Scientometrics (2021) 126:831–842
https://doi.org/10.1007/s11192-020-03675-3

Publishing volumes in major databases related to Covid-19

Jaime A. Teixeira da Silva[1] · Panagiotis Tsigaris[2] · Mohammadamin Erfanmanesh[3]

Received: 18 July 2020 / Published online: 28 August 2020

Abstract
The SARS-CoV-2 virus, which causes Covid-19, induced a global pandemic for which an effective cure, either in the form of a drug or vaccine, has yet to be discovered. In the few brief months that the world has known Covid-19, there has been an unprecedented volume of papers published related to this disease, either in a bid to find solutions, or to discuss applied or related aspects. Data from Clarivate Analytics' Web of Science, and Elsevier's Scopus, which do not index preprints, were assessed. Our estimates indicate that 23,634 unique documents, 9960 of which were in common to both databases, were published between January 1 and June 30, 2020. Publications include research articles, letters, editorials, notes and reviews. As one example, amongst the 21,542 documents in Scopus, 47.6% were research articles, 22.4% were letters, and the rest were reviews, editorials, notes and other. Based on both databases, the top three countries, ranked by volume of published papers, are the USA, China, and Italy while *BMJ*, *Journal of Medical Virology* and *The Lancet* published the largest number of Covid-19-related papers. This paper provides one snapshot of how the publishing landscape has evolved in the first six months of 2020 in response to this pandemic and discusses the risks associated with the speed of publications.

Keywords Acceptance and rejection · Biomedicine · Correction of the literature · Peer review · Open access · Preprints · Retractions · SARS-CoV-2 virus

✉ Jaime A. Teixeira da Silva
jaimetex@yahoo.com

✉ Panagiotis Tsigaris
ptsigaris@tru.ca

✉ Mohammadamin Erfanmanesh
amin.erfanmanesh@gmail.com

[1] P. O. Box 7, Miki-cho post office, Ikenobe 3011-2, Kagawa-ken 761-0799, Japan

[2] Department of Economics, Thompson Rivers University, 805 TRU Way, Kamloops, BC V2C 0C8, Canada

[3] Department of Library and Information Science, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

Introduction

The global Covid-19 pandemic has already infected 13,734,518 people, causing 588,149 deaths[1] in one of humanity's greatest challenges of modern time. For academics, it has provided an extremely rare opportunity to examine so many aspects, biomedical and other (economic, public health, psychology, social, historical, etc.), related to the SARS-CoV-2 virus. In just a few months, crudely tens of thousands of preprints and peer reviewed papers and other documents related to Covid-19 have been published. NCBI's LitCovid (Chen et al. 2020) shows that 31,360 documents had already been published on PubMed until July 14, 2020. The objective of this paper is to provide a snapshot of the publishing landscape to appreciate the volume of papers that have been published in indexed, peer reviewed journals in two major databases, Clarivate Analytics' Web of Science (WoS) and Elsevier's Scopus, from January 1 until June 30, 2020. More than 80% (83% in Scopus, 89% in WoS) of Covid-19 papers are open access (OA), but not all, despite a public agreement to make all papers OA.[2] We synthesize what we have discovered in these two databases. A brief discussion of some of the risks of the high publication volume of papers related to Covid-19 and the discrepancies between database findings is provided at the end of the letter.

Methods

A search was limited, in WoS (SCI, SSCI, A&HCI and ESCI) and Scopus, to any documents published in the past six months (January 1 to June 30, 2020) using the search query "SARS-CoV-2" OR "COVID-19" OR "Coronavirus 2019" OR "Corona Virus 2019" OR "novel coronavirus" OR "novel corona virus" OR "2019-nCoV", in titles and keywords. Data was collected on July 1, 2020. As a result, 12,331 and 21,602 documents were retrieved from WoS and Scopus, respectively. These documents were then cleaned up manually in order to remove any duplicates. In a few cases, 2.3% in WoS and 0.28% in Scopus, early access and published versions of a same document were published twice with different accession numbers (in WoS) or Scopus IDs. Consequently, the sample sizes decreased to 12,052 documents with unique digital object identifiers (DOIs) or titles for WoS and 21,542 documents for Scopus. These records were then analyzed by a variety of fields including subject areas, document types, organizations, funding sponsors, authors, source titles, countries, languages, and most cited documents.

