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

SUPPORTING INFORMATION

Yulia V. Sevryugina[1]* and Andrew J. Dicks[2]

[1] University of Michigan Library, Ann Arbor, Michigan, United States of America

[2] School of Information, University of Michigan, Ann Arbor, Michigan, United States of America

* Corresponding author: E-mail: yulias@umich.edu (YS)

1. Data sources

This paper examines data acquired from several sources, including the collection of COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv [1], Crossref [2], E-utilities [3], Dimensions [4], CORD-19 [5], and CADRE [6]. Chart 1 below shows the usage of various data platforms in this study.

Chart 1. Data Flow Chart.
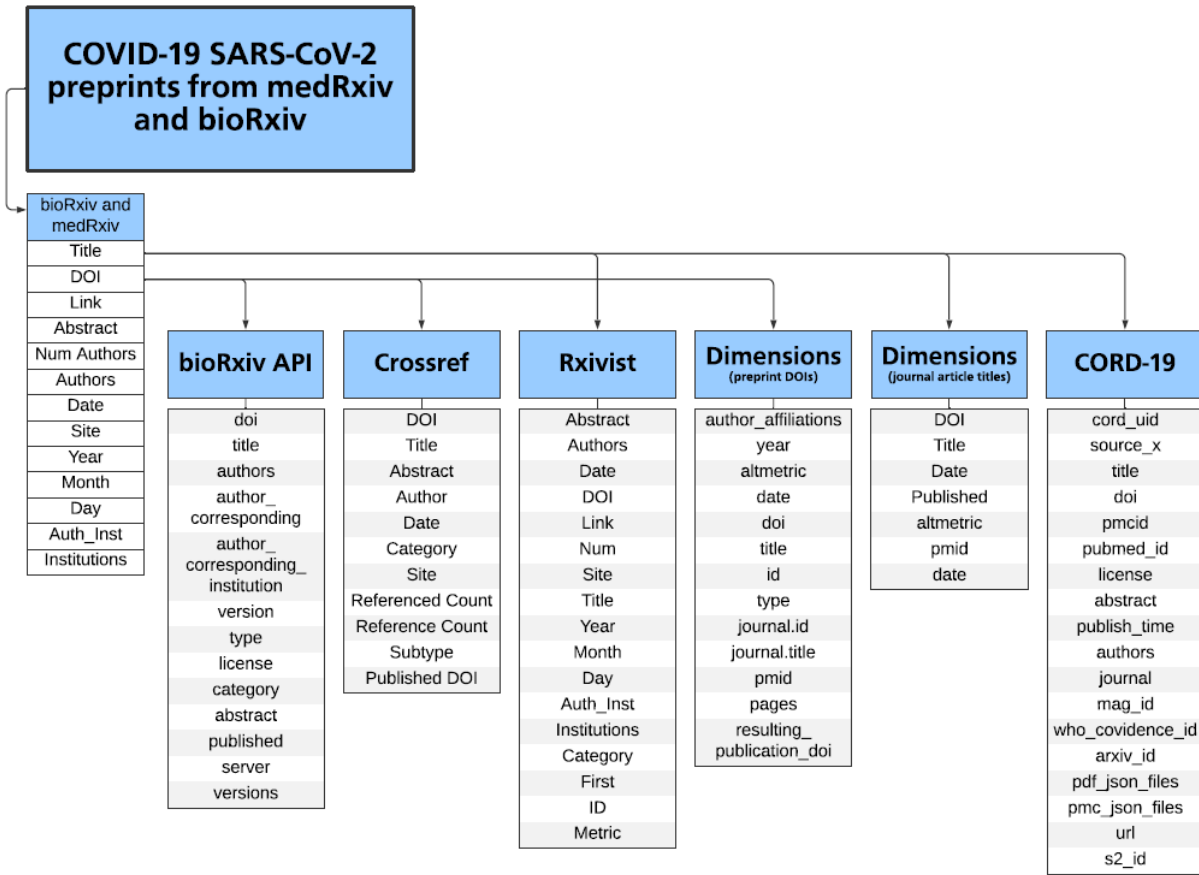
2. Published Collections

Shows how various data were combined to account for all published preprints (Chart 2).

