

Impact Factor polarization during the COVID-19 pandemic

Taekho You*, Jinseo Park**, June Young Lee**, and Jinhyuk Yun***

**th_you@postech.ac.kr*

<https://orcid.org/0000-0002-4406-5514>

Institute for Social Data Science, Pohang University of Science and Technology, Republic of Korea

***jayoujin@kisti.re.kr; road2you@kisti.re.kr*

Center for Global R&D Data Analysis, Korea Institute of Science and Technology Innovation, Republic of Korea

****jinhuk.yun@ssu.ac.kr*

<https://orcid.org/0000-0002-0218-1699>

School of AI Convergence, Soongsil University, Republic of Korea

Abstract

The COVID-19-related research field has emerged with a number of papers and citations in a very short time period. Journals published COVID-19-related works have increased their impact factor (IF), which reflects the attention on COVID-19. With publications of COVID-19-related works in Web of Science, we found that COVID-19-related papers increased IF of journals but more benefits were given to the high IF journals. Highly cited COVID-19-related papers were distributed in high IF journals. This increases the inequality of IF in research category. In conclusion, our findings imply that IF is vulnerable to external events, therefore it supports to warn the use of quantitative indicators in assessment.

1. Introduction

The significant number of COVID-19-related papers has been introduced to the academia, which purposed to overcome the recent pandemic. The expansion of this new research field has shown a substantial impact on the scholarly publishing ecosystem. COVID-19-related papers received a large number of citations in a short period (Ioannidis *et al.*, 2022), and journals benefited from publishing COVID-19-related research. For example, the *Lancet* increased its impact factor (IF) from 79.323 to 202.731, according to the 2021 Journal Citation Reports (JCR) released in June 2022 (McVeigh, 2022).

These changes of IF rises a question whether IF reflects the impact of scientific items. Even though the heavy-tailed nature of citation (Bornmann & Leydesdorff, 2017), which is sometimes referred to as the rich-get-richer effect, they rely solely upon mean citation counts in two-years time window (Pendlebury, 2009; Lozano, Larivière & Gingras, 2012). They failed to account for variations across disciplines. However, as already warned in the San Francisco Declaration on Research Assessment (DORA) and Leiden manifesto (San Francisco Declaration on Research Assessment, 2012; Hicks *et al.*, 2015), IFs misunderstood and widely used as the impact of individual item in assessment (Calcagno *et al.*, 2012; Rafols *et al.*, 2012).

In this study, we quantitatively exhibit the impact of COVID-19-related papers on the citation ecosystem to aid in resolving the long-lasting controversy on the IF metric. We calculated IFs using the complete Web of Science Core Collection and investigated the changes in IF by the publication of COVID-19-related papers. We found that the increase of IF is proportional to the prior IF of journals, not related to the number of COVID-19-related papers. The highly cited COVID-19-related papers tend to be published in high IF journals in the early 2020; it may provide a clue to the IF-oriented citation dynamics.

2. Data and methods

COVID-19-related papers were retrieved from Web of Science (WoS) database using search query provided by Dimensions (<https://dimensions.ai/covid19/>). We limit the publications

from 2019 and 251,718 COVID-19-related papers were collected on July 4, 2022, since they contain other coronaviruses. We consider all other papers in WoS that are not retrived as non-COVID-19 papers.

We reproduced the IF following the way of JCR impact factor with an in-house XML copy of Web of Science Core Collection: citations received by items divided by the number of citable items published in past 2 years. We limit the citable items belonging to journals indexed in SCI-Expanded, SSCI, and A&HCI. We consider only publications with the type of article, review, and proceedings paper as citable items, yet types are not considered when computing the number of citations received. We also considered early access publications as regular publications in 2020, following the policy of Clarivate Inc. The reproduced IFs are highly correlated with the IF provided by Clarivate Journal Citation Reports (Pearson $\gamma = 0.998$).

