Letter to the Editor

A bibliometric study of COVID-19 research in Web of Science

Since December 2019, a number of unexplained viral pneumonia cases have been found in some hospitals [1,2]. On 11 February, 2020, the International Committee on Taxonomy of Viruses (ICTV) announced the official name for novel coronavirus: severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) [3]. On the same day, the World Health Organization (WHO) announced that the novel coronavirus was officially named as "COVID-19" [4,5]. The COVID-19 pandemic has now emerged as one of the world's greatest health challenges. SARS-CoV-2 has spread to five continents. As of 6:31 p.m. CET on 28 January 2021, there were a total of 100,455,529 confirmed cases of COVID-19 involving 223 countries, areas or territories with cases worldwide, including 2,166,440 deaths, which have been reported to WHO [6]. Although most patients with mild infection had a good disease prognosis, some patients had a higher mortality rate from diffuse alveolar injury and acute respiratory distress syndrome (ARDS), which was the main cause of death [7,8], and it mainly occurred in elderly patients or patients with underlying diseases [9,10]. The most common symptoms of COVID-19 are cough, fever and fatigue. Other symptoms include expectoration, diarrhoea, headache, hemoptysis, dyspnea and lymphocytopenia. Nevertheless, COVID-19 also showed some special clinical characteristics, including targeting the lower respiratory tract with obvious upper respiratory symptoms such as runny nose, sneezing, and sore throat [1,9,11]. At present, patients infected with COVID-19 are treated by various strategies such as convalescent plasma therapy (CPT), monoclonal antibody therapy (such as tocilizumab) and traditional Chinese medicine [12–15]. As of December 7, 2020, a total of 214 vaccines are under development worldwide, 52 of which have entered the clinical trial stage, and 13 of which are in the phase III clinical stage [15]. Currently, several vaccines have been approved for full marketing and vaccination.

The goal of this work is to conduct a comprehensive bibliometric analysis using big data analysis to identify hot research topics in various institutions, journals, and active groups of academic institutions and physicians/scientists. This approach has proved to be of considerable value in quantitatively describing the output of various scientific publications in other areas of biomedical research [16–21]. This comprehensive bibliometric analysis can not only reveal the major countries, major institutions and research impact, but also obtain the main research direction and trend information. Researchers and non-researchers can use these data to quickly locate and better achieve goals in their respective fields of science (for example, to identify research topics or identify potential partners). Therefore, in this study, our goal was to analyse the literature related to COVID-19, and at the same time to identify, collate and analyse the 50 most frequently cited papers on COVID-19 and the 2000 most frequently cited papers on COVID-19, so as to better grasp the publication trend, research topics and other useful information.

A total of 51,047 publications have been retrieved in this literature search as of November 6, 2020. Most of them were original articles (n = 24,488, 47.971%) and reviews (n = 5287, 10.357%). The majority of papers were written in English (n = 48,719, 95.439%). The top 50 most cited papers related to COVID-19 are shown in Table S1. They mainly consist of articles (n = 33) and letters (n = 11), as well as Review (n = 3) and Editorial Material (n = 3). All 50 papers were written in English and published in the journal with an Impact Factor (IF) ranging from 1.553 to 74.699. Symptoms of infection and therapy seemed to be two research hotspots. The relevant research key words were China (n = 33), patient (n = 32), infection (n = 28), COVID (n = 31), SARS CoV (n = 23), and Wuhan (n = 21). The analysis of the top 2000 most cited papers related to COVID-19 therapy as of January 22, 2021 showed that the relevant research key words were respiratory distress syndrome (n = 123), acute respiratory syndrome (n = 68), antibody (n = 60) and structure (n = 54).

The top five WoS categories of the analysed COVID-19 publications were Medicine General Internal (n = 6483, 12.700%), Public Environmental Occupational Health (n = 3817, 7.477%), Infectious Diseases (n = 2643, 5.178%), Surgery (n = 2296, 4.498%), Immunology (n = 1770, 3.467%).

The top ten institutions are distributed in the US, UK, China, France, Canada with citations per publication (CPP) counts ranking from 8.19 to 41.48 (Table S2). The top ten countries/regions with the most in-depth research in the field of COVID-19 are the US, followed by China and Italy (Table S3). Not surprisingly, the dominance of the US is similar to the results of related bibliometric analyses in other research areas, such as neuroscience in general [19,20], neuroimaging [21]. In terms of CPP, China is far ahead of the other nine countries.

