Publishing of COVID-19 preprints in peer-reviewed journals, preprinting trends, public discussion and quality issues

Kodvanj, Ivan; Homolak, Jan; Virag, Davor; Trkulja, Vladimir

Repository / Repozitorij:

Dr Med - University of Zagreb School of Medicine Digital Repository

Check for updates

Publishing of COVID-19 preprints in peer-reviewed journals, preprinting trends, public discussion and quality issues

Ivan Kodvanj[1] · Jan Homolak[1] · Davor Virag[1] · Vladimir Trkulja[1]

Abstract

COVID-19-related (vs. non-related) articles appear to be more expeditiously processed and published in peer-reviewed journals. We aimed to evaluate: (i) whether COVID-19-related preprints were favored for publication, (ii) preprinting trends and public discussion of the preprints, and (iii) the relationship between the publication topic (COVID-19-related or not) and quality issues. Manuscripts deposited at bioRxiv and medRxiv between January 1 and September 27 2020 were assessed for the probability of publishing in peer-reviewed journals, and those published were evaluated for submission-to-acceptance time. The extent of public discussion was assessed based on Altmetric and Disqus data. The Retraction Watch Database and PubMed were used to explore the retraction of COVID-19 and non-COVID-19 articles and preprints. With adjustment for the preprinting server and number of deposited versions, COVID-19-related preprints were more likely to be published within 120 days since the deposition of the first version (OR = 1.96, 95% CI: 1.80–2.14) as well as over the entire observed period (OR = 1.39, 95% CI: 1.31–1.48). Submission-to-acceptance was by 35.85 days (95% CI: 32.25–39.45) shorter for COVID-19 articles. Public discussion of preprints was modest and COVID-19 articles were overrepresented in the pool of retracted articles in 2020. Current data suggest a preference for publication of COVID-19-related preprints over the observed period.

Keywords Preprint · COVID19 · Peer-review · Publishing

Introduction

The Coronavirus disease 2019 (COVID-19) pandemic prompted a need for rapid communication and discussion of scientific information resulting in an unprecedented surge in the number of publications on various aspects of the disease. Be it due to elementary greed or a praiseworthy desire to disseminate information, publishers have greatly shortened the peer-review process for COVID-19-related manuscripts (Homolak et al., 2020; Horbach, 2020; Kun, 2020). Although such a practice might have improved the speed of information

✉ Vladimir Trkulja
vladimir.trkulja@mef.hr

1   Department of Pharmacology, School of Medicine, University of Zagreb, Šalata 11, 10000 Zagreb, Croatia

exchange, it is not necessarily beneficial—concerns about the quality of articles published under such circumstances have been raised (Homolak et al., 2020). As an alternative or as a supplement to publishing in peer-reviewed journals, preprinting is clearly an additional option for rapid scientific communication. This practice of presenting scientific/scholarly papers to the scientific community before they are published in peer-reviewed journals dates back to the seventeenth century (Cobb, 2017). Modern attempts to create platforms for sharing unpublished works were at first heavily criticized, but nowadays preprinting is a standard practice in many scientific fields (Berg et al., 2016; Four Years of Information Exchange, 1966; Pasternack, 1966; Woodruff, 1966). In biomedicine, preprinting had been considered relatively uncommon, but increasing trends have been observed during the COVID-19 pandemic (Fraser et al., 2021; Fu & Hughey, 2019). Concerns have been expressed that preprinted manuscripts are generally of lower quality than journal-published articles due to the lack of peer-review (although, evidence of the effectiveness of the peer-review process is scarce) (Carneiro et al., 2020; Jefferson et al., 2002; Nabavi Nouri et al., 2020; Smith, 2006; Vercellini et al., 2016). On the other hand, lack of peer-review was at some point considered beneficial as it was thought that this way worthwhile ideas might be shared, that otherwise could have been unjustly disregarded and overlooked in the editorial process (Green, 1964). The fact stands that preprinting, in addition to a possibility of rapid research communication, also provides an opportunity to peers to comment and discuss, and an opportunity to authors to improve their work before submitting it to peer-reviewed journals.
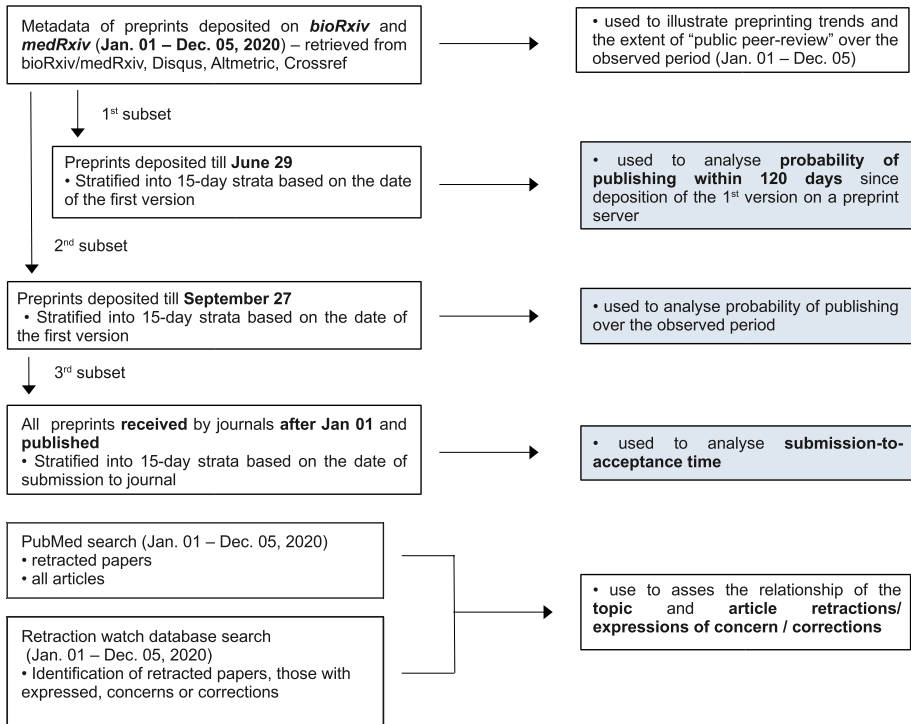
Within such a constellation of coinciding, but apparently unrelated topics, i.e., overflow of publications on various aspects of COVID-19 disease and suggested (Fraser et al., 2021; Fu & Hughey, 2019) concurrent changes in preprinting of biomedical manuscripts, we considered that three questions deserved to be addressed: (i) are COVID-19-related manuscripts preferred for journal publishing over the non-COVID-19-related manuscripts? In essence, the answer to this question comes to estimation of probability of publishing for these two types of manuscripts. In an attempt to do so, a proper denominator would need to be the total number of generated/written manuscripts of either kind. However, since such information is unavailable, the number of preprinted (where the fact of preprinting serves only as an evidence of existence) works seems to be a reasonable proxy; (ii) have the circumstances of the COVID-19 pandemic reflected on preprinting trends regarding biomedical research, and to what extent has the opportunity (provided by the fact of preprinting) to discuss the preprinted research been seized (i.e., what is the extent of "pre-submission public peer-review")?; (iii) is the concern that published [under the circumstances of shortened peer-review process (Homolak et al., 2020; Horbach, 2020; Kun, 2020)] COVID-19-related papers are more commonly flawed than their non-COVID counterparts justified?