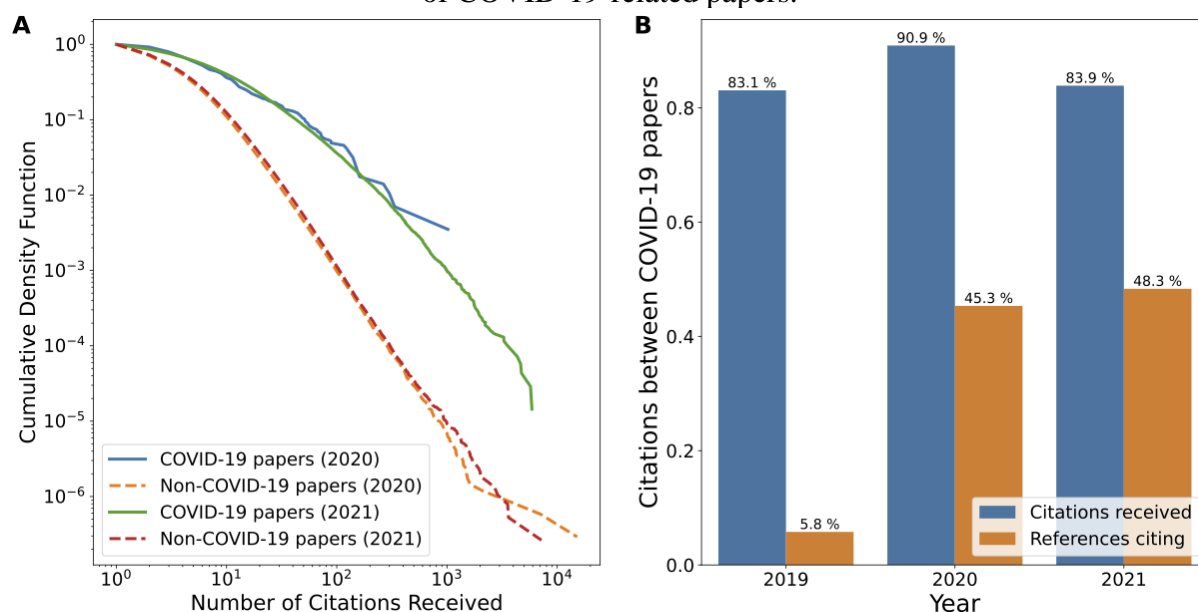
3. Results

3.1. Citation homophily between COVID-19-related papers

In 2019, only 350 papers (0.013% of all publications in 2019) were related to COVID-19, many of which were mainly focused on other coronaviruses, based on our search query. As the virus spread, their share increased to 2.004% of all publications in 2020 and 4.194% in 2021. Moreover, they occupied major fraction of all citations across academia. Papers published in 2020 received 2,654,613 citations until the end of 2021, which is 13.8% among the total citations in 2020 and 2021. As a result, they show a heavier-tailed citation distribution (Figure 1A) and high average citations. COVID-19-related papers received 22.6 citations on average, while non-COVID-19 papers received 4.9 citations in 2020.

We found a citation homophily between COVID-19-related papers. More than 40% of references in COVID-19-related papers cited other COVID-19-related papers, and more than 80% of citations were come from other COVID-19-related papers (Figure 1B). This implies that the rising amount of COVID-19-related works in 2020 and 2021 inflated citations of COVID-19-related papers.

Figure 1. Citations of COVID-19-related papers. **A** Citation distribution of COVID-19-related and non-COVID-19 papers that can contribute to the annual IF calculation. **B** Citation source of COVID-19-related papers.

