

Publication practices during the COVID-19 pandemic: Expedited publishing or simply an early bird effect?

 Yulia V. Sevryugina ^{1*} and Andrew J. Dicks ²


Yulia V. Sevryugina



Andrew J. Dicks

¹University of Michigan Library, Ann Arbor, Michigan, USA

²School of Information, University of Michigan, Ann Arbor, Michigan, USA

ORCID:

 Y. V. Sevryugina: [0000-0003-1844-952X](https://orcid.org/0000-0003-1844-952X)

 A. J. Dicks: [0000-0002-5824-8258](https://orcid.org/0000-0002-5824-8258)

*Corresponding author: Yulia V. Sevryugina, University of Michigan Library, Ann Arbor, MI, USA.

 E-mail: yulias@umich.edu; ajamesd@umich.edu

Abstract: This study explores the evolution of publication practices associated with the SARS-CoV-2 research papers, namely, peer-reviewed journal and review articles indexed in PubMed and their associated preprints posted on bioRxiv and medRxiv servers: a total of 4,031 journal article-preprint pairs. Our assessment of various publication delays during the January 2020 to March 2021 period revealed the early bird effect that lies beyond the involvement of any publisher policy action and is directly linked to the emerging nature of new and ‘hot’ scientific topics. We found that when the early bird effect and data incompleteness are taken into account, COVID-19 related research papers show only a moderately expedited speed of dissemination as compared with the pre-pandemic era. Medians for peer-review and production stage delays were 66 and 15 days, respectively, and the entire conversion process from a preprint to its peer-reviewed journal article version took 109.5 days. The early bird effect produced an ephemeral perception of a global rush in scientific publishing during the early days of the coronavirus pandemic. We emphasize the importance of considering the early bird effect in interpreting publication data collected at the outset of a newly emerging event.

Keywords: COVID-19, expedited publication practices, publication delays

INTRODUCTION

A pneumonia of unknown origin was first reported in Wuhan, China on 30 December 2019 (ProMED International Society for Infectious Diseases, 2019), and in about 3 months, the coronavirus was declared a pandemic (WHO, 2020). The response of the scientific community was outstanding; already in January, the first research reports appeared as preprints, clinical trials and journal articles (Fidahic et al., 2020). On 31 January 2020, various journal publishers, and research organizations signed the Statement on Data Sharing in Public Health Emergencies reaffirming the principles of rapid access to research data and publications relevant to the COVID-19 outbreak (Welcome Fund, 2020). Within a few

months of the outbreak, journal publishers partially or completely lowered their paywalls concerning the sharing of SARS-CoV-2 related research (Retta, 2021), supported and encouraged scientific communication through preprints (Eisen et al., 2020), and ensured a fast-track peer-review process for COVID-19 works (CFP for COVID-19 works, 2020). A successful example of the latter is the initiative Rapid Reviews: COVID-19 (RR:C19), launched on 27 April 2020, by Hindawi, the Royal Society, PLOS and PeerJ to create and share a pool of expert reviewers for COVID-19 manuscripts (EurekAlert, 2020; OASPA, 2020). These enhanced peer-review practices and publication policies, coupled with an increasing number of preprints, suggested the emergence of a new era in scientific communication prompted by the

COVID-19 pandemic (Callaway, 2020; Krumholz et al., 2020; Kupferschmidt, 2020).

The urgency and transparency of scientific communication were genuinely welcomed early in the pandemic by the public as well as the scientific community, but the unprecedented volume of research jeopardized the previously established standards for peer-review and publication policies (Dinis-Oliveira, 2020). For instance, *eLife* announced curtailing requests for additional experiments when reviewing the SARS-CoV-2 papers (Eisen et al., 2020). Further, early analyses of publication practices related to COVID-19 manuscripts reported a median peer-review time of 6 days (Kun, 2020; Palayew et al., 2020), which stands in stark contrast with a standard peer-review time that remained at around 100 days for the last 30 years (Powell, 2016). The daily rate of COVID-19 preprints posted on servers increased dramatically and many of them had a strong impact on the public health policy-making during the early pandemic (Fraser et al., 2021). The following wave of article retractions (Retraction Watch, 2020) summoned initial concerns that the fast publication speed of COVID-19 works could be incompatible with a rigorous peer-review process and may lead to damaging the integrity of science communication (Dreisbach, 2020; Steinberg, 2020).

Indeed, the year 2020 was fraught with intense debates on social media around several controversial, high-profile journal publications that affected COVID-19 related health policies. To cite a few, a published clinical trial describing the successful use of hydroxychloroquine in COVID-19 patients (Gautret et al., 2020) was actively promoted by the Trump administration (Baker et al., 2020) despite multiple concerns regarding the quality of its study design (Fauci et al., 2020; Servick, 2020; Voss, 2020). Another study, a letter published in the *Lancet Respiratory Medicine* journal early in the pandemic raising concerns about the use of ibuprofen to treat COVID-19 symptoms (Fang et al., 2020), was initially supported by WHO (Moffitt, 2020) and led doctors to advise against treating COVID-19 fever with nonsteroidal anti-inflammatory drugs (NSAIDs), like ibuprofen (Day, 2020), and the French Health Ministry to completely ban NSAIDs (DGS-urgent, 2020); all these actions being swiftly reversed once data insufficiency became apparent (Drake et al., 2021). These disputes demonstrate the pressure that the peer-review system and the entire scientific community experienced during the early coronavirus pandemic (Chirico et al., 2020).

In an effort to maintain high-quality standards in reviewing COVID-19 studies, in April 2020, EASE (European Association of Science Editors) encouraged all editors to enforce the previously established guidelines on authors and require a clear statement of study limitations (EASE, 2020). Throughout the first year of the pandemic, a number of editorials further reiterated the importance of maintaining a rigorous peer-review process during such a large-scale public health emergency (Sepúlveda-Vildósola et al., 2020; Smart, 2020). To cite a few, the editor of *JAMA* warned that ‘Rushing publication, if there

Key points

- The early bird effect is observed in extremely short publication delays for scientific manuscripts on new and ‘hot’ topics at the outset of a newly emerging event.
- The early bird effect produced an ephemeral perception of a global rush in scientific publishing during the early days of the coronavirus pandemic.
- Publication delays for SARS-CoV-2 research papers show only a moderate expediting as compared to the pre-pandemic era; a median peer-review and production stage delays were 66 and 15 days, respectively, during the 1 January 2020 to 31 March 2021 period.
- Early bird manuscripts and data incompleteness are both intrinsic features of publication data and shall be taken into account when interpreting the publishing landscape.

are mistakes, will ultimately undermine public trust in science’ (Bauchner et al., 2020); the editor of *Thorax* stressed that ‘...it is crucial that journals streamline, but maintain high-quality peer-review processes’ (Smyth et al., 2020); and *The Lancet Global Health* called for ‘a need to slow down’ and ‘resist pressure from researchers and their institutions to expedite every step’ because ‘When research, writing, and peer review are rushed, the consequences may be damaging’ (The Lancet Global Health, 2020).

