Review

Coronavirus Disease (COVID-19): A Machine Learning Bibliometric Analysis

FRANCESCA DE FELICE[1] and ANTONELLA POLIMENI[2]

[1]Department of Radiotherapy, Policlinico Umberto I "Sapienza" University of Rome, Rome, Italy;
[2]Department of Oral and Maxillo Facial Sciences, Policlinico Umberto I,
"Sapienza" University of Rome, Rome, Italy

Abstract. *Background/Aim: To evaluate the research trends in coronavirus disease (COVID-19). Materials and Methods: A bibliometric analysis was performed using a machine learning bibliometric methodology. Information regarding publication outputs, countries, institutions, journals, keywords, funding and citation counts was retrieved from Scopus database. Results: A total of 1883 eligible papers were returned. An exponential increase in the COVID-19 publications occurred in the last months. As expected, China produced the majority of articles, followed by the United States of America, the United Kingdom and Italy. There is greater collaboration between highly contributing authors and institutions. The "BMJ" published the highest number of papers (n=129) and "The Lancet" had the most citations (n=1439). The most ubiquitous topic was COVID-19 clinical features. Conclusion: This bibliometric analysis presents the most influential references related to COVID-19 during this time and could be useful to improve understanding and management of COVID-19.*

The coronavirus disease (COVID-19) pandemic puts unprecedented pressures on healthcare systems worldwide (1). Extensive measures to reduce and prevent its transmission have been implemented (1). As COVID-19 spreads rapidly, the research community has been active in publishing novel articles on this dreadful disease. In this scenario, a bibliometric analysis of the current COVID-19

landscape can improve clinicians' understanding of the large and growing body of evidence on the topic.

Bibliometric analysis is a statistical evaluation of scientific publications and represents an effective method to measure and quantitatively describe their influence in clinical practice over time (2).

Here we present a bibliometric analysis of the publications on COVID-19. The aim is to identify the trends of present studies and provide some directions for future research.

Materials and Methods

A comprehensive search of Scopus database was performed, using "COVID-19" as research criteria. Literature search was refined to include manuscripts published between December 1, 2019 and April 20, 2020. Statistical analysis was carried out using R-Studio 0.98.1091 software. A machine learning bibliometric methodology was applied to evaluate the distribution of each factor. The bibliometrix R package was used (2). The following variables were used: citation information – author(s), document title, year, source title, volume, issue, pages, citation count, source and document type, bibliographical information, affiliations, editor(s), keywords and funding details (number, acronym, sponsor).

Results

In total, 1883 documents matched the choice criteria across 9 document types. The 9 document types were article (n=821), letter (n=308), editorial (n=267), note (n= 28), review (n=209), short survey (n=26), erratum (n=21), conference paper (n=2) and data paper (n=1).

