.2
- 781 37. Fraser N, Momeni F, Mayr P, Peters I. The relationship between bioRxiv preprints, citations and
782 altmetrics. *Quant Sci Stud*. 2020; 1–21. doi:10.1162/qss_a_00043
- 783 38. Horbach SPJM. Pandemic Publishing: Medical journals drastically speed up their publication
784 process for Covid-19. *bioRxiv*. 2020; 2020.04.18.045963. doi:10.1101/2020.04.18.045963
- 785 39. Penfold NC, Polka JK. Technical and social issues influencing the adoption of preprints in the
786 life sciences. *PLOS Genet*. 2020;16: e1008565. doi:10.1371/journal.pgen.1008565
- 787 40. Inglis J. We've just put an additional, cautionary note about the use of preprints on every
788 @biorxivpreprint <https://t.co/08eSXL4dDi>. In: Twitter [Internet]. 1 Feb 2020 [cited 22 May
789 2020]. Available: <https://twitter.com/johringlis/status/1223598414493077505>
- 790 41. Mohammadi E, Thelwall M, Kwasny M, Holmes KL. Academic information on Twitter: A user
791 survey. *PLOS ONE*. 2018;13: e0197265. doi:10.1371/journal.pone.0197265
- 792 42. Robinson-Garcia N, Costas R, Isett K, Melkers J, Hicks D. The unbearable emptiness of
793 tweeting—About journal articles. *PLoS ONE*. 2017;12. doi:10.1371/journal.pone.0183551
- 794 43. Hausteijn S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: An analysis of
795 tweets and citations in the biomedical literature. *J Assoc Inf Sci Technol*. 2014;65: 656–669.
796 doi:10.1002/asi.23101
- 797 44. Mian A, Khan S. Coronavirus: the spread of misinformation. *BMC Med*. 2020;18.
798 doi:10.1186/s12916-020-01556-3
- 799 45. Lally C, Christie L. COVID-19 misinformation. *UK Parliam POST*. 2020 [cited 21 May 2020].
800 Available: <https://post.parliament.uk/analysis/covid-19-misinformation/>,
801 <https://post.parliament.uk/analysis/covid-19-misinformation/>
- 802 46. Pradhan P, Pandey AK, Mishra A, Gupta P, Tripathi PK, Menon MB, et al. Uncanny similarity of
803 unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *bioRxiv*. 2020;
804 2020.01.30.927871. doi:10.1101/2020.01.30.927871
- 805 47. Sheldon T. Preprints could promote confusion and distortion. *Nature*. 2018;559: 445–446.
- 806 48. Carneiro CFD, Queiroz VGS, Moulin TC, Carvalho CAM, Haas CB, Rayêe D, et al. Comparing
807 quality of reporting between preprints and peer-reviewed articles in the biomedical literature.
808 *bioRxiv*. 2020; 581892. doi:10.1101/581892
- 809 49. Johansson MA, Saderi D. Open peer-review platform for COVID-19 preprints. *Nature*.
810 2020;579: 29–29. doi:10.1038/d41586-020-00613-4
- 811 50. Brierley L. The role of research preprints in the academic response to the COVID-19 epidemic.
812 doi:10.22541/au.158516578.89167184