Judging by the subsequent analyses that demonstrated consistently longer publication delays for COVID-19 manuscripts as the pandemic evolved, these concerns seemed to be addressed. For example, early in the pandemic, from January to April 2020, median peer-review time for COVID-19 manuscripts was reported as 6 days (Kun, 2020; Palayew et al., 2020) and median elapsed time for COVID-19 preprints, which is how long it takes for a preprint to transform into a peer-reviewed journal article, was 22.5 days (Fraser et al., 2020). However, later in the pandemic, based on data we collected in October 2020 (Sevryugina & Dicks, 2021b), these median delays lengthened to 37 and 57 days, respectively; and within a month, the latter extended to 68 days (Fraser et al., 2021). Herein, we initially set out to determine if the same trend was maintained throughout the first 15 months of the pandemic—January 2020 to March 2021—by exploring the evolution of publication practices during this period. We also enquired on the origin for the abovementioned elongation of publication delays and found a novel phenomenon, addressed herein as the early bird effect, that produced an ephemeral perception of a global rush in scientific publishing during the early pandemic. We will show that this new effect lays beyond the involvement of any publisher policy action and is directly linked to the emerging nature of new and ‘hot’ scientific topic.

METHODS

Scope

The scope of this study is biomedical literature related to the SARS-CoV-2 research, namely, peer-reviewed journal and review articles indexed in PubMed and their associated preprints posted on bioRxiv and medRxiv servers, a total of 4,031 deduplicated journal article-preprint pairs. Since our study focuses on publication delays experienced by manuscripts as they transition from preprints to peer-reviewed journal articles, we do not include in our analysis “unpublished” preprints that do not have peer-reviewed articles associated with them. Based on the 70% publication rate for bioRxiv preprint server (Sever et al., 2019), we believe, we covered the majority of COVID-19 publications associated with two major preprint servers for biomedical literature.

Timeline

Two data sets are discussed herein. One of them was collected on 4 May 2021 and it includes a total of 4,031 deduplicated journal article-preprint pairs, where preprints were posted on preprint servers from 1 January 2020 to 31 March 2021. Another data set was collected on 19 October 2020 and it includes 1,099 journal article-preprint pairs, where preprints were posted on preprint servers from 1 January 2020 to 30 September 2020. The October 2020 data set has been discussed in our preprint (Sevryugina & Dicks, 2021b) and its data analyses were posted on Zenodo (Sevryugina & Dicks, 2021a).

Data sources

This paper examines data acquired from a number of sources, including the database of COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv (BioRxiv API, 2021), Crossref (Crossref REST API, 2021), E-utilities (Bethesda, 2010), Dimensions (Herzog et al., 2020), CORD-19 (Wang et al., 2020) and CADRE (Mabry et al., 2020).

Metadata for each individual COVID-19 preprint deposited to bioRxiv or medRxiv was gathered by accessing the bioRxiv database of COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv, to which we will further refer as BioRxiv API (BioRxiv API, 2021). Data were retrieved in JavaScript Object Notation (JSON) format. Data analysis and visualization was done in Python (pandas, numpy, requests, matplotlib, bokeh and seaborn) using Jupyter Notebook.

Crossref (Crossref REST API, 2021) is an official DOI registration agency of the International DOI Foundation that establishes a cross-publisher citation linking system for academic that include journals, conference proceedings, books, data sets, and so forth. It works with thousands of publishers to provide authorized access to their metadata including DOI, publication date and other basic information.

To search PubMed, we used Entrez Programming Utilities (E-utilities) (Bethesda, 2010), an application programming

interface (API) that allows searching 38 databases from the National Center for Biotechnology Information (NCBI). We used PubMed (through E-Utilities) to obtain metadata on peer-reviewed articles of ‘Journal Article’ and ‘Review’ article types as the most traditional types of scholarly output. According to Kun’s (2020) estimates, these two types constitute about 24% of all PubMed publications that include 187 different publication types (NLM, 2020). We used the single-term search query ‘COVID-19’ followed on the recommendations by Lazarus et al. (2020). From E-Utilities, data were downloaded via CSV and converted to Microsoft Excel for further analysis and visualization.

Dimensions (Herzog et al., 2020) is a comprehensive database that links scholarly outputs to a research analytics suite to track the impact of research across its life cycle. Dimensions tracks many preprint servers (Altmetric, 2020) but we only used it for bioRxiv and medRxiv preprints (Data Flow Chart in Data S1).

CORD-19 or COVID-19 Open Research Data set (Wang et al., 2020) is a free resource of over 200,000 scholarly articles about COVID-19, SARS-CoV-2 and related coronaviruses prepared by the Allen Institute for AI (AI2) in collaboration with many partners and released on 16 March 2020. We used its 4 March 2021 release downloaded on 26 April 2021 from CADRE (Mabry et al., 2020) for metadata associated with refereed journal articles.

Data availability

Source data for all figures have been provided in supporting files that were deposited in a Zenodo repository with DOI 10.5281/zenodo.6415280.

Analysis of published preprints

When a preprint is published in a peer-review journal, a reference to the new DOI of the journal article appears next to its title, and DOIs of a preprint and a published article are permanently linked in indexing platforms and tools, which pull from various APIs. We found that the most reliable method of extracting metadata about each individual preprint was by accessing the BioRxiv API (BioRxiv API, 2021). Using the Python library requests, we were able to extract information about each preprint based on DOI, which gave us a column called ‘published.’ Within this column, if the preprint was also published in a journal, the metadata provided the DOI that corresponded to the published version of the paper. To ensure we found all published preprints, we also accessed data from Crossref, Dimensions and CORD-19 APIs. To establish the linkage between the preprints and corresponding peer-reviewed journal articles we performed both, DOI and title matching. All channels were then combined and duplicates were dropped. For a detailed demonstration of data obtained by every data channel, see published collections in Data S1.

To validate whether we found all peer-reviewed preprint versions based on a combination of Crossref, CORD-19, Dimensions and BioRxiv API, we randomly selected a sample of 100 preprints

that our data returned as ‘unpublished’ from both bioRxiv and medRxiv, and searched Google Scholar by title. Our analysis of ‘unpublished’ preprints returned 10% of bioRxiv and 4% of medRxiv preprints as being published in refereed journals. All found journal publications had slight modifications in article titles or authors’ list, and the original ‘unpublished’ preprints were not linked on preprint servers to the corresponding published versions. In comparison, this false-negative rate is lower than the 37.5%, reported by Abdill and Blekhan (2019) and is similar to the 9.1% rate reported by Cabanac et al. (2021). All manually found journal article versions of ‘unpublished’ preprints were manually added to data discussed in this article.

Double DOIs

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

Analysis of publication delays

Preprints posting dates were extracted from Crossref. For journal articles received, accepted and published online dates, we used E-Utilities: PubMedPubDate@PubStatus = ‘received’; PubMedPubDate@PubStatus = ‘accepted’ and ArticleDate@DateType = ‘Electronic’. When ArticleDate@DateType = ‘Electronic’ from PubMed was not available, we substituted it with the ‘created-date’ from Crossref. A detailed description of our selection process for the appropriate dates was reported in our preprint (Sevryugina & Dicks, 2021b).

Pre-submission time (t_{α})

Interval between the date when a preprint is deposited to the server and the date when it is submitted to the journal. Pre-submission time = date the journal article was ‘received’–preprint deposition date.