Results

We discovered 23,634 unique documents in WoS and Scopus, i.e., documents with unique DOIs or titles. More specifically, 2092 documents were exclusively indexed in WoS (but not in Scopus), 11,582 documents were exclusively indexed in Scopus (but not in WoS), and 9960 documents were indexed in both WoS and Scopus (Fig. 1).

[1] https://coronavirus.jhu.edu/map.html (John Hopkins University; last accessed: July 17, 2020).

[2] https://wellcome.ac.uk/coronavirus-covid-19/open-data (John Hopkins University; last accessed: July 18, 2020).

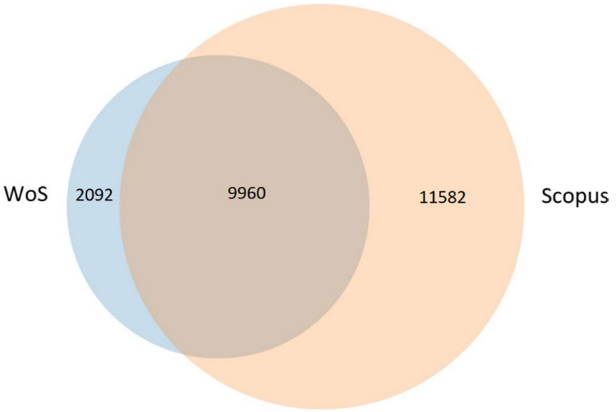

Fig. 1 Venn diagram showing the overlap in documents with unique DOIs or titles in two major citation databases (WoS and Scopus) and also unique documents in each database

Table 1 Type of documents related to Covid-19 indexed in Scopus and Web of Science (WoS): January 1 to June 30, 2020

Scopus		Relative %
Article	10,254	47.6
Letter	4817	22.4
Review	2056	9.5
Editorial	1986	9.2
Note	1972	9.2
Short survey	226	1.0
Erratum	114	0.5
Conference paper	98	0.5
Data paper	19	0.1
Total	21542	100.0%
WoS		Relative %
Article	4434	36.8
Editorial material	3281	27.2
Letter	2629	21.8
Review	1121	9.3
News item	478	4.0
Correction	101	0.8
Other document types	8	0.1
Total	12,052	100.0%

Table 1 indicates that articles, letters and editorials were generally ranked as the top three categories of published documents, although the rank and the relative percentage differed, depending on the database. For example, original papers accounted for 47.6% of all Covid-19-related documents in Scopus (vs. 36.8% in WoS). In general, editorials are not peer reviewed, while letters to the editors are generally only screened by editors.

Scopus reveals (Table 2) that the top three (by volume of unique documents) institutions were: Huazhong University of Science and Technology (China) (442), Tongji Medical

Table 2 Top ten organizations, authors, countries, funding sponsors, journals, languages, subject areas and most cited documents related to Covid-19 indexed in Scopus: January 1 to June 30, 2020

Organizations (country)	Number of unique documents	Relative %
Huazhong University of Science and Technology (China)	442	2.1
Tongji Medical College (China)	433	2.0
Harvard Medical School (USA)	395	1.8
INSERM (France)	327	1.5
University of Milan (Italy)	287	1.3
University College London (UK)	281	1.3
IRCCS Foundation (Italy)	253	1.2
Sapienza University of Rome (Italy)	249	1.2
University of Toronto (Canada)	228	1.1
Assistance Publique - Hopitaux de Paris (France)	213	1.0
Authors (country/countries; ORCID number, if available)	**Number of unique documents**	**Relative %)**
Wiwanitkit, Viroj (China, India, Thailand)	83	0.4
Mahase, Elisabeth (UK)	65	0.3
Iacobucci, Gareth (UK)	53	0.2
Lippi, Giuseppe (Italy; https://orcid.org/0000-0001-9523-9054)	37	0.2
Joob, Beuy (Thailand; https://orcid.org/0000-0002-5281-0369)	35	0.2
Rodriguez-Morales, Alfonso J. (Columbia, Peru; https://orcid.org/0000-0001-9773-2192)	30	0.1
Fabbrocini, Gabriella (Italy; https://orcid.org/0000-0002-0064-1874)	25	0.1
Goldust, Mohamad (Germany, Switzerland, Italy)	24	0.1
Hsueh, Po-Ren (Taiwan; https://orcid.org/0000-0002-7502-9225)	22	0.1
Yuen, Kwok-Yung (Honk Kong; https://orcid.org/0000-0002-2083-1552)	22	0.1
Countries	**Number of unique documents**	**Relative %**
United States of America (USA)	5033	23.4
China	3511	16.3
Italy	2590	12.0
United Kingdom (UK)	2286	10.6
India	1218	5.7
France	1031	4.8
Canada	873	4.1
Spain	838	3.9
Germany	814	3.8
Australia	763	3.5
Funding sponsors (country)	**Number of unique documents**	**Relative %**
National Natural Science Foundation of China (China)	560	2.6
National Institutes of Health (USA)	254	1.2
National Institute for Health Research (UK)	88	0.4
National Institute of Allergy and Infectious Diseases (USA)	85	0.4

