

The Pandemic of the COVID-19 Literature: A Bibliometric Analysis

Running Title: Bibliometric Analysis of the COVID-19 Literature

Elie A. Akl (✉ ea32@aub.edu.lb)

American University of Beirut Medical Center

Lokman I. Meho

American University of Beirut

Sarah H. Farran

American University of Beirut

Ali A. Nasrallah

American University of Beirut Medical Center

Bachir Ghandour

American University of Beirut Medical Center

Research

Keywords: Bibliometric analysis, Coronavirus disease 2019, COVID-19, Ebola, H1N1, MERS, SARS

Posted Date: December 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-129261/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

The pandemic of the COVID-19 literature: A bibliometric analysis

Running title: Bibliometric analysis of the COVID-19 literature

Lokman I. Meho¹, Sarah H. Farran^{2,3}, Ali A. Nasrallah², Bachir Ghandour², Elie A. Akl^{4,5,6*}

¹University Libraries, American University of Beirut, Lebanon

²American University of Beirut Medical Center, Beirut, Lebanon

³ Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

⁴Department of Internal Medicine, American University of Beirut Medical Center, Beirut, Lebanon

⁵Clinical Research Institute, Biostatistics Unit, American University of Beirut, Beirut, Lebanon.

⁶Department of Health Research Methods, Evidence, and Impact (HE&I), McMaster University, Hamilton, ON, Canada.

Email addresses: LIM: lmeho@aub.edu.lb; SHF: shf08@mail.aub.edu; AAN: aan28@mail.aub.edu; BG: bmg03@mail.aub.edu; EAA: ea32@aub.edu.lb

*Correspondence to:

Elie A. Akl, MD, MPH, PhD

Department of Internal Medicine American University of Beirut Medical Center

P.O. Box: 11-0236

Riad-El-Solh Beirut 1107 2020

Beirut, Lebanon

ea32@aub.edu.lb

Keywords: Bibliometric analysis; Coronavirus disease 2019; COVID-19; Ebola; H1N1;
MERS; SARS

Abstract

Background

The research interest in COVID-19, one of the most serious pandemics in recent human history, is unprecedented. This study aims to determine the volume of COVID-19 research and to assess the characteristics of its production and publication.

Methods

We searched Scopus, Embase, PubMed, and the Web of Science databases for publications, up to August 20, 2020. We included all types of documents except corrections, interviews, personal narratives, and retracted publications. We analyzed publication count, type, status, research themes, publication venues, authorship trends, language, institutions, countries, collaboration, and funding.

Results

Of 40,519 eligible documents, 49% were original articles. Forty-nine percent of the original articles and reviews were published in top quartile journals, and 19% were single-authored. More than half of the documents were produced in the United States, China, the United Kingdom, and Italy. Twenty-two percent of the documents involved international collaboration and 17% reported financial support by at least one agency, with the National Natural Science Foundation of China being the most frequently reported funding source (n=982). There are already more documents published on COVID19 than documents ever published on the Ebola, MERS, HIN1, and SARS combined.

Conclusions

The first few months' research output on COVID-19 is relatively large and originated mostly from four countries. Single-authored publications, international collaboration, and governmental funding activities were relatively common.

Background

On December 30, 2019, the Wuhan Municipal Health Committee announced multiple cases of pneumonia with severe acute respiratory syndrome (SARS) of unknown causes within several medical institutions in the city. The first case was traced back to November 17, 2019 (1). The novel causative virus was identified as the SARS-CoV-2 virus, and the respective disease was named as the Coronavirus disease 2019 (COVID-19) (2). The exponential rise in confirmed cases led to the lockdown of Wuhan in January 2020. Subsequently, the World Health Organization (WHO) declared COVID-19 as a global pandemic, with 782,456 deaths globally as of August 20, 2020 (3).

The scientific community's response to COVID-19 was swift in terms of generating and disseminating knowledge (4). Within a few weeks, the pathogen's genome was identified, diagnostic tools were developed (5), and epidemiological and clinical studies were conducted and published (6). The WHO's International Working Group on Ethics & COVID-19 declared that conducting research is an ethical obligation during the pandemic. The Working Group also highlighted the importance of establishing local and international collaborative partnerships (7). A critical aspect of conducting research during the pandemic is prompt and effective execution, a lesson the scientific community learned from the 2014-2016 Ebola experience (8).

Amid an upward spiral in information generation, possibly creating a "data deluge" (9), it is essential that we understand that information better (10). Accordingly, the objective of our study was to determine the volume and trends of COVID-19 research and assess the characteristics of its production and publication.

Methods

Overall design and literature eligibility

We analyzed bibliometric indicators related to the COVID-19 literature published from November 17, 2019, the date when the first COVID-19 infection was detected, up to August 20, 2020. We included documents that met the following eligibility criteria: any mention of COVID-19 in the bibliographic record or abstract of the document; all types of documents except corrections, interviews, personal narratives, and retracted publications; and all languages covered in the databases we used.

Database selection and search

We used the Scopus database to identify relevant documents and carry out our bibliometric analyses. We chose Scopus because it is the largest bibliographic and citation database in the world (11), indexing over 25,000 journals published in over 100 countries and 50 languages. We also used Embase, PubMed, and the Web of Science databases to identify eligible documents that we may have missed as a result of imperfections of our Scopus search query or because of indexing deficiencies by Scopus. To be as inclusive as possible, we used broad search queries in each database as shown in Additional file 1.

Data analysis

We captured data on publication count, type, publication status, research themes, publication venues, author count, language, institutions, countries, collaboration, and funding. We calculated publication trends over time. We also conducted descriptive analyses for the characteristics of the included documents: frequencies and percentages/proportions for categorical variables and median and Interquartile Range (IQR) for continuous variables. We calculated the document count and proportion per country and per capita for each country.

For population count information, we used the demographics provided by the 2019 United Nations report, and the updated World Population Review website (12) for any

missing information. The timely generation of scholarly knowledge is particularly relevant in the case of the COVID-19 pandemic; accordingly, we calculated the number of articles per country relative to the number of days since the first confirmed case as provided by the World Health Organization Dashboard. For the classification of the COVID-19 publications by research themes we used PubMed's Medical Subject Headings (MeSH) subheadings (13, 14) as well as Embase's Emtree subject headings. To assess trends in COVID-19 publication activity, we examined the number of documents added to PubMed every week starting with the first document published on the topic up to mid-August 2020.