We undertook the present study in an attempt to provide reasonable answers to these questions.

Materials and methods

Study outline and study outcomes

Types of datasets used and their purpose in the present study are outlined in Fig. 1. We defined three study objectives based on their relevance (objective 1—most relevant, objective 3—least relevant) and the anticipated level of susceptibility to bias/confounding

Metadata of preprints deposited on **bioRxiv** and **medRxiv** (**Jan. 01 – Dec. 05, 2020**) – retrieved from bioRxiv/medRxiv, Disqus, Altmetric, Crossref

→ • used to illustrate preprinting trends and the extent of "public peer-review" over the observed period (Jan. 01 – Dec. 05)

1st subset

Preprints deposited till **June 29**
• Stratified into 15-day strata based on the date of the first version

→ • used to analyse **probability of publishing within 120 days** since deposition of the 1st version on a preprint server

2nd subset

Preprints deposited till **September 27**
• Stratified into 15-day strata based on the date of the first version

→ • used to analyse probability of publishing over the observed period

3rd subset

All preprints **received** by journals **after Jan 01** and **published**
• Stratified into 15-day strata based on the date of submission to journal

→ • used to analyse **submission-to-acceptance time**

PubMed search (Jan. 01 – Dec. 05, 2020)
• retracted papers
• all articles

Retraction watch database search (Jan. 01 – Dec. 05, 2020)
• Identification of retracted papers, those with expressed, concerns or corrections

→ • use to asses the relationship of the **topic** and **article retractions/ expressions of concern / corrections**

Fig. 1 Outline of datasets used in the present analysis and their purpose. Outcomes used to achieve the primary study objective are shaded

(objective 1—least susceptible, objective 3—most susceptible), given the nature of the study (observational) and the type of data (metadata). The 1st objective was to investigate whether COVID-19-related preprints were favored (over non-related) for publication in peer-reviewed journals. Preprints were used as proxies of (hypothetical) "all generated manuscripts on the topics" since they were the only ones whose existence could be clearly verified and for which subsequent developments could be prospectively evaluated. We considered that the probability of publishing was the most informative outcome in this respect and that further insight would be obtained from analysis of submission-to-acceptance time. Therefore, we defined three outcomes of interest regarding the 1st study objective: (i) *The primary outcome* was defined as probability of publishing within 120 days since the deposition of the first preprint version. The time constraint was imposed to reduce the risk of bias arising from unequal "time-at-risk" (for publishing); hence,the analysis was restricted to preprints deposited till June 29 to allow all preprints to have 120 days "available" to be published between the deposition date and the date of data collection/analysis. Preprints that remained unpublished by day 120 were censored. We considered 120 days to be a reasonable period of time for the submission and review process to take place; (ii) *secondary outcome* was probability of publishing over the entire observed period (i.e., up to November 01, 2020) and it was assessed in a subset of preprints deposited before September 27, 2020. Manuscripts that remained unpublished by November 01, 2020, were censored. We considered this outcome to be complementary to the primary outcome, which

