The coronavirus pandemic introduced many changes to our society and deeply affected the academic research. For liaison librarians, it meant finding new routes of engagement with our constituents. In this talk, a STEM liaison librarian will address the quality of published COVID-19 publications. This comprehensive study will provide an overview of general trends and statistical facts about the biomedical preprints and peer-reviewed journal articles and changes that the 2020 crisis transpired to publication practices and a culture of preprints in life sciences. Namely, the emergence of a new category of authors, a storm of COVID-19 related publications, and the dangerous speed at which the science was moving during this pandemic that opened the door for opportunistic publication practices, including questioned editorial assessment and peer-review practices. It will be of interest to liaison librarians, librarians who work in scholarly communication, as well as editors, publishers, and open science enthusiasts.

INTRODUCTION

What does it mean to be a liaison librarian today? The answer to this question is complex but the guiding principles are straightforward. To be a liaison librarian is to serve as a communication bridge between the library and academic community. It is to become a true partner and a collaborator in all aspects of campus intellectual life. And importantly, the librarians’ role never stands still.

The coronavirus pandemic introduced many changes to our society and deeply affected the academic research through social distancing policies and stay-at-home orders. Scientific community quickly responded to the call for the SARS-CoV-2 research through restructuring lab spaces and engaging in collaborative projects. Journal publishers and research organizations signed the Statement on Data Sharing in Public Health Emergencies that resulted in partial or complete removal of paywalls associated with the sharing of SARS-CoV-2-related research. The principles of rapid access to research data and publications relevant to the COVID-19 outbreak were further reaffirmed by ensuring a fast-track peer-review process for COVID-19 journal publications.

For liaison librarians, the coronavirus pandemic modified many established communication practices. And yet, being fully engaged in research activities of our constituents has remained the...
most successful and reliable way to support the academic research. It is through this engagement that the author, a STEM liaison librarian, developed a project aimed at exploring publication practices in regard to preprints prior to and during the coronavirus pandemic. The focus of this study were biomedical preprints posted on bioRxiv and medRxiv servers between 1 January 2020 and 31 March 2021 and associated with them journal and review articles indexed in PubMed, a total of 4,031 deduplicated journal article-preprint pairs. Based on the results of this study, the author was able to offer consultations and teach-outs on evolution of publication practices during the critical times.

PREPRINTING PRIOR TO THE CORONAVIRUS PANDEMIC

The lifecycle for any research starts and ends with a scholarly communication, the golden standard of which is a publication in a peer-reviewed journal. Despite its benefits, the peer-review process has been notorious for reviewer bias, lack of agreement between reviewers, harsh criticism concealed by anonymity, multiple cycles of reviews and rejections by different journals, and associated delays and expenses. Over the last decade, preprint platforms have been widely explored as hubs for disseminating scientific findings, especially for early-stage career researchers. The first and the most well-known preprint server arXiv.org was launched in 1991 and, by 2020, it accommodated 1,585,117 preprints in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering, systems science, and economics. The bioRxiv was launched in November 2013 by Cold Spring Harbor Laboratory with the purpose of covering all aspects of life sciences research and, by 2020, it contained 64,879 preprints. The medRxiv preprint server launched in June 2019 by BMJ, Yale University, and Cold Spring Harbor Laboratory to cover all aspects of research in medical, clinical, and related health sciences hosted 788 preprints before the pandemic. BioRxiv and medRxiv are two major biomedical preprint servers and, thus, are the focus of this study.