Table 2 (continued)

Funding sponsors (country)	Number of unique documents	Relative %
Wellcome Trust (UK)	84	0.4
National Basic Research Program of China (China)	84	0.4
Fundamental Research Funds for the Central Universities (China)	61	0.3
Pfizer (USA)	54	0.3
National Heart, Lung, and Blood Institute (USA)	51	0.2
National Science Foundation (USA)	49	0.2
Journals (Country, publisher)	**Number of unique documents**	**Relative %)**
British Medical Journal (UK, BMJ Publishing Group)	574	2.7
Journal of Medical Virology (USA, John Wiley & Sons, Inc.)	317	1.5
The Lancet (UK, Elsevier)	230	1.1
JAMA (Journal of The American Medical Association) (USA, American Medical Association)	161	0.7
International Journal of Environmental Research & Public Health (Switzerland, MDPI)	157	0.7
Journal of Infection (UK, Elsevier)	158	0.7
Medical Hypothesis (US, Elsevier)	152	0.7
International Journal of Infectious Diseases (Netherlands, Elsevier)	138	0.6
New England Journal of Medicine (USA, Massachusetts Medical Society)	131	0.6
Nature (UK, Springer Nature)	125	0.6
Language	**Number of unique documents**	**Relative %**
English	20232	93.9
Chinese	510	2.4
Spanish	475	2.2
French	275	1.3
German	212	1.0
Italian	114	0.5
Portuguese	78	0.4
Dutch	27	0.1
Norwegian	25	0.1
Russian	26	0.1
Subject areas	**Number of unique documents**	**Relative %**
Medicine	17578	81.6
Biochemistry, genetics and molecular biology	2065	9.6
Immunology and microbiology	1722	8.0
Social sciences	1200	5.6
Nursing	792	3.7
Neuroscience	795	3.7
Pharmacology, toxicology and pharmaceutics	741	3.4
Environmental science	593	2.8

Table 2 (continued)

Subject areas	Number of unique documents	Relative %
Psychology	575	2.7
Health professions	418	1.9
Most cited documents		Number of citations
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., . . . Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet, 395*(10223), 497–506. https://doi.org/10.1016/s0140-6736(20)30183-5		3469
Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., . . . Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA - Journal of the American Medical Association, 323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585		2031
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., . . . Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine, 382*(8), 727–733. https://doi.org/10.1056/nejmoa2001017		1887
Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., . . . Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet, 395*(10223), 507–513. https://doi.org/10.1016/s0140-6736(20)30211-7		1788
Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., . . . Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, china: A retrospective cohort study. *The Lancet, 395*(10229), 1054–1062. https://doi.org/10.1016/s0140-6736(20)30566-3		1507
Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., . . . Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine, 382*(13), 1199–1207. https://doi.org/10.1056/nejmoa2001316		1376
Zhou, P., Yang, X., Wang, X., Hu, B., Zhang, L., . . . Shi, Z. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature, 579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7		1327
Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in china: Summary of a report of 72314 cases from the Chinese center for disease control and prevention. *JAMA - Journal of the American Medical Association, 323*(13), 1239–1242. https://doi.org/10.1001/jama.2020.2648		1324
Chan, J. F., Yuan, S., Kok, K., To, K. K., Chu, H., . . . Yuen, K. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet, 395*(10223), 514–523. https://doi.org/10.1016/s0140-6736(20)30154-9		1044
Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., . . . Tan, W. (2020). Genomic characterization and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet, 395*(10224), 565–574. https://doi.org/10.1016/s0140-6736(20)30251-8		980