To provide a more contextualized interpretation of our results, we compared the research output on COVID-19 with those of medicine in general during the same time period. Similarly, we compared the research output of COVID-19 with four other recent and important public health threats, namely, the 2002 severe acute respiratory syndrome (SARS), the 2009 novel H1N1 influenza (15), the 2012 Middle East respiratory syndrome (MERS), and the 2014 Ebola in West Africa (13, 15, 16). We considered the date of the first publication related to each emergency as a reference point for comparison.

Results

Volume and characteristics of documents

Our Scopus search identified 40,379 documents on COVID-19, and we found an additional 140 documents via Embase, PubMed, and the Web of Science databases, for a total of 40,519 documents. Table 1 shows the publication stage, access form, and type of these documents, and compares them to those of the field of medicine in general. Within the COVID-19 literature, nearly one-fourth of the documents are added to the Scopus database while they are still in press, compared with 18% for medicine overall. In terms of access form, 81% of the COVID-19 literature is published in Open Access compared to only 41% in

medicine. There are also significant differences between COVID-19 and medical literature in terms of document types. For example, the proportion of original research articles is 49% for COVID-19, but 74% for medicine overall.

Table 1: Characteristics of included COVID-19 and ‘Medicine’ documents in 2020 up to August 20, 2020.

	COVID-19	Medicine
Publication Stage		
Final	30 977 (76%)	479 943 (82%)
In Press	9 542 (24%)	105 353 (18%)
Access Form		
Open Access	32 780 (81%)	239 971 (41%)
Other	7 739 (19%)	345 325 (59%)
Document Type		
Articles	19 709 (49%)	433 119 (74%)
Letters	8 361 (21%)	35 118 (6%)
Reviews	4 232 (10%)	58 530 (10%)
Notes	3 906 (10%)	23 412 (4%)
Editorials	3 741 (9%)	20 331 (4%)
Short Surveys	463 (1%)	3 171 (1%)
Conference papers	52 (<1%)	7 368 (1%)
Data papers	40 (<1%)	5 (<1%)
Book chapters	15 (<1%)	4 242 (1%)
Total Included	40 519	585 296

Figure 1 shows the number of COVID-19 related documents added weekly to the PubMed database up to mid-August 2020. As the figure shows, the number of weekly documents added increased steadily since the beginning of the pandemic until the end of June; reaching around 2,700 documents per week. After that, however, there seems to be a plateau at around 2000 weekly document.

Figure 1. Number of COVID-19 documents added weekly to PubMed, including documents not yet published or accessible.

Figure 2 displays the number of publications found in Embase, PubMed, and Scopus within ten months after the start of COVID-19 and within a year in the case of West African Ebola, MERS, H1N1, and SARS. As shown, there are already more documents published on COVID19 than documents ever published on the four other outbreaks combined.

Figure 2. Total COVID-19, Ebola, MERS, H1N1, and SARS documents published within a year after their start.

Figure 3 displays the number of publications found in Embase, PubMed, and Scopus between 2003 and 2020 for West African Ebola, MERS, H1N1, and SARS, starting with the beginning of spread for each disease. An early peak over the first two years is particularly prominent for the Ebola, H1N1, and SARS outbreaks. The curve for MERS-CoV outbreak is more flat. Table 1 (found in Additional file 2) provides more information about the distribution of document type per disease, since the start of disease spread up to August 20, 2020. Original articles are consistently the most common document type, while the next most common varies between letters and reviews.

Figure 3. The number of documents published on Ebola, MERS, H1N1, and SARS per year since start of spread.

Note: Time 0 corresponds to the start of each and time 1 to the first year of disease spread.

Of the 40,519 COVID-19 documents identified in this study, NLM assigned MeSH subheadings to 13,403 (33%), as opposed to 20% for all topics in Medicine in 2020 (up to

August 20). Table 2 provides a ranked list of the subheadings assigned to these documents, limited to subheadings representing over 10% of the total COVID-19 documents. The top subheadings for COVID-19 and for all topics in Medicine varied remarkably, including epidemiology (62% and 16%, respectively), prevention and control (33% and 9%), diagnosis (26% and 12%), virology (25% and 5%), and methods (25% and 23%). Table 2 provides more examples.

Table 2: Main subheadings used in COVID-19 and in all topics in ‘Medicine’ in PubMed in 2020.

MeSH subheadings*	COVID-19 (%) N=13,403	Non-COVID-19 (%) N=116,334
Epidemiology	8 240 (62)	18 996 (16)
Prevention & Control	4 371 (33)	9 915 (9)
Diagnosis	3 513 (26)	13 472 (12)
Virology	3 324 (25)	5 440 (5)
Methods	3 307 (25)	26 208 (23)
Therapy	3 285 (25)	11 973 (10)
Complications	2 620 (20)	9 608 (8)
Transmission	2 168 (16)	2 764 (2)
Drug Therapy	2 043 (15)	12 825 (11)
Isolation & Purification	1 792 (13)	6 597 (6)
Organization & Administration	1 725 (13)	3 493 (3)
Statistics & Numerical Data	1 702 (13)	9 496 (8)
Therapeutic Use	1 673 (13)	10 803 (9)

*Subheadings were limited to those representing over 10% of the total COVID-19 documents

Figure 4 represents a bibliometric map of the subjects of COVID-19 documents as classified by Embase’s Emtree subject headings. Each subject is represented in the map by a node. A larger number of documents under a subject yields a larger node. Each color represents a

subject area. Subjects are inter-related if documents are assigned to both on Embase based on their keywords, and closer nodes represent a larger number of common documents between them. Figure 4 shows a large network of inter-connected subjects, the most prominent of which are “China”, “Pneumonia”, “Healthcare personnel”, “Telemedicine”, “Practice Guidelines” and “Clinical Features”.

Figure 4. Bibliometric map of the classification of COVID-19 documents based on Embase’s Emtree subject headings.

Table 3 shows the distribution of documents by journal quartile per Scopus’s CiteScore (14). We found that 49% of the documents were published in top quartile journals, compared to 43% for the field of medicine in general. About 2% of the documents on COVID-19 are published in journals not rated by CiteScore.

Table 3: Distribution of original research articles and reviews by journal quartiles for COVID-19 and ‘Medicine’ documents.

CiteScore quartile	COVID-19* (N=23,941)	Medicine (N=478,207)
Q1 (top 25)	49	43
Q2 (top 26 - 50)	25	30
Q3 (top 51 - 75)	16	17
Q4 (top 76 - 100)	9	9
Not rated	2	1

*Percent is greater than 100 due to overlap

Figure 5 and Table 2 of Additional file 2 provide information about authorship trends in COVID-19 literature and the other four outbreaks. Results show that 19% of the documents on COVID-19 were single-authored, 15% had two authors, and 52% had three to nine contributing authors. Ebola had the highest proportion of single-author documents (27%).