BioRxiv (as of August 20, 2020)

	BioRxiv API	Dimensions (Preprints)	Dimensions (Postprints)	CORD-19
BioRxiv API	**171**	41	69	56
Dimensions (Preprints)	41	**42**	14	15
Dimensions (Postprints)	69	14	**108**	56
CORD-19	56	15	56	**71**

MedRxiv (as of November 29, 2020)

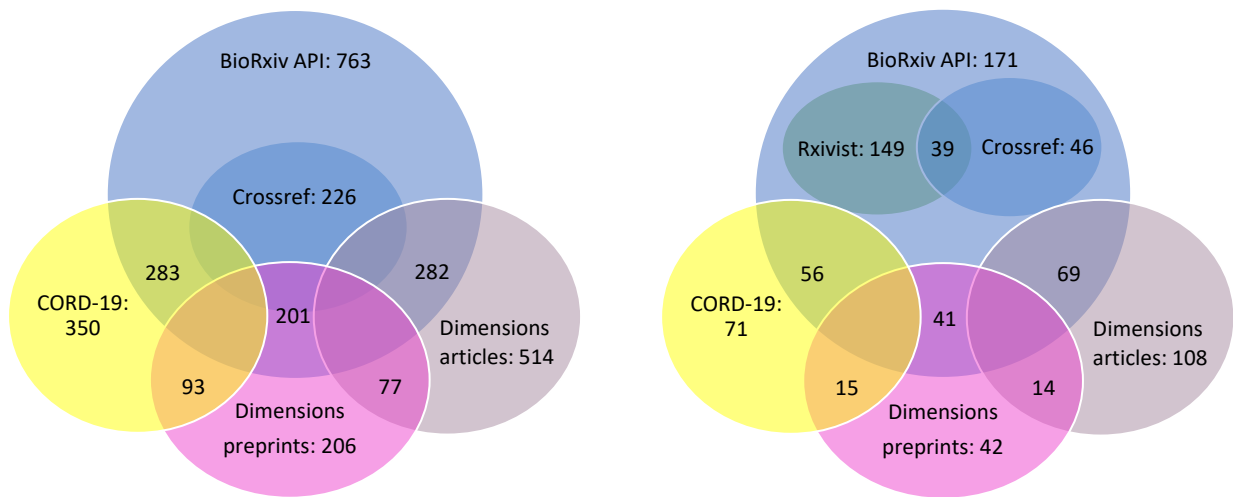|	medRxiv API	Crossref	Dimensions Preprints	Dimensions Postprints	CORD-19
medRxiv API	**763**	226	201	282	283
Crossref		**226**	201	82	105
Dimensions Preprints			**206**	77	93
Dimensions Postprints				**514**	223
CORD-19					**350**

Chart 2. Venn diagrams demonstrating the number of published preprints that were discovered *via* a combination of different sources: for medRxiv preprints (left, as of Nov 29, 2020) and for bioRxiv preprints (right, as of Aug 20, 2020).

