



# The Published Scientific Literature on COVID-19: An Analysis of PubMed Abstracts

Mohleen Kang<sup>1</sup> · Saumya S. Gurbani<sup>1</sup> · Jordan A. Kempker<sup>1</sup>

Received: 16 November 2020 / Accepted: 19 November 2020 / Published online: 25 November 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

To the Editor,

Since the first report of a cluster of unexplained pneumonia cases in Wuhan, China on December 31, 2019, there has been a veritable avalanche of research publications on COVID-19 to better understand the virus and its impact. We sought to quantify the pace of publication and describe characteristics of abstracts pertaining to COVID-19 in PubMed since the beginning of the calendar year using an automated approach.

We conducted a search of PubMed using R version 4.0.2 and the National Library of Medicine's (NLM) E-utilities application programming interface (API) and query string: ((wuhan[All Fields] AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])) AND 2019/12[PDAT]: 2030[PDAT]) OR 2019-nCoV[All Fields] OR 2019nCoV[All Fields] OR COVID-19[All Fields] OR SARS-CoV-2[All Fields]. All results from January 1, 2020 to November 02, 2020 were included. Details, including date of publication, country, language, publication type and journal name were extracted.

A total of 57,263 articles were included in our analysis. 19,469 (34.0%) were ahead of print, 14,383 (25.1%) were e-published, and 23,411 (40.9%) were published in print at the time of data extraction. Over the 43-week period, a median of 1682 articles were published per week. There was a peak of 2277 articles published the week of May 11th (Fig. 1). The United States accounted for the most publications (20,460 [35.7%]) followed by England (15,471 [27.0%]) and the Netherlands (4980 [8.70%]). Most publications were in English (56,114 [98.0%]) with small percentages in Spanish (379 [0.66%]), German (231 [0.40%]), and French (225

[0.39%]). The preprint servers, medRxiv (609 [1.1%]) and bioRxiv (479 [0.8%]), were among the top five sources with most publications.

The COVID-19 pandemic has been accompanied by an unprecedented rate of scientific publication that has overwhelmed frontline providers and the public health community. Our analysis found a total of 57,263 articles in PubMed on COVID-19, compared to only 3386 articles regarding Influenza H1N1 pandemic during the initial 43-week period from April 20, 2009 to February 15, 2010 [1].

Using the NLM's official search string in Medline, which is updated daily, we were able to capture the latest articles and ahead of print articles in our analysis. However, limiting our search to a single database also restricted our analysis to mostly English language articles, a majority of which were published in the US and a few European countries. While we did collect data on publication type, it was not adequately characterized for any meaningful interpretation. We also did not include all of the preprint servers in our search; however, PubMed now includes preprints from authors with either affiliation or support from the National Institute of Health as part of a new pilot program [2]. Preprint articles have become an increasingly popular avenue for researchers to share their findings before a formal peer review process and have emerged as significant drivers of discourse in the scientific community.

A number of free and easily downloadable article databases, notably from the Centers for Disease Control and Prevention and even NLM, have been created to help clinicians and researchers parse through the plethora of information on COVID-19 [3]. However, these databases also contain thousands of articles, still making it difficult to efficiently search for answers to specific queries. Since there are still many unanswered questions regarding COVID-19, additional research will be needed. Concurrently, there is an emergent need to evaluate and prioritize the quality of the literature that is being published at an astronomical rate to help ease the burden on the consumers of this information who are

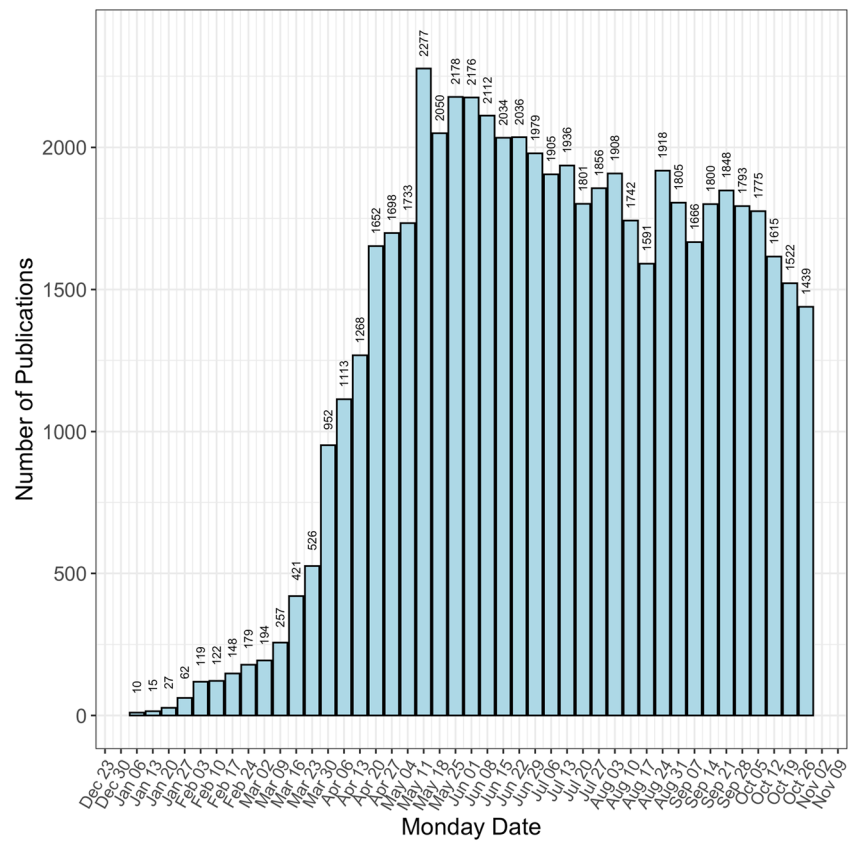
---

This article is part of the Topical Collection on *Education & Training*

✉ Mohleen Kang  
mkang30@emory.edu

<sup>1</sup> Emory University School of Medicine, 615 Michael St., NE Ste 205, Atlanta, GA 30322, USA

**Fig. 1** Weekly number of publications related to COVID-19 from January 1, 2020 to November 02, 2020 in PubMed. Legend: The x axis indicates the weeks from January 1, 2020 to November 02, 2020. Weeks are defined by the Monday to Sunday period with x-axis tick marks indicating the date of the Monday starting each week



trying to make important medical and public health decisions in a constantly changing environment.

**Funding** Dr. Kempker received support from the Agency for Healthcare Quality and Research under Award Number K08HS025240.

**Data availability** The datasets generated during and/or analyzed during the current study is available from the corresponding author.

### Compliance with ethical standards

**Conflict of interest** Dr. Kempker has served as a consultant for Grifols, Inc. Dr. Kang declares that she has no conflict of interest. Dr. Gurbani declares that he has no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Code availability** The custom R code used in the current study is available from the corresponding author.

### References

1. H1N1 citations in PubMed. <https://pubmed.ncbi.nlm.nih.gov/?term=%28%28%28swine+OR+h1n1%29+AND+%28flu+OR+influenza+OR+virus+OR+outbreak+OR+pandemic%29%29+AND+%28%28%222009%2F04%2F20%22%5BDate+-+Publication%5D+%3A+%222010%2F02%2F15%22%5BDate+-+Publication%5D%29%29&sort=date> Accessed November 3 2020.
2. NIH Preprint Pilot. <https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/> Accessed June 10 2020.
3. COVID-19 Research Articles Downloadable Database. <https://www.cdc.gov/library/researchguides/2019novelcoronavirus/researcharticles.html> Accessed June 5 2020.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.