The author count range of 3 to 9 authors per document was the most frequent for all five public health threats. The median and interquartile range (IQR) of the number of authors for all COVID-19 articles was 6 and 6 (IQR 3, 9). As Table 3 of Additional file 2 shows, the highest median number of authors per publication type was 4 for original articles, letters, reviews, and data papers. MERS had the highest median number of authors per original article (n=7).

Figure 5. Authors per document for COVID-19, Ebola, MERS, H1N1, and SARS in first year after spread.

Table 4 lists the top 10 languages of publication for COVID-19 related documents. The large majority were published in English (94%). The language in which COVID-19 documents were published followed a similar trend as documents in medicine in general. The majority (95%) of documents from all medical themes were published in English, followed by Spanish, Chinese, and German with equal proportion (1.5 of the total each). The full list of publication languages for COVID-19 documents can be found in Table 4 of Additional file 2.

Table 4: The list of top 10 languages of the published COVID-19 related and ‘Medicine’ documents.

Rank	Language of Publication	COVID-19 (%) N=40,519	Medicine (%) N=585,296
1	English	94.2	94.8
2	Spanish	2.2	1.5
3	Chinese	1.8	1.5
4	German	1.2	1.5
5	French	1.1	0.9
6	Italian	0.5	0.1

7	Portuguese	0.5	0.3
8	Russian	0.2	0.5
9	Dutch	0.1	0.1
10	Hungarian	0.1	<0.1

Research productivity

Table 5 represents the rankings of the top 15 institutions in terms of the number of documents related to COVID-19. The list of institutions with more than 150 published documents on COVID-19 can be found in the Additional files (Table 5 of Additional file 2). As of August 20, 2020, Harvard University in the United States generated the largest number of documents (n=1,102), followed by Huazhong University of Science and Technology (n=751) in China and Institut national de la santé et de la recherche médicale (n=563) in France. Out of the 83 top published institutions (defined as having more than 150 published documents), 25 (30%) were in the United States and 13 (16%) in China.

Table 5: Names and locations of the top 15 institutions in terms of the number of documents related to COVID-19.

Rank	Institution	Country/Region	Number of documents
1	Harvard University	United States	1 102
2	Huazhong University of Science and Technology	China	751
3	Institut national de la santé et de la recherche médicale	France	563
4	Wuhan University	China	530
5	University of Toronto	Canada	516
6	Johns Hopkins University	United States	473
7	University of Milan	Italy	455
8	University College London	United Kingdom	438
9	University of Rome La Sapienza	Italy	415
10	Imperial College London	United Kingdom	413

11	University of Oxford	United Kingdom	398
12	National University of Singapore	Singapore	371
13	Assistance publique – Hôpitaux de Paris	France	349
14	University of Washington	United States	349
15	University of Pennsylvania	United States	347

Table 6 represents the rankings of the top 20 countries in terms of the number of COVID-19 documents. The number of COVID-19 documents by country ranged from 1 to 10,369 with a median of 19 and an interquartile range (IQR) of 132. The country with the largest number of documents was the United States. After adjusting to the population size in million, Singapore ranked first with 108 documents per million-population. The United States had the highest number of documents relative to the days since the first detected case (48.7). The full table for worldwide performance can be found in the Additional files (Table 6 of Additional file 2).

Table 6: Research productivity for COVID-19 by country, for the top 20 countries.

Rank	Country	Documents Count	Proportion of International Output (%)	Population (in Million)	Count/Population (in Million)	Count/Days since first case
1	United States	10 369	25.6	329.1	31.5	48.7
2	China	5 549	13.7	1 420.1	3.9	20
3	United Kingdom	4 493	11.1	67	67.1	22.2
4	Italy	4 388	10.8	59.2	74.1	21.7
5	India	2 485	6.1	1 368.7	1.8	12.2
6	France	1 756	4.3	65.5	26.8	8.4
7	Spain	1 692	4.2	46.4	36.5	8.4
8	Canada	1 663	4.1	37.3	44.6	8
9	Germany	1 568	3.9	82.4	19.0	7.6
10	Australia	1 488	3.7	25.1	59.3	7.2
11	Brazil	1 085	2.7	212.4	5.1	6.2
12	Iran	965	2.4	82.8	11.7	5.3
13	Switzerland	865	2.1	8.6	100.6	4.9

14	Netherlands	769	1.9	17.1	45.0	4.4
15	Turkey	681	1.7	84.3	8.1	4.2
16	Singapore	638	1.6	5.9	108.1	3
17	Hong Kong	568	1.4	7.5	75.7	2.7
18	Belgium	564	1.4	11.6	48.6	2.6
19	Japan	544	1.3	126.9	4.3	2.5
20	South Korea	469	1.2	51.3	9.1	2.2

Collaboration and funding

We found that, on average, 21% of all documents in all topics under “Medicine” and 22% specifically on COVID-19 involved collaboration among authors from two or more countries in 2020 (Figure 1 in Additional file 2). Moreover, academic researchers and researchers from corporations collaborated in 2% of published documents in all topics in “Medicine” and in 1% of those on COVID-19 in 2020 (Figure 2 in Additional file 2).

Table 7 shows the list of funding agencies and the number of studies reporting funding by each of these agencies. Out of the 40,519 documents examined for COVID-19, 17% reported financial support by at least one agency (compared to 32% for medicine overall and 47% for all fields together in 2020). The top three funding agencies for COVID-19 research were the National Natural Science Foundation in China (n=928), the United States National Institutes of Health (n=659), and the United Kingdom National Institute for Health Research (n=190).

Table 7: List of the top 15 funding agencies and their respective number of funded studies.

Rank	Funding Source	COVID-19	Medicine
1	National Natural Science Foundation of China	982	21 920
2	National Institutes of Health	659	20 769
3	National Institute for Health Research	190	3 380
4	National Institute of Allergy and Infectious Diseases	179	1 958

5	National Basic Research Program of China (973 Program)	169	2 903
6	Wellcome Trust	161	2 463
7	Fundamental Research Funds for the Central Universities	134	1 852
8	Conselho Nacional de Desenvolvimento Científico e Tecnológico	124	528
9	National Science Foundation	123	1 664
10	National Heart, Lung, and Blood Institute	113	2 925
11	Pfizer	108	2 878
12	Novartis	105	2 538
13	European Commission	103	2 242
14	Bill and Melinda Gates Foundation	96	2 538
15	Canadian Institutes of Health Research	93	3 187

Discussion

Largely due to the disastrous consequences of COVID-19 on all aspects of life all over the world, it was not surprising to discover that there has been much more literature published on this pandemic than Ebola, MERS, H1N1, and SARS combined. The research output on Ebola, H1N1, and SARS started to decline sharply after the second year. Whether COVID-19 will experience a similar fate is highly unlikely largely because of the scale of impact that it had on the world economy, society, and environment and because a cure/vaccine for the virus is yet to be found. COVID-19 research activity is more likely to resemble MERS and other diseases such as HIV.