Publications came from 94 countries worldwide. The most productive country in the Covid-19 research field was China, with a publication share of 30 % (n=565). The United States of America ranked second (n=368, 19.5%), followed by the United Kingdom (n=218, 11.6%), Italy (n=143, 7.6%) and France (n=74, 3.9%). The remaining top 20 most productive countries were mostly in Europe and Asia. Figure 1 displays country collaborations. The network diagram – each node
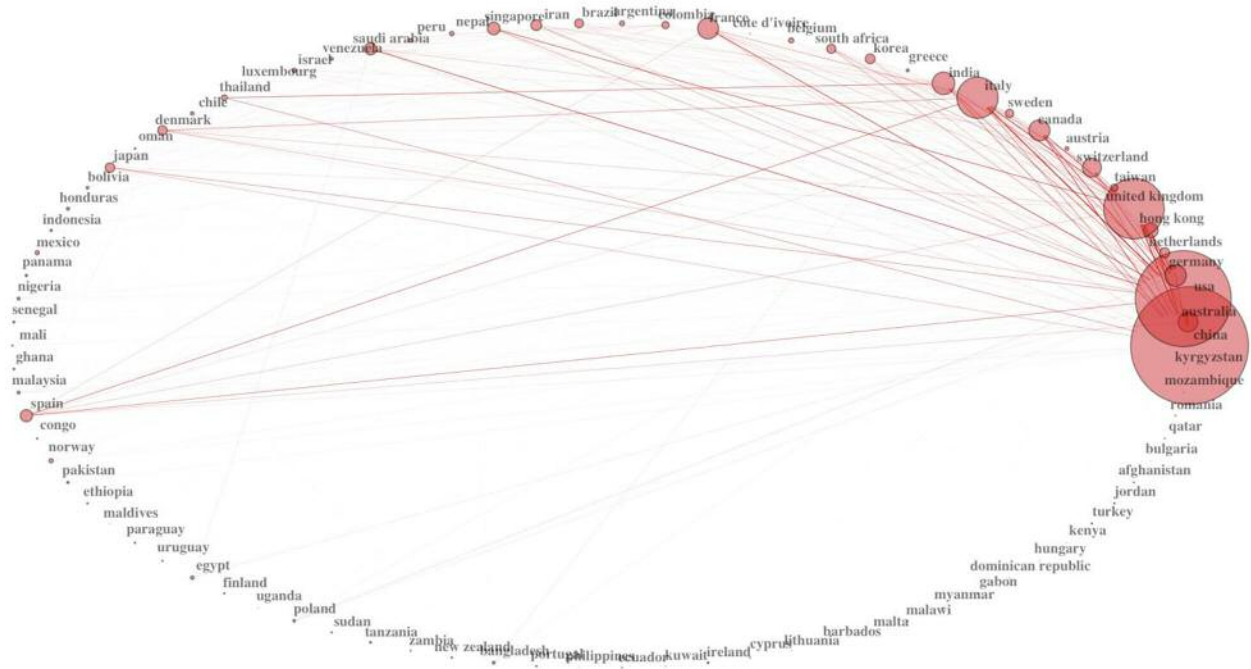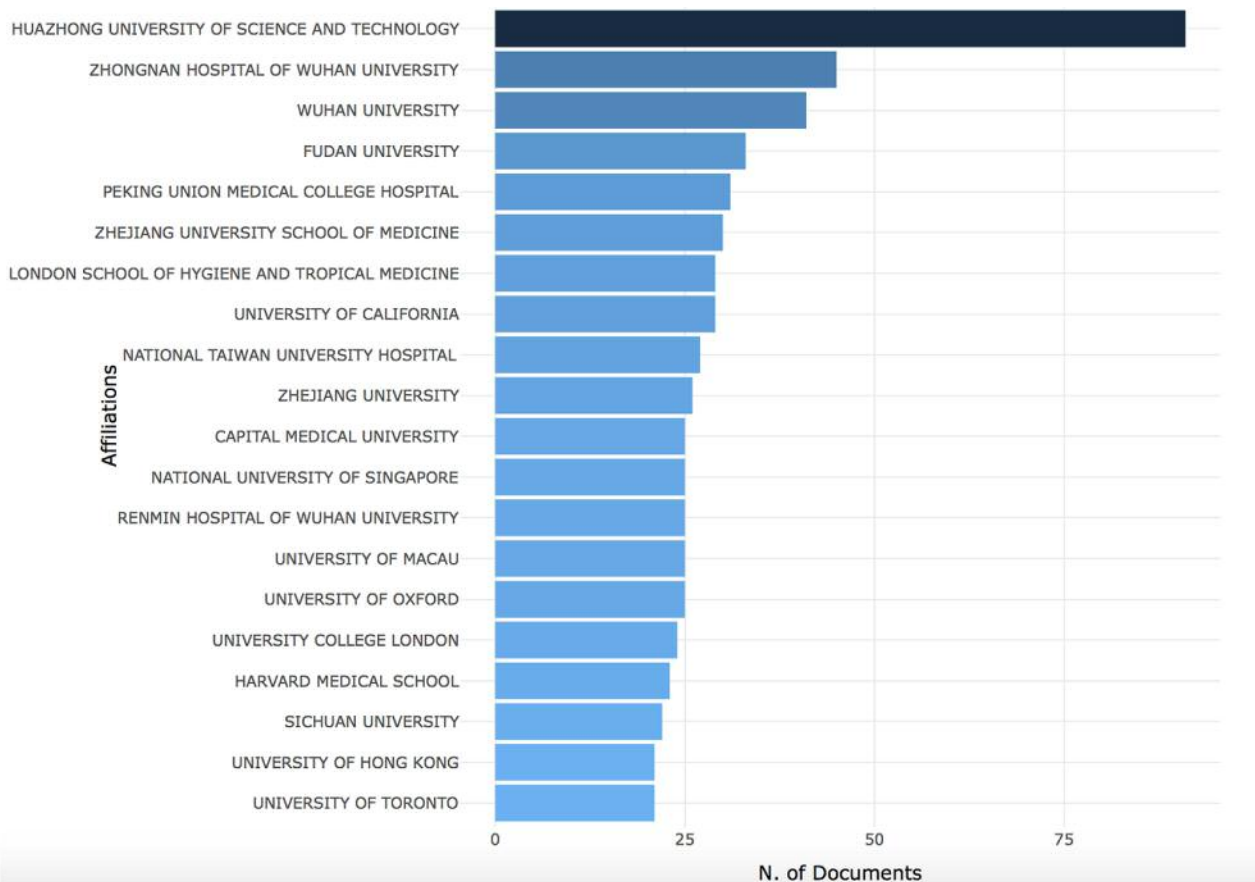
Figure 1. *Country collaboration*.