The top ten journals are *British Medical Journal (BMJ)* (n = 999, 1.957%), *Journal of Medical Virology* (n = 658, 1.289%), *International Journal of Environmental Research and Public Health* (n = 416, 0.815%), *Cureus* (n = 407, 0.797%), *Lancet* (n = 385, 0.754%) and so on. Among them, the CPP counts of *New England Journal of Medicine* (n = 253, 0.496%) was the highest, reaching 107.82 (Table S4).

The top ten research areas are General Internal Medicine (n = 7458, 14.610%), Public Environmental Occupational Health (n = 3817, 7.477%), Infectious Disease (n = 2643, 5.178%), Surgery (n = 2296, 4.498%), Cardiovascular System Cardiology (n = 2208, 4.325%), Neurosciences Neurology (n = 1973, 3.865%), Pharmacology Pharmacy (n = 1826, 3.577%), Immunology (n = 1770, 3.467%), Psychiatry (n = 1731, 3.391%) and Oncology (n = 1718, 3.366%) (Table S5).

SARS-CoV-2 is a single-stranded RNA positive strand enveloped β-coronavirus [13,22]. Its genome encodes nonstructural proteins, structural proteins, and helper proteins. Non-structural proteins include 3-chymotrypsin-like protease, papain-like protease, helicase and RNA-dependent RNA polymerase (RdRp), while structural proteins consist of the membrane protein, the envelope protein, the spicule, and the

spike glycoprotein [13,23]. The spike glycoprotein is an indispensable part in mediating virus invasion [24], while the other four non-structural proteins make a critical difference in virus proliferation. Therefore, these five proteins are considered as significant multi-targets for the development of antiviral drugs.

Nucleoside inhibitors targeting the key enzyme of viral replication RdRp [13], including favipiravir [25,26], ribavirin [27,28], remdesivir [14,26,27,29], galidesivir [27] and so on. In a study of 237 patients with COVID-19, 158 received remdesivir and the rest received placebo. The results showed that patients treated with remdesivir in the first few days of symptoms improved more quickly than those treated with placebo [29]. Protease inhibitors for HIV-1 and HIV-2 are also effective, including kaletra (lopinavir/ritonavir) [26,28,30], darunavir [26,31], atazanavir [32], indinavir [31,33], nelfinavir [34], and so on.

Coronaviruses specifically bind to cell receptors mainly through the receptor binding domain (RBD) of S protein. After binding, the conformational change of S protein is induced, and the fusion of virus envelope and cell membrane is mediated, so that the nucleocapsid of virus enters the cell for subsequent viral replication [35]. Angiotensin converting enzyme 2 (ACE2) is a receptor of SARS-CoV. ACE2 may interact with S protein, may also be mediated SARS-CoV-2 infecting alveolar type II epithelial cells [23]. The first strategy is to give patients a drug, which binds to the ACE2 receptor. The advantage of this approach is that the ACE2 protein of host does not change, so there is less concern about the binding of drugs. Another approach is to design a class of antibody-like molecules that combine with the coronavirus itself and express a soluble ACE2 receptor protein that binds to the S protein of SARS-CoV-2 to neutralize the virus [36].

Clinical studies have shown that COVID-19 is correlated with IFN levels (including IFNA-2a, IFNA-2b, IFNB-1a, IFNB-1b, etc.), and the disease symptoms of patients have been alleviated to a certain extent after further treatment with IFN, but the specific mechanism and long-term treatment effect remain to be observed and studied. IFN has the potential to become a therapeutic cornerstone for COVID-19 [13,28,37,38].

Small-molecule drugs already approved to treat other human diseases may also modulate the interplay between SARS-CoV-2 and its host. Such as chloroquine [39,40], chloroquine phosphate [14,41] and hydroxychloroquine [26,40] have been approved.

Severe complications of COVID-19 are thought to be driven by a "cytokine storm" (or "cytokine release syndrome") [42], particularly by signaling molecules called interleukin 1β (IL-1β) and interleukin 6 (IL-6), which can rapidly cause single or multiple organ failure when they occur. Tocilizumab (TCZ) is a monoclonal antibody against IL-6 receptor that improves the prognosis of patients by blocking the inflammatory storm that prevents progression to severe and critical illness [43]. In a research of 64 patients admitted for COVID-19, 49 patients (76.6%) responded well to TCZ early on [42]. Canakinumab is an IL-1β inhibitor that targets to bind and neutralize IL-1β, thereby inhibiting the inflammatory response by blocking its action. Nevertheless, a multicenter, randomized, double-blind, placebo-controlled phase 3 trial failed [44]. In addition, a study involving 8 patients in Greece showed that anakinra (essentially an IL-1 inhibitor), an anti-inflammatory drug used to treat rheumatoid arthritis, improved symptoms in critically ill patients with COVID-19 [45].