Nasrallah et al. found that a large number of COVID-19 interventional clinical trials were registered soon after the pandemic onset (17). This relatively high production of research is partly explained that the COVID-19 outbreak, and unlike the other four outbreaks, has evolved into a pandemic, so a very large number of countries were affected. Also, the public health effects (in terms of morbidity, mortality, and effects on healthcare systems) of the

COVID-19 pandemic have been unparalleled, at least in our modern times(16, 18). Previous experiences, including the Ebola epidemic, have highlighted the importance of timely conduct of clinical research in decreasing human losses during public health emergencies (8). The World Health Organization launched a number of initiatives to improve coordination between scientists and global health professionals and accelerate the research and development process. One of the initiatives was the activation of the World Health Organization Research and Development Blueprint to “accelerate diagnostics, vaccines and therapeutics for this novel coronavirus” (19). In parallel, the WHO developed a policy brief summarizing the key universal ethical standards to ensure ethical conduct of research during the COVID-19 outbreak (7).

Funding is often a topic of debate in scientific research, particularly when the research can potentially generate profits (20) or ground-breaking scientific discoveries. We sought to investigate the behavior of funding agencies with regards to COVID-19 research, as an indicator of international, governmental and academic interest in the topic. Governments have invested significant efforts to enhance research productivity and have been a major source of support for COVID-19 research. We found that the top three funding sources were governmental agencies. This highlights the urgency with which funding agencies, including governments, have supported research in response to the crisis. Interestingly, however, we found that 17% of documents reported financial support by at least one agency, compared to 32% for medicine overall. This could be related to the urge with which many researchers set to conduct studies and publish before seeking or obtaining funding

We sought to compare the productivity of different institutions and countries to better understand differences in attitudes towards the COVID-19 pandemic, and their current

capacity to conduct research and the direction of future research efforts. With a better understanding of the current status of research productivity, efforts can better be guided to lead to more efficient and better quality research (21). Our results show large productivity at the levels of researchers, institutions, and countries. The most productive institutions were located in the United States and in China. Within a period of a few months, average international collaboration in COVID-19 research for all publication types (22%) was close to and even higher than the average for all subjects (21%), reflecting productivity and widespread engagement on an international level. The number of authors per document often reflects collaboration, and the field of infectious diseases has been shown to be associated high levels of collaboration (22). There are multiple benefits to collaborative work. At the level of scientific competencies, collaborating with others with complementary skills allows better performance in areas of specialization, providing technical, competitive and financial advantages (23, 24). These advantages are especially evident at the level of international collaboration, with evidence that international publications are increasing in number and improving in quality (24), and that research teams with higher numbers of reported collaborators have larger grants (23). At the level of organizations, collaboration allows better use of resources and interdisciplinary skills (24). Interestingly, we found that academic researchers and researchers in corporations collaborated together in 2% of documents in all research areas combined. In contrast, such collaborations were 1% for COVID-19 publications. We believe this is largely because research on COVID-19 is still in its early stages. We predict that academic-corporate collaborations will increase in intensity in the future, especially with the race for finding appropriate medications and vaccinations for the disease.

Introduced as an alternative to the widely used Journal Citation Reports (and its metric Journal Impact Factor), CiteScore reflects the impact and performance of a publication venue (25), which helps evaluate the quality of COVID-19 publications as deemed by journals. We found that close to half (49%) of the documents were published in top quartile journals. It is unclear whether this fact reflects a high quality of the research work or a high interest in COVID-19 research by the top quartile journals. It is worth noting that the majority (49%) of COVID-19 documents were original articles, and while letters corresponded to 6% of all documents in medicine in 2020 up to August 20, they accounted for 21% of COVID-19 documents. This finding might have been driven by two factors: first, the size and rapidity of the pandemic onset which led to a rush of researchers and clinicians attempting to comment, in order to establish guidance. Second, multiple management aspects of COVID-19 have been publicized and possibly even politicized, for example, testing (26), distancing guidelines (27), preventive measures (28), and hydroxychloroquine administration (29), which may have led to a large number of experts providing their opinions in the form of letters.

The distribution of languages on a specific topic within a database can serve as a general indicator of the geographical and institutional concentration of research efforts and the audience informed. We found that for COVID-19 related documents and for all publications in medicine in 2020, 94.2% and 94.8% of documents, respectively, were in English. The percent of non-English language documents on COVID-19 was, however, around 6%, similar to all topics in medicine (5%). However, we predict that if COVID-19 is to follow a similar trend as that of SARS-CoV and the H1N1 influenza, during which the number of documents peaks during the second year after spread, the number of non-English language documents would increase to over 10%, especially given the increasing impact of COVID-19 on most countries of the world.

This study has a number of strengths. To our knowledge, it is the most recent and comprehensive study of the COVID-19 literature using multiple bibliographic tools (30-32). We used Scopus, a largely inclusive database, as well as data from Embase, Pubmed, and the Web of Science to avoid missing many documents. In addition to quantitatively exploring the characteristics of scholarly data, we have adjusted our analysis to best represent productivity trends. For example, for the number of documents per country, we report our findings according to population size as done in previous bibliometric analyses (articles per million of population) (21) and time (since first confirmed COVID-19 case). Moreover, to adequately evaluate our results, we compared the COVID-19 indicators to the indicators of publications from all themes, and to the research output of other public health emergencies. We carry out our comparisons during the same time-frame to avoid time-dependent bias (33). Since value of citation numbers is debated in the literature, we opted out of reporting citation numbers. Highly-cited articles are thought to have an impact on the direction in which a medical field is growing, and analyses of these trends have frequently been reported in the literature (34-36). For example, the number of citations a document receives is a quantitative index, that does not reflex the quality of the work or its impact on scientific research and clinical practices (33, 37). Other strengths of bibliometric analyses in general apply to our study: the method we followed is simple and our results are easily replicable. Our analysis is also minimally subject to bias, which enhances its validity and makes it a valuable addition to peer-reviews in evaluating the literature (37).