was susceptible to bias arising from a possibility that not all preprinted manuscripts were actually submitted to journals within the same/similar timeframe, i.e., that some might have been purposely left in a preprint form over a longer period of time. To further reduce bias/confounding arising from unequal "time-at-risk", preprints in both datasets were stratified into 15-day strata regarding the date of the first preprinted version. Due to the limited number of COVID-19 related preprints, the first two strata (Jan. 01–15 and Jan. 16–30) were merged into one stratum; (iii) *tertiary outcome* was submission-to-acceptance time, considered a proxy of the length of the peer-review process, and was assessed in a subset of preprints that were published in peer-reviewed journals during the observed period and submitted to journals after January 01, 2020. Here, we anticipated potential confounding arising from varying interest in COVID-19-related and non-related topics over time, so manuscripts were stratified into 15-day strata regarding the journal submission dates (Fig. 1). The 2nd objective was to illustrate preprinting trends of COVID-19-related and non-related manuscripts on bioRxiv and medRxiv, their usage statistics and to estimate the extent of the public peer-review (i.e., pre-submission peer-review) using the number of posted comments and Altmetric data as proxies. This was assessed using all preprints deposited at the two platforms between January 01 and December 05, 2020 (Fig. 1). The 3rd objective was to evaluate a possible association between the publication topic (COVID-19-related or non-related) and quality issues related to the published papers using notifications on retraction or issuance of concerns or corrections as proxies. This was assessed using data on all published papers indexed in PubMed and all notifications issued in the Retraction Watch Database (Retraction Watch Database, n.d.) between January 01 and December 05, 2020 (Fig. 1).

Data collection

All data retrieval and management were done in *R* (version 4.0.2) (R Core Team, 2020). *Rbiorxiv* and *medrxivr* packages were used to access bioRxiv and medRxiv application programming interfaces (APIs) and to collect metadata and usage statistics on the preprinted manuscripts (January 01–December 05, 2020). COVID-19 preprints were identified using the search terms COVID-19 OR SARS-CoV-2 OR Coronavirus disease 19 OR 2019-nCoV. Furthermore, CORD-19 dataset was used to identify additional COVID-19 related preprints. All other preprints were classified as non-COVID-19 preprints. Publication status was retrieved from bioRxiv and medRxiv services. Publication dates of articles with bioRxiv preprints were provided by the bioRxiv API. *Rcrossref* was used to gather publication dates for journal-published preprints, initially deposited on the *medRxiv*. The *DisqusR* and *rAltmetric* packages were used to access the Disqus and Altmetric API and identify the number of comments and retrieve Altmetric data for each preprint deposited on bioRxiv and medRxiv servers. Submission and acceptance dates for published preprints were retrieved from PubMed with the *RISmed* package. The Retraction Watch Database was used to retrieve the number of COVID-19-related articles with issued retraction notice, expression of concern, or correction during 2020 (till December 05, 2020). Controls to COVID-19-related manuscripts in this analysis were manuscripts pertaining to four different viruses and their associated diseases: human immunodeficiency virus, hepatitis virus (any), herpes virus (any) and influenza virus. Four search phrases were constructed to retrieve the number of retraction notices, expression of concerns or corrections pertaining to these topics; two search phrases were used to retrieve the numbers pertaining to two related topics (immunology and epidemiology of viral infectious diseases); one search

phrase was used to retrieve the number of retraction notices, expressions of concerns or corrections issued for all COVID-19-unrelated articles (obtained as a difference between the total number of retrieved items and the number retrieved for COVID-19-related papers).

The exact search methodology is depicted under Table 3. All numbers are expressed as a proportion of the total number of articles indexed by PubMed. To minimize the "time-at-risk" bias, only articles published during 2020 were compared. However, some risk of bias still remains due to the possibility of different publishing rates of articles during 2020. To avoid bias due to possible over- or underrepresentation of the COVID-19 articles in PubMed, in comparison with the Retraction Watch Database, we repeated the analysis using only data provided by the PubMed database. The search was conducted using the search term (Retracted Publication[PT]) AND ("2020/01/01"[Date—Publication]: "2020/12/05"[Date—Publication]), with and without "AND (COVID-19 OR SARS-CoV-2 OR "Coronavirus disease 19" OR 2019-nCoV)", to identify the number of retractions of COVID-19 and non-COVID-19 articles during 2020.

Data analysis

All data visualization and analysis were performed in *R* (version 4.0.2) (R Core Team, 2020). Data on preprinting trends over time, preprint usage statistics, Altmetric data and Disqus comments on the manuscripts preprinted during the observed period, and Retraction Watch Database data on the published papers, were summarized by the preprinting platform and topic (COVID-19-related or not).

The probability of publishing within 120 days since the 1st preprint version and the probability of publishing over the entire observed period were analyzed by fitting stratified (in respect to preprint date) logistic regression (package *survival*, function clogit). Submission-to-acceptance time was analyzed by fitting a hierarchical (mixed) model with the submission date stratum as a random effect. Articles with submission dates before 2020 were excluded from the dataset that was used for the analysis submission-to-acceptance time. Fixed effects in all analyses were topic (COVID-19-related or non-related), preprinting platform (bioRxiv or medRxiv), and the number of preprinted versions (dichotomized as one or ≥ 2). The latter adjustment was introduced to account for potential bias arising from different intentions of the preprinting authors. For example, a preprint might have been submitted to a journal at the time of preprinting, or it might have been a work in progress with several versions and purposely kept (only) as a preprint over a longer period of time; more preprinted versions might have improved the quality of the final submitted version, hence peer-review process might have been shorter.

We conducted a supplemental analysis of preprinting/publishing trends and other aspects of bioRxiv and medRxiv preprints over a longer period of time that we found informative for discussion of the main study results (see Supplemental Material, Supplemental methods).