CPT is a treatment method that collects antibody-rich blood from recovered patients and transfuses it to other patients after treatment, which belongs to passive immunotherapy. More and more evidence confirms that CPT is safe and reliable in critically ill patients currently receiving treatment [46]. The US Food and Drug Administration (FDA) has also approved CPT for patients with COVID-19, but more clinical researches are needed [47]. In 5 critically ill patients with COVID-19 and ARDS, clinical outcomes were improved after the administration of convalescence plasma containing neutralizing antibodies [48].

Dexamethasone is a glucocorticoid. As a treatment option, it usually suppresses inflammation and immune responses. British scientists conducted clinical trials and found that dexamethasone can help patients survive under certain conditions [49]. However, the balance between the risks and benefits of hormone therapy needs to be carefully assessed.

Baricitinib is an oral JAK1/JAK2 inhibitor approved for the therapy of moderate to severe stage of patients with rheumatoid arthritis (RA) in more than 65 countries [50]. Considering the cytokine storm of COVID-19, the anti-inflammatory activity of baricitinib is considered to be potentially beneficial in the treatment of COVID-19. Nevertheless, in one study, baricitinib is not a perfect treatment option for COVID-19 [51] because of the observed increase in creatine kinase in patients receiving baricitinib, this could potentially pose a therapeutic risk [52]. Ruxolitinib is a pioneering orally bioavailble JAK1/JAK2 inhibitor. Overactivation of the JAK-STAT pathway has been associated with a variety of cancers and other severe immune-mediated diseases. Many patients with severe respiratory disease (e.g., pneumonia) caused by COVID-19 have characteristics consistent with cytokine storms and enhanced activation of the JAK-STAT pathway, and many studies have suggested that ruxolitinib possibly play an important part in treating patients with COVID-19 [53–55].

Fig. S1 shows the term maps of the titles and abstracts of the 50 most cited COVID-19 papers as of November 6, 2020. The largest bubbles represent a few keywords, such as China (n = 33; CPP = 1657.91), patient (n = 32; CPP = 1724.72), infection (n = 28; CPP = 1359.71), COVID (n = 31; CPP = 1068.10), SARS CoV (n = 23; CPP = 1310.17), and Wuhan (n = 21; CPP = 2123.86). In the meantime, examples of keywords with highest CPPs included researcher (n = 2; CPP = 5694.00), family (n = 2; CPP = 5360.00), real time RT PCR (n = 2; CPP = 5273.00), affected patient (n = 2; CPP = 4335.50), clinical characteristics (n = 2; CPP = 4335.50), Huanan seafood market (n = 2; CPP = 4155.67), and chest (n = 4; CPP = 4068.00). Among these 281 terms, there were many symptoms related to COVID-19 and complications caused by COVID-19, such as acute kidney injury (AKI; n = 2; CPP = 1034.50), acute respiratory distress syndrome (ARDS; n = 3; CPP = 2216.33), diabetes (n = 3; CPP = 1871.67), diarrhea (n = 2; CPP =2631.00), respiratory failure (n = 2; CPP = 666.00), acute respiratory syndrome (ARS; n = 2; CPP = 1761.50), severe acute respiratory syndrome (SARS; n = 7; CPP = 1306.29), acute cardiac injury (ACI; n = 2; CPP = 3781.50), dyspnoea (n = 2; CPP = 3781.50), myalgia (n = 2; CPP = 3779.50), sore throat (n = 2; CPP = 2192.00), and so on.

