Publishing of COVID-19 Preprints in Peer-reviewed Journals, Preprinting Trends, Public Discussion and Quality Issues

Ivan Kodvanj, Jan Homolak, Davor Virag, Vladimir Trkulja

Department of Pharmacology, University of Zagreb School of Medicine, Zagreb, Croatia

Correspondence: Vladimir Trkulja
Department of Pharmacology, Zagreb University School of Medicine
Šalata 11
10000 Zagreb, Croatia
E-mail: vladimir.trkulja@mef.hr

ABSTRACT

Introduction: 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 favoured for publication, (ii) preprinting trends and public discussion of the preprints and (iii) relationship between the publication topic (COVID-19-related or not) and quality issues.

Methods: Manuscripts deposited at bioRxiv and medRxiv between January 1 and October 21 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.

Results: 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=1.99, 95%CI 1.76-2.25) as well as over the entire observed period (OR=1.49, 95%CI 1.36-1.62). Submission-to-acceptance was by 41.69 days (95%CI 46.56-36.80) 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.

Conclusion: Current data suggest a preference for publication of COVID-19-related preprints over the observed period.

Keywords: preprint; COVID19; peer-review; publishing

1. INTRODUCTION

The concepts of preprinting, a practice of presenting scientific/scholarly papers to the scientific community before they are published in peer-reviewed journals, date back to the 17th century[1]. In the modern era, one of the first large-scale attempts to share unpublished works was the Information Exchange Group (IEG) experiment. It was carried out by the National Institute of Health (NIH) in the 1960s[2], with the goal of speeding up information dissemination and promoting discussion[3]. Although the project had numerous members, it was brought to an early and infamous end after several discouraging editorials had been published[4,5]. The proposal for the Physics Information Exchange (PIE)[2] elicited similar unfavourable reactions. It was called "offensive"[6], and declared "a serious threat to physics communication and to the physics research community" by the editor of the Physical Review[7]. After this setback, it took years for PIE to start, but finally, it evolved into the first major preprint server[8] – arXiv - currently distributing articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics[9].

While preprinting is a standard practice in some fields[10], it is still uncommon in biomedicine[11]. Today this seems like a paradox since the IEG, one of the first attempts to establish a platform for sharing preprints, was focused on the biomedical field. It took years for biomedical preprints servers to be established, even though the initiative was present but often disregarded. For instance, Harold Varmus, Nobel prize winner and director of NIH, proposed development of a preprint repository to "accelerate much-needed public discussion of electronic publication"[12] in 1999. After several failed attempts to establish preprints servers (Nature Precedings and ClinMed Netprints)[2,13], bioRxiv was finally started[14] - currently one of the largest preprint servers in the field.

Through preprints, authors can rapidly share their findings and receive feedback from other scientists. Both sides can profit from this - authors by improving their manuscripts before submitting them to journals, based on the received feedback; whilst others could steer their research in accordance with the newest findings that would otherwise be unknown for months. Furthermore, some authors even believe that the lack of peer-review is beneficial since it does not restrain the sharing of worthwhile ideas that might be falsely disregarded and overlooked in the editorial process[3]. On the other hand, there is a concern that without peer-review, preprints might be of lower quality. Both bioRxiv and medRxiv explicitly declare that preprints "should not be regarded as conclusive, guide clinical practice/health-related behaviour, or be reported in news media as established information". However, one should keep in mind that peer-review is (also) not flawless and the evidence of its effectiveness is scarce[15,16]. Anecdotally described by Smith[17], former editor for the British Medical Journal, editors can also be biased and under pressure when they receive articles from distinguished authors.

The Coronavirus disease 2019 (COVID-19) pandemic prompted a need for rapid communication and discussion of scientific information and resulted in an unprecedented surge in the number of publications on various aspects of the disease. Publishers greatly shortened the peer-review process for COVID-19-related manuscripts[18]. Be it due to elementary greed or to a praiseworthy desire to disseminate information, such a practice is not necessarily beneficial – concerns about the quality of articles published under these circumstances have been raised[18–20]. Preprinting apparently became a necessity – not only by enabling rapid dissemination of information but also by encouraging public discussion and providing an

opportunity for peers to intervene early if methodological fallacies are identified and limit the spread of misinformation to improve the overall quality of the manuscripts. We aimed: a) to evaluate preprinting trends for COVID-19-related and non-related manuscripts on two key biomedical preprint platforms *bioRxiv* and *medRxiv* during the COVID-19 pandemic; b) to estimate the extent to which the preprints were commented on (as an indicator of the extent of "pre-submission peer-review"); c) to estimate whether COVID-19 manuscripts are favoured over non-COVID-19 manuscripts for journal publishing; d) to probe the feasibility of the assumption that published COVID-19 papers are (under the circumstances of shortened peer-review process) more commonly flawed than the published non-COVID-19 papers.

2. MATERIALS AND METHODS

2.1. Study outline and study outcomes

Types of datasets used and their purpose in the present study are outlined in **Figure 1**. We defined three objectives based on their relevance (objective 1 – most relevant, objective 3 – least relevant) and the anticipated level of susceptibility to bias/confounding (objective 1 – least susceptible, objective 3 – most susceptible), given the nature of the study and the type of data. The **1st objective** was to investigate whether COVID-19-related preprints were favoured (over non-related) for publication in peer-reviewed journals. We considered that the probability of publishing was the most informative outcome in this respect and that further insight would be obtained from submission-to-acceptance data. Therefore, we defined three outcomes for the purpose: (i) *primary outcome* was the probability of publishing within 120 days since the deposition of the first preprint version and was assessed in a subset of preprints deposited till May 29, 2020. Since the time-period during which manuscripts could have been preprinted and published was bounded (January 1 – October 21, 2020), to reduce the risk of bias arising from unequal "time-at-risk" (for publishing), the analysis was restricted to preprints deposited till May 29 and a time-window for the publishing of 120 days. We considered it to be a reasonable period of time for submission and review process to take place, and it was "available" to all preprints in this subset; (ii) *secondary outcome* was the probability of publishing over the entire observed period (i.e., before October 21, 2020) and it was assessed in a subset of preprints deposited before August 27, 2020. 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 at the same time, i.e., that they 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, first two strata (Jan 01 - Jan 15 and Jan 16 - Jan 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 a peer-reviewed journal 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 (**Figure 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 1 and October 21, 2020 (**Figure 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[21] between January 1 and October 26, 2020 (**Figure 1**).