As part of the submission process, a preprint usually undergoes basic screening for its scope, plagiarism, ethical issues, and compliance, often performed as quickly as within 24 hours, after which it is published online with a digital object identifier (DOI) that allows it to be citable and trackable. Once posted on a preprint server, preprint can be read, commented on-site or by email, and further shared on the Web and through social media (e.g., Twitter). Additionally, any revision of a preprint's content or status (e.g., publication in a peer-reviewed journal) is time-stamped. The top four concerns with preprint servers include scooping of scientific results, poor data quality, spread of misinformation, and non-obvious deposition policies. Additionally, many authors refrain from depositing preprints due to uncertainty on whether a preprint will be accepted for a journal publication after being posted on a server. In spite of those concerns, it has been widely recognized that submitting a preprint:

1. Allows authors to establish the scientific priority by timestamping the first public record of the research study;
2. Provides authors with a community feedback that can help them to further improve the manuscript quality;
3. Expedites research sharing;
4. Increases research visibility;
5. Streamlines the journal submission process;
6. Allows sharing studies that are difficult to publish in traditional journals (works-in-progress, negative results, replications, contradictions);
7. Provides an open-access publication record.

It is those benefits that catalyze the adoption of preprint servers as suitable platforms for timely, open, and transparent scientific communication.

PANDEMIC-INITIATED EDITORIAL AND FUNDER POLICY CHANGES

With the embrace of preprints as means of scholarly communication, publishers, funders, and scientific organizations set out to improve three aspects of preprinting: promotion, credibility, and discoverability. Even prior to the pandemic, library system vendors such as EBSCO, ProQuest, Ex Libris, and OCLC WorldCat invested substantial resources into the development of discovery tools for a variety of open access sources, including
preprints.\textsuperscript{26} With the onset of the coronavirus pandemic, in June 2020, the National Institutes of Health (NIH) launched the Preprint Pilot to index in PubMed those preprints that come from the NIH-funded projects.\textsuperscript{27} In 2023, Clarivate Plc added the Preprint Citation Index to the Web of Science platform.\textsuperscript{28} While inviting journal submissions from preprint servers was not unusual prior to the pandemic,\textsuperscript{29} during it, this practice helped to ensure the credibility of published research. Thus, in June 2020, MIT Press and the Berkeley School of Public Health launched a new COVID-19 journal, Rapid Reviews: COVID-19, which, after a thorough peer-review, published preprint articles with good research and discredited those with bad.\textsuperscript{30} In response to public concerns about the quality of un-refereed preprints,\textsuperscript{31} the Sinai Immunology Review Project initiated review and validation of the COVID-19 related preprints posted to medRxiv and bioRxiv servers.\textsuperscript{32} Furthermore, the Review Commons platform, launched in December 2019, has been offering preprint authors to request journal-independent portable peer reviews.\textsuperscript{33} Lastly, it is worth mentioning the editorial policies from eLife that resulted in the launch of Preprint Review\textsuperscript{34} and made the preprint deposition mandatory prior to a journal submission.\textsuperscript{35}

RESULTS AND DISCUSSION

Pandemic Publishing

Since an outbreak of pneumonia of unknown origin in Wuhan, China, on December 2019, over 45,000 journal articles on COVID-19 have been indexed in PubMed and over 26,000 preprints posted to medRxiv and bioRxiv servers. Submissions of new preprints on the topic of COVID-19 increased from 10-20 a day in February 2020 to about 150 a day in May 2020 (Fig. 1). Overall, in 2020, the medRxiv server collected 9,152 COVID-19-related preprints and bioRxiv – 2,894. Moreover, the fraction of posted bioRxiv and medRxiv preprints with respect to journal and review articles indexed in PubMed increased from 11% in January 2020 to 43% in November 2020. These figures attest an outstanding adoption of preprints in scholarly communication practices of the biomedical community in response to the COVID-19 outbreak.

\textbf{FIGURE 1}


The majority of COVID-19 related preprints were posted in categories most relevant to the coronavirus research: \textit{microbiology}, \textit{bioinformatics}, and \textit{immunology} on bioRxiv; and \textit{infectious diseases}, \textit{epidemiology}, and \textit{public and global health} on medRxiv (Fig. 2). Notably, these categories were largely unpopulated prior to the pandemic,\textsuperscript{36} suggesting an emergence of a new community of preprint authors. This agrees with an earlier report of Fraser et al. who found that 83% of COVID-19 preprint authors were posting a preprint for the first time.\textsuperscript{37} It is worth noting that the same study also found that most corresponding authors were not switching to COVID-19 research from other fields.
FIGURE 2

Daily COVID-19 Preprint Submissions by a Preprint Category: (A) bioRxiv and (B) medRxiv
(1 Jan – 30 Sep 2020)
**Exploring Authors’ Motives toward Preprinting**

Among the motives for accelerated preprinting during the pandemic, the author set to explore three: (i) opportunity to publish works-in-progress that would otherwise be difficult to publish in traditional peer-reviewed journals; (ii) access to public feedback that could further improve the manuscript prior to a journal submission, and (iii) rapid dissemination of research results.