Submission-to-acceptance dataset validation

After all articles with submission dates before January 01, 2020, had been removed, we identified and excluded one article with the submission-to-acceptance time of—1 days (an obvious mistake). To validate the dataset, submission and acceptance dates were checked for 597 (10% of all) randomly selected articles. For 12 articles we could not find the submission and acceptance dates on the journal website, nor published pdf of the article. Five
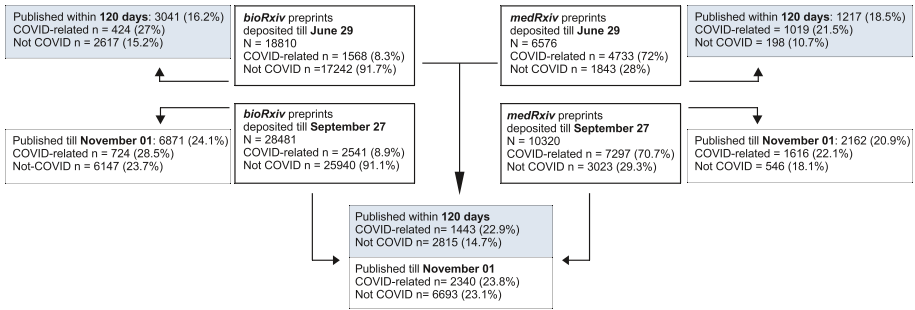
Fig. 2 Structure of preprints deposited by June 29 and by September 27, 2020, used to evaluate the probability of publishing by platform-by topic-by publishing outcome. Shading marks the primary outcome related to the 1st study objective

Table 1 Summary of the analysis of the probability of publishing within 120 days since the first preprint version (a subset of preprints deposited till June 29) and time-to-publishing considering the entire observed period (a subset of preprints deposited till September 27)

Predictors	Published within 120 days		Published before November 01	
	OR (95% CI)	p	OR (95% CI)	p
COVID-19 vs Not COVID-19	1.96 (1.80–2.14)	< 0.001	1.39 (1.31–1.48)	< 0.001
bioRxiv vs medRxiv	1.32 (1.21–1.44)	< 0.001	1.24 (1.17–1.32)	< 0.001
≥ 2 vs 1 version	0.72 (0.67–0.78)	< 0.001	0.90 (0.85–0.94)	< 0.001

Effects are expressed as odds ratios (OR). Stratified (by preprinting date of the first version) logistic regression was fitted to the probability of being published within 120 days and over the entire observed period

articles with erroneous date values were identified. We considered that the number and size of errors were acceptable and not likely to affect the conclusions of the study.

Results

Are COVID-19-related preprints favored for publishing in peer-reviewed journals?

The subset of preprints deposited till June 29, 2020, and used to evaluate the probability of publishing within 120 days since the 1st preprint version comprised a total of 18,810 preprints on bioRxiv (8.3% COVID-19-related and 91.7% non-related) and a total of 6576 preprints on medRxiv (72.0% COVID-19-related and 28.0% non-related) (Fig. 2). The subset of preprints deposited till September 27, 2020, and used to evaluate the probability of publishing over the entire observed period (up to November 01, 2020) comprised 28,481 preprints on bioRxiv (8.9% COVID-19-related and 91.1% non-related) and 10,320 preprints on medRxiv (70.7% COVID-19-related and 29.3% non-related) (Fig. 2). Raw proportions of published papers by the topic and preprinting platform are depicted in Fig. 2. In multivariate analyses (Table 1), COVID-19-related preprints were associated with higher odds of publishing within 120 days than non-COVID-19 preprints (OR = 1.96, 95% CI: 1.80–2.14), and with a higher probability of publishing over the observed period

than non-COVID-19 preprints (OR = 1.39, 95% CI: 1.31–1.48). The probability for both outcomes was higher for preprints deposited on bioRxiv and lower for preprints with ≥ 2 versions than for those with only one version (Table 1). Journal submission-to-acceptance time was identified for preprints published before November 01, 2020. In multivariate analysis, the COVID-19-related topic was associated with approximately 36 days shorter submission-to-acceptance time (mean difference − 35.85, 95% CI: − 39.45 to − 32.25) (Table 2). The preprinting platform and number of preprint versions did not appear associated with the outcome (Table 2).

Preprinting trends, usage statistics, and indicators of public pre-submission peer-review

For the observed period (January 01–December 05, 2020) we identified a total of 13,257 preprints newly deposited at medRxiv [9228 (69.6%) COVID-19-related)], and 36,267 preprints deposited at bioRxiv [3305 (9.11%) COVID-19-related]. There was a clear increasing trend in the number of newly deposited preprints (Fig. 3A), but on medRxiv, the increase was largely due to the increasing number of COVID-19-related preprints, while the number of newly deposited preprints on bioRxiv appeared comparable for COVID-19-related and not related topics (Fig. 3A). Usage statistics of bioRxiv preprint server indicated an increase in the number of abstract views, full-text views and PDF downloads (Fig. 3B).

The overall proportion of preprints that have been commented on is rather low (5.7%), but somewhat higher for COVID-19-related preprints (Fig. 4A): 17.5 and 3.2% of the COVID-19-related and not related preprints respectively, on bioRxiv; and 12.3 and 1.4% respectively, on medRxiv. By far, the most preprints that were commented on, received only one comment (Fig. 4B).