2.2. Data collection

All data retrieval and management were done in *R* (version 4.0.2)[22]. *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 1 – October 21, 2020). COVID-19 preprints were identified using the search terms COVID-19 OR SARS-CoV-2 OR Coronavirus OR 2019-nCoV. All other preprints were classified as non-COVID-19 preprints. 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 to retrieve Altmetric data for each preprint deposited on bioRxiv and medRxiv servers. Submission and acceptance dates for published preprints was fetched from PubMed with the *RISmed* package. The Retraction Watch database was used to retrieve the number of articles with issued retraction notice, expression of concern or correction during 2020. Notices for COVID-19 articles or preprints were identified using the search terms COVID-19 OR SARS-CoV-2 OR Coronavirus OR 2019-nCoV. Results were additionally filtered to retrieve the number of journal research articles and preprints separately. The number of retrieved notices was expressed as a proportion of the total number of COVID-19 and non-COVID-19 articles indexed by PubMed. Additionally, to avoid bias due to possible over- or under-representation 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] OR "Retraction:"[title]) AND ("2020/01/01"[Date - Publication] : "2020/10/26"[Date - Publication]), with and without “AND (COVID-19 OR SARS-CoV-2 OR Coronavirus OR 2019-nCoV)”, to identify the number of retractions of COVID-19 and non-COVID-19 articles during 2020.

2.3. Data analysis

All data visualization and analysis were performed in *R* (version 4.0.2)[22]. 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).

Probability of publishing within 120 days since the 1st preprint version and the probability of publishing over the entire observed period were analysed by fitting stratified (in respect to preprint date) logistic regression (package *survival*, function clogit). Submission-to-acceptance

time was analysed by fitting a hierarchical (mixed) model with submission date stratum as a random effect. Articles with submission dates before 2020 were excluded from the dataset that was used for analysing 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 that were identified (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 (only) kept 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 results of the main study (see Supplemental Material, Supplemental methods).

2.4. Submission-to-acceptance dataset validation

After all articles with submission dates before January 1, 2020, were removed, we identified and excluded one article with the submission-to-acceptance time of -107 days (an obvious mistake). To validate the dataset, submission and acceptance dates were checked for 100 randomly selected articles. Three articles with erroneous values were found - one with a large and two with what we consider to be a rather small deviation from the correct value (≅2.5%). We considered that the number and size of errors were acceptable and not likely to affect the conclusions of the study.

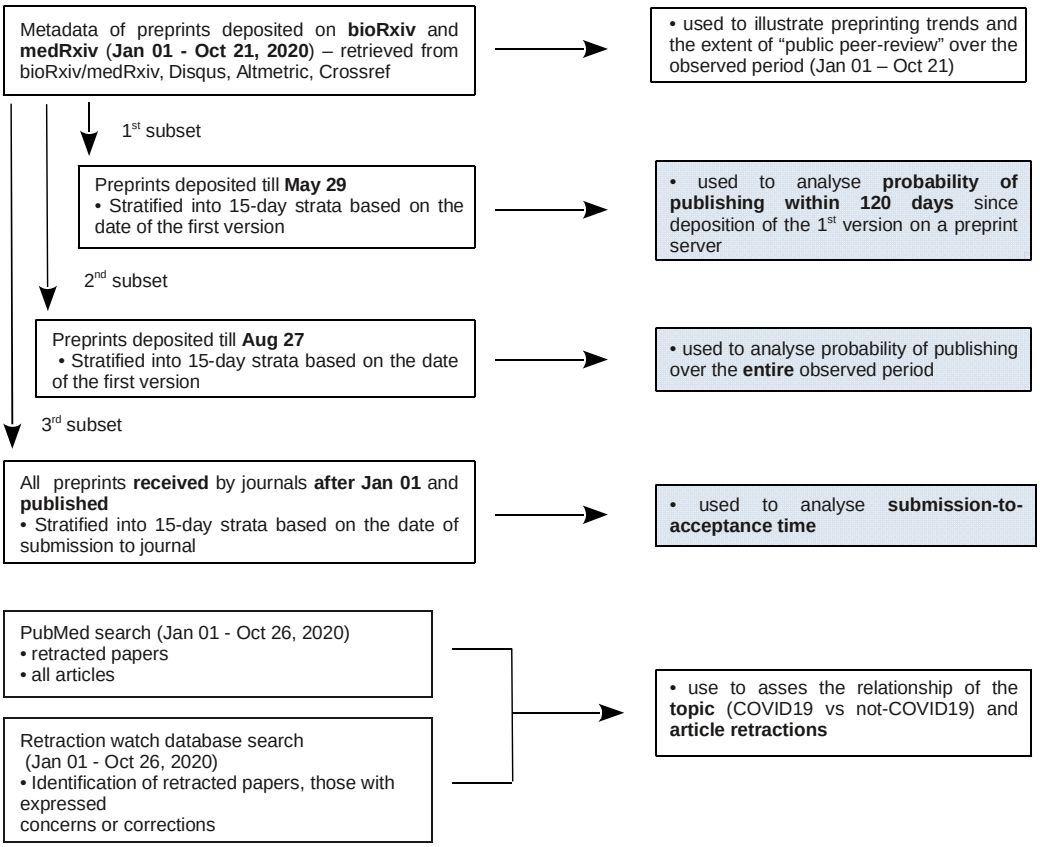


Figure 1. **Study outline.**

3. RESULTS

3.1. Are COVID-19-related preprints favoured for publishing in peer-reviewed journals?

The subset of preprints deposited till May 29, 2020, and used to evaluate the probability of publishing within 120 days since the 1st preprint version comprised a total of 15239 preprints on bioRxiv (5.6% COVID-19-related and 94.4% non-related) and a total of 5078 preprints on medRxiv (64.7% COVID-19-related and 35.3% non-related) (**Fig. 2**). The subset of preprints deposited till August 27, 2020, and used to evaluate the probability of publishing over the entire observed period (before October 21, 2020) comprised 25892 preprints on bioRxiv (6.8% COVID-19-related and 93.2% non-related) and 9210 preprints on medRxiv (64.4% COVID-19-related and 35.6% non-related) (**Fig. 2**). Raw proportions of published papers by the topic and preprinting platform are depicted in **Figure 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=1.99, 95%CI 1.76-2.25), and with a higher probability of publishing over the observed period than non-COVID-19 preprints (OR=1.49, 95%CI 1.36-1.62). Probability of both outcomes was higher for bioRxiv than for medRxiv preprints, and lower for preprints with ≥2 versions than for those with only one preprinted version (**Table 1**). Journal submission-to-acceptance time was identified for preprints published before October 21, 2020.