Review time (t_R)

Interval between the date when manuscript is submitted to the journal and the date it is accepted for publication. Review time = date the journal article was ‘accepted’–date the journal article was ‘received’.

Production stage time (t_{β})

Interval between the acceptance date for a manuscript and the date the peer-reviewed journal article appears online. Production stage time = date the journal article was posted online (‘Electronic’)–date the journal article was ‘accepted’.

Elapsed time (T_{Σ})

Interval between the date when a preprint was deposited to the server and publication date for its journal article analogue. Elapsed time = date the journal article was posted online (‘Electronic’)–preprint deposition date.

Altmetric data

Altmetric Attention Scores were retrieved from Dimensions by querying for articles published during the 1 January to 30 September 2020 period using the recommended query for COVID-19 (DSL, 2020). We matched DOI’s of articles found by Dimensions to DOI’s of articles we identified earlier as published preprints.

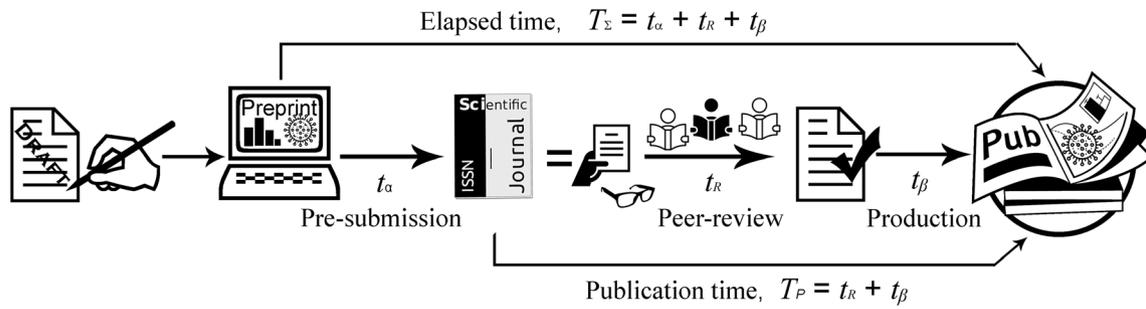
Statistical analysis

Descriptive analysis of the data and Student’s *t* test were conducted on the Statistical Package for Social Sciences version 27 (SPSS).

RESULTS AND DISCUSSION

Our study is based on data we collected on 4 May 2021 that includes peer-reviewed journal articles indexed in PubMed as related to the SARS-CoV-2 research and their associated preprints posted on bioRxiv and medRxiv servers from 1 January 2020 to 31 March 2021, a total of 4,031 deduplicated journal article-preprint pairs. The inclusion of preprint data in our study allows for a glimpse into the pre-submission history of draft manuscripts through their timestamping on preprint servers. In discussing the various publication delays involved in scholarly publishing, we will further refer to the simplified model depicted in Scheme 1, where the elapsed time (T_{Σ}) is made up of three components: pre-submission time (t_{α}), peer-review time (t_R), and production stage time (t_{β}), where the combined ($t_R + t_{\beta}$) represents the publication time (T_P).

The 15-month assessment of publication delays yielded medians of 67 days for peer-review time (t_R), 82 days for publication time (T_P), and 104 days for elapsed time (T_{Σ}), in agreement with the abovementioned trend of the overall elongation of publication delays as the pandemic evolved. To shed some light onto the observed trend, we dissected our data into monthly intervals and noted that publication delays averaged by month followed a convex-shaped curve (Fig. 1). The curvature of the plot revealed two striking features: (i) the earliest as well as the most recent dates both displayed short publication



SCHEME 1 Preprint publication process (Icons, 2021). Pre-submission time (t_α)—interval between the date when a preprint is posted on the preprint server and the date when it is submitted to the journal. Review time (t_R)—interval between the date when manuscript is submitted to the journal and the date it is accepted for publication. Production stage time (t_β)—interval between the acceptance date for a manuscript and the date the peer-reviewed journal article appears online. Publication time (T_P)—interval between the publication date for a journal article and the date when manuscript is submitted to the journal. Elapsed time (T_Σ)—interval between the date when a preprint was deposited to the server and publication date for its peer-reviewed journal article analogue.

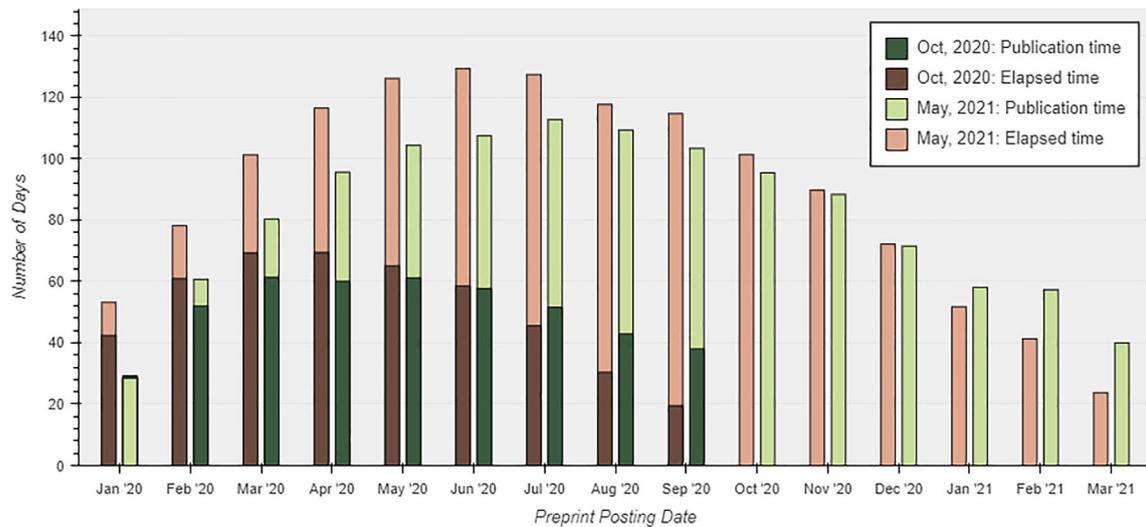


FIGURE 1 Average publication delays by month for COVID-19 journal articles that were posted as preprints since 1 January 2020: T_Σ in shades of brown, T_P in shades of green, and $t_\alpha = T_\Sigma - T_P$. Two data sets are overlapped: one collected in October 2020 includes all preprints posted before 1 October 2020 (dark shades), the other collected in May 2021 includes all preprints posted before 1 April 2021 (light shades).

delays, and (ii) publication delays were longer as compared to the previously reported values. To exemplify (i), the 3-month period, 1 January–31 March, in 2020 and in 2021, yielded mean publication delays much shorter than the 3-month period in-between ($T_P = 72.7$ days and 54.1 days, respectively vs. the 109.6 days during 1 June–31 August 2020). To exemplify (ii), we obtained a median peer-review time of 13 days for COVID-19 works published during the 30 January to 23 April 2020 period versus the previously reported 6 days for the same period (Palayew et al., 2020). Intrigued by these features, we retrieved the data collected in an identical way on 19 October 2020 (Sevryugina & Dicks, 2021b) and confirmed that it displayed similar key characteristics (Fig. 1).