College (China) (433), and Harvard Medical School (USA) (395); publishing authors: Viroj Wiwanitkit (83), Elisabeth Mahase (65), and Gareth Iacobucci (53); countries: USA (5033), China (3511), and Italy (2590); funding sponsors: National Natural Science Foundation of China (560), National Institutes of Health (USA) (254), and National Institute for Health Research (UK) (88); source titles: *BMJ*[3] (BMJ Publishing Group Ltd.) (574),

[3] Including "BMJ Clinical Research Edition".

Journal of Medical Virology (John Wiley & Sons, Inc.) (317), and *The Lancet* (Elsevier) (230); languages[4]: English (20,232), Chinese (510), and Spanish (475); subject areas: medicine (17,578), biochemistry, genetics and molecular biology (2,065), and immunology and microbiology (1722); top cited papers: 3469 citations (https://doi.org/10.1016/s0140-6736(20)30183-5; Huang et al.; *The Lancet*), 2031 citations (https://doi.org/10.1001/jama.2020.1585; Wang et al. *JAMA*), and 1887 citations (https://doi.org/10.1056/nejmoa2001017; Zhu et al. *New England Journal of Medicine*).

WoS reveals (Table 3) that the top three (by volume of unique documents) institutions were: University of London (UK) (370), Harvard University (USA) (292), and University of California (USA) (250); publishing authors: Elisabeth Mahase (55), Gareth Iacobucci (43), Viroj Wiwanitkit (39); countries: USA (2999), China (2131), and Italy (1513); funding sponsors: National Natural Science Foundation of China (472), United States Department of Health and Human Services (329), and National Institutes of Health (USA) (317); source titles: *BMJ* (BMJ Publishing Group Ltd.) (456), *Journal of Medical Virology* (John Wiley & Sons, Inc.) (248), and *The Lancet* (Elsevier) (183); Languages: English (11,447), Chinese (155), and German (150); research areas: general internal medicine (2,178), public environmental occupational health (959), and surgery (701); top cited papers: 2513 citations (https://doi.org/10.1016/s0140-6736(20)30183-5; Huang et al.; *The Lancet*), 1484 citations https://doi.org/10.1001/jama.2020.1585; Wang et al.; *JAMA*), and 1356 citations (https://doi.org/10.1016/s0140-6736(20)30211-7; Chen et al.; *The Lancet*).

Lack of disambiguation of Chinese names in WoS

While analyzing the top-ranked authors (in terms of publishing volume), it was noticed that several Chinese authors were highly ranked. For example, in WoS, Wang Y, Zhang Y, Li Y and Liu Y were ranked 1st, 2nd, 3rd and 5th in the top 10 authors. Similarly, Wang L, Li L, Wang J and Liu J were ranked 6th to 9th. As one example, Wang Y, who should have been a "unique" author, and who had 76 papers attributed to his name, was in fact found, after manual disambiguation and analysis of the original records, be several Wang Y.[5]

The importance of accurate, culturally-sensitive indexing of Chinese names cannot be over-emphasized as incorrectly indexed names can drastically alter author-based metrics (Teixeira da Silva 2020a, b) and, as can be observed in the statistics in the previous paragraph, metrics specifically related to Covid-19 literature. There are several techniques available to improve author name disambiguation (Hussain and Hasghar 2017). One solution to solve the issue of name ambiguity is to inspect authors' names and their background information (e.g., the institution in which they worked or the research areas or topics in which they are active) prior to any further bibliometric study.

[4] More than one language is assigned to some documents in both databases.

[5] A non-exhaustive list of examples: Wang Yu from Changzhi mental Health Center, Wang Yan from Chinese Academy of Science, Wang Yun from Sichuan University, Wang Yi from Huazhong University of Science & Technology, Wang Ying from Fenyang Hospital, Wang Ying from Sun Yat Sen University, Wang Yong from Hubei University of Medical Sciences, Wang Yin from Huazhong University of Science & Technology, Wang Yang from Nanjing University, Wang Yiquan from University of Hong Kong or Wang Yuan from Xiameng Chan Hospital.