- 813 51. Markus A, Oransky I, Retraction Watch. Eye for Manipulation: A Profile of Elisabeth Bik. In: The
814 Scientist Magazine® [Internet]. 7 May 2019 [cited 21 May 2020]. Available: [https://www.the-](https://www.the-scientist.com/news-opinion/eye-for-manipulation--a-profile-of-elisabeth-bik-65839)
815 [scientist.com/news-opinion/eye-for-manipulation--a-profile-of-elisabeth-bik-65839](https://www.the-scientist.com/news-opinion/eye-for-manipulation--a-profile-of-elisabeth-bik-65839)
- 816 52. Wickham H, RStudio. rvest: Easily Harvest (Scrape) Web Pages. 2019. Available:
817 <https://CRAN.R-project.org/package=rvest>
- 818 53. Abdill RJ, Adamowicz EM, Blekhman R. International authorship and collaboration in bioRxiv
819 preprints. bioRxiv. 2020; 2020.04.25.060756. doi:10.1101/2020.04.25.060756
- 820 54. Chamberlain S, Zhu H, Jahn N, Boettiger C, Ram K. rcrossref: Client for Various “CrossRef”
821 “APIs.” 2020. Available: <https://CRAN.R-project.org/package=rcrossref>
- 822 55. Jahn N, rOpenSci TS (Tuija S reviewed the package for,
823 <https://github.com/ropensci/onboarding/issues/115>) see, rOpenSci RM (Ross M reviewed the
824 package for, <https://github.com/ropensci/onboarding/issues/115>) see. roadoi: Find Free
825 Versions of Scholarly Publications via Unpaywall. 2019. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=roadoi)
826 [project.org/package=roadoi](https://CRAN.R-project.org/package=roadoi)
- 827 56. Kearney M. rtweet: Collecting and analyzing Twitter data. J Open Source Softw. 2019;4: 1829.
828 doi:10.21105/joss.01829
- 829 57. Rinker T. sentimentr: Calculate Text Polarity Sentiment. 2019. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=sentimentr)
830 [project.org/package=sentimentr](https://CRAN.R-project.org/package=sentimentr)
- 831 58. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-Based Methods for Sentiment
832 Analysis. Comput Linguist. 2011;37: 267–307. doi:10.1162/COLI_a_00049
- 833 59. Steve Royle. Screenager: screening times at bioRxiv. In: quantixed [Internet]. 30 Mar 2020
834 [cited 22 May 2020]. Available: [https://quantixed.org/2020/03/30/screenager-screening-](https://quantixed.org/2020/03/30/screenager-screening-times-at-biorxiv/)
835 [times-at-biorxiv/](https://quantixed.org/2020/03/30/screenager-screening-times-at-biorxiv/)
- 836 60. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York: Springer-Verlag;
837 2002. doi:10.1007/978-0-387-21706-2
- 838 61. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J
839 Biom Z. 2008;50: 346–363. doi:10.1002/bimj.200810425