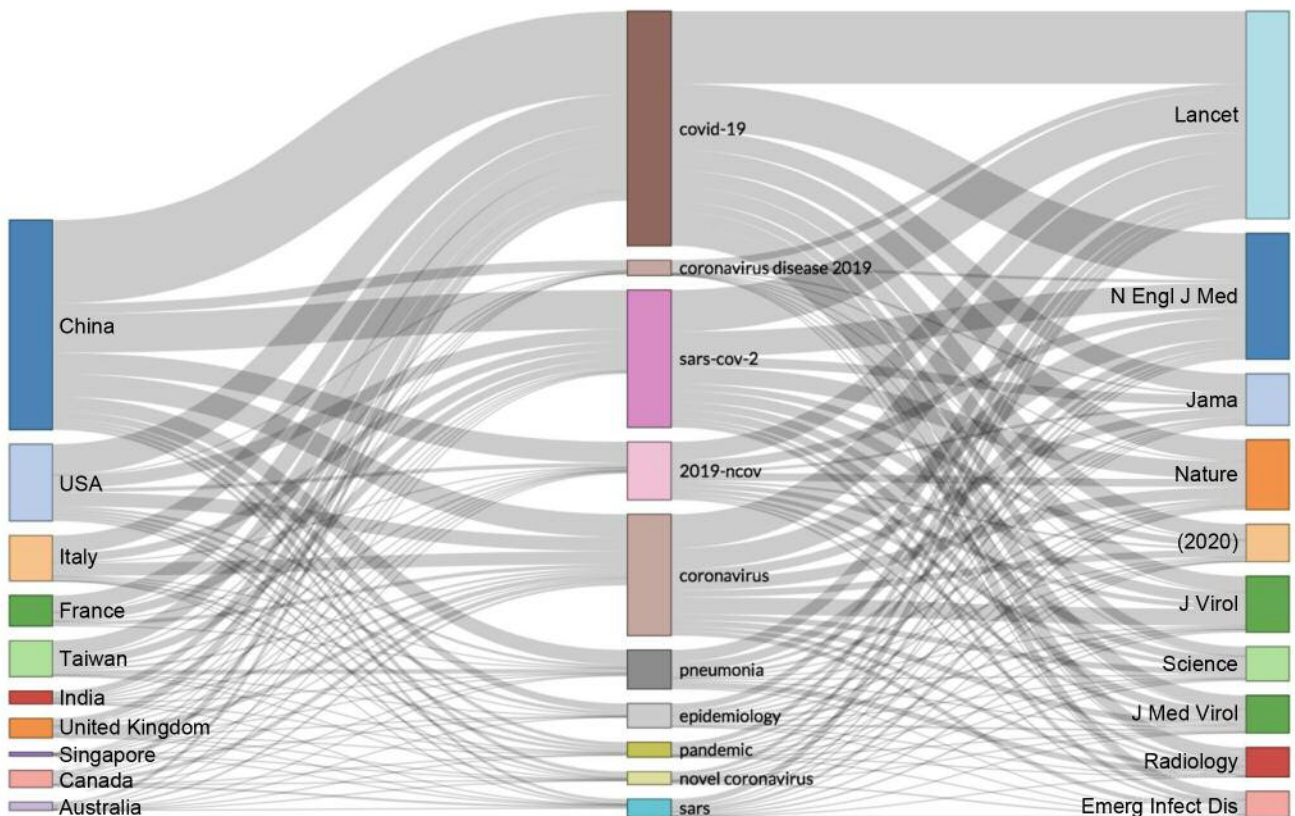Figure 2. *The top twenty relevant affiliations*.

