Network of institutions, source journals, and keywords on COVID-19 by Korean authors based on the Web of Science Core Collection in January 2021

Kyung Won Kim¹, Geum Hee Jeong²

¹Department of Nursing, Daegu Haany University, Daegu; ²School of Nursing, Hallym University Chuncheon, Korea

Abstract

Purpose: The aim of this study was to characterize the network of institutions, journals, and topics of coronavirus disease 2019 (COVID-19) literature by Korean authors in the Web of Science Core Collection. The specific goals were to identify the collaborative relationships between Korean authors and international authors and to explore clusters of institutions, journals, and topics.

Methods: Literature was searched in the Web of Science Core Collection on January 30, 2021. The search terms were “SARS-CoV-2” or “COVID” or “novel coronavirus” in the subject field. The search results were limited again to “South Korea” as the country and the publication type of “article.” The measurement tool was Biblioshiny, an app version tool for Bibliometrix.

Results: Korean authors published 3.2 times more COVID-19-related articles in journals outside of Korea than in Korean journals. The journals showed three clusters by bibliographic coupling. In contrast, the co-citation network showed four clusters. Only a few journals were included in the clusters in both analyses. The conceptual structure of Keywords Plus by factorial analysis showed two clusters: “pathology and clinical treatment” and “knowledge and attitudes.” Institutions’ collaborative network consisted of four clusters. Korean researchers actively collaborated with international researchers, especially those in the United States.

Conclusion: Because only a few Korean journals were included in the journal clusters by both coupling and co-citation network, more active citation of Korean journals is recommended. The identification of human behavior as a distinct theme in COVID-19 research suggests a different focus in this area besides clinical studies.

Keywords

Bibliometrics; Coronavirus; COVID-19; Journal publishing; Republic of Korea

Received: February 5, 2021
Accepted: February 10, 2021

Correspondence to Geum Hee Jeong
ghjeong@hallym.ac.kr

ORCID
Kyung Won Kim
<https://orcid.org/0000-0002-2498-728X>
Geum Hee Jeong
<https://orcid.org/0000-0002-2254-1962>

Introduction

Background/rationale: After the first report of an imported case of coronavirus disease 2019 (COVID-19) in January 2020 in South Korea (hereafter, Korea), infections have continued for a year, although there have been daily fluctuations in the number of reported cases. Korean researchers have also published papers on COVID-19 to provide information on this condition, with topics including its biology, diagnosis, treatment, prevention, prognosis, and epidemiology. Some studies have presented bibliometric analyses of the COVID-19-related literature, including networks of authors, affiliations, countries, source journals, and keywords. The analysis methods have included citation analysis, clustering by coupling, co-occurrence networks, co-citation networks, and collaboration networks. Bibliographic coupling occurs when two articles cite a third article together, indicating that both articles are likely to address the same topic [1].

The bibliometric analysis of COVID-19 has usually focused on global research in a specific field [2], and little data have been published on country-level analyses, except Iran [3], Peru [4], and India [5,6]. Therefore, we conducted a bibliometric analysis of COVID-19-related literature authored only by Korean researchers. The results of this analysis will provide insights into the diversity of research topics related to COVID-19, as well as networks of institutions, journals, and countries.

Objectives: This study investigated the networks of institutions, source journals, and keywords of COVID-19 literature published by Korean authors based on the Web of Science Core Collection on January 30, 2021. Bibliographic coupling and conceptual, intellectual, and social structures were analyzed using Biblioshiny. Furthermore, the topics of the literature were grouped into clusters to clarify trends in research. The specific goals were as follows: first, to identify collaborative relationships between Korean authors and authors in other countries; second, to compare the number of articles that Korean authors have published in Korean and international journals; third, to identify the cluster of Korean institutions that published numerous COVID-19 articles; fourth, to explore the Korean journal clusters that published numerous COVID-19 papers; and fifth, to identify topics primarily covered by Korean researchers and clusters of research areas.