Altmetric score, indicative of the public attention received by the preprints, appeared somewhat higher for COVID-19-related than for non-related preprints, particularly those posted on bioRxiv (vs. medRxiv) (Fig. 4C). Closer examination revealed that all preprints were mentioned on the Twitter platform (likely because of bots that tweet all deposited articles), a smaller percentage were mentioned in blog posts and news outlets, and a negligible number of preprints were mentioned on other venues (Fig. 4D). Grouped by the topic and preprint server, mentioning of the preprints on the Twitter platform closely reflected the overall Altmetric score. COVID-19-related preprints appeared more commonly shared on Facebook and blogs than non-COVID-19-related

Table 2 Analysis of *submission-to-acceptance* time for published preprints (in days)

Predictors	Estimates (95% CI)	p
COVID-19 vs Not COVID-19	− 35.85 (− 39.45 to − 32.25)	< 0.001
bioRxiv vs medRxiv	1.04 (− 2.58 to 4.67)	0.573
≥ 2 vs 1 preprint version	0.48 (− 2.48 to 3.44)	0.750

Linear mixed-effects model (submission date specified as a random effect) was fitted to the number of days between the date of receipt by the journal and date of acceptance. Submission to publication time was identified for 533 COVID-19/bioRxiv (median = 53 days), 4079 not-COVID-19/bioRxiv (median = 95 days), 1107 COVID-19/medRxiv (median = 49 days) and for 348 not-COVID-19/medRxiv preprints (median = 104 days)

A

B
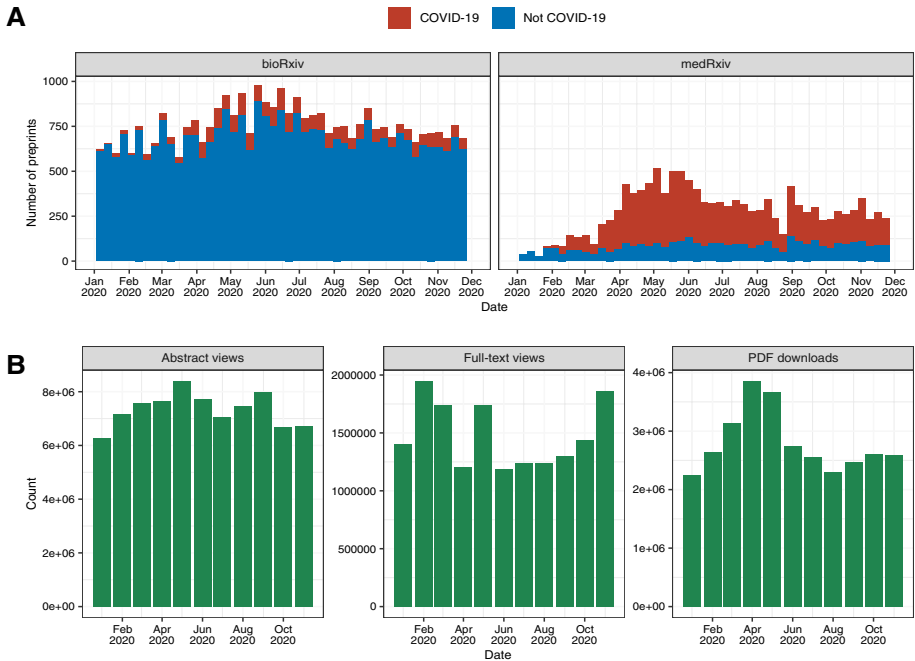
Fig. 3 Posting of preprints and bioRxiv server usage statistics. **A** The number of new preprints posted on bioRxiv and medRxiv over the observed period. **B** Monthly abstract views, full-text views, and PDF downloads on bioRxiv server (not available for medRxiv)

preprints (Fig. 4E). Preprints posted on the more clinical-oriented medRxiv were more reported in news outlets compared to those posted on bioRxiv (Fig. 4E).