Table 3 Top ten organizations, authors, countries, funding sponsors, journals, languages, subject areas and most cited documents related to Covid-19 indexed in Web of Science (WoS) (January 1 to June 30, 2020)

Organizations (country)	Number of unique documents	Relative %
University of London (UK)	370	3.1
Harvard University (USA)	292	2.4
University of California (USA)	250	2.1
Huazhong University of Science Technology (China)	204	1.7
Wuhan University (China)	173	1.4
Harvard Medical School (USA)	173	1.4
University College London (UK)	166	1.4
INSERM (France)	158	1.4
University of Milan (Italy)	156	1.3
University of Toronto (Canada)	150	1.2

Authors (country/countries)	Number of unique documents	Relative %
Mahase, Elisabeth (UK)	55	0.5
Iacobucci, Gareth (UK)	43	0.4
Wiwanitkit, Viroj (China, India, Thailand)	39	0.3
Lippi, Giuseppe (Italy; https://orcid.org/0000-0001-9523-9054)	33	0.3
Rimmer, Abi (UK)	30	0.2
Henry, Brandon Michael (USA)	24	0.2
Rodriguez-Morales, Alfonso J. (Columbia, Peru; https://orcid.org/0000-0001-9773-2192)	23	0.2
Joob, Beuy (Thailand; https://orcid.org/0000-0002-5281-0369)	20	0.2
Fabbrocini, Gabriella (Italy; https://orcid.org/0000-0002-0064-1874)	20	0.2
Dhama, Kuldeep (India; https://orcid.org/0000-0001-7469-4752)	19	0.2

Countries	Number of unique documents	Relative %
United States of America (USA)	2999	24.9
China	2131	17.7
Italy	1513	12.6
United Kingdom (UK)	1234	10.2
India	576	4.8
Canada	557	4.6
Germany	544	4.5
France	516	4.3
Australia	478	4.0
Spain	360	3.0

Funding sponsors (country)	Number of unique documents	Relative %
National Natural Science Foundation of China (China)	472	3.9
United States Department of Health Human Services (USA)	329	2.7
National Institutes of Health (USA)	317	2.6
National Key Research and Development Program (China)	69	0.6
Wellcome Trust (UK)	54	0.4

Table 3 (continued)

Funding sponsors (country)	Number of unique documents	Relative %
Fundamental Research Funds for the Central Universities (China)	51	0.4
Canadian Institutes of Health Research (Canada)	46	0.4
National Institute of Allergy and Infectious Diseases (USA)	45	0.4
National Heart, Lung, and Blood Institute (USA)	43	0.4
National Council for Scientific and Technological Development (Brazil)	43	0.4
Journals (country, publisher)	**Number of unique documents**	**Relative %**
British Medical Journal (UK, BMJ Publishing Group)	456	3.8
Journal of Medical Virology (USA, John Wiley & Sons, Inc.)	248	2.1
The Lancet (UK, Elsevier)	183	1.5
Cureus (USA, Cureus Inc.)	135	1.1
Critical Care (UK, BioMed Central Ltd.)	109	0.9
Head and Neck (USA, John Wiley & Sons, Inc.)	90	0.7
Otolaryngology Head and Neck Surgery (UK, Sage Publishing Group)	89	0.7
JAMA (Journal of The American Medical Association) (USA, American Medical Association)	88	0.7
New England Journal of Medicine (USA, Massachusetts Medical Society)	88	0.7
Journal of Infection (UK, Elsevier)	83	0.7

Language	Number of unique documents	Relative %
English	11447	95.0
Spanish	155	1.3
German	150	1.2
Italian	86	0.7
French	79	0.7
Portuguese	35	0.3
Norwegian	27	0.2
Hungarian	23	0.2
Turkish	23	0.2
Russian	15	0.1

Subject areas	Number of unique documents	Relative %
General internal medicine	2178	18.1
Public environmental occupational health	959	8.0
Surgery	701	5.8
Infectious diseases	697	5.8
Cardiovascular system cardiology	496	4.1
Oncology	448	3.7
Virology	403	3.3
Radiology nuclear medicine medical imaging	387	3.2
Pharmacology pharmacy	380	3.2
Research experimental medicine	366	3.0