The first option was explored through the analysis of the so-called “published” preprints—those that, in addition to being posted on preprint servers, are also submitted to peer-reviewed journals and later emerge as journal articles. A percentage of “published” COVID-19 preprints was used as an indicator for their completeness. Data analysis showed a variation in publication rates across preprint categories (Fig. 3). The mean publication rates for COVID-19 preprints were 34% and 29% for bioRxiv and medRxiv servers, respectively (as of December 7, 2020). These publication rates were lower than those reported for bioRxiv preprints prior to the pandemic (42% or 70%). Furthermore, publication rates for the COVID-19 preprints showed to be lower than for preprints associated with previous health crises (e.g., 60% for Ebola and 48% for Zika). This finding suggests that many of the COVID-19 preprints posted during the early pandemic reported incomplete, in-progress studies, which is also in agreement with Fraser *et al.*

![Published vs. Not Published COVID-19 Preprints by Category: (A) bioRxiv and (B) medRxiv](image)

(1 Jan – 30 Sep 2020). Preprints that have journal article analogues — in yellow, and those that do not — in blue. The total number of preprints in the category is the sum of yellow and blue bars. Only categories with at least 100 preprints are displayed for clarity purposes.
To assess the second motivational factor, namely, whether coronavirus researchers used bioRxiv and medRxiv preprint servers to gather public feedback, we examined the pre-submission time for the COVID-19 preprints. Pre-submission delay is a time interval between the date when a preprint is posted on the server and the date when it is submitted to a journal, and it is positive when preprint is posted on a server prior to the journal submission. Data analysis indicated that authors of the COVID-19 preprints posted them on preprint servers 7 days (median) prior to journal submissions (Fig. 4). A more detailed analysis showed that only 35% of preprints were posted more than 10 days in advance of their journal submissions, and 30% of the COVID-19 preprints were deposited to bioRxiv or medRxiv servers after being submitted to journals (negative pre-submission delay). These results resonate with the earlier reports of short pre-submission delays for bioRxiv and arXiv preprints.

This behavior was explained by a widespread fear for “getting scooped” when making details of research publicly available. In this regard, we see no changes in authors’ publication strategy—during the pandemic authors preferred posting COVID-19 manuscripts as preprints concurrently with their submissions to journals. This also implies that authors of the COVID-19 preprints did not specifically pursue collecting pre-submission feedback.

Considering that commenting of preprints was designed as a tool for collecting public feedback, one would expect that preprints with large positive pre-submission delays would collect a fair amount of constructive criticism that would in-turn facilitate the following peer-review process. In other words, if all preprints collected feedback and this feedback was useful, we would find a positive correlation between a preprint’s pre-submission time and the corresponding peer-review time for its journal article version. However, we established no such correlation (Fig. 5). Furthermore, for selected journals we found that the peer-review time was similar for manuscripts that had or did not have preprint versions (Student’s t-test non-significant, Table 1). The explanation was provided by Malicki’s study, which found that less than 10% of all bioRxiv preprints ever receive public comments and only 12% of non-author comments are full review reports that could help in preparing a manuscript for journal publication. Faced with the uncertainty of receiving the meaningful feedback and a prospect of their studies being scooped, authors of the COVID-19 preprints, unsurprisingly, chose posting their preprints along with journal submissions.

Note, that a pre-submission delay may also include delays associated with submissions to multiple journals. Thus, a short pre-submission time means that the majority of published COVID-19 preprints were accepted to journals of authors’ first choice. Only 10% of COVID-19 preprints in this study had pre-submission delays over an average review time of 70 days.