Potential quality issues

A total of 345 notices related to the articles published between January 01 and December 05, 2020, were identified in the Retraction Watch database, 46 of which were related to COVID-19. Comparison of the number of issued notices for COVID-19-related articles and articles pertaining to four different viruses and viral diseases and 2 different topics is depicted in Table 3. Briefly, only articles pertaining to herpes viruses had a higher approximated retraction rate than the COVID-19-related articles. COVID-19 articles had more retractions/expressions of concern/corrections than COVID-19-unrelated articles, articles pertaining to three other viruses and their associated diseases (HIV, influenza, and hepatitis virus) and two topics related to COVID-19 (epidemiology of infectious diseases and immunology). Similarly, by searching PubMed we discovered that the retraction rate for COVID-19 articles was higher than for COVID-19-unrelated articles, although the difference is not so prominent (0.15 vs. 0.13‰).
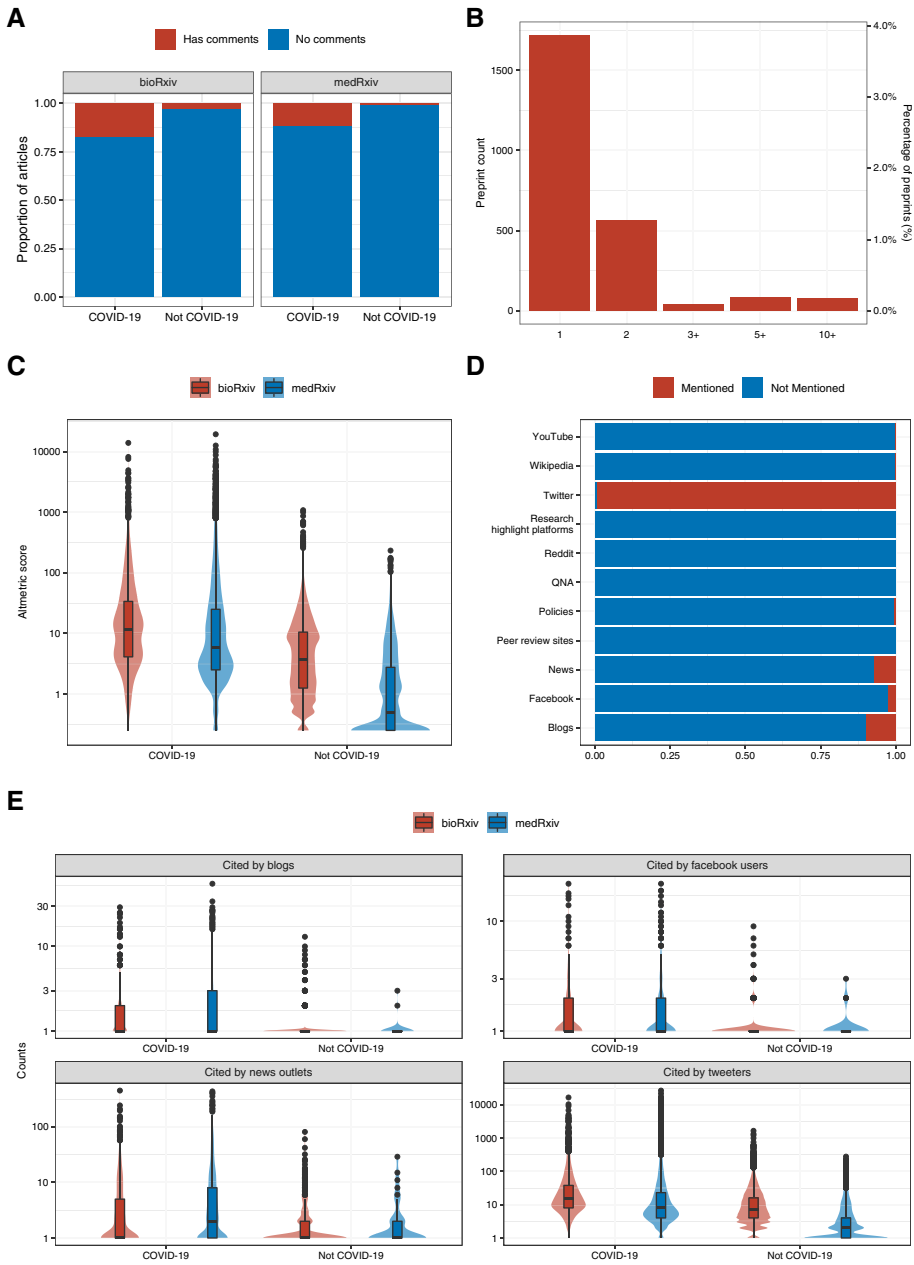
Fig. 4 Disqus comments and Altmetric data for the identified preprints. **A** Percentage of preprints that received comments on bioRxiv or medRxiv website. **B** Distribution of comments number for preprints with comments. **C** Overall Altmetric score for all preprints on bioRxiv and medRxiv, with the first version released in 2020. **D** Percentage of preprints mentioned at various sources. **E** Counts of mentions by Facebook users, blog posts, news outlets, and Twitter. All data was retrieved on December 05–06, 2020

Table 3 The number of retractions/expressions of concern/corrections identified at Retraction Watch Database (RWD), the total number of PubMed articles, and retraction rates (notices/total number of articles) for COVID-19-related articles and articles related to four other viruses/associated diseases, and two research fields (epidemiology, immunology)

Topic	RWD count[a]	Pubmed count[b]	Retraction rate (‰)
COVID-19	36	64,915	0.55
Not COVID-19	306	1,428,252	0.21
HIV	1	10,453	0.10
Herpes virus	3	2663	1.13
Influenza	0	3411	–
Hepatitis virus	0	1321	–
Epidemiology	6	40,279	0.15
Immunology	8	63,577	0.13

[a]Search phrases used for searching RWD: COVID-19 OR SARS-CoV-2 OR Coronavirus disease 19 OR 2019-nCoV; HIV OR AIDS OR human immunodeficiency virus; Herpes OR Herpes simplex OR HSV1 OR HSV2 OR varicella-zoster OR HZV OR Cytomegalovirus OR CMV OR Epstein-Barr OR EBV; Influenza OR Influenzavirus OR flu OR *influenza* OR H1N1 OR H5N1; Hepatitis virus OR HepA OR HepB OR HepC OR HepD OR HepE OR HAV OR HBV OR HCV OR HDV OR HEV OR Viral hepatitis; epidemiol* OR infect* OR virology OR virus OR viral; immun* OR inflamma* OR cytokine* OR leukocy* OR neutroph* OR lymphoc* OR antigen* OR antibod*. Search was limited to articles published from January 01, 2020 till December 05, 2020, with notice published during the same period. All preprints were omitted and all COVID-19 articles were excluded where the topic was not related to COVID-19

[b]Same search phrases were used in conjunction with "AND ("2020/01/01"[Date—Publication]: "2020/12/05"[Date—Publication]) NOT (COVID-19 OR SARS-CoV-2 OR "Coronavirus disease 19" OR nCoV-19) NOT Preprint[Publication Type]", and with "[title]" added next to each term