Table 3 (continued)

Most cited documents	Number of citations
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., . . . Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet, 395*(10223), 497–506. https://doi.org/10.1016/s0140-6736(20)30183-5	2513
Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., . . . Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA - Journal of the American Medical Association, 323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585	1484
Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., . . . Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet, 395*(10223), 507–513. https://doi.org/10.1016/s0140-6736(20)30211-7	1356
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., . . . Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine, 382*(8), 727–733. https://doi.org/10.1056/nejmoa2001017	1327
Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., . . . Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The Lancet, 395*(10229), 1054–1062. https://doi.org/10.1016/s0140-6736(20)30566-3	940
Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., . . . Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine, 382*(13), 1199–1207. https://doi.org/10.1056/nejmoa2001316	911
Chan, J. F., Yuan, S., Kok, K., To, K. K., Chu, H., . . . Yuen, K. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet, 395*(10223), 514–523. https://doi.org/10.1016/s0140-6736(20)30154-9	806
Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in china: Summary of a report of 72314 cases from the Chinese center for disease control and prevention. *JAMA - Journal of the American Medical Association, 323*(13), 1239–1242. https://doi.org/10.1001/jama.2020.2648	764
Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., . . . Tan, W. (2020). Genomic characterization and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet, 395*(10224), 565–574. https://doi.org/10.1016/s0140-6736(20)30251-8	707
Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., … & Diaz, G. (2020). First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine, 382*(10), 929–936. https://doi.org/10.1056/nejmoa2001191	531

Errors in assignment of DOIs

Some errors in assignment of DOIs were observed in WoS and Scopus during data analysis. The incorrect assignment of a single DOI to multiple papers is one of these issues. For example, each of the following DOIs is mistakenly assigned to two different articles in Scopus: "https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.001", "https://doi.org/10.4414/smw.2020.20247", "https://doi.org/10.1001/jama.2020.6122", "https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.001", and "https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.002". To exemplify this, Fig. 2 shows a screenshot of the Scopus database response to the query "DOI = https://doi.org/10.4414/smw.2020.20247". A single paper with two different DOIs was another error observed in both databases. For example,

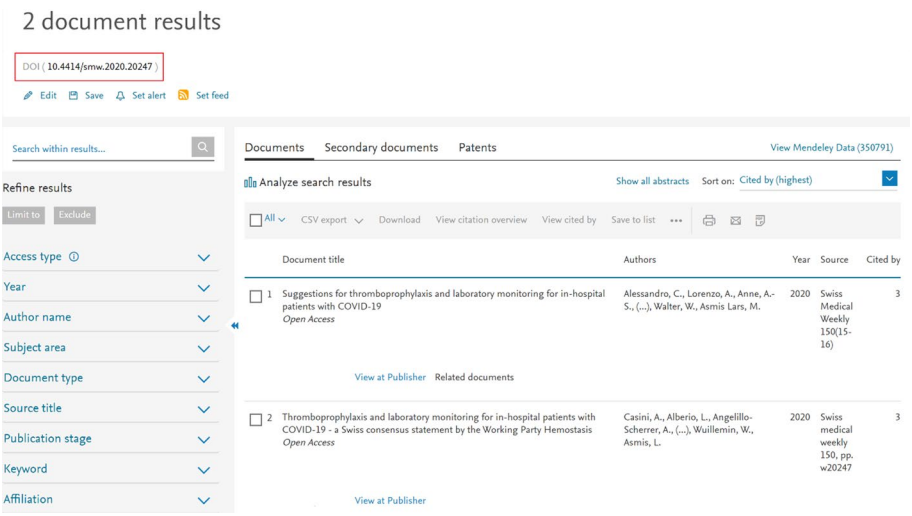

Fig. 2 Example of a Scopus query, returning the same DOI mistakenly assigned to two papers (retrieved on July 14, 2020)

both the DOIs "https://doi.org/10.15252/embj.2020105114" and "https://doi.org/10.15252/embj.20105114" can be used to search a unique paper in WoS.