Furthermore, by overlaying the October 2020 and May 2021 data sets, the variation between our two estimates became apparent (Fig. 1); for example, the estimated average publication time (T_P) for COVID-19 works posted on preprint servers in July 2020 is 51.5 versus 112.8 days, respectively. We then subtracted the data we collected in October 2020 from the data we collected in May 2021 and analysed the remainder data (added data (D) in Table 1). We identified two major factors contributing to the observed discrepancy between the two data sets. One is the emergence of previously invisible manuscripts that were undergoing review or production stages at the time of analysis, all of which have a publication date after 30 September 2020. The second factor is related to a persistent problem of unreliable linking

TABLE 1 Average publication delays by month for COVID-19 preprints and corresponding peer-reviewed articles

Month	October data set (d)			Added data (D)			Completeness, %	Student's t test for T_p			
	N	T_p	t_a	N	T_p	t_a		df	t	p	Cohen's d
Jan 2020 ^a	14	29.2	3.3	3	23.5	8.5	82				
Feb 2020	70	51.9	9.5	36	78.0	39.0	66	34.4	1.66	0.11	
Mar 2020	174	61.3	8.8	108	113.1	50.5	62	103.8	5.18	<0.001	0.85
Apr 2020	298	60.0	10.0	256	136.8	42.5	54	313.5	13.70	<0.001	1.35
May 2020	295	61.2	5.2	488	131.0	35.8	58	656.7	17.87	<0.001	1.28
Jun 2020	151	57.7	0.4	401	126.8	32.1	27	417.6	16.01	<0.001	1.27
Jul 2020	75	51.5	-2.1	392	127.4	20.8	16	290.1	15.87	<0.001	1.22
Aug 2020	20	42.9	-12.9	317	113.2	18.1	6	20.1	8.82	<0.001	1.38
Sep 2020 ^a	2	38.0	-18.5	310	103.9	12.9	1				
Oct 2020				205	95.4	6.8					
Nov 2020				193	88.4	4.3					
Dec 2020				107	71.5	2.1					
Jan 2021				71	58.1	-2.5					
Feb 2021				24	57.3	-11.1					
Mar 2021				22	40.0	-16.2					
Overall	973	58.4	5.8	2,467	113.5	22.9		2,690.4	25.01	<0.001	0.98

Note: October data set (d) includes preprint posting dates 1 January–30 September 2020. The added data is a difference between two data sets, one collected in May 2021 and the other collected in October 2020. The Student's t test is used to compare publication delays (T_p) between the October data set (d) and added data (D). N corresponds to the number of analysed preprint-journal article pairs.

^a Not enough papers to support the Student's t test; p is two-sided. % Completeness = $d/(d + D)$.

between preprints and their peer-reviewed journal article versions (Lin et al., 2020). Despite our best efforts, title matching across multiple databases was only partially effective in identifying the missing links (Sevryugina & Dicks, 2021b). During the 6-month period between our two data collections, preprint servers continued to link preprints and their journal article versions enabling us to re-discover older manuscripts that previously appeared as lacking preprint analogs. This second factor accounts for the unexpectedly short publication delays found for the added data, such as the mean publication time of 78 days in February 2020 that implies those articles were already published in October 2020, but were not included in our October 2020 data set. Both aforementioned factors are well-known culprits for introducing systematic errors in publication data (Abdill & Blekman, 2019; Sevryugina & Dicks, 2021b). The data incompleteness resulting from these two factors is the main reason for observing the aforementioned elongation of publication delays with widening of the analysis window.

The most intriguing feature of the data plot in Fig. 1 is its convex shape. Taking into account that the most recent publication delays are subject to data incompleteness, we focused on publication delays during the early pandemic. We propose their low values can be explained through a phenomenon related to what we will further refer herein to as early bird manuscripts.

Those are fast-tracked articles published within up to 2 months from their submission date ($T_p < 60$ days) and in the journal of their first choice (median $t_a = 3$ days). Despite being inherent entities of the publishing landscape, early bird manuscripts represent only a small fraction of the total published works; for example, about 24% of manuscripts in our May 2021 data set could be considered early birds. We found that the mean Altmetric Attention Score for early bird manuscripts related to the SARS-CoV-2 research exceeds threefold that for the remainder of manuscripts on the same topic (393.4 vs. 145.3, respectively). The latter explains our observation of negative pre-submission delays for early bird manuscripts (Fig. 1), an indication that their authors prefer posting them on preprint servers after their submission to journals, likely, in an attempt to avoid scooping (Anderson, 2019); for example, among the most recently posted preprints (1 February–31 March 2021), 51% have $t_a < 0$ as compared with 24% of preprints posted during the 1 April–31 December 2020 period.

The practice of promoting the visibility of important research findings by publishing them rapidly, certainly, is not new and has been exploited by a number of medical (CMAJ, 2001; Ghali & Cornuz, 2000; Goldbeck-Wood & Robinson, 1999; Kassirer & Angell, 1997; McNamee & Horton, 1997; Winker & Fontanarosa, 1999) and multidisciplinary high-profile journals,

among which are *Science* (Science, 2021) and *Nature* (AOP Nat Immunol, 2001; AOP Nat Med, 2002a; AOP Nat Struct Biol, 2002b), in as early as 1997. For example, prior to the coronavirus pandemic, selected articles submitted to *JAMA* were published in 10 to 12 days (Bauchner et al., 2020); and a study identifying the cellular receptor for anthrax toxin submitted to *Nature* was reviewed in 19 days and posted on the Web in 13 days (Bradley et al., 2001); to name a few. During the early months of the coronavirus pandemic, these pre-existing fast-track publication practices proved extremely useful in disseminating the critical knowledge on what was in January simply known as the ‘novel coronavirus’.