In multivariate analysis, COVID-19-related topic was associated with around 40 days shorter submission-to-acceptance time (mean difference -41.69 95%CI -46.59 to -36.80) (**Table 2**). The preprinting platform and number of preprint versions did not appear associated with the outcome (**Table 2**).

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

For the observed period (Jan 1 – Oct 21, 2020) we identified a total of 12377 preprints newly deposited at medRxiv [7239 (58.5%) COVID-19-related)], and 100496 preprints deposited at bioRxiv [2164 (2.15%) 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%), but somewhat higher for COVID-19-related preprints (**Fig. 4A**): 23% and 4% of the COVID-19-related and not related preprints respectively, on bioRxiv; and 12% and 6% 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 preprints (**Fig. 4E**). Preprints posted on the more clinically-oriented medRxiv were more reported in news outlets compared to those posted on bioRxiv (**Fig. 4E**).

3.4. Potential quality issues

A total of 290 retracted papers were identified in the Retraction Watch database over the observed period, of which 48 were related to COVID-19. **Figure 5** summarizes these findings by article type (all article types vs original research articles), in respect to the nature of notice, separately for COVID-19-related and not-related papers (as a number of events/1000 publications). Additionally, we found 11 preprints that were retracted - all related to COVID-19. Similar results were obtained by searching the PubMed - we identified 0.358‰ retracted articles related to COVID-19 and 0.110‰ unrelated to COVID-19.

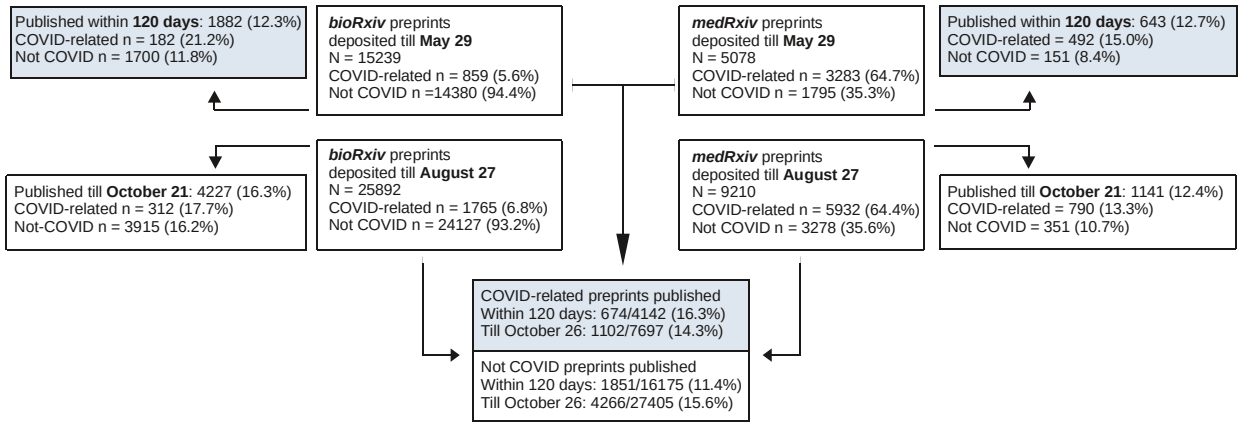


Figure 2. Structure of preprints deposited by May 29 and August 27, used to evaluate the probability of publishing by platform-by topic-by publishing outcomes.

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 May 29) and time-to-publishing considering the entire observed period (a subset of preprints deposited till 27 August)[1]. Effects are expressed as odds ratios (OR).

	Published within 120 days		**Published before October 21**	
Predictors	*OR (95% CI)*	*p*	*OR (95% CI)*	*p*
COVID-19 vs Not COVID-19	1.99 (1.76 to 2.25)	**<0.001**	1.49 (1.36 to 1.62)	**<0.001**
bioRxiv vs medRxiv	1.38 (1.23 to 1.56)	**<0.001**	1.34 (1.23 to 1.46)	**<0.001**
≥2 vs 1version	0.80 (0.72 to 0.88)	**<0.001**	0.95 (0.89 to 1.01)	0.120

[1] 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.

Table 2. Analysis of *submission-to-acceptance* time for published preprints (in days)[1].

Predictors	*Estimates (95% CI)*	*p*
COVID-19 vs Not COVID-19	-41.69 (-46.59 to -36.80)	**<0.001**
bioRxiv vs medRxiv	0.33 (-4.42 to 5.09)	0.890
≥2 vs 1 preprint version	1.51 (-2.15 to 5.16)	0.420

[1] 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 237/312 COVID-19/bioRxiv (median=44 days), 2542/3915 not-COVID-19/bioRxiv (median=92 days), 550/790 COVID-19/medRxiv (median=39 days), 222/351 not-COVID-19/medRxiv preprints (median=100 days).

Figure 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).