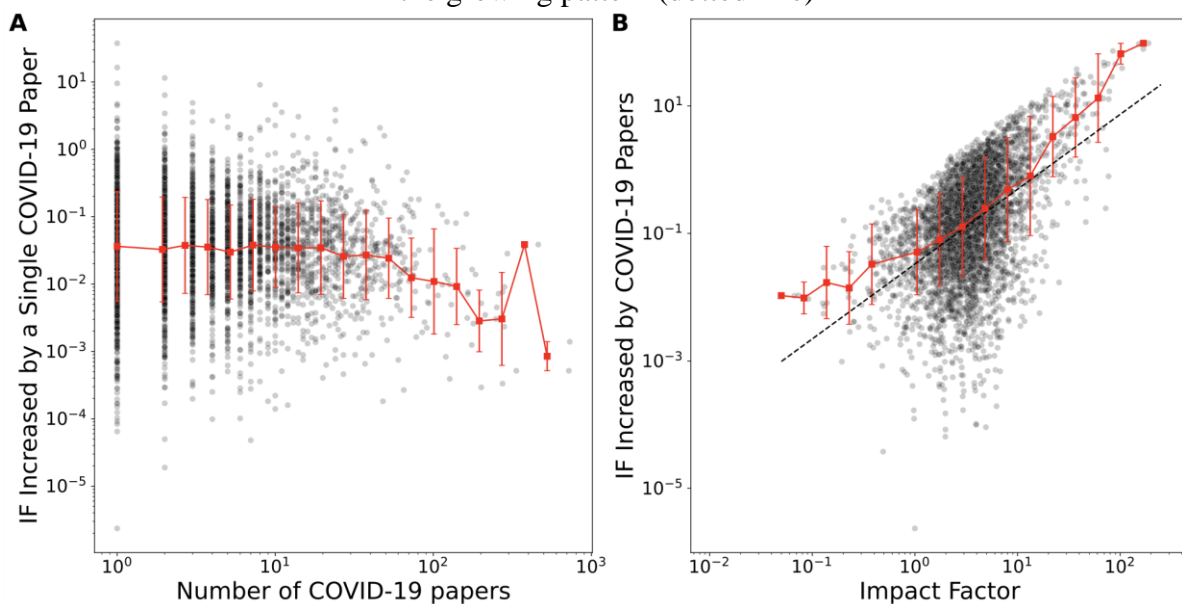


3.2. Contribution of COVID-19-related works to IF inflation

One can simply presume that publishing COVID-19-related papers increases IF because they received more citations than others. Roughly, publishing COVID-19-related papers, although they received large number of citations, may not guarantee the increase of IF. IF will increase only if they receive more citations than the average citations of the journal's papers. To estimate the advantage of publishing COVID-19-related papers we compared two types of IFs: IF excluding COVID-19-related papers and IF including them. We observed that 4,004 journals (84% among those publishing one or more COVID-19-related papers) enhanced their IFs through COVID-19-related publication. Other 763 Journals that decrease their IFs less than 1, except *CA-A Cancer Journal for Clinicians* that decreased 15.78 in our computation. We found that publishing numerous COVID-19-related papers does not necessarily increase the IF. Instead, the gain of IF from a single paper decreased by the number of COVID-19-related papers (Figure 2A). Journals that publish only one COVID-19-related paper have increased their IF by 0.12 on average, whereas journals that publish over 500 papers have increased their IF only by 0.0009. One journal increased IF up to 37 while publishing only 1 COVID-19-related papers. In short, publishing a large number of COVID-19-related papers did not provide more benefits (Pearson $\gamma = 0.110$).

We found that the prior IF provide more benefits to the surplus IF (Pearson $\gamma = 0.670$). The superlinear relationship ($y=x^{1.7}$) between the prior IF and the surplus IF is observed (Figure 2B). To confirm that COVID-19-related papers have legitimately increased the IF, we examine the correlation between IFs across years. The correlations between IFs excluding COVID-19-related papers (Pearson $\gamma = 0.959$ in 2019 and 2020, 0.925 in 2020 and 2021) is higher than IFs including COVID-19-related papers (Pearson $\gamma = 0.850$ in 2020 and 2021). The correlation between IFs excluding COVID-19-related papers in 2020 and IFs including COVID-19 papers in 2021 is also relatively low (Pearson $\gamma = 0.849$), which confirms the contribution of COVID-19-related papers to the increase of IF. To summarize, the publication of COVID-19-related papers increased IF of journals but more benefits were given to the higher IF journals, which may contribute to the polarization of IF.