In summary, our analysis showed that using preprints to quickly disseminate the critical COVID-19-related research showed successful, and yet, posting a preprint did not facilitate its peer-review process when it was submitted to a journal. Even though most of the early preprints on COVID-19 were works-in-progress, their

**FIGURE 4**

Pre-submission Time for COVID-19 Published Preprints (in days): bioRxiv (red) and medRxiv (blue) (1 Jan 2020 – 31 Mar 2021). The 0 date is the date the preprint was submitted to the peer-reviewed journal.
authors did not pursue collecting feedback through commenting options available on preprint servers. The most likely reasons for accelerated adoption of preprints during the early pandemic are: (i) facilitated process of posting preprints along with journal submissions; and (ii) genuine inspiration of biomedical community with a newly emerged ‘hot’ topic.

**Publication Delays**

Another important change in scientific publishing in response to the COVID-19 pandemic was expedited publication process for the COVID-19-related articles. According to the dataset we collected in May 2021, COVID-19 articles associated with preprints were reviewed 50% faster (66 days median) and prepared for the final version 90% faster (15 days median) than biomedical articles were prior to the pandemic (141 days and
147 days, respectively, in 2013). Further, COVID-19 related preprints transformed into peer-reviewed journal publications (a so-called elapsed time) in 109.5 days, significantly faster than prior to the pandemic (155 or 166 days during 2013-2018, $Z = 32-37$, $p < 0.001$). This was not observed for Zika or Ebola related preprints where the median elapsed time for both outbreaks (150 days) was reported to be no different than the normal publication timeline.

Publication delays varied across journals. This is demonstrated below for ten selected journals from the top five publishers that published the majority of the COVID-19 medRxiv and bioRxiv preprints. Among them, *Journal of Clinical Virology* had the shortest peer-review and production stage delays (29.8 and 3.6 days, respectively), while *Scientific Reports* had the longest peer-review of 143.2 days and *Nature Communications* had the longest production stage delay of 29.4 days (Table 2, Fig 6).

**FIGURE 6**

Mean Publication Delays for Selected Journals Publishing COVID-19 Manuscripts. Review time (dark blue bar) and production stage (light blue bar) are displayed in days (1 Jan 2020 – 31 Mar 2021). The entire two-colored bar represents the total publication time.
### TABLE 2

Mean Publication Delays (in days) for Selected Journals Publishing COVID-19 Manuscripts
(1 Jan 2020 – 31 Mar 2021)

<table>
<thead>
<tr>
<th>Journal</th>
<th>Fraction of COVID-19 Articles (%)</th>
<th>N</th>
<th>Pre-submission</th>
<th>Peer-review</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>µ</td>
<td>SD</td>
<td>µ</td>
</tr>
<tr>
<td>J Clin Virol</td>
<td>52</td>
<td>32</td>
<td>10.2</td>
<td>25.1</td>
<td>29.8</td>
</tr>
<tr>
<td>Int J Infect Dis</td>
<td>10</td>
<td>38</td>
<td>40.5</td>
<td>53.0</td>
<td>35.3</td>
</tr>
<tr>
<td>Science</td>
<td>12</td>
<td>53</td>
<td>0.2</td>
<td>18.4</td>
<td>56.5</td>
</tr>
<tr>
<td>Int J Infect Dis</td>
<td>43</td>
<td>59</td>
<td>24.2</td>
<td>43.3</td>
<td>56.8</td>
</tr>
<tr>
<td>Cell</td>
<td>18</td>
<td>39</td>
<td>-2.0</td>
<td>17.5</td>
<td>60.0</td>
</tr>
<tr>
<td>J Med Virol</td>
<td>64</td>
<td>43</td>
<td>12.3</td>
<td>28.0</td>
<td>52.1</td>
</tr>
<tr>
<td>PLOS ONE</td>
<td>6</td>
<td>375</td>
<td>13.0</td>
<td>40.0</td>
<td>111.0</td>
</tr>
<tr>
<td>Nat Commun</td>
<td>3</td>
<td>102</td>
<td>8.5</td>
<td>39.1</td>
<td>117.5</td>
</tr>
<tr>
<td>Sci Rep</td>
<td>2</td>
<td>144</td>
<td>19.5</td>
<td>61.1</td>
<td>143.2</td>
</tr>
</tbody>
</table>

*N = 59; *N = 44.