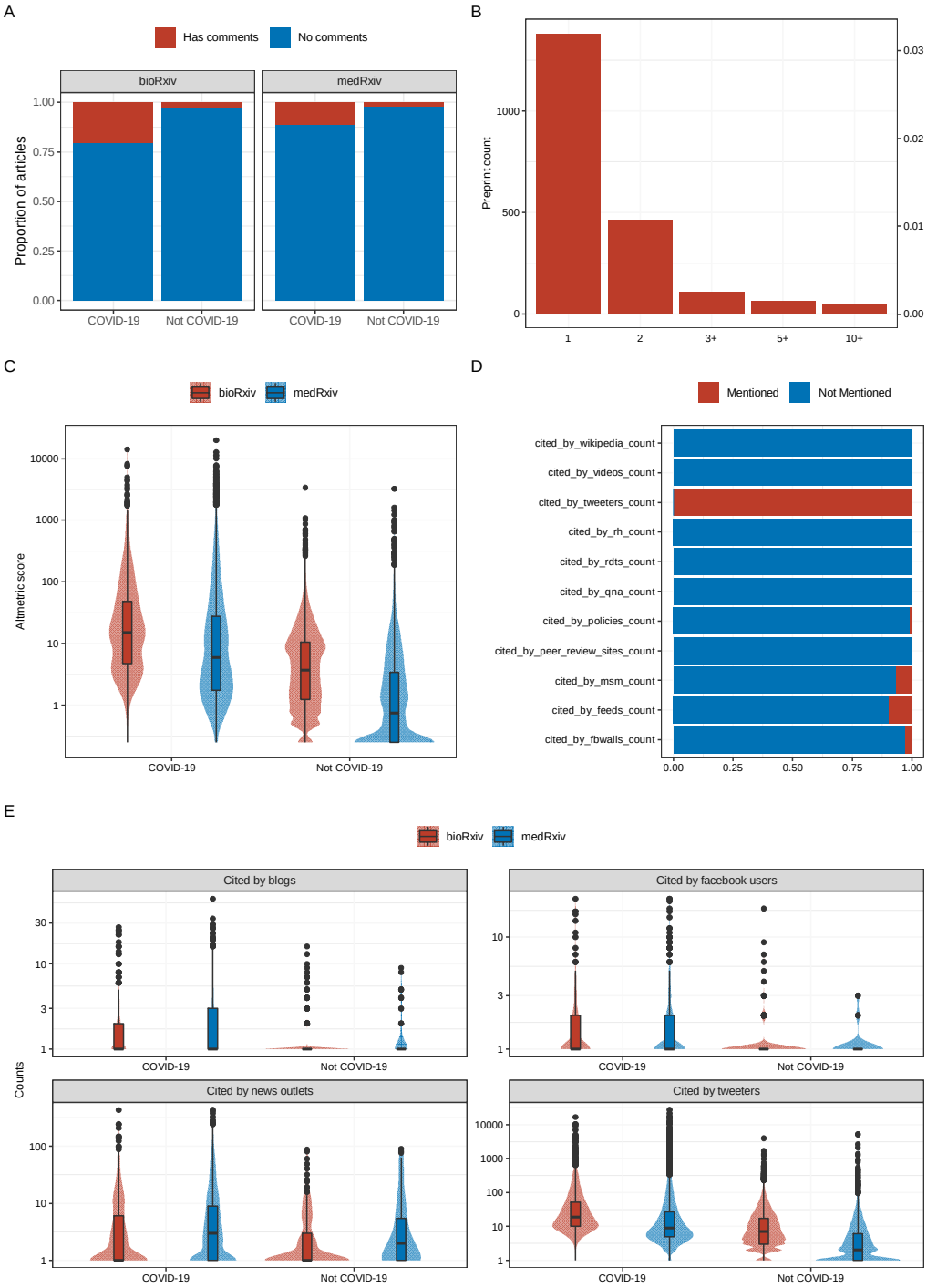


Figure 4. Disqus comments and Altmetric data for the identified preprints. (A) Percentage of preprints that received comments on bioRxiv or medRxiv website. **(B)** Distribution of comments number for preprints with comments. **(C)** Overall Altmetric score for all preprints on bioRxiv and medRxiv, with 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.

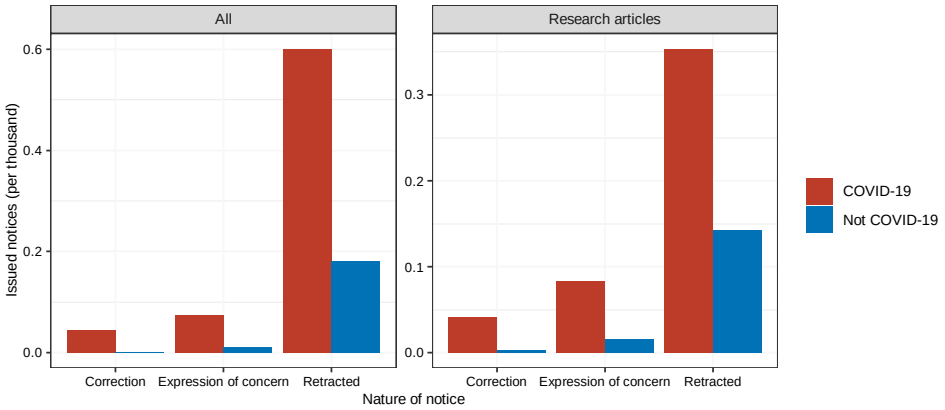


Figure 5. Corrections, expressions of concern and retractions of articles published in 2020.

4. 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[18-20], 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, i.e., 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 1 to October 21, 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 pandemics. 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[18-20]. 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)[23]. 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 a 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[24], 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[25], 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 pandemics[26]. 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 time period before the pandemic) also seem to be intensified (see Supplemental Figure 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 in 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"[27]. Thus, the consequences of publishing bad research can be tremendous and hard to rectify (e.g. opposing vaccination and the case of Andrew Wakefield's article).

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.

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 the 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 1 and October 26, 2020, but the incidence of retractions/concerns was higher for published COVID-19-related than for non-related papers. Overall data support a view that so far, COVID-19-related manuscripts were favoured for publishing in peer-reviewed journals, commonly with very short peer-review.