Fig. 1 shows the term maps of the titles and abstracts of the 2000 most cited COVID-19 therapy papers as of January 22, 2021. The largest bubbles represent keywords, such as respiratory distress syndrome (n = 123; CPP = 296.36), acute respiratory syndrome (ARS; n = 68; CPP = 122.15), antibody (n = 60; CPP = 114.07) and structure (n = 54; CPP = 115.22). At the same time, examples of keywords with highest CPPs included haemoptysis (n = 2; CPP = 5058.50), IL-10 (n = 2; CPP = 4895.50), hospital patient (n = 2; CPP = 4883.50), abnormal finding (n = 2; CPP = 4777.00), Huanan seafood market (n = 4; CPP = 3766.75), higher plasma level (n = 3; CPP = 3351.75), and sputum production (n = 3; CPP = 3284.00). The 3651 terms covered a wide range of therapies, such as antiretroviral therapy (n = 3; CPP = 56.00), radiation therapy (n = 5; CPP = 45.00), kidney replacement therapy (n = 2; CPP = 720.50), high flow oxygen therapy (n = 3; CPP = 1934.00), convalescent plasma transfusion (n = 10; CPP = 121.00), corticosteroid therapy (n = 4; CPP = 55.50), pharmacotherapy (n = 3; CPP = 93.67), and antithrombotic therapy (n = 2; CPP = 324.00). They also include common drugs such as anakinra (n = 5; CPP = 83.00), lopinavir/ritonavir (n = 2; CPP = 734.50), tocilizumab (n = 6; CPP = 83.33), hydroxychloroquine (n = 2; CPP = 67.50), dexamethasone (n = 2; CPP = 289.00), remdesivir (n = 5; CPP = 382.80), and so on.

We recognize that this study has several limitations. One important limitation of this article is that all the data in this article was calculated and presented only by extracting data from the WoS database, which means that manuscripts not listed in the WoS are not included in this article. These include Science Citation Index Expanded (SCI-Expanded),
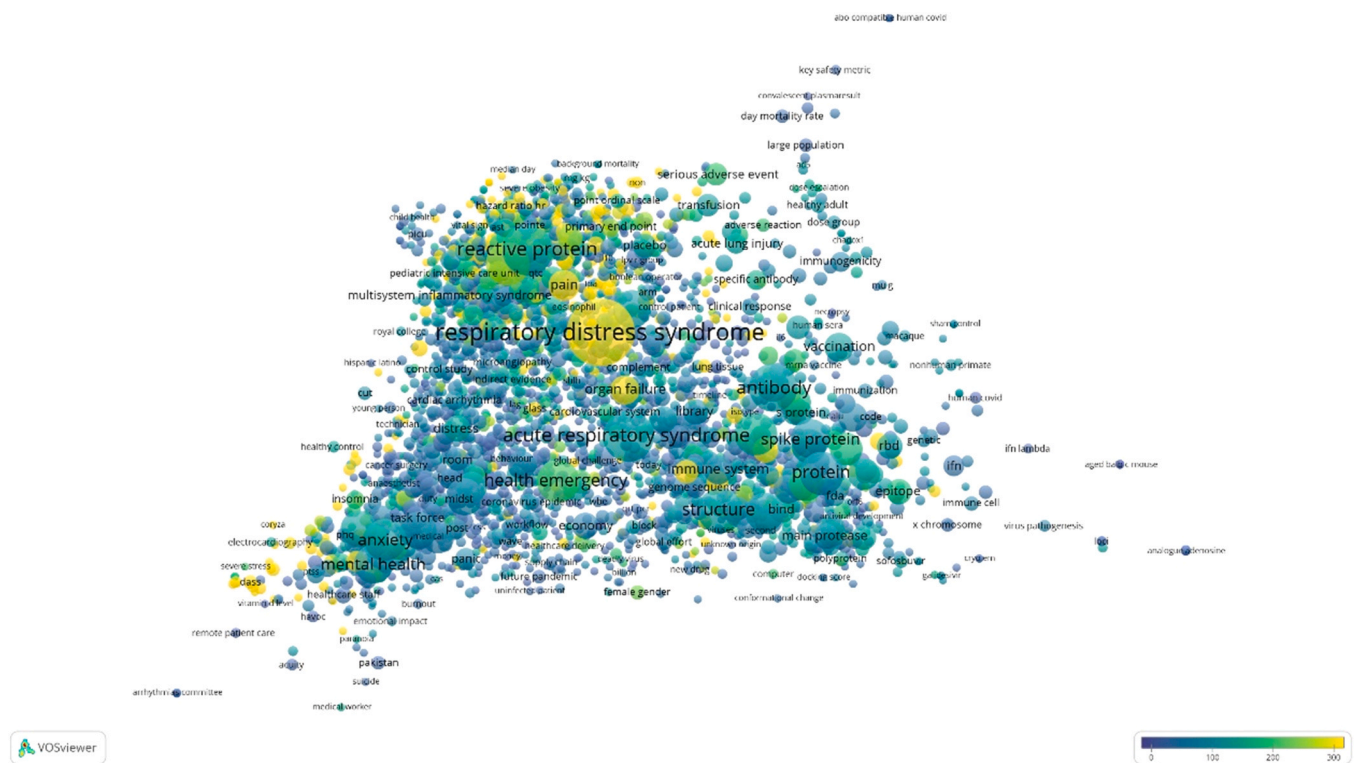
Fig. 1. Term maps of the titles and abstracts of the 2000 most cited COVID-19 therapy papers as of January 22, 2021. There were 3651 keywords or phrases that appeared in at least two papers, therefore they were included in the visualization. Complications caused by COVID-19 seem to be the focus of research, such as the largest bubble being respiratory distress syndrome. Therapies of COVID-19 not only include dexamethasone, remdesivir, but also include CPT, monoclonal antibody therapy.