840

841

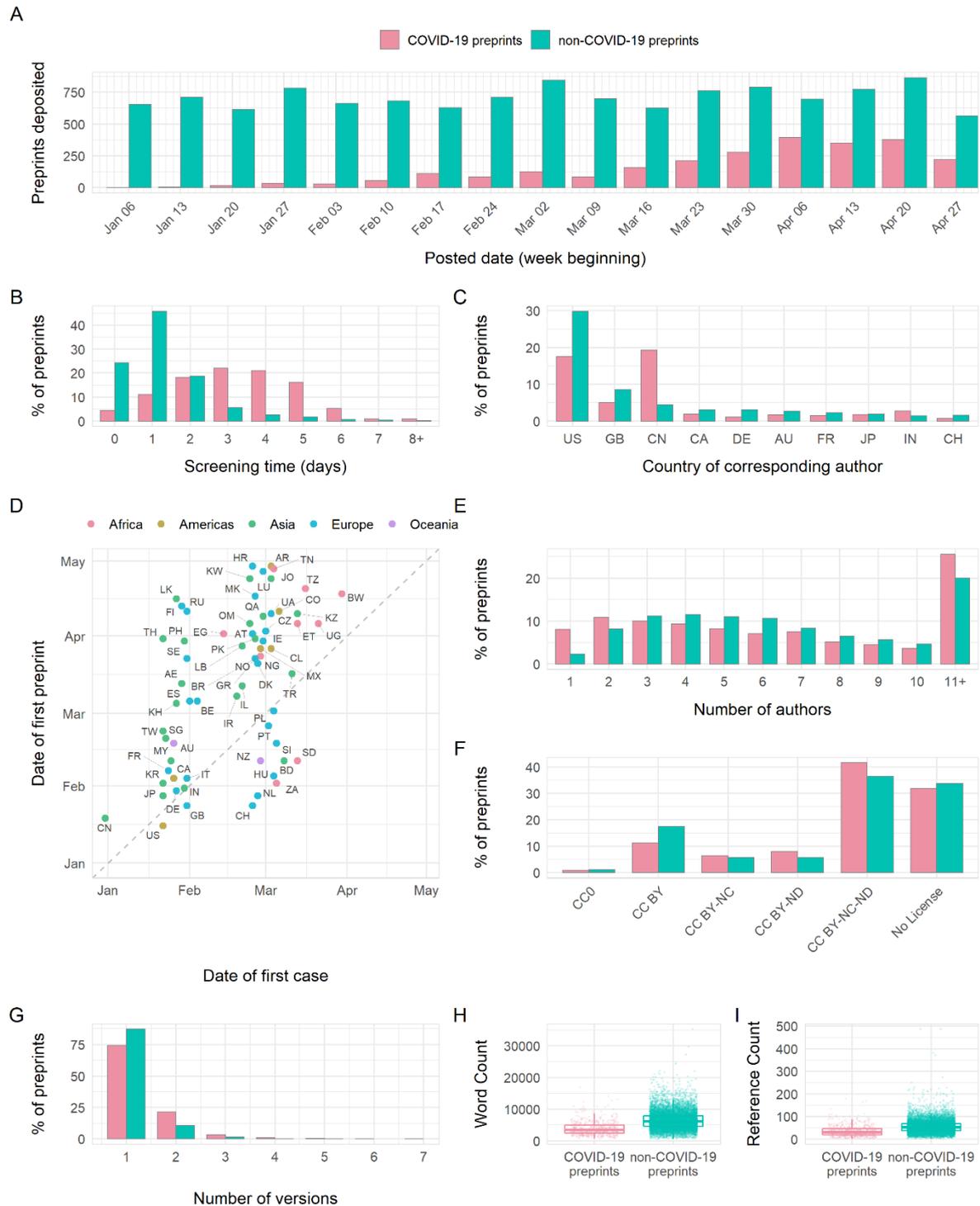
842 **Figures**



843

844 **Figure 1. Development of COVID-19 and publication response between January 2020 and April 2020.**

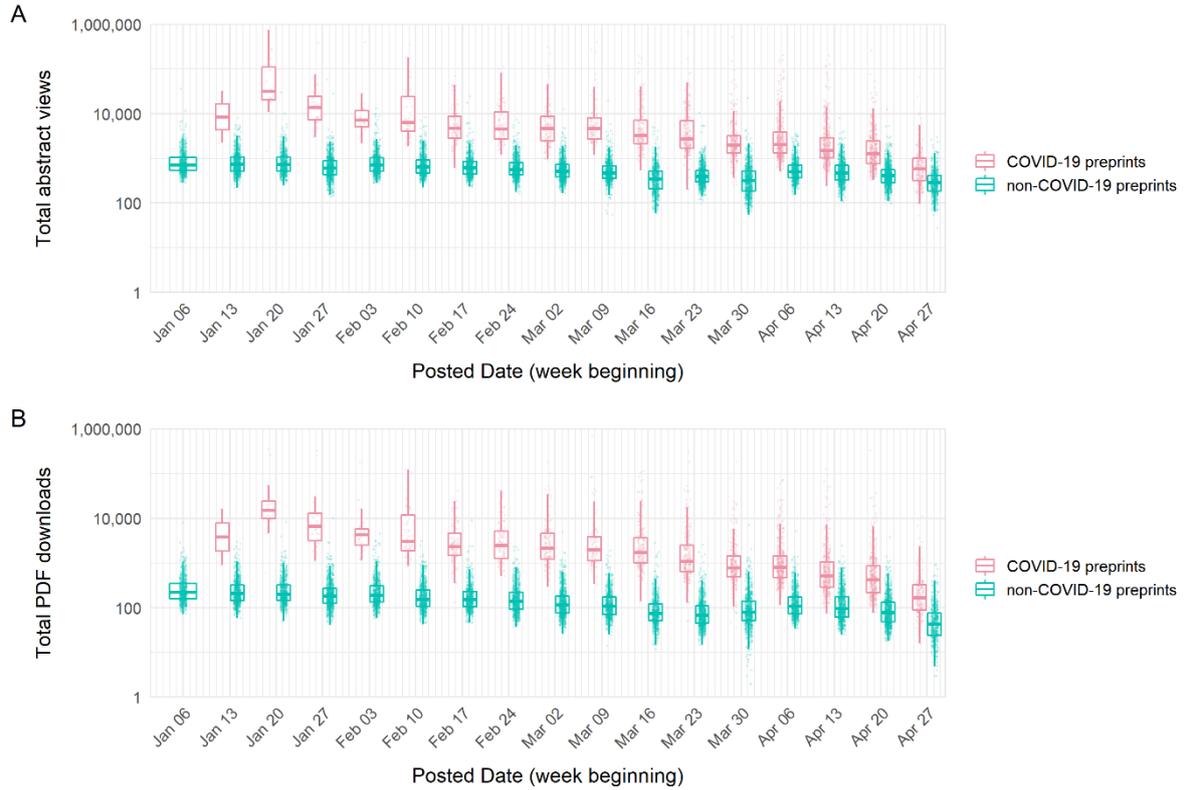
845 (A) Number of COVID-19 confirmed cases and reported deaths. Data is sourced from
 846 <https://github.com/datasets/covid-19/>, based on case and death data aggregated by the Johns
 847 Hopkins University Center for Systems Science and Engineering (<https://systems.jhu.edu/>). (B)
 848 Cumulative growth of journal articles and preprints containing COVID-19 related search terms. (C)
 849 Cumulative growth of preprints containing COVID-19 related search terms, broken down by individual
 850 preprint server. (D) Timeline representing significant changes made by traditional publishers as they
 851 adopt journal policies relating to COVID-19 research. Journal data in (B) is based upon data extracted
 852 from Dimensions (<https://www.dimensions.ai>), preprint data in (B) and (C) is based upon data
 853 gathered by Fraser and Kramer (2020; <https://doi.org/10.6084/m9.figshare.12033672>).