5. DECLARATIONS

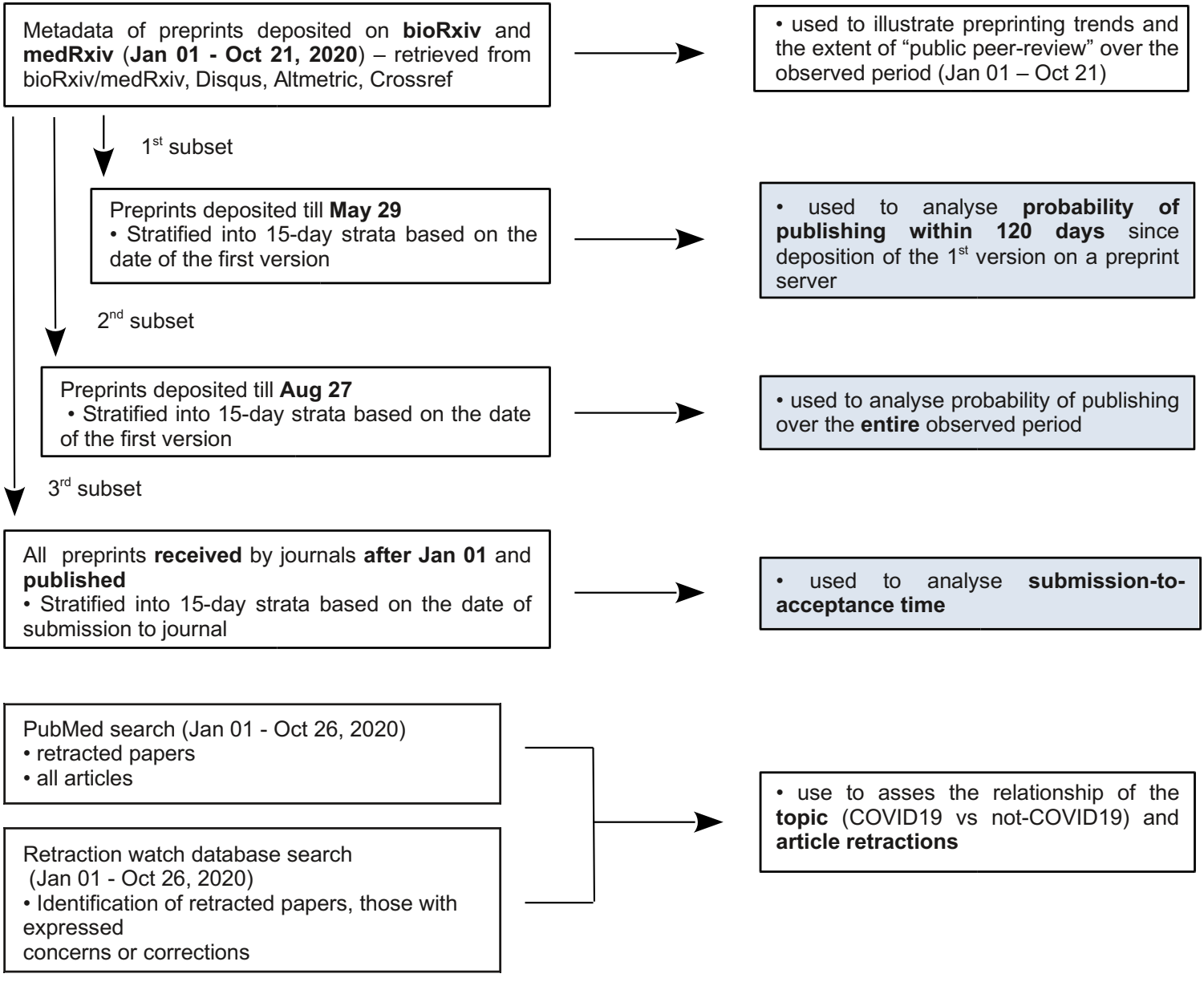
Acknowledgements: Nothing to acknowledge.
Ethics approval: None applicable.
Competing interest: None.
Funding: None.
Authors' contributions: IK and VT designed the study. IK gathered data. IK and VT analysed the data. DV and JH provided valuable suggestions to analysing data. All authors participated in the writing of the manuscript.