3. Analysis of publication delays

It is worth noting that none of the publication delays display a standard Gaussian distribution (Fig. 1), thus we discuss both their medians and means.
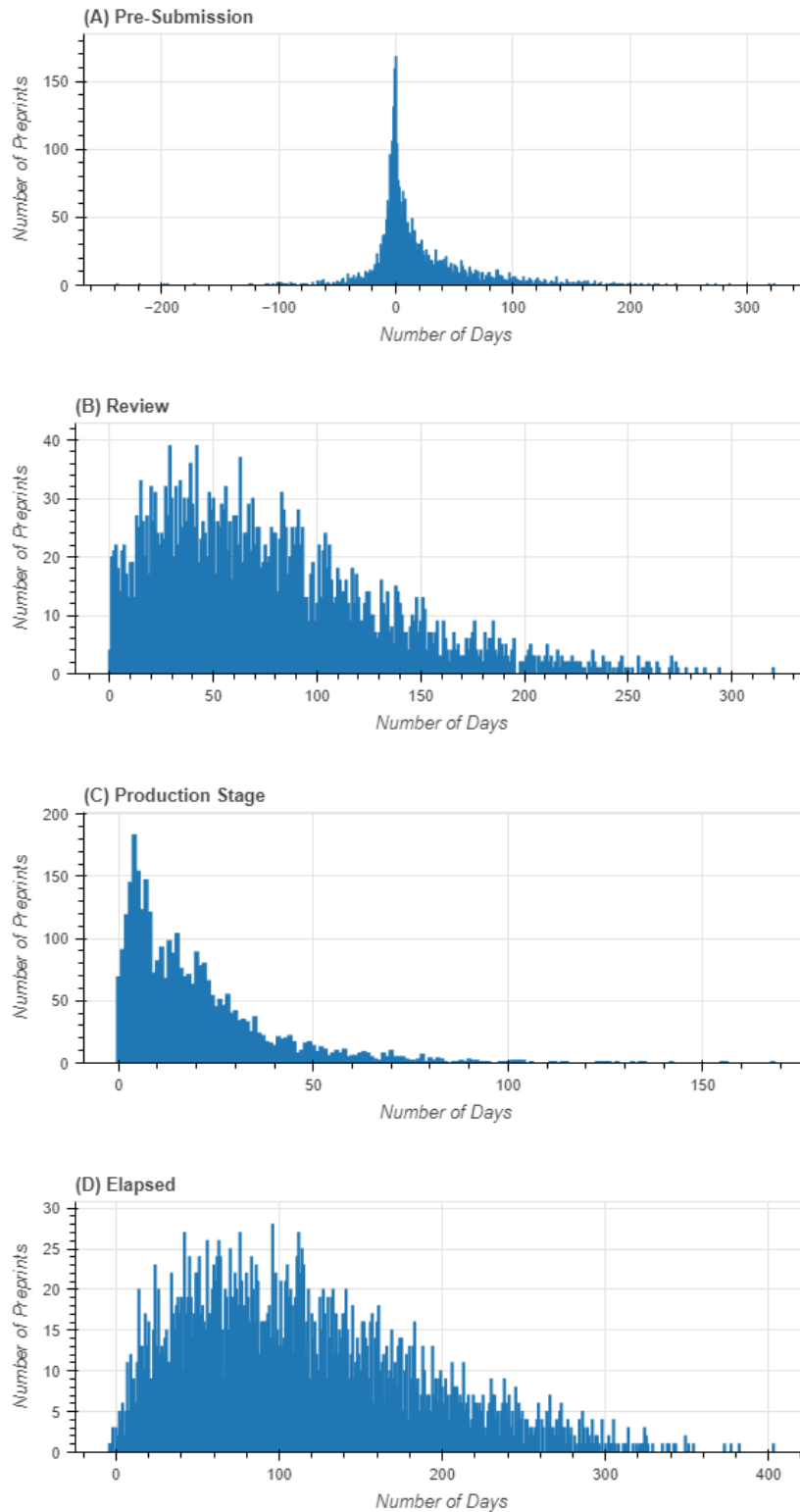
Fig. 1. Publication delays from preprint to journal publication. Number of COVID-19 combined medRxiv and bioRxiv preprints with given number of days that correspond to various

publication delays: (A) *Pre-submission time*; (B) *Review time*; (C) *Production stage time*; and (D)

Elapsed Time.

References:

[1] Database of COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv [Internet]. BioRxiv.org [cited 2020 Oct 7]. Available from: https://connect.biorxiv.org/relate/content/181

[2] Crossref [Internet]. Crossref REST API [cited 2020 Oct 7]. Available from: https://www.crossref.org/education/retrieve-metadata/rest-api/

[3] Entrez Programming Utilities Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2010 [cited 2020 Oct 7]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK25501/.

[4] Mouratidis RW. Dimensions. Journal of the Medical Library Association: JMLA. 2019; 107(3): 459–461. Available from: https://doi.org/10.5195/jmla.2019.695

[5] Lu Wang 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 AD, Wang K, Wilhelm C, Xie B, Raymond D, Weld DS, Etzioni O, Kohlmeier S. CORD-19: The Covid-19 Open Research Dataset. ArXiv [Preprint]. 2020 Apr 22;arXiv:2004.10706v2. PubMed PMID: 32510522.

[6] Mabry PL, Yan X, Pentchev V, Van Rennes R, McGavin SH, Wittenberg JV. CADRE: A Collaborative, Cloud-Based Solution for Big Bibliographic Data Research in Academic Libraries. Frontiers in Big Data. 2020 Nov 20;3:42