The study also has some limitations. Scopus covers all 37,000,000 documents indexed in Embase and Medline—the two largest biomedical databases in the world—and covers an additional 43,000,000 documents. However, Scopus is still largely a database that focuses on

English language literature and, as such, misses many journals published in other languages. Moreover, COVID-19 is a fast-growing research area and, as a result, our study misses many of the manuscripts accepted for publication but not yet indexed in the Scopus database. Upon examining these missing manuscripts, we found that they count a total of 2,135, of which 9.6% were in Chinese. This percentage contrasts with the findings from the dataset used in the study, in which only 1.8% of the overall documents published were in Chinese. Unpublished documents can, therefore, remarkably change the characteristics of the data, even when a comprehensive database such as Scopus is used. There are also inherent limitations in some of the indicators used in bibliometric analyses, mostly related to the qualitative nature of some tools (38). Due to the recency of the pandemic and the short time-frame during which we carried out the analysis, we could not use tools like citation mining, which captures the impact of a scientific work (38). However, we included a variety of indices to adequately evaluate the quality of documents. The number of authors in a study, the amount of collaboration, and the publication journals' Cite-Score represent the amount of effort invested in scholarly work and the quantitative trends in the literature compared to the citation number received per document. One limitation is the rapidly changing pool of COVID-19 literature. A bibliometric analysis of Scopus, Pubmed and Web of Science retrieved a total of 227 documents as of March 10, 2020. China, the United States and Canada had the highest document count (39). Only months later, we found over 40,000 documents, and the United Kingdom had replaced Canada as third most productive country in terms of document number. The number of documents is increasing on a daily basis, as would the evidence that the pooled data can provide.

Conclusion

The initial literature on COVID-19 is promising when compared to the literature of similar previous outbreaks, at least in terms of quantity and in collaboration. This is in line with calls for prompt and effective conduct of research during pandemics to help patients and contain catastrophic global consequences as soon as possible (7, 8). It is also in line with recommendations made by WHO that conducting research, both individually and collaboratively, is an ethical obligation during the pandemic and should be considered a measure of public health response (7). Whether this research output will end up impacting the progress of the pandemic and of its consequences, is still to be seen, and will depend to a large extent on its judicious use in decision making.

List of abbreviations

COVID-19= Coronavirus disease 2019

SARS-CoV-2= Severe acute respiratory syndrome coronavirus 2

SARS= Severe acute respiratory syndrome

WHO= World Health Organization

MERS= Middle East respiratory syndrome

IQR= Interquartile Range (IQR)

MeSH= Medical Subject Headings

Declarations

Authors' contributions

LIM and EAA conceptualised and conceived the project. LIM collected the data. EAA was the principle investigator for the project. LIM, SHF, AAN and BG analysed the data. All the authors interpreted the data. LIM, SHF and AAN wrote the first draft. All authors contributed to the final draft and finally approved it to be published. All authors revised the final draft and revised the work for critical intellectual content. All authors agreed to be accountable for all aspects of the work for any issue related to the accuracy or integrity of any part of the work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data supporting the conclusions of the article is available on the Scopus database (URL: <https://www.scopus.com/home.uri>), which is produced by Elsevier (URL: <https://www.elsevier.com/en-xm>). Results can be reproduced using the queries in Additional file 1.

Competing interests

None to declare.

Funding

Not applicable.

Acknowledgements

Not Applicable.

References

1. ProMED-mail. UNDIAGNOSED PNEUMONIA - CHINA (HUBEI): REQUEST FOR INFORMATION. 2019.
2. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Hu Y, et al. Complete genome characterisation of a novel coronavirus associated with severe human respiratory disease in Wuhan, China. *bioRxiv*. 2020:2020.01.24.919183.
3. WHO COVID-19 Dashboard: World Health Organization; [cited 2020. Available from: <https://who.sprinklr.com/>].
4. Kalil AC. Treating COVID-19-Off-Label Drug Use, Compassionate Use, and Randomized Clinical Trials During Pandemics. *JAMA*. 2020.
5. Lake MA. What we know so far: COVID-19 current clinical knowledge and research. *Clin Med (Lond)*. 2020;20(2):124-7.
6. Edwards KM, Kochhar S. Ethics of Conducting Clinical Research in an Outbreak Setting. *Annu Rev Virol*. 2020.
7. World Health O. Ethical standards for research during public health emergencies: distilling existing guidance to support COVID-19 R&D. Geneva: World Health Organization; 2020 2020. Contract No.: WHO/RFH/20.1.
8. Busta ER, Mancher M, Cuff PA, McAdam K, Keusch G. Integrating Clinical Research into Epidemic Response: The Ebola Experience. Washington (DC)2017.
9. Lawton A, Burns J. A review of competencies needed for health librarians--a comparison of Irish and international practice. *Health Info Libr J*. 2015;32(2):84-94.
10. Rodrigues SP, van Eck NJ, Waltman L, Jansen FW. Mapping patient safety: a large-scale literature review using bibliometric visualisation techniques. *BMJ Open*. 2014;4(3):e004468.
11. Masic I, Milinovic K. On-line biomedical databases-the best source for quick search of the scientific information in the biomedicine. *Acta Inform Med*. 2012;20(2):72-84.
12. World Population Review [Available from: <https://worldpopulationreview.com/>].
13. Wang L, Wu JT. Characterizing the dynamics underlying global spread of epidemics. *Nat Commun*. 2018;9(1):218.
14. Kamat PV. Impact Factor, CiteScore, and Citation Analysis. *ACS Energy Letters*. 2020;5(7):2452-3.
15. World Health Organization International health regulations committees and expert roster: World Health Organization; [Available from: https://www.who.int/ihr/procedures/ihr_committees/en/].
16. Balkhair AA. COVID-19 Pandemic: A New Chapter in the History of Infectious Diseases. *Oman Med J*. 2020;35(2):e123.
17. Nasrallah AA, Farran SH, Nasrallah ZA, Chahrour MA, Salhab HA, Fares MY, et al. A large number of COVID-19 interventional clinical trials were registered soon after the pandemic onset: a descriptive analysis. *Journal of Clinical Epidemiology*. 2020;125:170-8.
18. Srivastava N, Baxi P, Ratho RK, Saxena SK. Global Trends in Epidemiology of Coronavirus Disease 2019 (COVID-19). *Coronavirus Disease 2019 (COVID-19)*. 2020:9-21.
19. R&D Blueprint and COVID-19: World Health Organization; 2020 [Available from: <https://www.who.int/blueprint/priority-diseases/key-action/novel-coronavirus/en/>].
20. Resnik DB. Science and money: problems and solutions. *J Microbiol Biol Educ*. 2014;15(2):159-61.
21. El Rassi R, Meho LI, Nahlawi A, Salameh JS, Bazarbachi A, Akl EA. Medical research productivity in the Arab countries: 2007-2016 bibliometric analysis. *J Glob Health*. 2018;8(2):020411-.