### FIGURE 7

Journal Scope Categories for Ten Selected Journals Publishing COVID-19 Manuscripts
To analyze the scope of journals that published the COVID-19 preprints, we plotted journal vs. journal’s scope categories, as defined by Scopus, for all published preprints and their article analogues (Fig. 7). Among the selected journals, we observed shorter peer-review delays for journals, which scope was highly relevant to the coronavirus pandemic, such as infectious diseases, virology, microbiology, and immunology. In contrary, journals with general or multidisciplinary scope displayed longer review times. We hypothesized that the advantage of the former group may stem from the availability of expert-reviewers with relevant disciplinary backgrounds. Indeed, according to Scopus, the Journal of Clinical Virology and Clinical Infectious Diseases have disciplinary scope in infectious diseases, while PLOS ONE or Nature Communications have scope in general disciplines (Fig. 7). To validate this observation, we ranked 32 journals that published at least 15 preprints by the relevance of their scope to the COVID-19 topic (based on Scopus categories). This analysis confirmed that journals with scope highly relevant to the coronavirus pandemic were able to provide faster peer-review for COVID-19 manuscripts (41.5 days median) than journals with general or irrelevant scope (84.8 days median). A Mann-Whitney test indicated that this difference was statistically significant (Z = -3.46, p = .001).

In order to assess whether publication time varied between COVID-19 related and unrelated manuscripts, we selected four journals that represented a specific group in terms of the volume of published COVID-19 articles by the journal and the length of a publication process (Table 3).

Our analysis showed that manuscripts unrelated to COVID-19 were published at much slower speed (Table 4, 5). For example, the mean review time in four selected journals varied between 104 and 221 days for articles unrelated to COVID-19 as compared to 37-111 days for COVID-19-related manuscripts (Table 4). At the same time, the mean production stage delays for COVID-19 unrelated manuscripts remained largely unchanged during the pandemic (6-27 days vs. 23.6 days, respectively, Tables 4 and 5).

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Characteristics of Journals Selected for Comparing the Publication Delays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal</td>
<td>COVID-19 Articles as Fraction of Total Publication Volume</td>
</tr>
<tr>
<td>J Clin Virol</td>
<td>high</td>
</tr>
<tr>
<td>Cell</td>
<td>medium</td>
</tr>
<tr>
<td>J Med Virol</td>
<td>high</td>
</tr>
<tr>
<td>PLOS ONE</td>
<td>small</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Descriptive Statistics for the Peer-review Time (in days) for COVID-19-related and Unrelated Manuscripts. Student’s t-test is used for comparing means. All papers and not only those associated with preprints are included.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal</td>
<td>COVID-19 Publications</td>
</tr>
<tr>
<td>N</td>
<td>μ</td>
</tr>
<tr>
<td>J Clin Virol</td>
<td>171</td>
</tr>
<tr>
<td>Cell</td>
<td>107</td>
</tr>
<tr>
<td>J Med Virol</td>
<td>652</td>
</tr>
<tr>
<td>PLOS ONE</td>
<td>1,333</td>
</tr>
</tbody>
</table>

μ: Mean; M: median; SD: standard deviation; N: number of analyzed publications; d: Cohen’s.
Among the lessons learned during the pandemic is that preprints quickly filled an information void created by the emergence of new coronavirus. The majority of the COVID-19 preprints were immediate in-progress findings, not suitable for being published in refereed journals; this conclusion was based on the low publication rate for the COVID-19 preprints. The authors of “published” preprints undervalued the pre-submission feedback and posted preprints on preprint servers within a week from submitting them to journals. The disciplinary journals with a scope highly relevant to COVID-19 were very successful in further accelerating the publication process for the COVID-19 works. It is our hope that as the pandemic recedes, journal publishers of biomedical and related sciences will keep on improving the policies that were most helpful in expediting, revising, and sharing critical research output during the pandemic.