Discussion and limitations of this study

Our analysis and data mining of two major citation databases indicates that 23,634 unique documents related to Covid-19 were published between January 1 and June 30, 2020. Fraser et al. (2020) also indicate that thousands of preprints have been published, accounting for roughly one third of the total volume of published papers related to Covid-19 in the January 1 to June 30 period. This sheer volume of papers related to a single topic may be unprecedented. These astonishing volumes of indexed papers and other documents related to Covid-19 in reputed databases are already placing pressure on academics to quickly publish their findings and their thoughts, and on editors and journals to rapidly release potentially medically important information that could be of value to health practitioners and policy makers alike. However, the publishing system is under pressure, and this may result in analytical errors (Ioannidis 2020).

The Covid-19 literature will require separate and thorough analyses to appreciate how this pandemic shaped the academic publishing sector. For example, a tiny fraction of the documents are corrections and retractions. These numbers may increase as Covid-19 papers, including preprints and papers in potentially predatory venues (Teixeira da Silva 2020a, b) are increasingly scrutinized. Results of the current research showed that 0.8% of the Covid-19 related documents in WoS (101 documents) and 0.5% of documents in Scopus (114 documents) have been corrected or retracted within a short time after publication.

One notable aspect of the data summary in Tables 1, 2 and 3 (also see Fig. 1) are the sometimes-stark differences between Scopus and WoS, i.e., the choice of database for such analyses might bias the findings. Scopus tends to offer a more comprehensive coverage of the scientific and scholarly literature than WoS, and since Scopus collects a huge number

of "secondary documents", by doing so, it can capture citations from documents that are not indexed in Scopus (Falagas et al. 2008; Martín-Martín et al. 2018). The most important reason is that the content coverage of these two databases are substantially different. Consequently, the results of bibliometric analyses may vary depending on the database used. There are also biases in each of these citation databases, e.g., WoS has been criticized for favoring North American-based, English-language journals (Mongeon and Paul-Hus 2016). There are a few reasons for the predominance of documents in English, and apparent under-representation of Chinese: Scopus covers more Chinese-language documents than WoS, there are not many Chinese-language journals indexed in WoS, China has its own citation index (Chinese Science Citation Database) (Vera-Baceta et al. 2019), and most publications of authors affiliated with Chinese institutions were published in English-language journals.

We recognize that the current study, which provides a six-month snapshot of two major databases, only gives a limited time- and database-based perspective, of the published Covid-19 literature. As the pandemic tails-off, the volume of papers may also begin to decline, but a thorough reanalysis will be required to ascertain this trend.

Author contributions JATdaS, PT, and ME, who are co-corresponding authors, contributed equally to all parts of the study analysis, development, writing and interpretation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Chen, Q., Allot, A., & Lu, Z. (2020). Keep up with the latest coronavirus research. *Nature, 579*(7798), 193. https://doi.org/10.1038/d41586-020-00694-1.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal, 22*(2), 338–342. https://doi.org/10.1096/fj.07-9492LSF

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., & Coates, J. A. (2020). Preprinting a pandemic: the role of preprints in the COVID-19 pandemic. *bioRxiv* (preprint) https://doi.org/10.1101/2020.05.22.111294.

Hussain, I., & Hasghar, S. (2017). A survey of author name disambiguation techniques: 2010 to 2016. *The Knowledge Engineering Review, 32*, e22. https://doi.org/10.1017/S0269888917000182.

Ioannidis, J. P. A. (2020). Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. *European Journal of Clinical Investigation, 50*(4), e13223. https://doi.org/10.1111/eci.13223

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics, 106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Teixeira da Silva, J. A. (2020a). Chinese names in the biomedical literature: Suggested bibliometric standardization. *Publishing Research Quarterly, 36*(2), 254–257. https://doi.org/10.1007/s12109-020-09725-1.

Teixeira da Silva, J. A. (2020b). An alert to COVID-19 literature in predatory publishing venues. *The Journal of Academic Librarianship, 46*(5), 102187. https://doi.org/10.1016/j.acalib.2020.102187.

Vera-Baceta, M., Thelwall, M., & Kousha, K. (2019). Web of science and Scopus language coverage. *Scientometrics, 121*(3), 1803–1813. https://doi.org/10.1007/s11192-019-03264-z.