Discussion

The present study was motivated primarily by previous observations of shorter submission-to-acceptance time for published COVID-19-related vs. non-related manuscripts, a phenomenon suggestive of preference of COVID-19-related manuscripts for publication in the peer-reviewed journals during the COVID-19 pandemic. The present data support such a view by demonstrating an independent association between COVID-19-related topic and a higher probability of publishing but have two major limitations that preclude straightforward generalizations: (a) the analysis was limited only to a subset of all manuscripts "produced" during the observed period and those that were preprinted. This, however, was the only reasonable choice—these are the only manuscripts whose existence could be clearly verified and for which the risk (probability) of publishing could be (prospectively) estimated; (b) the observed period was bounded (January 01 to November 01, 2020), which might have affected the outcomes: our supplemental analysis indicates that it could take up to around 500 days for a preprint to get published (Supplemental Figure S1), hence the present observations might simply reflect a certain lag-time present for non-COVID-19-related preprints. Therefore, the present results pertain and should be interpreted specifically with respect to preprinted manuscripts and the observed period which almost completely

overlaps with the duration of the COVID-19 pandemic. We particularly accounted for potential bias arising from unequal "time at risk" by definition of two complementary outcomes differentially affected by the bounded observational period, and by stratification of preprints in respect to preprinting date. With adjustment for several other (albeit not all potentially relevant) potential sources of bias that could be captured in this kind of a study, the present estimates should be considered accurate. The observed considerably shorter submission-to-acceptance time for the published COVID-19-related vs. non-related preprints further supports the conclusion about the preference of COVID-19-related topics and is in line with observations pertaining to all published papers (Homolak et al., 2020; Horbach, 2020; Kun, 2020). In this respect, the present data should be viewed as reasonably indicative for all papers (during the COVID-19 pandemics) in general.

Increased preference (due to any reason) for the publishing of COVID-19-related (vs. non-related) papers combined with shorter submission-to-acceptance time, indicative of a shorter peer-review (and thus unlikely to be thorough and meaningful), creates a situation that could be reasonably considered susceptible to the impulsive release of publications of inadequate quality, i.e., susceptible to publishing bad or incorrect science, or just nonsense (e.g. article reporting a link between 5G and SARS-CoV-2) (Fioranelli et al., 2020). At least theoretically, preprinting provides a (possible) way to ameliorate this problem by opening a time window for public pre-submission peer-review that could complement the journal peer-review. The extent, quality and relevance of any peer-review, and in particular the public pre-submission peer-review, is difficult to quantify. The present data, using the number of comments pertaining to preprinted manuscripts as a proxy, do not suggest that such a practice is actually common: only around 5% of the preprints were commented on, typically with only one comment. On the positive side, COVID-19-related papers received more comments (than non-related), suggesting that these preprints are (at least) publicly discussed. Just as observed by others (Yeo-Teh & Tang, 2020), the present results indicate that published COVID-19 articles, at the present state, have a higher retraction rate than non-COVID-19 articles. Our data, however, should be viewed with additional caution—reporting retractions, corrections, and expression of concern as fractions with numbers obtained from PubMed is not optimal and is only an approximation. Furthermore, as Abritis et al. have warned (Abritis et al., 2020), it is hard to compare the number of retracted articles given that it takes years for article retractions. There is a possibility that the retraction numbers of non-COVID-19 articles are only lagging and will eventually catch up with the COVID-19 articles. Additionally, higher retraction rates might reflect greater public scrutiny, not necessarily lower quality.

While not the primary focus of this article, our limited data supports claims that preprinting in the biomedical field has increased during the COVID-19 pandemic (Fraser et al., 2021). Several other preprint-related activities, like deposition of more than one preprint version, shortened time between versions, or changes of preprint titles (as compared to the time period before the pandemic) also seem to be intensified (see Supplemental Figures S2 and S3). We hold this fact to be much needed and long overdue. We believe that unfeigned quality concerns due to lack of peer-review are (more or less) surmountable by the audiences' critical approach. On the other hand, lack of peer-review and editorial screening might be advantageous with respect to the speed of information sharing, open discussion, and lack of "censorship". This could reduce the need for hasty publication of inadequate papers in scientific journals. In addition to causing a distrust towards science, the damage done by the poor journal-published articles is hard to rectify, largely due to the false sense of unquestionable credibility assigned to articles published in peer-reviewed journals. This concern in the context of COVID-19 pandemic has already been recognized

and brought up by Serge Horbach declaring that "nonsense or incorrect science in one of these papers is potentially much more harmful" (Kwon, 2020). One illustrative example from the past is the infamous case of the article reporting a link between MMR vaccine and autism that drives distrust towards vaccination even today, years after retraction (Omer, 2020)).

Finally, in addition to the advantages for the profession and science, preprints might be a valuable "tool" for researching science itself. Preprinting provides an opportunity to study what happens "behind the curtains" of the hidden journal submission process (i.e. provides an opportunity to study the peer-review itself). We hope this will be recognized and utilized in the future. However, incorrect identification of the journal published preprints is one of the obstacles that have yet to be overcome. A significant number of preprints has not been identified as published by bioRxiv/medRxiv services (Abdill & Blekhman, 2019), even though the declaration is posted on their website stating that this happens "on rare occasions" because authors or titles have changed (Frequently Asked Questions (FAQ), n.d.). This is potentially (if the error rate is not similar in COVID-19-related vs -unrelated preprints) the most significant limitation of this study.

Limitations

The most important limitations of the study, some already mentioned, include the following: (1) conclusions should not be generalized outside the observed period. The patterns are prone to change and may be different in the future; (2) in this study, we relied on information provided by bioRxiv and medRxiv about the publication status of preprints. This is potentially a source of bias because concerns have been raised that some journal-published preprints were not identified by bioRxiv/medRxiv (Abdill & Blekhman, 2019). However, it is reasonable to assume that error frequency is similar for both COVID-19-related and -unrelated preprints. Thus, we do not expect this to pose a significant problem; (3) quality assessment was not performed by investigating individual publications and checking their methodological soundness but by comparing the number of issued concerns, corrections, and retractions relative to the total number of publications. This measure is not perfect since the process of issuance of such notices is often slow and may not necessarily reflect only the quality of the articles.