The following hypotheses were set: first, there is a concentration effect among the institutions in Korea related to COVID-19 research; and second, Korean researchers have published more articles in international journals than Korean journals if the target journals are limited to those in the Web of Science Core Collection.

Methods

Ethics statement: This study did not involve human subjects, so neither approval by the institutional review board nor obtaining of informed consent was required.

Study design: This was a bibliometric study based on the literature in the Web of Science Core Collection.

Setting: On January 30, 2021, the literature was searched from the Web of Science Core Collection. The search term was “SARS-CoV-2” OR “COVID” OR “novel coronavirus” in the subject field. The search results were limited to authors from South Korea and publications from 2020 to January 2021. The number of results was 1,082, out of which only the publication type “article” was selected. The number of studies was 727, which included 667 articles, 59 early-access articles, and one proceedings paper. Data in plain text format were downloaded for analysis. There was no need for data cleaning. The downloaded plain text format data were converted to the R data format by the Biblioshiny app.

Variables: Variables were not required.

Data sources/measurement: Articles were selected after searching the Web of Science Core Collection as described above. The measurement tool was Biblioshiny, an app version tool of Bibliometrix (an R tool for comprehensive science mapping analysis available at <https://bibliometrix.org/Biblioshiny.html>) [7]. This tool was used because it is freeware, and it provides various analysis methods.

Selection of target and analysis methods through the Biblioshiny function interface menu: Biblioshiny provides a multifunctional interface according to the tutorial available from the above website. The following functions were selected for the present data analysis: first, the main information and three-fields plot from the Dataset menu; second, the most relevant source journals from the Sources menu; third, the most relevant affiliation (institution) from the Authors menu; fourth, the most frequent word and word cloud from the Document menu; fifth, clustering by coupling from the Coupling menu; sixth, the co-occurrence network and factorial network for Keywords Plus from the Conceptual Structure menu; seventh, the co-citation network for source journals from the Intellectual Structure menu; and eighth, the collaborative network of institutions and collaborative world map from the Social Structure menu. Only institutions, journals, and keywords were included in the analysis.

Bias: There was no bias in searching and selecting the target literature.

Study size: The sample size could not be estimated before the study. It was not required to estimate the sample size.

Statistical methods: Descriptive statistics were applied.

Results

Dataset 1 is the exported bibliometrix file of the 727 articles used for the Biblioshiny.

Main information and three-fields plot: The corresponding data are presented in Suppl. 1. The number of authors was 3,473. The number of single-authored documents was 46. The average number of authors per document was 4.78. Relationships among the top 20 institutions (intellectual root), top 20 journals, and top 20 Keywords Plus (research content) were summarized by a Sankey plot (a diagram used for the flow of input and output of given characteristics or objects) (Fig. 1).

Top 20 most relevant source journals: The most relevant journal title was *Journal of Korean Medical Science*, succeeded by *International Journal of Environmental Research and Public Health*, *Sustainability*, *International Journal of Infectious Diseases*, and *Journal of Clinical Medicine* (Fig. 2, Suppl. 2). In the top 20, *Journal of Korean Medical Science*, *Korean Journal of Internal Medicine*, *Epidemiology and Health*, *Infection and Chemotherapy*, and *Journal of the Korean Medical Association* were included as Korean journals.

Most relevant affiliations: The top 25 most relevant affiliations are presented in Fig. 3 (Suppl. 3). Seoul National University,

Yonsei University, Kyungpook National University, Korea University, and Sungkyunkwan University were the top five most relevant institutions (Fig. 3, Suppl. 3).

Most frequent words and word cloud: The most frequent word list is given in Suppl. 4. The word cloud generated from this list is presented in Fig. 4. “Pneumonia,” “outbreak,” “risk,” “models,” “China,” “health,” and “Wuhan” were the most frequent words besides “coronavirus,” “COVID-19,” and “SARS.”