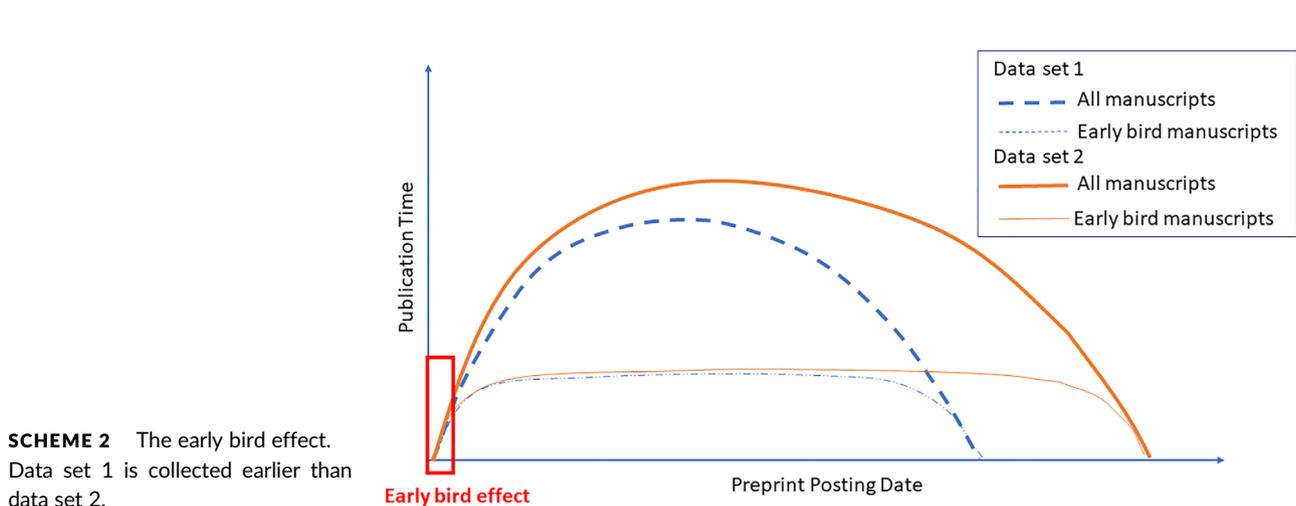
The subtle presence of early bird manuscripts in the publishing landscape becomes most apparent either at the outset of a newly emerging event or in the most recent data; the latter explains the shortening of publication delays as data analysis approaches the most recent dates (Scheme 2). The temporal bias that early bird manuscripts introduce into publication data, observed in shorter than usual publication delays, is normally continuously compensated by new manuscripts that took a long time to be published. However, the sudden onset of the coronavirus pandemic meant that prior to 1 January 2020, no COVID-19 publications existed and early bird manuscripts became then the dominant species of the publishing landscape; indeed, the number of COVID-19 manuscripts reviewed in less than a week was 59% in April 2020 (Palayew et al., 2020), but only 4% by the end of 2020. Along this line of thought, we suggest that short publication delays for COVID-19 manuscripts during the early pandemic are primarily the result of the inadvertent bias introduced by early bird manuscripts, namely the early bird effect (Scheme 2). The latter is a permanent feature of the publishing landscape as early bird COVID-19 manuscripts will forever remain the only publication entities during the early months of the pandemic.

The early bird effect is not unique to the coronavirus pandemic but, we believe, is associated with any new and ‘hot’ topic. As an example, we verified that Zika-related manuscripts were published much faster during the onset of Zika outbreak (mean

(T_p) < 54 days during February–April 2016) as compared to the following 18 months of the Zika virus epidemic (66 days < mean (T_p) < 182 days) (Fig. 2).

Our study further suggests that the observed steady growth of publication delays following the early months of the coronavirus pandemic is not specifically related to a slowdown in scientific publishing, rather it demonstrates a natural evolution of the publishing landscape upon the emergence of a new high-impact topic. Over time, new arrivals, together with the initially present early bird manuscripts, form a complete publication data set capable of representing the diverse collection of journal articles. To evaluate the length of this stabilization period for COVID-19 publication data set, we referred to the data completeness calculated in Table 1. We found that in the October 2020 data set (Sevryugina & Dicks, 2021b), the last 9 months of data, February to October 2020, were subject to data incompleteness with the amount of missing data exceeding 20% (Table 1). Similarly, our May 2021 data set displays a reduction in publication delays starting in August 2020, 9 months prior to the data collection date on 4 May 2021 (Fig. 2). Taking into account the previous studies that showed that as the amount of missing data approaches 20%, the time-dependent bias becomes likely inconsequential (Derrick et al., 2017; Schlomer et al., 2010), we trimmed off the last 9 months of publication data—1 August 2020 to 31 March 2021—which maintained 68.5% of the initial May 2021 data set. The resulting data are time-independent and bias-free as it now represents a collection of manuscripts with various publication experiences.

The revised publication delays (Table 2 and Fig. 3) display only a moderate expediting in publishing COVID-19 works as compared to the pre-pandemic period. While the median peer-review time for COVID-19 manuscripts ($t_R = 66$ days) is smaller as compared to the 100-day benchmark established prior to the pandemic (Powell, 2016), it is significantly longer than that reported in April 2020 ($t_{R,04,20} = 6$ days; Kun, 2020; Palayew et al., 2020) or in September 2020 ($t_{R,09,20} = 37$ days; Sevryugina & Dicks, 2021b). Likewise, while the mean production stage for COVID-19 manuscripts ($t_\beta = 20.4$ days) is 86% faster



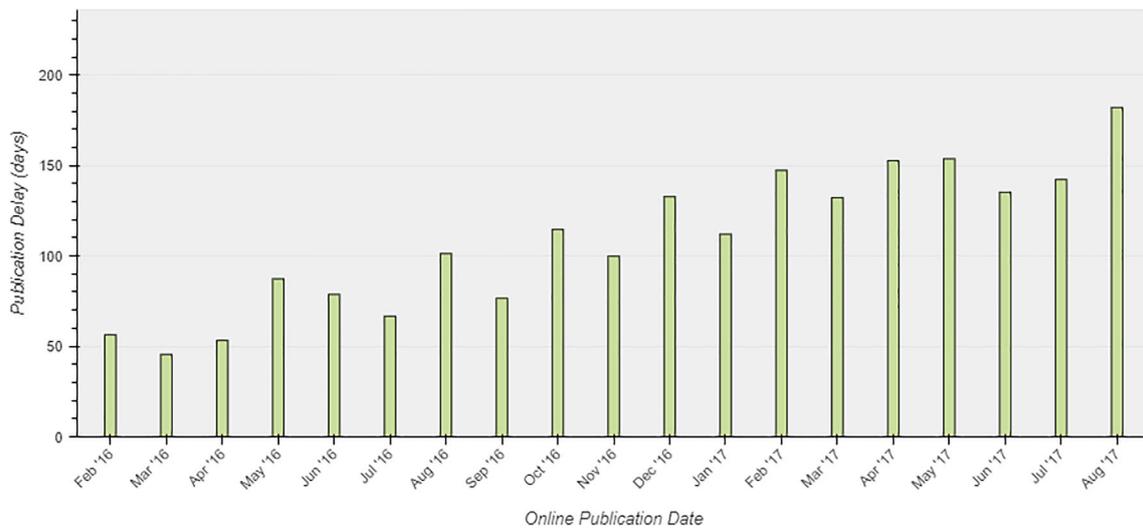


FIGURE 2 Average publication delays (T_P) by month ($N > 4$) for a total of 482 peer-reviewed Zika manuscripts (of which 20% had preprint analogs) published between 1 February 2016 and 31 August 2017.

TABLE 2 Revised publication delays (in days) for COVID-19 manuscripts and their preprint versions

Publication delays	Symbol	Mean	SD	Median	IQR	Mode	N
Pre-submission	t_α	22.6	43.9	7	37	0	2,367
Review	t_R	79.3	59.2	66	79	15	2,312
Production	t_β	20.4	20.1	15	21	4	2,338
Elapsed time	T_Σ	120.3	77.2	109.5	116	42 and 51	2,760

Note: Descriptive statistics reflects the preprint posting period 1 January–31 July 2020.