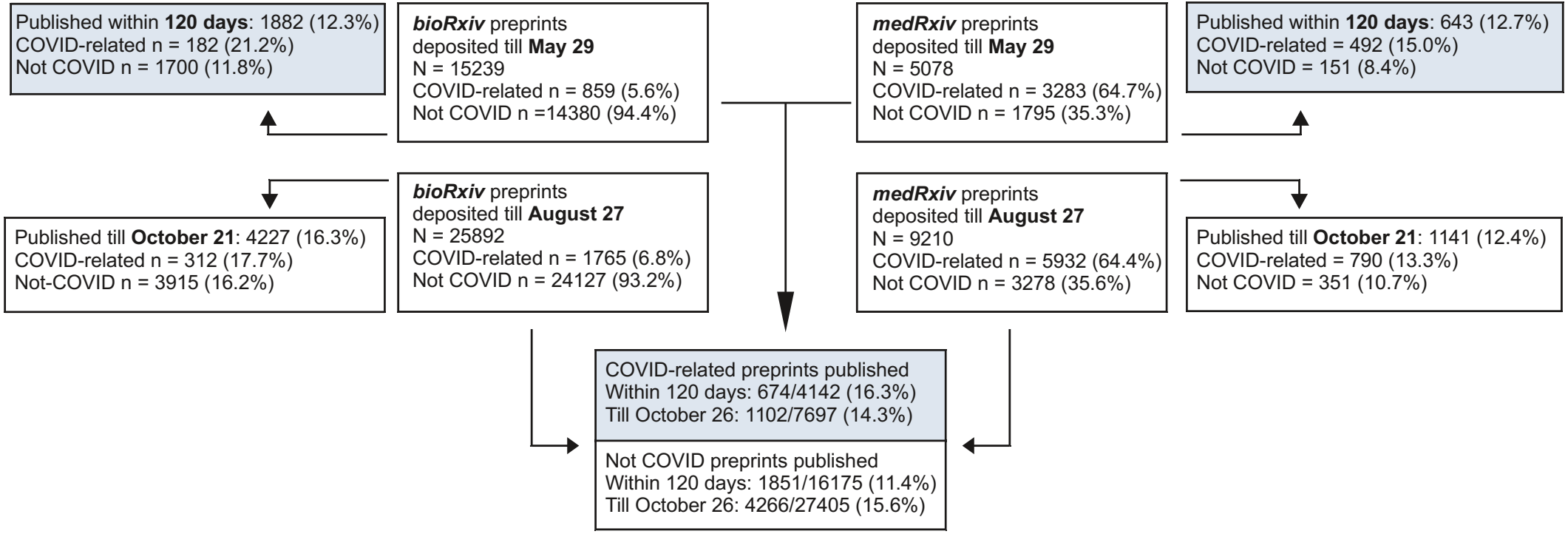
REFERENCES

1. Moore, C. A. Preprints. An Old Information Device with New Outlooks. *J. Chem. Doc.* **5**, 126–128 (1965).
2. Cobb, M. The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLoS Biol.* **15**, e2003995 (2017).
3. Green, D. E. An Experiment in Communication: The Information Exchange Group. *Science* **143**, 308–309 (1964).
4. Abelson, P. H. Information exchange groups. *Science* **154**, 727 (1966).
5. Four years of information exchange. *Nature* **211**, 904–905 (1966).
6. Unpublished Literature. *Nature* **211**, 333–334 (1966).
7. Pasternack, S. A debate on preprint exchange—Con: Criticism of the proposed Physics Information Exchange. *Phys. Today* **19**, 63–69 (1966).
8. Till, J. E. Predecessors of preprint servers. (2001) doi:10.1087/09531510125100214.
9. arXiv. https://arxiv.org/about.
10. Berg, J. M. *et al.* SCIENTIFIC COMMUNITY. Preprints for the life sciences. *Science* **352**, 899–901 (2016).
11. Fu, D. Y. & Hughey, J. J. Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *Elife* **8**, (2019).
12. Varmus, H. [E-Biomed: A Proposal for Electronic Publications in the Biomedical Sciences (Draft and Addendum)]. (1999).
13. About Precedings. https://www.nature.com/npre/about.
14. Callaway, E. & Powell, K. Biologists urged to hug a preprint. *Nature* **530**, 265 (2016).
15. Vercellini, P., Buggio, L., Viganò, P. & Somigliana, E. Peer review in medical journals: Beyond quality of reports towards transparency and public scrutiny of the process. *Eur. J. Intern. Med.* **31**, 15–19 (2016).
16. Jefferson, T., Alderson, P., Wager, E. & Davidoff, F. Effects of editorial peer review: a systematic review. *JAMA* **287**, 2784–2786 (2002).
17. Smith, R. Peer Review: A Flawed Process at the Heart of Science and Journals. *J. R. Soc. Med.* **99**, 178–182 (2006).
18. Homolak, J., Kodvanj, I. & Virag, D. Preliminary analysis of COVID-19 academic information patterns: a call for open science in the times of closed borders. *Scientometrics* 1–15 (2020) doi:10.1007/s11192-020-03587-2.
19. Horbach, S. P. J. M. Pandemic publishing: Medical journals strongly speed up their

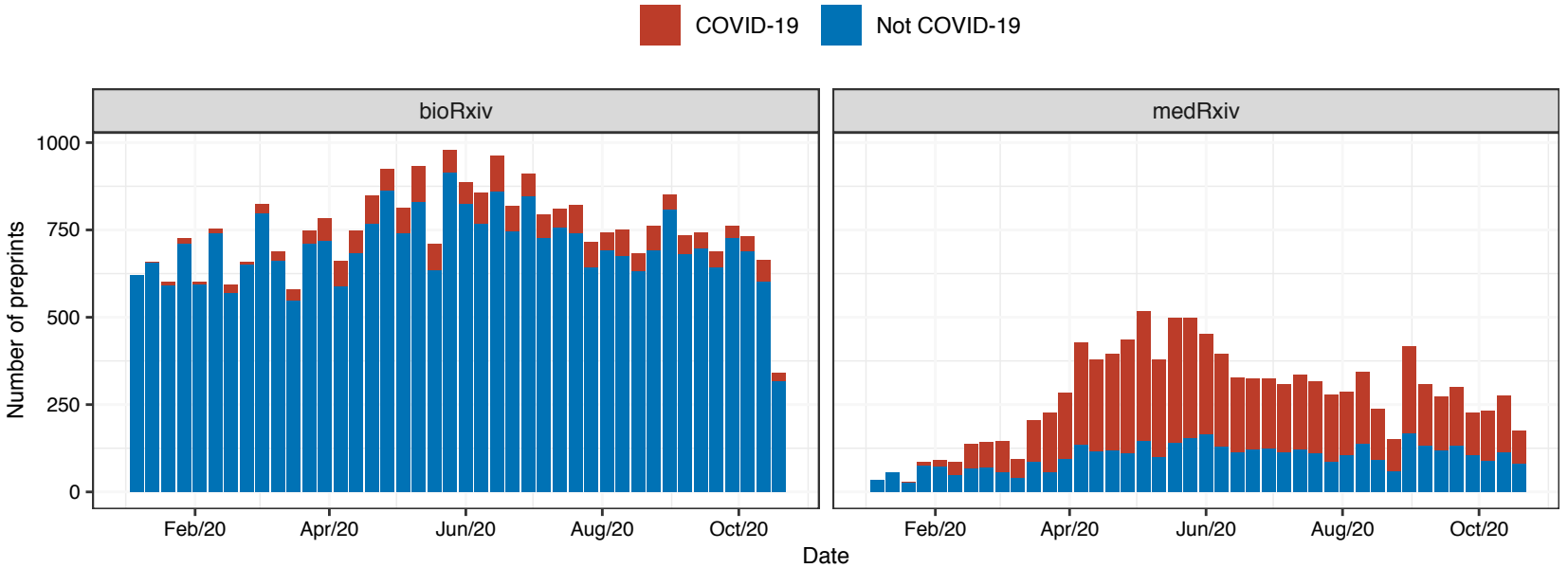
publication process for COVID-19. *Quantitative Science Studies* **1**, 1056–1067 (2020).

20. Kun, Á. Time to Acceptance of 3 Days for Papers About COVID-19. *Publications* **8**, 30 (2020).
21. Retraction Watch Database. http://retractiondatabase.org/RetractionSearch.aspx?
22. The R Project for Statistical Computing. https://www.r-project.org/.
23. Fioranelli, M. *et al.* RETRACTED: 5G Technology and induction of coronavirus in skin cells. *J. Biol. Regul. Homeost. Agents* (2020) doi:10.23812/20-269-E-4.
24. Yeo-Teh, N. S. L. & Tang, B. L. An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). *Account. Res.* 1–7 (2020) doi:10.1080/08989621.2020.1782203.
25. Abritis, A., Marcus, A. & Oransky, I. An ‘alarming’ and ‘exceptionally high’ rate of COVID-19 retractions? *Account. Res.* 1–2 (2020) doi:10.1080/08989621.2020.1793675.
26. Fraser, N. *et al.* Preprinting the COVID-19 pandemic. *Scientific Communication and Education* (2020).
27. Kwon, D. How swamped preprint servers are blocking bad coronavirus research. *Nature* **581**, 130–131 (2020).