Study Limitations

One limitation to this study is that articles that are undergoing the peer-review or various editorial processes are invisible for data analysis. In the past, Blekhman et al. showed that publication rate for preprints was 67% during the comprehensive study that covered the period of several years (2013-2016) but only 20% during the last year of their study (2018). Indeed, the reported herein publication rates are higher than those reported by two independent studies of the COVID-19 preprints at the end of April 2020 (6.1% for COVID-19 preprints retrieved from Dimensions and 4% for bioRxiv and medRxiv preprints). Similarly, we showed that many early reports on publication practices during the year 2020 were subject to data bias due to the early-bird effect and data incompleteness. We urged researchers to allow a minimum of nine months before starting to interpret publication data.

Data availability. Source data for all figures and tables have been provided in supporting files that were deposited in a Zenodo repository with DOI 10.5281/zenodo.6415280. See also a corresponding publication in Learned Publishing, https://onlinelibrary.wiley.com/doi/10.1002/leap.1483.

Scope. Unless stated otherwise, this study is based on data retrieved from bioRxiv and medRxiv preprint servers and PubMed on 4 May 2021; a total of 4,031 deduplicated journal article-preprint pairs with preprint posting dates between 1 January 2020 and 31 March 2021. For journal articles, we considered only journal and review articles indexed in PubMed that were either linked to the above-mentioned preprints or published between 1 January 2020 and 4 May 2021. Data were retrieved using API access to COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv, Crossref, E-utilities, and Dimensions.

REFERENCES

1. (a) Kai Kupferschmidt, "Preprints bring 'firehose' of outbreak data,” Science 367, no. 6481 (February 2020):963-964; (b) Ewen

---

**TABLE 5**

<table>
<thead>
<tr>
<th>Journal</th>
<th>COVID-19 Publications</th>
<th>Non-COVID-19 Publications</th>
<th>t-test w/ vs. w/o COVID-19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>μ</td>
<td>SD</td>
</tr>
<tr>
<td>J Clin Virol</td>
<td>171</td>
<td>4.1</td>
<td>3.1</td>
</tr>
<tr>
<td>Cell</td>
<td>104</td>
<td>7.7</td>
<td>6.1</td>
</tr>
<tr>
<td>J Med Virol</td>
<td>652</td>
<td>28.0</td>
<td>26.3</td>
</tr>
<tr>
<td>PLOS ONE</td>
<td>1333</td>
<td>21.6</td>
<td>12.7</td>
</tr>
</tbody>
</table>

μ - Mean; M - median; SD - standard deviation; N - number of publications analyzed; d - Cohen’s.


25. Ibid.


29. (a) Gregory S. Barsh, Casey M. Bergman, Christopher D. Brown, Nadia D. Singh, and Gregory P. Copenhaver, “Bringing PLOS


32. Nicolas Larivière, Cassidy R. Sugimoto, Benoit Macaluso, Stasa Milojevic, Blaise Cronin, and Mike Thelwall, “arXiv E-Prints and...”


37. Serge P. J. M. Horbach, “Pandemic Publishing. ”

38. Richard J Abdill and Ran Blekhman, “Tracking the popularity and outcomes of all bioRxiv preprints.”


40. Nicholas Fraser, et al., “Preprinting the COVID-19 pandemic.”

41. Yulia V. Sevryugina and Andrew J. Dicks, “Publication practices during the COVID-19 pandemic. ”

42. Serge P. J. M. Horbach, “Pandemic Publishing.”


48. Adam Palayew, Ole Norgaard, Kelly Safreed-Harmon, Tue Helms Andersen, Lauge Neimann Rasmussen, and Jeffrey V. Lazarus, “Tracking the popularity and outcomes of all bioRxiv preprints.”

49. Nicholas Fraser, et al., “Preprints: An underutilized mechanism to accelerate outbreak science.”