854

855 **Figure 2. Attributes of COVID-19 and non-COVID-19 preprints deposited on bioRxiv and medRxiv**
 856 **between January and April 2020.** (A) Number of preprints deposited per week. (B) Screening time
 857 for bioRxiv and medRxiv. (C) Percentage of preprints deposited by country of corresponding author.
 858 (D) Correlation between date of the first preprint originating from a country (according to the
 859 affiliation of the corresponding author) and the date of the first confirmed case from the same
 860 country for COVID-19 preprints. (E) Distribution of the number of authors per preprint. (F)
 861 Distribution of preprint licence chosen by the author. (G) Distribution of the number of deposited
 862 preprint versions. (H) Word counts per preprints. (I) Reference counts per preprint. Data for (H) and

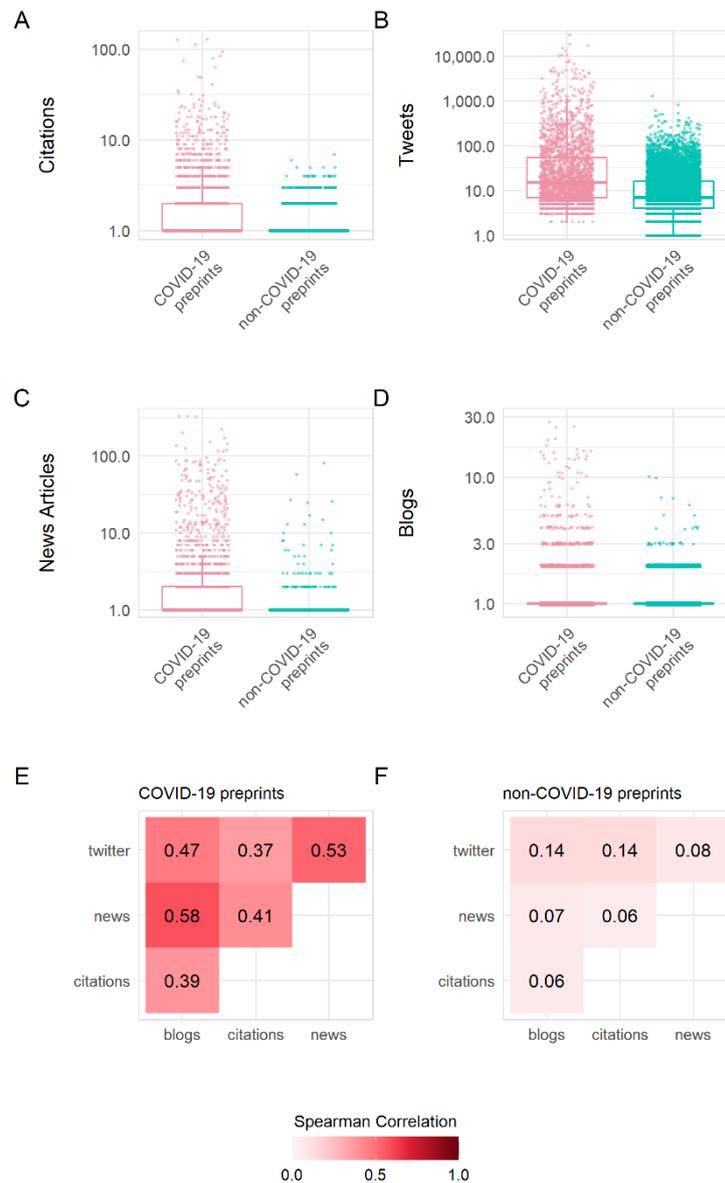
863 (I) were from bioRxiv only. Boxplot horizontal lines denote lower quartile, median, upper quartile,
864 with whiskers extending to $1.5 \times \text{IQR}$. All boxplots additionally show raw data values for individual
865 preprints with added horizontal jitter for visibility.