Figure 2. Surplus impact factor (IF) increased by COVID-19-related publications. **A** Increase in IF per single COVID-19-related paper in the journal. **B** Increase of IF of journals by publishing COVID-19-related papers. The simple superlinear growth $y \sim x^{1.7}$ can characterize the growing pattern (dotted line)



3.3. The Matthew effect of publications

The proportion of highly cited articles in prestigious journals has the potential to exacerbate polarization. Indeed, we found that the share of COVID-19-related papers decreases as the journal ranking in the research category falls (Figure 3). While the top 10% ranked journals published 26.3% of all COVID-19-related papers, the 90% to 100% ranked journals published only 3.5% until 2021. The disparity increases for highly cited COVID-19-related papers. 84% of 102 papers with over 1000 citations were published in the top 10% journals, while no papers were published in the 50% to 100% journals. Only 7.8% of papers with over 100 citations were published in the 10% to 20% journals.

Figure 3. Distribution of COVID-19-related papers and their disparities. Journal percentile is the relative ranking in the research category.

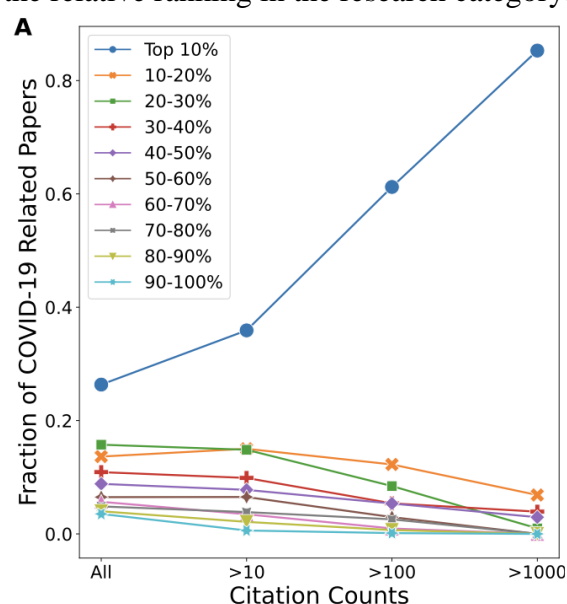
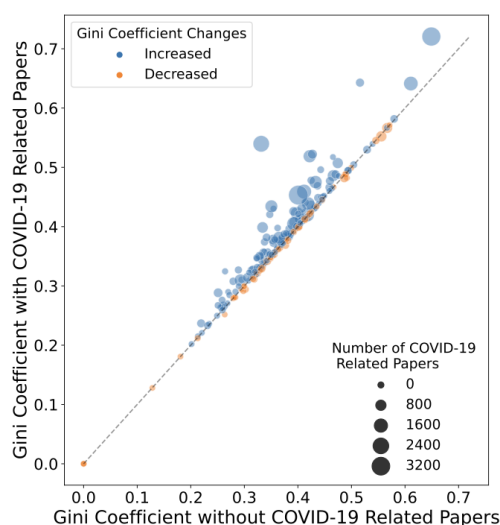


Figure 4. The Gini coefficient of the IF distribution within the JCR category. Each dot represents a research category in JCR.



In summary, by publishing COVID-19-related work, the IF of journals increases, but journals with higher IFs received greater benefits with possessing number of highly cited COVID-19-related papers. Individual scientists have a greater tendency to cite prestigious, widespread,

and popular journals than less popular journals due to psychological, sociological, and economic factors, leading to the rich-get-richer phenomenon of citations (Wang, 2014). These citation dynamics may worsen the polarization. The Gini coefficient, which quantifies the inequality of IFs in the research category, shows that publishing COVID-19-related papers increases the inequality (Figure 4).

4. Discussion

Because of the rich-get-richer nature of citations, papers published in prestigious journals tend to receive more citations. The effect predominantly benefited prestigious journals, while others did not experience benefits to the same extent. The fluctuations in IF under the rich-get-richer nature may not well reflect the actual impact of academic publications. The homophily between COVID-19-related papers reveal that the high citation counts are the result of inflation by the number of COVID-19-related papers. The high correlations of IFs excluding COVID-19-related papers show that the majority of journals may revert to their pre-pandemic IF levels when the pandemic is over. Therefore, our findings imply that IFs are vulnerable to external events.