Social Sciences Citation Index (SSCI), Conference Proceedings Citation Index-Science (CPCI-S), etc. The reason we used WoS Core Collection is because it is a professional database, which can track, reference and analyse data, while another professional database is Scopus. The two databases are different and the calculation methods are different, but the evaluation indexes of the two have good consistency and high correlation [56]. Since COVID-19 is still an ongoing pandemic, the number of publications is increasing and the research focus is changing dynamically. In particular, many COVID-19 related papers were deposited in preprint servers, such as BioRxiv and MedRxiv, future analysis by integrating publications in these preprint servers will provide a more precise picture of the COVID-19 research trend. Secondly, focusing on a small number of most cited publications (top 50 or top 2000) would potentially provide a biased picture of early COVID-19 research, instead of major trends in the literature. Thirdly, in this study, only the most cited publications are analyzed and discussed, there is a potential bias of countries (e.g., US, China, UK etc.) and institutions known to publish the largest share of research papers, as well as dominate globally in high-impact journals, i.e., logically linked to high numbers of publications.

By counting COVID-19 related papers retrieved in WoS, the *British Medical Journal* was the focus and priority of publishing channel. The 50 most cited COVID-19 papers were identified by bibliometric analysis. The *Lancet* or the *New England Journal of Medicine* was the preferred choice of publishing channel. Infection symptoms, complications, and the SARS-CoV-2 virus seem to be the main topics. According to the results of the analysis of the top 2000 papers related to COVID-19 therapy, treatment methodsare the focus of research. Timely bibliometric and global analysis of all COVID-19 research not only reveals the major countries, institutions and research impacts of COVID-19 research, but also provides information on major research directions and trends, which will equip researchers with more knowledge to combat this devastating pandemic.

Code availability

Not applicable.

Authors' contributions

Jianping Weng and Suowen Xu designed the study. Meiming Su wrote the manuscript. All authors read, and revised the manuscript, and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.phrs.2021.105664.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (10223) (2020) 497–506, https://doi.org/10.1016/S0140-6736(20)30183-5. Epub 2020 Jan 24.

[2] H.Z. Lu, C.W. Stratton, Y.W. Tang, Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle, J. Med Virol. 92 (4) (2020) 401–402, https://doi.org/10.1002/jmv.25678. Epub 2020 Feb 12.

[3] A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B. L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L.L. M. Poon, D. Samborskiy, I.A. Sidorov, I. Sola, J. Ziebuhr, Severe acute respiratory syndrome-related coronavirus: the species and its viruses–a statement of the coronavirus study group, BioRxiv (2020), 937862, https://doi.org/10.1101/2020.02.07.937862.

[4] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, R. Agha, World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19), Int J. Surg. 76 (2020) 71–76, https://doi.org/10.1016/j.ijsu.2020.02.034. Epub 2020 Feb 26.

[5] M.K. Daga, N. Kumar, J. Aarthi, G. Mawari, S. Garg, I. Rohatgi, From SARS-CoV to coronavirus disease 2019 (COVID-19) – a brief review, J. Adv. Res Med. 6 (4) (2019) 1–9. https://doi.org/10.24321/2349.7181.201917.

[6] World Health Organization . Coronavirus disease (COVID-2019) situation reports [Internet]. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/.

[7] C. Wu, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, S. Wang, Y. Chao, Z. Yang, J. Xu, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, Y. Song, Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China, JAMA Intern. Med. 180 (7) (2020) 934–943, https://doi.org/10.1001/jamainternmed.2020.0994.

Meiming Su, Suowen Xu[*], Jianping Weng[*]

Institute of Endocrine and Metabolic Diseases, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230001, China

[*] Corresponding authors.

E-mail addresses: sxu1984@ustc.edu.cn (S. Xu), wengjp@ustc.edu.cn (J. Weng).