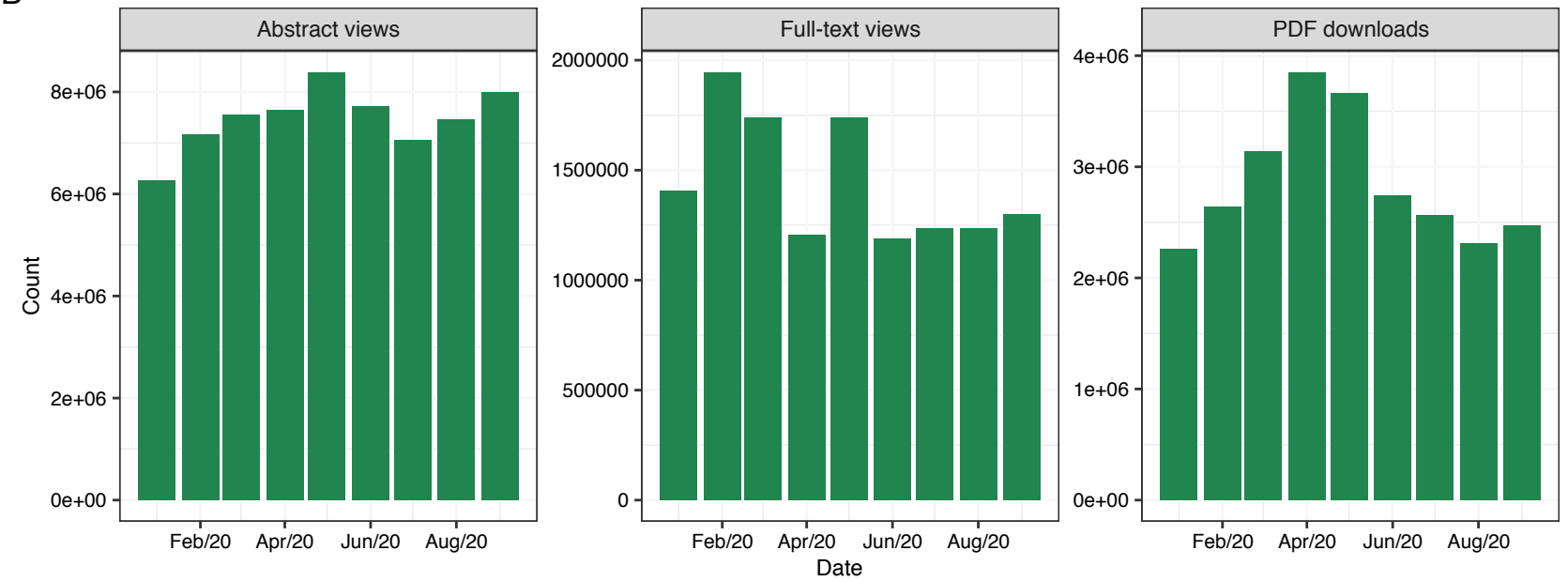

Metadata of preprints deposited on **bioRxiv** and **medRxiv** (**Jan 01 - Oct 21, 2020**) – retrieved from bioRxiv/medRxiv, Disqus, Altmetric, Crossref
• used to illustrate preprinting trends and the extent of "public peer-review" over the observed period (Jan 01 – Oct 21)
1st subset
Preprints deposited till **May 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 **Aug 27**
• Stratified into 15-day strata based on the date of the first version
• used to analyse probability of publishing over the **entire** 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 - Oct 26, 2020)
• retracted papers
• all articles
Retraction watch database search
(Jan 01 - Oct 26, 2020)
• Identification of retracted papers, those with expressed
concerns or corrections
• use to asses the relationship of the **topic** (COVID19 vs not-COVID19) and **article retractions**