866

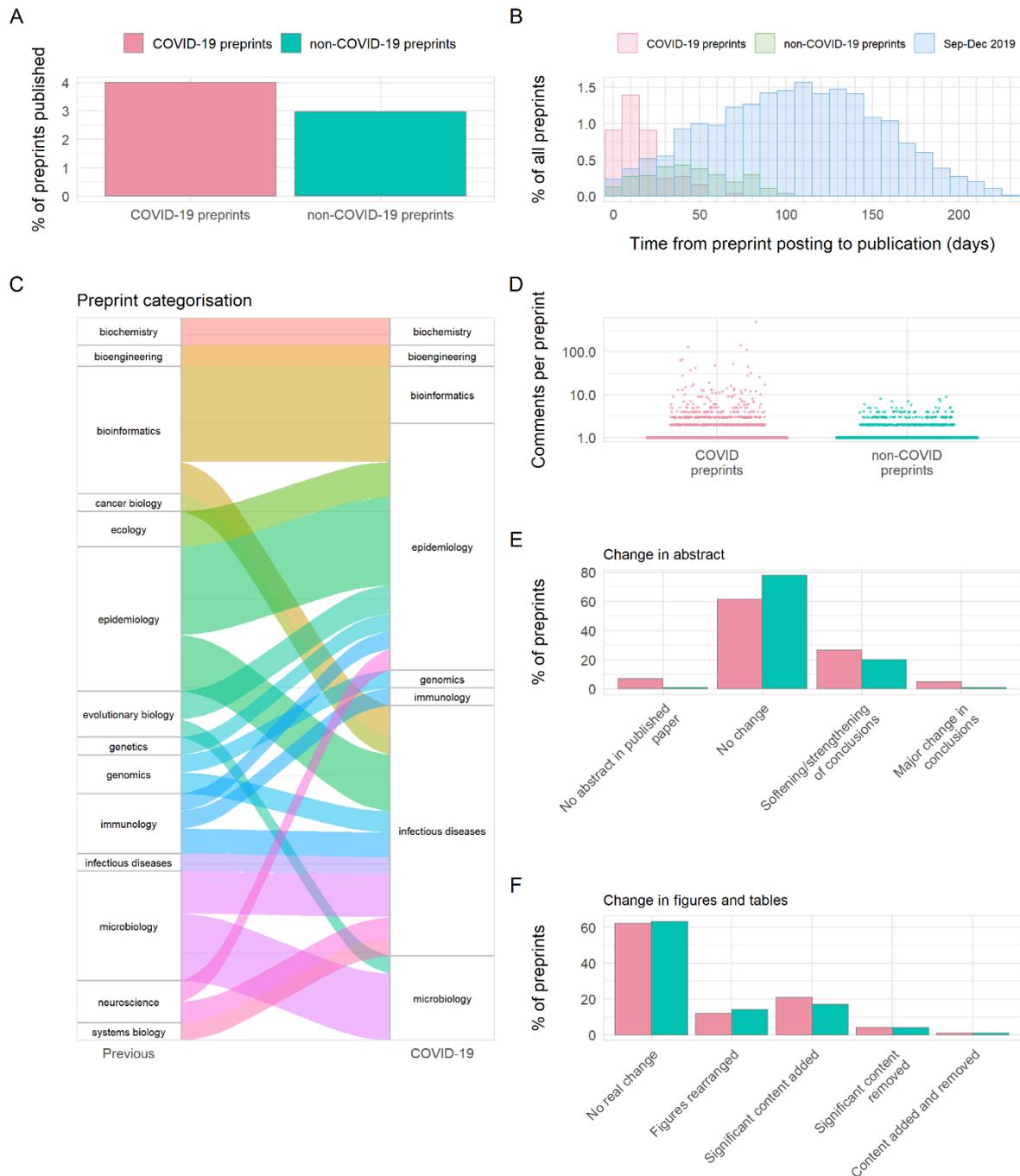
Figure 3. Distribution of access statistics for COVID-19 and non-COVID-19 preprints posted on bioRxiv and medRxiv. (A) Total abstract views. (B) Total PDF downloads.