Clustering and coupling of source journals: Given that the number of units was 250, the minimum cluster frequency per 1,000 units was 10, and the number of labels for each cluster was five, source journals’ coupling map measured by references and local citation score formed three clusters (Fig. 5, Suppl. 5). Out of the three clusters, *Annals of Laboratory Medicine*, *Korean Journal of Radiology*, and *Diabetes & Metabolism Journal* were in the same cluster. *Epidemiology and Health* was in another cluster. No other journal was included in the third coupling clusters.

Co-occurrence network and factorial network for Keywords Plus for the conceptual structure: The Keywords Plus co-occurrence network in conceptual structure is presented in Fig. 6 (Suppl. 6) under the following options: Keywords Plus field, automatic layout for network layout, association for normalization, no node color by year, Louvain for the clustering algo-

Fig. 1. Affiliation-journal-keyword flow of COVID-19-related articles of Korean authors in the Web of Science Core Collection generated from the Dataset menu of Biblioshiny on January 30, 2021.

rithm, 50 nodes, removal of isolated nodes, a minimum number of edges of 3, and a number of labels of 50. Six clusters are shown in Fig. 6. The main keywords of the three principal clusters were “pneumonia,” “coronavirus,” and “risk.”

The factorial analysis map of Keywords Plus in conceptual

structure is presented in Fig. 7 under the following options: multiple correspondence analysis, 50 terms, an automatic number of clusters, and 5 documents. Fig. 7 shows two clusters: one relates to pathogenesis and clinical care; the other is knowledge and attitudes (Suppl. 7). The dendrogram also

Fig. 2. Top 20 most relevant journal titles of the COVID-19–related articles of Korean authors in the Web of Science Core Collection on January 30, 2021.

Fig. 3. Top 20 most relevant affiliations of the COVID-19–related articles of Korean authors in the Web of Science Core Collection on January 30, 2021.

Co-citation network for source journals for the intellectual structure: Journals were clustered into four groups by the co-citation network of intellectual structure, given the options of an automatic layout, Louvain for the clustering algorithm, 50 nodes, no removal of isolated nodes, a minimum of two edges, and 50 labels (journals) (Fig. 9, Suppl. 8). Three Korean journals were included in one cluster, including the *Journal of Korean Medical Science*, *Osong Public Health and Research Perspectives*, and *Epidemiology and Health*. No other Korean journals were included in other three clusters.

The collaborative world map showed that internationally

Fig. 7. Conceptual structure map of Keywords Plus based on the factorial analysis of the COVID-19–related articles of Korean authors in the Web of Science Core Collection on January 30, 2021.

Fig. 8. Topic dendrogram of Keywords Plus based on the factorial analysis of the COVID-19–related articles of Korean authors in the Web of Science Core Collection on January 30, 2021.

Discussion

Korean authors published 3.2 times more COVID-19-related articles in journals outside Korea than in Korean journals in this analysis of data from the Web of Science Core Collection. The coupling of source journals showed three clusters, and the major journal of one cluster was *Annals of Laboratory Medicine* (local citation score 5.99). In contrast, the co-citation network showed four clusters of journals. The conceptual

Fig. 10. Collaborative network of institutions of the COVID-19-related articles of Korean authors in the Web of Science Core Collection on January 30, 2021.

structure of Keywords Plus by factorial analysis showed two clusters: one was pathogenesis and clinical care, and the other was knowledge and attitudes. The collaborative network of the institutions consisted of four clusters. The United States was the country with the most collaborations with Korean researchers for COVID-19 studies.

Interpretation

Network of institutions: The relevance was reflected in four institutional clusters. Yonsei University, Kyungpook National University, and Yeungnam University were in the same cluster (Fig. 10). Sungkyunkwan University, Soonchunhyang University, and Hanyang University were in the same cluster. Korea University and the University of Ulsan were in the same cluster. Seoul National University, Hallym University, and the Catholic University of Korea were in the same cluster. Those clusters reflect collaborative work between institutions. The cluster containing Korea University and the University of Ulsan showed stronger collaboration with international institutions, including the University of Toronto, Stanford University, Harvard University, the University of California at Los Angeles, and the University of Michigan. The three-field plot also showed flow from the most relevant institutions to the most relevant journals and to the most frequent Keywords Plus (Fig. 1).