Figure 3. *Summary plot matching the most productive countries (on the left), the most relevant author keywords (in the middle) and the most relevant sources (on the right).*

represented a country, node size corresponded to publications number, connecting lines represented country cooperation and line thickness indicated collaboration frequencies – showed that China was the leader in COVID-19 research in cooperation with other countries. Overall, developed countries had greater collaboration networks than developing territories.

The distribution of the most productive authors reiterated the predominance of China in the COVID-19 research field. The most productive institution was the Huazhong University of Science and Technology, accounting for 4.8% (n=91) of the entire documents on this topic (Figure 2). The 1883 articles were published in 644 journals, with 400 journals publishing only one article. The "BMJ" headed the list with a total number of 129 publications, followed by "The Lancet" journal (n=3) and the "Journal of medical virology" (n=60). When analyzed publications' reference lists, the most cited sources were "The Lancet" and "The New England Journal of Medicine" in 1439 articles and 1014 articles, respectively. Overall, the three most common author keywords were COVID-19 with 475 occurrences,

coronavirus with 219 occurrences and SarsCov2 with 217 occurrences. Figure 3 summarizes these results matching the most productive countries (on the left), the most relevant author keywords (in the middle) and the most relevant sources (on the right).

Table I lists the top 10 cited papers (3-12). The number of citations ranged from 533 for Huang *et al*. – "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." (3) – to 88 for Wu *et al*. – "A new coronavirus associated with human respiratory disease in China" (12) –. The Lancet had the most citations (n=4 and 1253 citations). Overall, clinical characteristic was the topic most widely studied (n=8) (3-7, 9-10, 12). One manuscript examined the genomic features and one study analyzed infection transmission from asymptomatic contact (8, 11).

The National Natural Science Foundation of China was the most productive funding agency with 85 funded articles, followed by the National Basic Research Program of China and the National Institutes of Health, with 21 funded articles and 20 funded articles, respectively.

Table I. *The most cited papers on COVID-19.*

Title	Authors	Year	Journal	Citations (n)
Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China (3)	Huang *et al.*	2020	The Lancet	533
A novel coronavirus from patients with pneumonia in China, 2019 (4)	Zhu *et al.*	2020	New England Journal of Medicine	346
Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study (5)	Chen *et al.*	2020	The Lancet	295
A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster (6)	Chan *et al.*	2020	The Lancet	235
A pneumonia outbreak associated with a new coronavirus of probable bat origin (7)	Zhou *et al.*	2020	Nature	211
Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding (8)	Lu *et al.*	2020	The Lancet	190
First case of 2019 novel coronavirus in the United States (9)	Holshue *et al.*	2020	New England Journal of Medicine	147
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 (10)	Wu *et al.*	2020	JAMA	128
Transmission of 2019-ncov infection from an asymptomatic contact in Germany (11)	Rothe *et al.*	2020	New England Journal of Medicine	122
A new coronavirus associated with human respiratory disease in China (12)	Wu *et al.*	2020	Nature	88

Discussion

The aim of this bibliometric analysis was to summarize and analyze the evolution of the immediate impact of the COVID-19 pandemic on scientific production. We presented a body of publications that quantified country-specific and worldwide coverage of COVID-19 literature at present. The volume and focus of publications showed a notable increase in the overall number of documents during the study period. There were extensive intra- and inter-collaboration networks between highly productive authors and institutions. Results confirmed a rapid sharing of data from China and defined the clinical features of COVID-19 patients as the most common topic.