The rapid emergence of COVID-19-related research field provide a chance to observe the current citation dynamics and its impact to IF changes. However, this study has limitations because of the lack of the motivation of citations, longitudinal observation of COVID-19-related research field, and quality of individual research. The citation counts of COVID-19-related papers do not quantitatively gauge the quality of works. The homophily between COVID-19-related papers may have a possibility to change since it only computes citations in a short period. In addition, there may be a chance that the benefits in high IF journals are just the result of well-qualified peer review system in these journals, although the retraction of COVID-19-related publications happened also in prestigious journals due to the rapid release of COVID-19 works during pandemic (El-Menyar *et al.*, 2021; Quinn *et al.*, 2021).

Nevertheless of limitations, this study provides supporting information to warn the use of quantitative indicators such as the IF in assessment. The DORA, which serves as the starting point for the long-lasting IF controversy, explicitly states that the use of journal-based measures should be avoided to act as a proxy for the quality of individual research publications, to evaluate the contributions of an individual scientists, or to make hiring, promotion, or funding choices. However, in practice, funders and institutions employs the impact of journals or the citation counts as markers of evaluations (Stephan, Veugelers & Wang, 2017; Quan, Chen & Shu, 2017). This motivates authors to publish in journals with high IF, and unawarely cite papers in high IF journals more. Combined with the limitations of citation indexes, we believe that responsible action is essential for the entire member of academia, as opposed to merely producing popular research to boost citation impact and subordinal reputation.

Open science practices

The data used in this study have been retrieved from Web of Science Core Collection licensed by Korea Institute of Science and Technology Innovation.

Acknowledgments

Author contributions

All authors have made substantial contributions to the conceptualization, investigation, analysis, interpretation of data.

Competing interests

The authors declare no competing interests.

Funding information

This research was supported by the MSIT (Ministry of Science and ICT), Republic of Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2022-RS-2022-00156360) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). This work was also supported by the National Research Foundation of Korea (NRF) funded by the Korean government (grant No. NRF-2022R1C1C2004277 (T.Y.) and 2022R1A2C1091324 (J.Y.)). The Korea Institute of Science and Technology Information (KISTI) also supported this research with grant No. K-23-L03-C01 (J.Y.L., J.P.) and by providing KREONET, a high-speed Internet connection. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Bornmann, L., & Leydesdorff, L. (2017). Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data. *Journal of Informetrics*, 11(1), 164-175.
- Calcagno, V., Demoinet, E., Gollner, K., Guidi, L., Ruths, D., & de Mazancourt, C. (2012). Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science*, 338(6110), 1065-1069.
- El-Menyar, A., Mekkodathil, A., Asim, M., Consunji, R., Rizoli, S., Abdel-Aziz Bahey, A., & Al-Thani, H. (2021). Publications and retracted articles of COVID-19 pharmacotherapy-related research: A systematic review. *Science progress*, 104(2), 00368504211016936.
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431.
- Ioannidis, J. P., Bendavid, E., Salholz-Hillel, M., Boyack, K. W., & Baas, J. (2022). Massive covidization of research citations and the citation elite. *Proceedings of the National Academy of Sciences*, 119(28), e2204074119.
- Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology*, 63(11), 2140-2145.
- McVeigh, M. (2022). Journal Citation Reports 2022: A preview. <https://clarivate.com/blog/journal-citation-reports-2022-a-preview/>.
- Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum immunologiae et therapiae experimentalis*, 57, 1-11.
- Quan, W., Chen, B., & Shu, F. (2017). Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). *Aslib Journal of Information Management*, 69(5), 486-502.

Quinn, T. J., Burton, J. K., Carter, B., Cooper, N., Dwan, K., Field, R., ... & Xin, Y. (2021). Following the science? Comparison of methodological and reporting quality of covid-19 and other research from the first wave of the pandemic. *BMC medicine*, 19(1), 1-10.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research policy*, 41(7), 1262-1282.

San Francisco Declaration on Research Assessment (2012), <https://sfdora.org/read/>.

Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature*, 544(7651), 411-412.

Wang, J. (2014). Unpacking the Matthew effect in citations. *Journal of Informetrics*, 8(2), 329-339.