867



868

869 **Figure 4. Comparison of citations, tweets, mentions in news articles and blogs for COVID-19 and**
 870 **non-COVID-19 preprints posted on bioRxiv and medRxiv between January and April 2020. (A)**
 871 **Citations per preprint. (B) Tweets per preprint. (C) News article mentions per preprint. (D)**
 872 **Blog mentions per preprint. (E) Spearman's correlation matrix between all indicators for COVID-19**
 873 **preprints. (F) Spearman's correlation matrix between all indicators for non-COVID-19 preprints. (A-D)**
 874 **as log scale, with +1 added for visualisation. Boxplot horizontal lines denote lower quartile, median,**
 875 **upper quartile, with whiskers extending to 1.5*IQR. All boxplots additionally show raw data values for**
 876 **individual preprints with added horizontal jitter for visibility.**



877

878 **Figure 5. Publishing and peer-review of COVID-19 preprints.** (A) percentage of COVID-19 and non-
 879 COVID-19 preprints published between Jan-April. (B) Time taken from depositing a preprint on bioRxiv
 880 or medRxiv and subsequent publication for COVID-19 preprints (red), non-COVID-19 preprints posted
 881 between January - April 2020 (green) and non-COVID-19 preprints posted between September –
 882 December 2019 (blue). (C) Change in preprint category for COVID-19 preprint authors compared to
 883 their previous preprint (COVID-19 or non-COVID-19), for combinations with $n \geq 5$ authors. (D)
 884 Numbers of comments for COVID-19 preprints and non-COVID-19 preprints, log scale. (E) Abstract
 885 changes between version 1 of a preprint and the associated published paper. (F) Qualitative overall
 886 changes (Information within main text figures and tables) between version 1 of a preprint and the
 887 associated published paper.

888

Table 1. Top 10 cited COVID-19 preprints

Rank	Source	doi	Title	Posted date	Citations
1	biorxiv	10.1101/2020.02.07.937862	Severe acute respiratory syndrome-related coronavirus - The species and its viruses, a statement of the Coronavirus Study Group	11/02/2020	127
2	medrxiv	10.1101/2020.02.06.20020974	Clinical characteristics of 2019 novel coronavirus infection in China	09/02/2020	126
3	medrxiv	10.1101/2020.01.23.20018549	Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions	24/01/2020	112
4	biorxiv	10.1101/2020.01.22.914952	Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin	23/01/2020	93
5	biorxiv	10.1101/2020.01.26.919985	Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV	26/01/2020	83
6	biorxiv	10.1101/2020.01.31.929042	The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells	31/01/2020	79
7	biorxiv	10.1101/2020.01.30.927806	The digestive system is a potential route of 2019-nCoV infection: a bioinformatics analysis based on single-cell transcriptomes	31/01/2020	74
8	medrxiv	10.1101/2020.02.10.20021675	Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China	11/02/2020	62
9	biorxiv	10.1101/2020.02.03.931766	Specific ACE2 Expression in Cholangiocytes May Cause Liver Damage After 2019-nCoV Infection	04/02/2020	49
10	medrxiv	10.1101/2020.03.03.20028423	Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts	04/03/2020	48