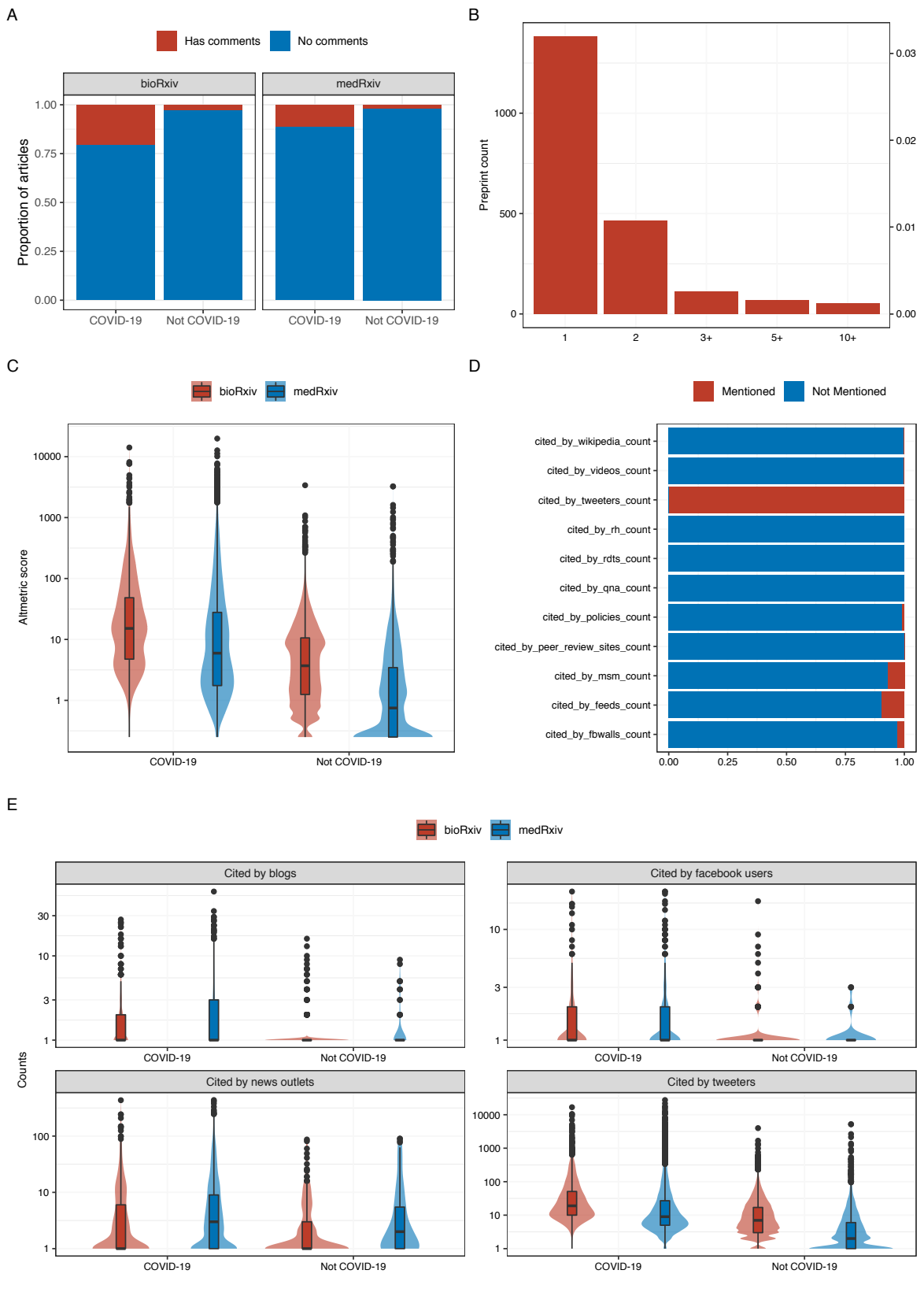

Published within 120 days: 1882 (12.3%)
COVID-related n = 182 (21.2%)
Not COVID n = 1700 (11.8%)
bioRxiv preprints
deposited till May 29
N = 15239
COVID-related n = 859 (5.6%)
Not COVID n =14380 (94.4%)
medRxiv preprints
deposited till May 29
N = 5078
COVID-related n = 3283 (64.7%)
Not COVID n = 1795 (35.3%)
Published within 120 days: 643 (12.7%)
COVID-related = 492 (15.0%)
Not COVID = 151 (8.4%)
Published till October 21: 4227 (16.3%)
COVID-related n = 312 (17.7%)
Not-COVID n = 3915 (16.2%)
bioRxiv preprints
deposited till August 27
N = 25892
COVID-related n = 1765 (6.8%)
Not COVID n = 24127 (93.2%)
medRxiv preprints
deposited till August 27
N = 9210
COVID-related n = 5932 (64.4%)
Not COVID n = 3278 (35.6%)
Published till October 21: 1141 (12.4%)
COVID-related = 790 (13.3%)
Not COVID = 351 (10.7%)
COVID-related preprints published
Within 120 days: 674/4142 (16.3%)
Till October 26: 1102/7697 (14.3%)
Not COVID preprints published
Within 120 days: 1851/16175 (11.4%)
Till October 26: 4266/27405 (15.6%)

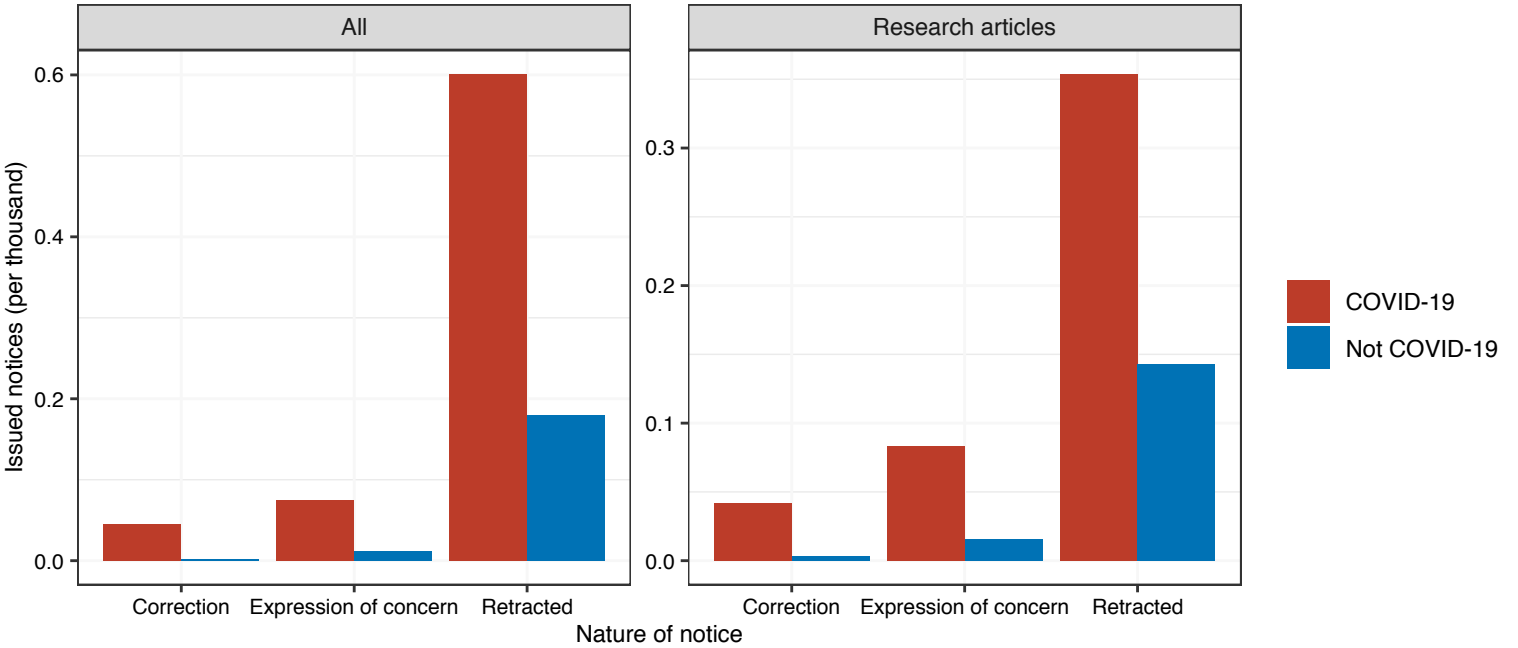
A


COVID-19
Not COVID-19
bioRxiv
medRxiv
Number of preprints
1000
750
500
250
0
Feb/20
Apr/20
Jun/20
Aug/20
Oct/20
Date

B


Abstract views
Full-text views
PDF downloads
Count
8e+06
6e+06
4e+06
2e+06
0e+00
2000000
1500000
1000000
500000
0
4e+06
3e+06
2e+06
1e+06
0e+00
Feb/20
Apr/20
Jun/20
Aug/20
Date


A
Has comments
No comments
bioRxiv
medRxiv
Proportion of articles
1.00
0.75
0.50
0.25
0.00
COVID-19
Not COVID-19
B
Preprint count
1000
500
0
0.03
0.02
0.01
0.00
1
2
3+
5+
10+
C
bioRxiv
medRxiv
Altmetric score
10000
1000
100
10
1
COVID-19
Not COVID-19
D
Mentioned
Not Mentioned
cited_by_wikipedia_count
cited_by_videos_count
cited_by_tweeters_count
cited_by_rh_count
cited_by_rdts_count
cited_by_qna_count
cited_by_policies_count
cited_by_peer_review_sites_count
cited_by_msm_count
cited_by_feeds_count
cited_by_fbwalls_count
0.00
0.25
0.50
0.75
1.00
E
bioRxiv
medRxiv
Cited by blogs
Cited by facebook users
Cited by news outlets
Cited by tweeters
Counts
30
10
3
1
100
10000
1000
COVID-19
Not COVID-19


All
Research articles
Issued notices (per thousand)
0.6
0.4
0.2
0.0
0.3
0.2
0.1
0.0
Correction
Expression of concern
Retracted
Nature of notice
COVID-19
Not COVID-19