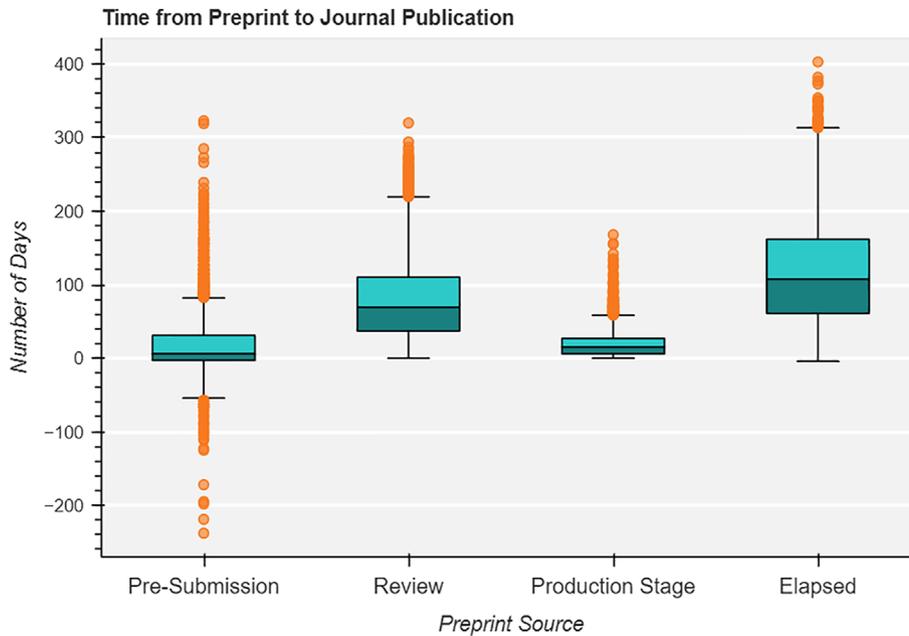


FIGURE 3 Revised publication delays (in days) for COVID-19 manuscripts and their preprint versions. Box plot displays data for the preprint posting period 1 January–31 July 2020.

than that almost a decade ago ($t_{\beta} = 146.6$ days; Björk & Solomon, 2013), it is 40% slower than that in September 2020 ($t_{\beta_{09_20}} = 14.6$ days; Sevryugina & Dicks, 2021b) and 55% slower than that in March 2020 ($t_{\beta_{03_20}} = 9.3$ days; Horbach, 2020). Our current data show that the entire transformation from a COVID-19 preprint to a peer-reviewed journal article, median elapsed time, took 109.5 days, which is 1.6 times longer than the 68 days found in October 2020 (Fraser et al., 2021), and almost five times longer than the 22.5 days reported in April 2020 (Fraser et al., 2020), but significantly shorter than the 150 days reported for Zika or Ebola related preprints (Johansson et al., 2018). Clearly, biomedical journal publishers succeeded in expediting publishing of COVID-19 works, but the magnitude of this success was rather mild.

CONCLUSION

In summary, we observed that publication delays for COVID-19 manuscripts were definitely improved, thanks to various initiatives launched after a few months into the pandemic, but this advance was only incremental. We showed that the alarmingly short publication delays reported during the early pandemic are a result of the early bird effect, an unrecognized bias in publication data by fast-tracked manuscripts that dominated the publishing landscape during the first 2 months of the pandemic. These delays, however, are in no case representative of the whole period of the coronavirus pandemic and do not reflect the publication delays of the entire collection of the published COVID-19 research articles. Our revised publication delays present no reasons for concern in regard to the deterioration of scientific literature, at least based on how fast the COVID-19 works were published. In fact, we believe, the scientific community successfully withheld just another test by skillfully employing the previously developed tools of scholarly communication and data sharing. Finally, we urge scientometrics researchers to consider the early bird effect and data completeness when analysing the publication data.

AUTHOR CONTRIBUTIONS

YS conceived the project and wrote the article; YS and AD developed the methodology, performed searches, study assessments, data extraction, and created the images.

ACKNOWLEDGEMENTS

We thank the University of Michigan Library for the financial support of this project and CADRE Fellowship from the Research Cohort for the Study of Coronaviruses Program for providing access to the COVID-19 data set. We appreciate consultations in statistics by Dr Craig Smith. We are very grateful to Dr Oscar Tutusaus for his assistance with manuscript editing.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Supporting Information.

REFERENCES

- Abdill, R. J., & Blehman, R. (2019). Tracking the popularity and outcomes of all bioRxiv preprints. *eLife*, 8, e45133. <https://doi.org/10.7554/eLife.45133>
- Altmetric. Repositories and preprint servers tracked by altmetric (2020). Retrieved from <https://help.altmetric.com/support/solutions/articles/6000242541-repositories-and-preprint-servers-tracked-by-altmetric>
- Anderson, K. R. (2019). bioRxiv: Trends and analysis of five years of preprints. *Learned Publishing*, 33, 104–109. <https://doi.org/10.1002/leap.1265>
- AOP. (2001). Advance online publication. *Nature Immunology*, 2(11), 979. <https://doi.org/10.1038/ni1101-979>
- AOP. (2002a). Advance online publication. *Nature Medicine*, 8(7), 641. <https://doi.org/10.1038/nm0702-641b>
- AOP. (2002b). Advance online publication. *Nature Structural Biology*, 9(4), 233. <https://doi.org/10.1038/nsb0402-233>
- Baker, P., Rogers, K., Enrich, D., & Haberman, N. (2020, 6 April). Trump's aggressive advocacy of malaria drug for treating coronavirus divides medical community. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/04/06/us/politics/coronavirus-trump-malaria-drug.html>.
- Bauchner, H., Fontanarosa, P. B., & Golub, R. M. (2020). Editorial evaluation and peer review during a pandemic: How journals maintain standards. *JAMA*, 324(5), 453–454. <https://doi.org/10.1001/jama.2020.11764>
- Bethesda, M.D. (2010). Entrez Programming Utilities Help. National Center for Biotechnology Information (US). Retrieved on 1 June 2021 from <https://www.ncbi.nlm.nih.gov/books/NBK25501/>.
- BioRxiv API (2021). COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. Retrieved 4 May 2021 from <https://connect.biorxiv.org/relate/content/181>.
- Björk, B. C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 7(4), 914–923. <https://doi.org/10.1016/j.joi.2013.09.001>
- Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J., & Young, J. A. (2001). Identification of the cellular receptor for anthrax toxin. *Nature*, 414(6860), 225–229. <https://doi.org/10.1038/n35101999>
- Cabanac, G., Oikonomidi, T., & Boutron, I. (2021). Day-to-day discovery of preprint–publication links. *Scientometrics*, 126(6), 5285–5304. <https://doi.org/10.1007/s11192-021-03900-7>
- Callaway, E. (2020). Will the pandemic permanently alter scientific publishing? *Nature*, 582(7811), 167–168. <https://doi.org/10.1038/d41586-020-01520-4>
- CFP for COVID-19 works (2020). Retrieved from (a) COVID-19 impact on peer review. *Int. J. Public Health* (Springer), <https://www.springer.com/journal/38/updates/17991816>; (b) COVID-19 research rapidly peer-reviewed and published in JMIR journals. *JMIR*, <https://www.jmir.org/announcements/202>; (c) PLOS COVID-19 research. *PLOS*, <https://plos.org/covid-19/>