Network of source journals: The most relevant journals from Korea were *Journal of Korean Medical Science*, *Korean Journal of Internal Medicine*, *Epidemiology and Health*, *Infection and Chemotherapy*, and *Journal of the Korean Medical Association* (Fig. 2). Two of these are general medicine journals, and the other two are journals in the category of epidemiology and infections. This finding reflects the fact that COVID-19 is an infectious and transmissible disease. Out of the top seven most relevant international journals, four were large journals, with 10,000 or more early-access publications: *International Journal of Environmental Research and Public Health*, *Sustainability*, *PLoS One*, and *Scientific Reports*. Korean authors published more articles in journals outside Korea than those published in Korea in this analysis limited to the Web of Science Core Collection. In Korean journals, 172 articles were published, while 555 articles were published in international journals.

The coupling of source journals showed three clusters. The journals with a high impact measure through the local citation score in each cluster were ranked as follows: first, *Annals of Laboratory Medicine*; second, *Disaster Medicine and Public Health Preparedness*; and third, *Computational Structural Biotechnology Journal* (Fig 5). This result is different from the most relevant journals (Fig. 2), because it reflects the citation score from 727 articles. In the first cluster, the *Korean Journal of Radiology and Diabetes & Metabolism Journal* were included.

In the second cluster, *Epidemiology and Health* was included. In the third cluster, no Korean journal was included in the top five ranking. Although there were other Korean journals with a higher number of articles, the citations were focused on the above-mentioned journals.

The co-citation network analysis of source journals for the intellectual structure generated four other clusters (Fig. 9). These findings are based on a co-citation network, and are therefore different from those obtained by coupling through a measure of high impact (Fig. 5). *Journal of Korean Medical Science*, *Osong Public Health and Research Perspectives*, and *Epidemiology and Health* were in the first cluster, and no Korean journals were in the second, third, and fourth clusters when the top-ranking 50 source journals were labeled. Those three journals dealt with the same topics. Both results showed that there were insufficient citations among articles published in Korean journals.

Keywords Plus network: In this analysis, Keywords Plus was used. This method is different from an analysis of the author's keywords, as described as follows: "The data in Keywords Plus are words or phrases that frequently appear in the titles of an article's references but do not appear in the title of the article itself. Based upon a special algorithm that is unique to Clarivate Analytics databases, Keywords Plus enhances the power of cited-reference searching by searching across disciplines for all the articles that have cited references in common" [8]. The Keywords Plus word cloud showed that besides the search terms—COVID-19, coronavirus, and SARS—"pneumonia," "outbreak," "risk," "model," and "China" frequently appeared. The word cloud presents terms in an easy-to-visualize format (Fig. 4). The conceptual structure of Keywords Plus based on the co-occurrence network showed 14 clusters (Fig. 6); however, the structure based on the factorial analysis showed two clusters (Figs. 7, 8). In the factorial analysis, multiple correspondence analysis was used as a dimensionality reduction technique. It has been described as follows: "Multiple correspondence analysis (MCA) is a data analysis technique for nominal categorical data, used to detect and represent underlying structures in a data set. It does this by representing data as points in a low-dimensional Euclidean space" [9]. The conceptual structure of Keywords Plus based on the factorial analysis indicated that the content of the COVID-19 studies focused primarily on the disease process. Nonetheless, there were research clusters that focused on human behavior, including intentions, attitudes, risk perception, information, knowledge, and the epidemic (Fig. 7). The elucidation of those two clusters were possible because dimensionality reduction technique is used in factorial analysis.