22. Parish AJ, Boyack KW, Ioannidis JPA. Dynamics of co-authorship and productivity across different fields of scientific research. *PloS one*. 2018;13(1):e0189742-e.
23. Bozeman B, Corley E. Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*. 2004;33(4):599-616.
24. Iglič H, Doreian P, Kronegger L, Ferligoj A. With whom do researchers collaborate and why? *Scientometrics*. 2017;112(1):153-74.
25. McCullough HZaR. CiteScore: a new metric to help you track journal performance and make decisions Elsevier Connect2016 [Available from: <https://www.elsevier.com/editors-update/story/journal-metrics/citescore-a-new-metric-to-help-you-choose-the-right-journal>].
26. CDC was pressured 'from the top down' to change coronavirus testing guidance, official says: CNN; 2020 [Available from: <https://edition.cnn.com/2020/08/26/politics/cdc-coronavirus-testing-guidance/index.html>]
27. Rothgerber H, Wilson T, Whaley D, Rosenfeld D, Humphrey M, Moore A, et al. Politicizing the COVID-19 Pandemic: Ideological Differences in Adherence to Social Distancing2020.
28. In a pandemic, political polarization could kill people: The Washington Post; 2020 [Available from: https://www.washingtonpost.com/outlook/2020/03/23/coronavirus-polarization-political-exaggeration/?fbclid=IwAR1oQe_PsJFSfso7cRr5OANzFskm202e_HrL_vnwdOzPoHbWHP_h0fY1CcNY#comments-wrapper].
29. FDA cautions against use of hydroxychloroquine or chloroquine for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems: U. S. Food and Drug Administration; 2020 [Available from: <https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or>].
30. Al-Zaman MS. Bibliometric analysis of COVID-19 literature. medRxiv. 2020:2020.07.15.20154989.
31. Aristovnik A, Ravšelj D, Umek L. A Bibliometric Analysis of COVID-19 across Science and Social Science Research Landscape. Preprints.org; 2020.
32. Zyoud SeH, Al-Jabi SW. Mapping the situation of research on coronavirus disease-19 (COVID-19): a preliminary bibliometric analysis during the early stage of the outbreak. *BMC Infectious Diseases*. 2020;20(1):561.
33. Rosas SR, Kagan JM, Schouten JT, Slack PA, Trochim WMK. Evaluating Research and Impact: A Bibliometric Analysis of Research by the NIH/NIAID HIV/AIDS Clinical Trials Networks. *PLOS ONE*. 2011;6(3):e17428.
34. Iftikhar PM, Ali F, Faisaluddin M, Khayyat A, De Gouvias De Sa M, Rao T. A Bibliometric Analysis of the Top 30 Most-cited Articles in Gestational Diabetes Mellitus Literature (1946-2019). *Cureus*. 2019;11(2):e4131-e.
35. Loonen MPJ, Hage JJ, Kon M. Value of citation numbers and impact factors for analysis of plastic surgery research. *Plast Reconstr Surg*. 2007;120(7):2082-91; discussion 92-4.
36. O'Sullivan KE, Kelly JC, Hurley JP. The 100 most cited publications in cardiac surgery: a bibliometric analysis. *Ir J Med Sci*. 2015;184(1):91-9.
37. Belter CW. Bibliometric indicators: opportunities and limits. *J Med Libr Assoc*. 2015;103(4):219-21.
38. Wallin JA. Bibliometric Methods: Pitfalls and Possibilities. *Basic & Clinical Pharmacology & Toxicology*. 2005;97(5):261-75.
39. Rafiei Nasab F, rahim F. Bibliometric Analysis of Global Scientific Research on SARSCoV-2 (COVID-19). medRxiv. 2020:2020.03.19.20038752.

Additional Files

File name: Additional file 1

File format: .pdf

Title of data: Search queries

Description of data: Search queries used to produce the results in each database.

File name: Additional file 2

File format: .pdf

Title of data: Table 1, Table 2, Table 3, Table 4, Table 5, Table 6, Figure 1, Figure 2.

Description of data: Additional tables and figures providing information that could provide details supplementary to the tables reported in the main manuscript, and tables that are too lengthy to include in the main text.