- (d) COVID-19, epidemics and policies: launch of a Special Section for quick knowledge dissemination. *Eur. Econ. Rev.* (Elsevier, Oct 2020), <https://www.journals.elsevier.com/european-economic-review/announcements/covid-19-epidemics-and-policies-launch-of-a-special-section>.
- Chirico, F., Teixeira da Silva, J. A., & Magnavita, N. (2020). "Questionable" peer review in the publishing pandemic during the time of COVID-19: Implications for policy makers and stakeholders. *Croatian Medical Journal*, 61(3), 300–301. <https://doi.org/10.3325/cmj.2020.61.300>
- CMAJ. (2001). With all due haste: Fast tracking at CMAJ [editorial]. *Canadian Medical Association Journal*, 164(8), 1117. Retrieved from www.cmaj.ca/cgi/content/full/164/8/1117
- Crossref REST API (2021). Retrieved from <https://www.crossref.org/education/retrieve-metadata/rest-api/>
- Day, M. (2020). COVID-19: Ibuprofen should not be used for managing symptoms, say doctors and scientists. *The BMJ*, 368, m1086. <https://doi.org/10.1136/bmj.m1086>
- Derrick, B., Russ, B., Toher, D., & White, P. (2017). Test statistics for the comparison of means for two samples that include both paired and independent observations. *Journal of Modern Applied Statistical Methods*, 16(1), 137–157. <https://doi.org/10.22237/jmasm/1493597280>
- DGS-urgent (2020). Service pour les professionnels de la santé. Retrieved from <https://dgs-urgent.sante.gouv.fr/dgsurgent/inter/detailsMessageBuilder.do?id=30500&cmd=visualiserMessage>.
- Disis-Oliveira, R. J. (2020). COVID-19 research: Pandemic versus "paperdemic", integrity, values and risks of the "speed science". *Forensic Sciences Research*, 5(2), 174–187. <https://doi.org/10.1080/20961790.2020.1767754>
- Drake, T. M., Fairfield, C. J., Pius, R., Knight, S. R., Norman, L., Girvan, M., Hardwick, H. E., Docherty, A. B., Thwaites, R. S., Openshaw, P. J. M., Baillie, J. K., Harrison, E. M., & Semple, M. G. (2021). Non-steroidal anti-inflammatory drug use and outcomes of COVID-19 in the ISARIC clinical characterisation protocol UKcohort: A matched, prospective cohort study. *The Lancet Rheumatology*, 3(7), E498–E506. [https://doi.org/10.1016/S2665-9913\(21\)00104-1](https://doi.org/10.1016/S2665-9913(21)00104-1)
- Dreisbach, E.N. (2020, 19 June). Journal retractions 'due to a rush to publish' information on COVID-19 [Infect. Dis. News]. Retrieved from <https://www.healio.com/news/infectious-disease/20200618/qa-journal-retractions-may-be-due-to-rush-to-publish-amid-pandemic>.
- DSL (Dimensions Search Language) (2020). Real-world query used to extract publications related to COVID-19. Retrieved from <https://api-lab.dimensions.ai/cookbooks/1-getting-started/5-Deep-dive-DSL-language.html>.
- EASE (European Association of Science Editors) (2020). EASE statement on quality standards. Retrieved from <https://ease.org.uk/publications/ease-statements-resources/ease-statement-on-quality-standards/>
- Eisen, M. B., Akhmanova, A., Behrens, T. E., & Weigel, D. (2020). Publishing in the time of COVID-19. *eLife*, 9, e57162. <https://doi.org/10.7554/eLife.57162>
- EurekaAlert. AAAS (2020). C19 Rapid Review Initiative expands to include 20 publishers and organizations. Retrieved from https://www.eurekaalert.org/pub_releases/2020-08/hl-crr081220.php.
- Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet Respiratory Medicine*, 8(4), e21. [https://doi.org/10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8) Erratum in: *Lancet Respir. Med.* 2020, 8(6), e54.
- Fauci, A. S., Lane, H. C., & Redfield, R. R. (2020). Covid-19 - navigating the uncharted. *The New England Journal of Medicine*, 382(13), 1268–1269. <https://doi.org/10.1056/NEJMe2002387>
- Fidahic, M., Nujic, D., Runjic, R., Civljak, M., Markotic, F., Lovric Makaric, Z., & Puljak, L. (2020). Research methodology and characteristics of journal articles with original data, preprint articles and registered clinical trial protocols about COVID-19. *BMC Medical Research Methodology*, 20(1), 161–169. <https://doi.org/10.1186/s12874-020-01047-2>
- Fraser, N., Brierley, L., Dey, G., Polka, J.K., Pálffy, M., Nanni, F., & Coates, J.A. (2020). Preprinting the COVID-19 pandemic. *BioRxiv* 111294. Retrieved from <https://doi.org/10.1101/2020.05.22.111294v2>
- Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálffy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biology*, 19(4), e3000959. <https://doi.org/10.1371/journal.pbio.3000959>
- Gautret, P., Lagier, J. C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., Doudier, B., Courjon, J., Giordanengo, V., Vieira, V. E., Tissot Dupont, H., Honoré, S., Colson, P., Chabrière, E., La Scola, B., Rolain, J. M., Brouqui, P., & Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: Results of an open-label non-randomised clinical trial. *International Journal of Antimicrobial Agents*, 56(1), 105949. <https://doi.org/10.1016/j.ijantimicag.2020.105949>
- Ghali, W. A., & Cornuz, J. (2000). Early uptake of research findings after fast-track publication. *Lancet*, 355, 579–580. [https://doi.org/10.1016/S0140-6736\(05\)73234-7](https://doi.org/10.1016/S0140-6736(05)73234-7)
- Goldbeck-Wood, S., & Robinson, R. (1999). BMJ introduces a fast track system for papers. *BMJ*, 318(7184), 620. <https://doi.org/10.1136/bmj.318.7184.620>
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), 387–395. https://doi.org/10.1162/qss_a_00020
- Horbach, S. P. G. M. (2020). Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19. *Quantitative Science Studies*, 1(3), 1056–1067. https://doi.org/10.1162/qss_a_00076
- Icons (2021) (a) Borewicz S. Scientific journal icon, CC-BY-SA (2014) [Internet]. Available from: Wikimedia Commons; (b) Designed by Freepik from Flaticon. Coronavirus, CC-BY [Internet]. Available from [Flaticon.com](https://www.flaticon.com/free-vector/coronavirus); (c) Adioma. Document Review, CC-BY [Internet]. Available from [Adioma.com](https://www.flaticon.com/free-vector/document-review); (d) Kyle Scott. Glasses, CC-BY [Internet]. Available from [Thenounproject.com](https://www.flaticon.com/free-vector/glasses); (e) Publication, CC-NC [Internet]. Available from [HiClipArt.com](https://www.flaticon.com/free-vector/publication); (f) Designed by srip from Flaticon. Read, CC-BY [Internet]. Available from [Flaticon.com](https://www.flaticon.com/free-vector/read).
- Johansson, M. A., Reich, N. G., Meyers, L. A., & Lipsitch, M. (2018). Preprints: An underutilized mechanism to accelerate outbreak science. *PLoS Medicine*, 15(4), e1002549. <https://doi.org/10.1371/journal.pmed.1002549>
- Kassirer, J. P., & Angell, M. (1997). Prepublication release of journal articles. *The New England Journal of Medicine*, 337, 1762–1763. <https://doi.org/10.1056/NEJM199712113372409>
- Krumholz, H.M., Bloom, T., & Ross, J.S. (2020). Preprints can fill a void in times of rapidly changing science [STAT]. Retrieved from