Comparison with previous country-level studies: No studies have yet presented bibliometric analyses of Korean research-

ers' publications on COVID-19, although country-level bibliometric analyses have been published for Iran, Peru, and India. The analysis of research from Iranian institutions comprised 849 papers on COVID-19 published in the Web of Science, Scopus, and PubMed until July 10, 2020. The number of papers by country ranked 13th in Scopus and 12th in Web of Science. An analysis of the co-authors' matrix showed that they frequently collaborated with researchers in the United States, Italy, United Kingdom, and Canada in descending order. Five clusters were identified in the co-occurrence network of keywords, indicating that "epidemiological studies and public health" and "clinical studies" were the largest clusters [3].

Twenty-four Peruvian authors' COVID-19 papers were selected from the PubMed/MEDLINE and SciELO databases and a direct search of the *Revista Peruana de Medicina Experimental y Salud Pública* archives up to May 21, 2020. Out of them, 29.7% were original articles or brief reports. The topic was primarily epidemiology. The articles were mainly published by researchers at an institution located in Lima, the capital city of Peru. Therefore, it was deemed necessary to conduct COVID-19 research in collaborate with other institutions [4].

The COVID-19 papers published in India from March 2 to May 12, 2020 were selected from the World Health Organization COVID-19 database. The papers on virology, diagnosis and treatment, and clinical features were more cited than papers dealing with epidemiological or pandemic-related topics [5]. On May 10, 2020, Indian authors' literature was searched from Google Scholar, Microsoft Academic, Lens, Dimensions, Scopus, PubMed, and Web of Science. The number of articles from India was within the top 10 ranking countries. India's top-ranking institutions, journals, and authors were listed. The keywords of the Indian authors could be grouped into 22 clusters [4].

The keyword analysis in India did not provide detailed data other than the keyword network diagram, which was not suitable for comparison with the present study. The co-occurrence network for author keywords from Iran showed five clusters: epidemiological and public health studies, clinical studies, signs and symptoms of the disease, the virus, and underlying diseases. It is difficult to compare those findings directly with present Keywords Plus clusters, which were grouped into disease processes and human behavior.

Limitations: The literature was limited to the Web of Science Core Collection. There were 2,301 articles on COVID-19 listed in the Korea Citation Index, the main scholarly journal abstract database in Korea, on February 2, 2021 (<https://www.kci.go.kr>). Therefore, the above results do not reflect the entire spectrum of Korean researchers' achievements on COVID-19.

In particular, there are many articles on social science and humanities topics.

Suggestions on COVID-19 bibliometric studies in Korea: Analyses of articles in local journals not indexed in international databases may provide information on the conceptual, intellectual, and social structures of literature on COVID-19. Furthermore, KoreaScience (<https://www.koreascience.or.kr/>) and KoreaMed (<https://koreamed.org/>) can serve as other excellent abstract databases for bibliometric analysis. Analyses of the literature in the above local databases will provide new information for bibliometric studies of the COVID-19 literature.

Generalizability: This study identified research topics related to COVID-19 in studies published by Korean researchers. These results may inspire researchers to engage with less frequently addressed research topics in the future. The presentation of a three-fields plot also makes it easy to understand the clustering of research topics among research institutions (Fig. 1).

Conclusion: The results provided sufficient answers for the research objectives. Korean researchers engaged in active collaborative work with international researchers, especially with those in the United States. The result of journal clusters by coupling was different from the journal clusters by the co-citation network. In both analyses, only a few Korean journals were included. Therefore, more active citations among Korean journals is recommended. The topics clustered by factorial analysis into two groups: disease processes and human behavior. The cluster of human behavior studies was small, but could be differentiated from other biomedical topics. Finally, two hypotheses set were accepted. First, the institutions that produced COVID-19 research in Korea were centralized among top-ranking institutions and grouped in four clusters. Second, Korean researchers published 3.2 times more articles in international journals than in Korean journals in the Web of Science Core Collection.