Figures

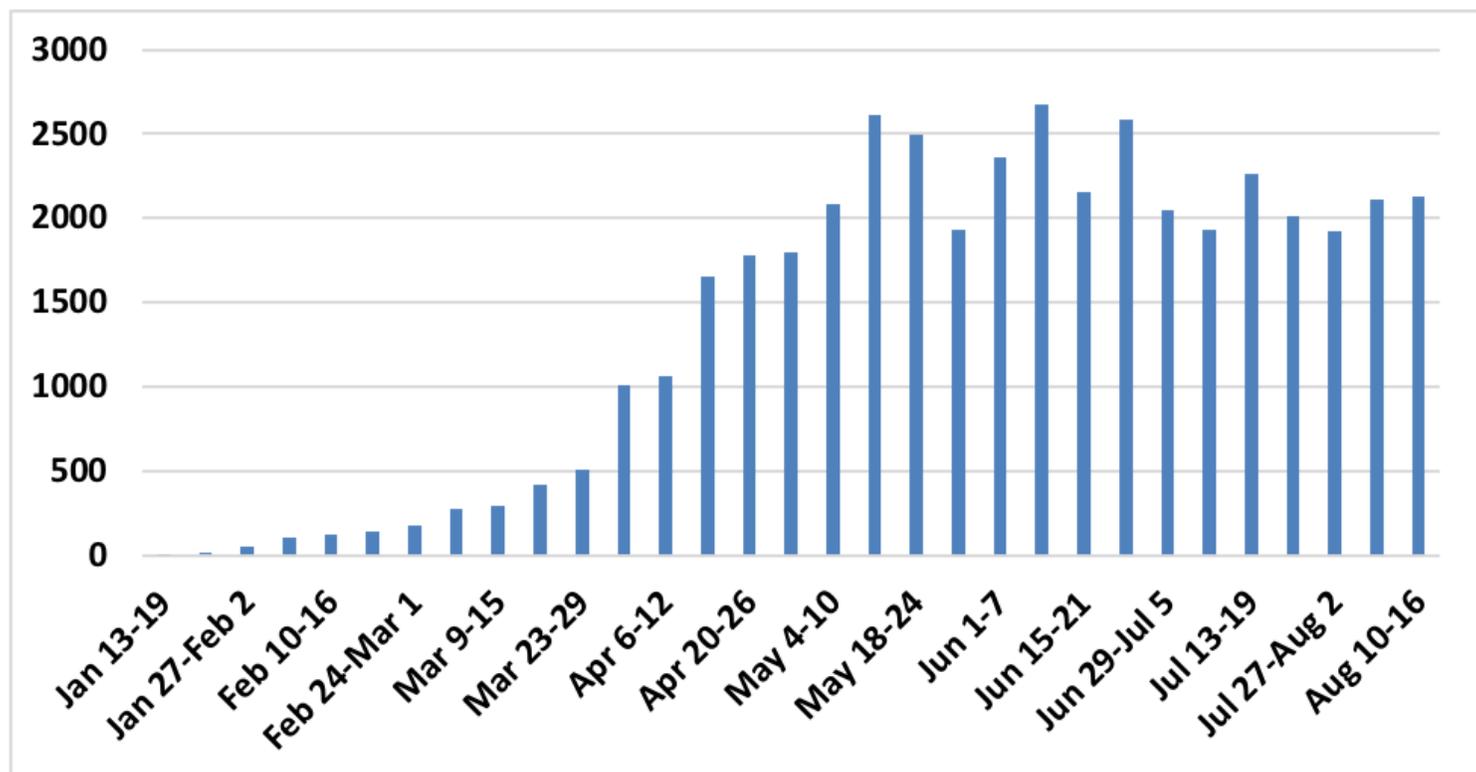


Figure 1

Number of COVID-19 documents added weekly to PubMed, including documents not yet published or accessible.

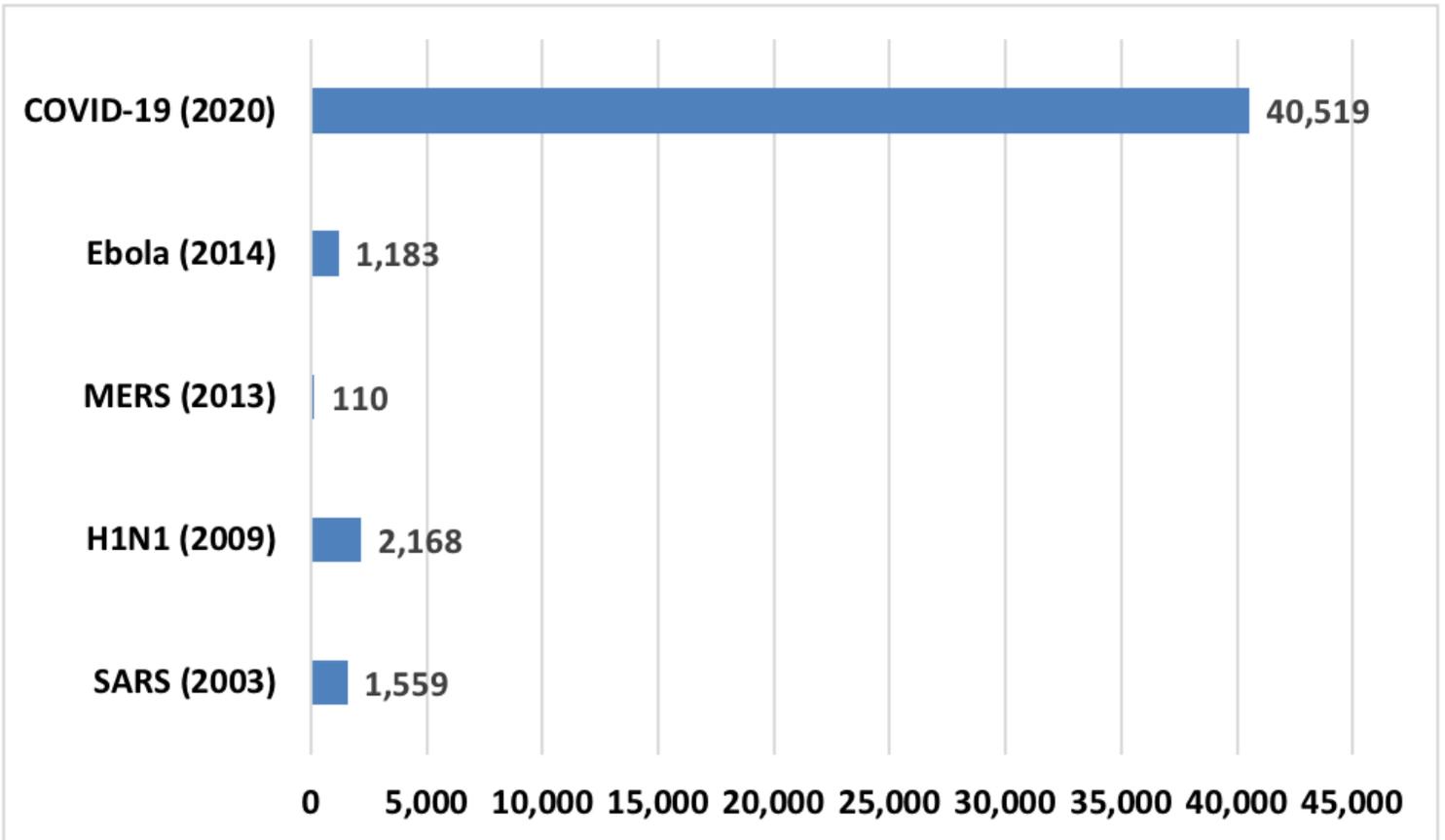


Figure 2

Total COVID-19, Ebola, MERS, H1N1, and SARS documents published within a year after their start.