Conclusions

In conclusion, during the COVID-19 pandemic there appeared an increasing preprinting trend on bioRxiv or medRxiv. In the case of the latter platform, the trend is primarily due to the preprinting of COVID-19-related manuscripts. COVID-19-related preprints are more likely to be published in peer-reviewed journals and their submission-to-acceptance time (a proxy for the peer-review process) is considerably shorter than for the COVID-19 non-related manuscripts. COVID-19-related preprints received more comments on the preprinting platforms, but the proportion of preprints commented on is generally modest. This suggests that the opportunity of public pre-submission peer-review, inherent to the concept of preprinting, is not seized to any relevant extent. Retractions and issued concerns/corrections were sporadic regarding the papers published (and indexed in PubMed) between January 01 and December 05, 2020, but the incidence of retractions/concerns was higher for

published COVID-19-related than for non-related papers. To sum-up: COVID-19-related preprints were more publicly discussed and favored for publishing in peer-reviewed journals, typically with a shorter peer-review process, which might have possible repercussions on the quality of journal-published articles.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval None applicable.

References

Abdill, R. J., & Blekhman, R. (2019). Tracking the popularity and outcomes of all bioRxiv preprints. *eLife*. https://doi.org/10.7554/eLife.45133

Abritis, A., Marcus, A., & Oransky, I. (2020). An "alarming" and "exceptionally high" rate of COVID-19 retractions? *Accountability in Research, 28*(1), 58–59.

Berg, J. M., Bhalla, N., Bourne, P. E., Chalfie, M., Drubin, D. G., Fraser, J. S., et al. (2016). Scientific community. Preprints for the life sciences. *Science, 352*(6288), 899–901.

Carneiro, C. F., Queiroz, V. G., Moulin, T. C., Carvalho, C. A., Haas, C. B., Rayêe, D., Henshall, D. E., De-Souza, E. A., Amorim, F. E., Boos, F. Z., & Guercio, G. D. (2020). Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Research Integrity and Peer Review, 5*(1), 1–19. Accessed December 09 2020.

Cobb, M. (2017). The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLoS Biology, 15*(11), e2003995.

Fioranelli, M., Sepehri, A., Roccia, M. G., Jafferany, M., Olisova, O. Y., Lomonosov, K. M., & Lotti, T. (2020). RETRACTED: 5G technology and induction of coronavirus in skin cells. *Journal of Biological Regulators and Homeostatic Agents*. https://doi.org/10.23812/20-269-E-4

Four Years of Information Exchange. (1966). *Nature*, *211*(5052), 904–905. https://www.nature.com/articles/211904a0.pdf

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 resarch and their impact on the science communication landscape. *Plos Biology*, *19*(4), e3000959. https://doi.org/10.1371/journalpbio.3000959.

Frequently Asked Questions (FAQ). (n.d.). *bioRxiv*. Retriebved 27 November, 2020, from https://www.biorxiv.org/about/FAQ

Fu, D. Y., & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*. https://doi.org/10.7554/eLife.52646

Green, D. E. (1964). An experiment in communication: The information exchange group. *Science, 143*(3604), 308–309.

Homolak, J., Kodvanj, I., & Virag, D. (2020). Preliminary analysis of COVID-19 academic information patterns: A call for open science in the times of closed borders. *Scientometrics, 25*, 1–15. https://doi.org/10.1007/s11192-020-03587-2.

Horbach, S. P. J. M. (2020). Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19. *Quantitative Science Studies, 1*(3), 1056–1067.

Jefferson, T., Alderson, P., Wager, E., & Davidoff, F. (2002). Effects of editorial peer review: A systematic review. *JAMA: the Journal of the American Medical Association, 287*(21), 2784–2786.

Kun, Á. (2020). Time to acceptance of 3 days for papers about COVID-19. *Publications, 8*(2), 30.

Kwon, D. (2020). How swamped preprint servers are blocking bad coronavirus research. *Nature, 581*(7807), 130–131.

Nabavi Nouri, S., Cohen, Y. A., Madhavan, M. V., Slomka, P. J., Iskandrian, A. E., & Einstein, A. J. (2020). Preprint manuscripts and servers in the era of coronavirus disease 2019. *Journal of Evaluation in Clinical Practice*. https://doi.org/10.1111/jep.13498

Omer, S. B. (2020). The discredited doctor hailed by the anti-vaccine movement. *Nature, 586*(7831), 668–669.

Pasternack, S. (1966). A debate on preprint exchange—Con: Criticism of the proposed physics information exchange. *Physics Today, 19*(6), 63–69.

Retraction Watch Database. (n.d.). http://retractiondatabase.org?

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178–182.

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundations for statistical computing, Vienna, Austria*. Retrieved October 30, 2020, from https://www.r-project.org/

Woodruff, M. (1966). Unpublished Literature. *Nature, 211*(5047), 333–334.

Vercellini, P., Buggio, L., Viganò, P., & Somigliana, E. (2016). Peer review in medical journals: Beyond quality of reports towards transparency and public scrutiny of the process. *European Journal of Internal Medicine, 31*, 15–19.

Yeo-Teh, N. S. L., & Tang, B. L. (2020). An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). *Accountability in Research, 28*(1), 47–53.