Research on COVID-19 will continue until the end of the pandemic. It is difficult to estimate when this pandemic will end completely, although vaccination and chemotherapies may be urgent solutions. Further regular follow-up studies on the conceptual, intellectual, and social structures of the literature will be necessary as research topics change.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Funding

This research was supported by Hallym University Research

Fund 2020 (HRF-202010-008).

Data Availability

Data are available from the author upon reasonable request.

Dataset 1. Exported bibliometrix data file of the 727 articles used for the Biblioshiny

Supplementary Material

Supplementary file is available from: the Harvard Dataverse at: <https://doi.org/10.7910/DVN/BKSGHR>

Suppl. 1. Main data of COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Dataset menu of Biblioshiny on January 30, 2021.

Suppl. 2. Most relevant source journals of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Sources menu of Biblioshiny on January 30, 2021.

Suppl. 3. Most relevant affiliation (institution) of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Authors menu of Biblioshiny on January 30, 2021.

Suppl. 4. Most frequent words of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Document menu of Biblioshiny on January 30, 2021.

Suppl. 5. Journal clustering by coupling of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Coupling menu of Biblioshiny on January 30, 2021.

Suppl. 6. Co-occurrence network for Keywords Plus of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Conceptual Structure menu of Biblioshiny on January 30, 2021.

Suppl. 7. Factorial network for Keywords Plus of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Conceptual Structure menu of Biblioshiny on January 30, 2021.

Suppl. 8. Co-citation network for source journals of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Intellectual Structure menu of Biblioshiny on January 30, 2021.

Suppl. 9. Collaborative network of institutions of the COVID-19–related articles of Korean authors in the Web of Science Core Collection generated from the Social Structure menu of Biblioshiny on January 30, 2021.

Suppl. 10. World collaboration of the COVID-19–related articles of Korean

authors in the Web of Science Core Collection generated from the Social Structure menu of Biblioshiny on January 30, 2021.

References

1. Boyack KW, Klavans R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? *J Am Soc Inf Sci Technol* 2010;61:2389-404. <https://doi.org/10.1002/asi.21419>
2. Oh J, Kim A. A bibliometric analysis of COVID-19 research published in nursing journals. *Sci Ed* 2020;7:118-24. <https://doi.org/10.6087/kcse.205>
3. Shamsi A, Mansourzadeh MJ, Ghazbani A, Khalagi K, Fahimfar N, Ostovar A. Contribution of Iran in COVID-19 studies: a bibliometrics analysis. *J Diabetes Metab Disord* 2020; 14:1-10. <https://doi.org/10.1007/s40200-020-00606-0>.
4. Tellez VA, Tellez WA. Trends in Peruvian scientific publications on COVID-19: a bibliometric analysis. *Sao Paulo Med J* 2020;138: 352-4. <https://doi.org/10.1590/1516-3180.2020.035322072020>
5. N VR, Patil SB. Indian publications on SARS-CoV-2: a bibliometric study of WHO COVID-19 database. *Diabetes Metab Syndr* 2020;14:1171-8. <https://doi.org/10.1016/j.dsx.2020.07.007>
6. Pathak M. COVID-19 research in India: a quantitative analysis. *Indian J Biochem Biophys* 2020;57:351-5.
7. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr* 2017;11:959-75. <https://doi.org/10.1016/j.joi.2017.08.007>
8. Clarivate Analytics. KeyWord Plus generation, creation, and changes [Internet]. Philadelphia, PA: Clarivate Analytics; 2018 [cited 2021 Feb 4]. Available from: https://support.clarivate.com/ScientificandAcademicResearch/s/article/KeyWords-Plus-generation-creation-and-changes?language=en_US
9. Wikipedia. Multiple correspondence analysis [Internet]. San Francisco, CA: Wikimedia Foundation; 2020 [cited 2021 Feb 4]. Available from: https://en.wikipedia.org/wiki/Multiple_correspondence_analysis