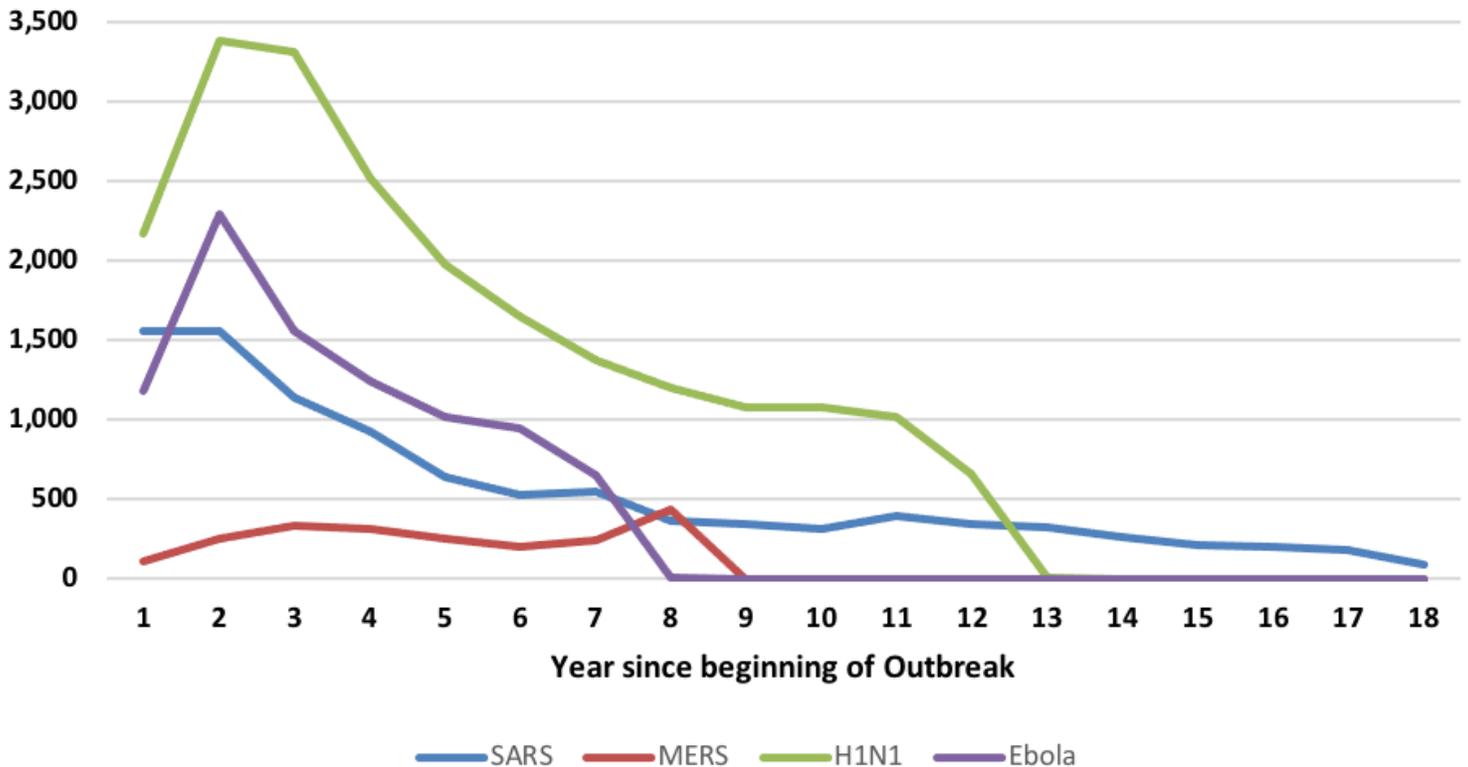


Figure 3

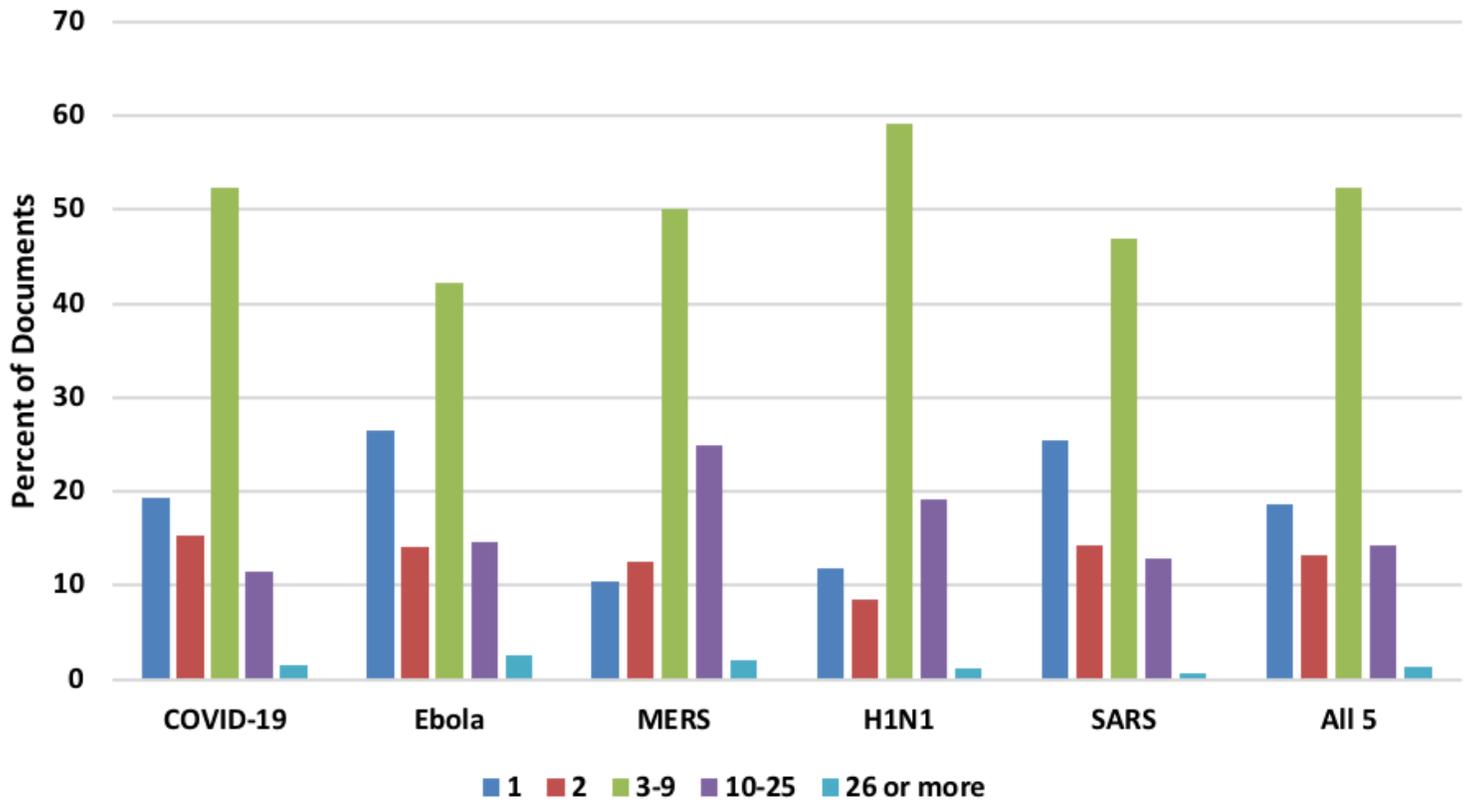


Figure 5

Authors per document for COVID-19, Ebola, MERS, H1N1, and SARS in first year after spread.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFiles.pdf](#)