Table 2. Top 10 tweeted COVID-19 preprints

Rank	Source	doi	Title	Posted date	Tweets	News articles	Blogs
1	medrxiv	10.1101/2020.04.14.20062463	COVID-19 Antibody Seroprevalence in Santa Clara County, California	17/04/2020	29984	328	24
2	biorxiv	10.1101/2020.01.30.927871	Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag	31/01/2020	18587	92	17
3	medrxiv	10.1101/2020.04.04.20053058	Indoor transmission of SARS-CoV-2	07/04/2020	17494	67	9
4	medrxiv	10.1101/2020.03.22.20040758	Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial	30/03/2020	15337	117	15
5	medrxiv	10.1101/2020.03.09.20033217	Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1	10/03/2020	13407	333	27
6	biorxiv	10.1101/2020.03.13.990226	Reinfection could not occur in SARS-CoV-2 infected rhesus macaques	14/03/2020	10870	225	19
7	medrxiv	10.1101/2020.04.16.20065920	Outcomes of hydroxychloroquine usage in United States veterans hospitalized with Covid-19	21/04/2020	10512	329	15
8	medrxiv	10.1101/2020.03.30.20048165	Association of BCG vaccination policy with prevalence and mortality of COVID-19	06/04/2020	10435	3	0
9	medrxiv	10.1101/2020.03.17.20037713	A serological assay to detect SARS-CoV-2 seroconversion in humans	18/03/2020	8094	153	13
10	medrxiv	10.1101/2020.03.24.20042937	Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study	28/03/2020	7427	77	5

Table 3. Top 10 COVID-19 preprints covered by news organisations

Rank	Source	doi	Title	Posted date	Tweets	News articles	Blogs
1	medrxiv	10.1101/2020.03.09.20033217	Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1	10/03/2020	13407	333	27
2	medrxiv	10.1101/2020.04.16.20065920	Outcomes of hydroxychloroquine usage in United States veterans hospitalized with Covid-19	21/04/2020	10512	329	15
3	medrxiv	10.1101/2020.04.14.20062463	COVID-19 Antibody Seroprevalence in Santa Clara County, California	17/04/2020	29984	328	24
4	biorxiv	10.1101/2020.03.13.990226	Reinfection could not occur in SARS-CoV-2 infected rhesus macaques	14/03/2020	10870	225	19
5	biorxiv	10.1101/2020.03.30.015347	Susceptibility of ferrets, cats, dogs, and different domestic animals to SARS-coronavirus-2	31/03/2020	4399	201	24
6	medrxiv	10.1101/2020.03.23.20039446	Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center	26/03/2020	4460	172	13
7	medrxiv	10.1101/2020.03.17.20037713	A serological assay to detect SARS-CoV-2 seroconversion in humans	18/03/2020	8094	153	13
8	medrxiv	10.1101/2020.04.07.20056424	Chloroquine diphosphate in two different dosages as adjunctive therapy of hospitalized patients with severe respiratory syndrome in the context of coronavirus (SARS-CoV-2) infection: Preliminary safety results of a randomized, double-blinded, phase IIb clinical trial (CloroCovid-19 Study)	11/04/2020	4503	146	15
9	biorxiv	10.1101/2020.03.08.982637	Aerodynamic Characteristics and RNA Concentration of SARS-CoV-2 Aerosol in Wuhan Hospitals during COVID-19 Outbreak	10/03/2020	972	138	12
10	medrxiv	10.1101/2020.03.11.20031096	Relationship between the ABO Blood Group and the COVID-19 Susceptibility	16/03/2020	3963	127	13

Table 4. Top 10 commented on COVID-19 preprints

Rank	source	doi	title	posted date	comments count
1	medrxiv	10.1101/2020.04.14.20062463	COVID-19 Antibody Seroprevalence in Santa Clara County, California	17/04/2020	508
2	medrxiv	10.1101/2020.03.24.20042937	Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study	28/03/2020	141
3	biorxiv	10.1101/2020.01.30.927871	Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag	31/01/2020	127
4	medrxiv	10.1101/2020.04.16.20065920	Outcomes of hydroxychloroquine usage in United States veterans hospitalized with Covid-19	21/04/2020	114
5	medrxiv	10.1101/2020.03.11.20031096	Relationship between the ABO Blood Group and the COVID-19 Susceptibility	16/03/2020	66
6	medrxiv	10.1101/2020.03.27.20043752	Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months	30/03/2020	61
7	medrxiv	10.1101/2020.03.22.20040758	Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial	30/03/2020	53
8	medrxiv	10.1101/2020.04.05.20054361	Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters	08/04/2020	47
9	biorxiv	10.1101/2020.01.26.919985	Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov	26/01/2020	44
10	medrxiv	10.1101/2020.03.09.20033217	Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1	10/03/2020	41