This is not surprising considering that COVID-19 infection was first reported in Wuhan City (China) and mainly caused serious respiratory disease (13). Chinese experience represented the scientific basis for the development of management strategies all over the world. This approach was powerful in helping to train local staff and bring clinical services online. It is important to recognize the interdisciplinary usefulness – involving medical science, social science, engineering and economics – to solve the problem. A comprehensive knowledge is crucial for the development of an adequate approach that could be deployed quickly and operated with ease. For instance, with the development of the pandemic, the asymptomatic presentation has become one of the focus of research to prevent the COVID-19 transmission. The more information on virus pathogenesis can be shared, the more rapidly specific treatments can be developed.

The main limitation of this paper is its intrinsic bias. It was based on Scopus database and we recognized that results may differ according to other databases or the inclusion of

other search terms. But, to our knowledge, there has been no recent comprehensive bibliometric analysis of COVID-19 publications. In our opinion, there was an urgent need to combine this information to derive the best benefit for the human population. Our bibliometric analysis provides not only a descriptive quantitative analysis, but also evidence that implementation and scale-up of networks are feasible. To conclude, a similar analysis can be useful to empower countries to develop and sustain better health care by establishing a more robust collaboration network, including low-income countries. All the important aspects in terms of COVID-19 epidemiology, virology, diagnosis and clinical characteristics should be actively shared to stimulate researchers to work with local and international groups of professionals.

Conflicts of Interest

The Authors have no conflicts of interest to disclose in relation to this study.

Authors' Contributions

FDF and AP have made substantial contributions to the conception of the work, the acquisition and interpretation of data and have drafted the work. All Authors have approved the submitted version.

References

1 WHO Health Emergency Dashboard. Available at: https://extranet. who.int/publicemergency [Last accessed 21 April, 2020]

2 Aria M and Cuccurullo C: Bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetrics *11(4)*: 959-975, 2017. DOI: 10.1016/j.joi.2017.08.007

3 Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J and Cao B: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395(10223)*: 497-506, 2020. PMID: 31986264. DOI: 10.1016/S0140-6736(20)30183-5

4 Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF and Tan W; China Novel Coronavirus Investigating and Research Team: A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med *382(8)*: 727-733, 2020. PMID: 31978945. DOI: 10.1056/NEJMoa2001017

5 Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X and Zhang L: Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet *395(10223)*: 507-513, 2020. PMID: 32007143. DOI: 10.1016/S0140-6736(20)30211-7

6 Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK and Yuen KY: A familial cluster of pneumonia associated with the 2019

novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet *395(10223)*: 514-523, 2020. PMID: 31986261. DOI: 10.1016/S0140-6736(20)30154-9

7 Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF and Shi ZL: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579(7798)*: 270-273, 2020. PMID: 32015507. DOI: 10.1038/s41586-020-2012-7

8 Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF10, Wu G, Chen W, Shi W and Tan W: Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395(10224)*: 565-574, 2020. PMID: 32007145. DOI: 10.1016/S0140-6736(20)30251-8

9 Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A, Diaz G, Cohn A, Fox L, Patel A, Gerber SI, Kim L, Tong S, Lu X, Lindstrom S, Pallansch MA, Weldon WC, Biggs HM, Uyeki TM and Pillai SK; Washington State 2019-nCoV Case Investigation Team: First case of 2019 novel coronavirus in the United States. N Engl J Med *382(10)*: 929-936, 2020. PMID: 32004427. DOI: 10.1056/NEJMoa2001191

10 Wu Z and McGoogan JM: 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, 2020. PMID: 32091533. DOI: 10.1001/jama.2020.2648

11 Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, Zimmer T, Thiel V, Janke C, Guggemos W, Seilmaier M, Drosten C, Vollmar P, Zwirglmaier K, Zange S, Wölfel R and Hoelscher M: Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med *382(10)*: 970-971, 2020. PMID: 32003551. DOI: 10.1056/NEJMc2001468

12 Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC and Zhang YZ: A new coronavirus associated with human respiratory disease in China. Nature *579(7798)*: 265-269, 2020. PMID: 32015508. DOI: 10.1038/s41586-020-2008-3

13 Ahn DG, Shin HJ, Kim MH, Lee S, Kim HS, Myoung J, Kim BT and Kim SJ: Current status of epidemiology, diagnosis, therapeutics, and vaccines for novel coronavirus disease 2019 (COVID-19). J Microbiol Biotechnol *30(3)*: 313-324, 2020. PMID: 32238757. DOI: 10.4014/jmb.2003.03011