- <https://www.statnews.com/2020/01/31/preprints-fill-void-rapidly-changing-science/>.
- Kun, A. (2020). Time to acceptance of 3 days for papers about COVID-19. *Publication*, 8(2), 30. <https://doi.org/10.3390/publications8020030>
- Kupferschmidt, K. (2020). Preprints bring 'firehose' of outbreak data. *Science*, 367(6481), 963–964. <https://doi.org/10.1126/science.367.6481.963>
- Lazarus, J. V., Palayew, A., Rasmussen, L. N., Andersen, T. H., Nicholson, J., & Norgaard, O. (2020). Searching PubMed to retrieve publications on the COVID-19 pandemic: Comparative analysis of search strings. *Journal of Medical Internet Research*, 22(11), e23449. <https://doi.org/10.2196/23449>
- Lin, J., Yu, Y., Zhou, Y., Zhou, Z., & Shi, X. (2020). How many preprints have actually been printed and why: A case study of computer science preprints on arXiv. *Scientometrics*, 124, 555–574. <https://doi.org/10.1007/s11192-020-03430-8>
- Mabry, P. L., Yan, X., Pentchev, V., Van Rennes, R., McGavin, S. H., & Wittenberg, J. V. (2020). CADRE: A collaborative, cloud-based solution for big bibliographic data research in academic libraries. *Frontiers in Big Data*, 3, 42. <https://doi.org/10.3389/fdata.2020.556282>
- McNamee, D., & Horton, R. (1997). Fast-track to publication in *The Lancet*. *Lancet*, 349(9057), 970. [https://doi.org/10.1016/S0140-6736\(05\)62891-7](https://doi.org/10.1016/S0140-6736(05)62891-7)
- Moffitt, M. (2020). WHO reverses advice on ibuprofen and COVID-19. Should you take it? [SFGATE]. Retrieved from <https://www.sfgate.com/science/article/Should-you-take-ibuprofen-if-you-have-COVID-19-15143646.php>.
- NLM National Library of Medicine. Publication Types (2020). Retrieved from <https://www.nlm.nih.gov/mesh/pubtypes.html>.
- OASPA (2020). COVID-19 Publishers Open Letter of Intent – Rapid Review. Retrieved from <https://oaspa.org/covid-19-publishers-open-letter-of-intent-rapid-review/>.
- Palayew, A., Norgaard, O., Safreed-Harmon, K., Andersen, T. H., Rasmussen, L. N., & Lazarus, J. V. (2020). Pandemic publishing poses a new COVID-19 challenge. *Nature Human Behaviour*, 4(7), 666–669. <https://doi.org/10.1038/s41562-020-0911-0>
- Powell, K. (2016). Does it take too long to publish research? *Nature*, 530, 148–151. <https://doi.org/10.1038/530148a>
- ProMED International Society for Infectious Diseases (2019). Undiagnosed Pneumonia – China (Hubei): RFI. Retrieved from <https://promedmail.org/promed-post/?id=6864153%20#COVID19>.
- Retraction Watch (2020). Retracted coronavirus (COVID-19) papers. Retrieved from <https://retractionwatch.com/retracted-coronavirus-covid-19-papers/>.
- Retta, M. (2021). What the pandemic means for paywalls [Columbia Journalism Review]. Retrieved from https://www.cjr.org/covering_the_pandemic/what-the-pandemic-means-for-paywalls.php.
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, 57(1), 1–10. <https://doi.org/10.1037/a0018082>
- Science (AAAS). 2021 About First Release. Retrieved from <https://www.science.org/toc/science/0/0>.
- Sepúlveda-Vildósola, A. C., Mejía-Aranguré, J. M., Barrera-Cruz, C., Fuentes-Morales, N. A., & Rodríguez-Zeron, C. (2020). Scientific publications during the COVID-19 pandemic. *Archives of Medical Research*, 51(5), 349–354. <https://doi.org/10.1016/j.arcmed.2020.05.019>
- Servick, K. (2020). COVID-19 data scandal prompts tweaks to elite journal's review process [Science]. Retrieved from <https://doi.org/10.1126/science.abe8656>.
- Sever, R., Roeder, T., Hindle, S., Sussman, L., Black, K.J., Argentine, J., Manos, W., & Inglis, J.R. (2019). *BioRxiv* 833400 [Preprint]. Retrieved from <https://doi.org/10.1101/833400v1>
- Sevryugina, Y., & Dicks A. (2021b). Publication practices during the COVID-19 pandemic: Biomedical preprints and peer-reviewed literature. *BioRxiv* 427563 <https://doi.org/10.1101/2021.01.21.427563>.
- Sevryugina, Y. V., & Dicks, A. (2021a). *Analysis of biomedical preprints and peer-reviewed literature during the COVID-19 pandemic*. Zenodo. <https://doi.org/10.5281/zenodo.4450311>
- Smart, P. (2020). Publishing during pandemic: Innovation, collaboration, and change. *Learned Publishing*, 33(3), 194–197. <https://doi.org/10.1002/leap.1314>
- Smyth, A. R., Rawlinson, C., & Jenkins, G. (2020). Preprint servers: A 'rush to publish' or 'just in time delivery' for science? *Thorax*, 75(7), 532–533. <https://doi.org/10.1136/thoraxjnl-2020-214937>
- Steinberg, I. (2020, 9 April). Coronavirus research done too fast is testing publishing safeguards, bad science is getting through. *The Conversation*. Retrieved from <https://theconversation.com/coronavirus-research-done-too-fast-is-testing-publishing-safeguards-bad-science-is-getting-through-134653>.
- The Lancet Global Health. (2020). Publishing in the time of COVID-19. *The Lancet Global Health*, 8(7), e860. [https://doi.org/10.1016/S2214-109X\(20\)30260-6](https://doi.org/10.1016/S2214-109X(20)30260-6)
- Voss, A. (2020, 3 April). Statement on IJAA paper. *International Society of Antimicrobial Chemotherapy*. Retrieved from <https://www.isac.world/news-and-publications/official-isac-statement>.
- Wang, L.L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z., Stilson, B., Wade, A.D., Wang, K., Wilhelm, C., Xie, B., Raymond, D., Weld, D.S., Etzioni, O., & Kohlmeier, S. (2020). COVID-19: The COVID-19 open research dataset. *ArXiv* [Preprint]. Retrieved from <https://arxiv.org/abs/2004.10706>.
- Welcome Fund (2020). Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Retrieved from <https://wellcome.org/coronavirus-covid-19/open-data>.
- WHO (2020). Listings of WHO's response to COVID-19. Retrieved from <https://www.who.int/news-room/detail/29-06-2020-covidtimeline>.
- Winker, M. A., & Fontanarosa, P. B. (1999). JAMA-EXPRESS: Rapid peer review and publication. *JAMA*, 281, 1754–1755. <https://doi.org/10.1001/jama.281.18.1754>