Retracted Covid-19 articles: significantly more cited than other articles within their journal of origin

Trenton Taros[1] (ORCID) · **Christopher Zoppo**[1] · **Nathan Yee**[1] · **Jack Hanna**[1] · **Christine MacGinnis**[2]

Abstract

With the expansion of research volume, coinciding with the age of the internet, the retraction of published papers from scientific journals has become crucial to preserving scientific integrity. Since the beginning of the COVID-19 pandemic, both public and professional interest in scientific literature has grown as people attempt to educate themselves on the virus. The Retraction Watch Database COVID-19 blog was accessed in June and November of 2022 and analyzed to ensure articles met inclusion criteria. Articles were then accessed on Google Scholar and the Scopus database to find number of citations and SJR/CiteScore. The average SJR and CiteScore for a journal that published one of the articles was 1.531 and 7.3 respectively. The retracted articles were cited an average of 44.8 times, which was significantly higher than the average CiteScore ($p = 0.01$). Between June and November, retracted COVID-19 articles gained a total of 728 new citations, presence of "withdrawn" or "retracted" before article title did not affect citation rates. COPE guidelines for retraction statements were not met for 32% of articles. We believe retracted COVID-19 publications may have been more likely to include bold claims that garnered a disproportionately high amount of attention within the scientific community. Additionally, we found many journals were not forthright with explanations for why articles had been retracted. Retractions could be a tool used to add to the scientific discourse, but currently we are only getting half the data, the what and not the why.

✉ Trenton Taros
Trenton.taros@umassmed.edu

1 University of Massachusetts TH Chan School of Medicine, 55 N Lake Ave, Worcester, MA 01605, USA

2 Department of Internal Medicine, University of Massachusetts TH Chan School of Medicine, Worcester, MA, USA

Introduction

In an age of expansion for modern medicine, coinciding with the continued growth of the internet, there has been a rapid expansion in the volume of scientific research activity and rate of publication. A 2021 study indicated a 4% annual growth in global publication output in peer-reviewed science and engineering journal articles as well as conference papers from 2010 to 2020 (National Science Board, 2021). The same study reports a 3% annual growth rate for biomedical and health sciences from 2009 to 2019, but an astounding 15% and 16% growth rate for each respectively in 2020. With this growth, new sets of academic standards for publication arise, along with increasingly important sets of ethical standards (Brand et al., 2004; Coats, 2009). As those standards are investigated and enforced, retraction of published papers from scientific journals becomes crucial to preserve scientific integrity and properly educate the general public and scientific community.

As the volume of scientific literature expands, retraction volume and rate have also increased (Nagella & Madhugiri, 2020; Steen, 2011). These retractions represent a self-corrective function built into the scientific research process to protect the integrity of the scientific method. Retraction of a published article can be initiated by the author or publisher for a variety of reasons: Institutional Review Board (IRB) violations, data errors, forgery, plagiarism, and inappropriate authorship, among others (Budd et al., 1998). A major concern for those articles which are retracted is their rate of citation prior to, and after retraction. This is particularly important if they are found in high-impact, prestigious, or influential journals, as this propagates the dissemination of false, inaccurate, or inconclusive information (Dinh et al., 2019).

Since the beginning of the COVID-19 pandemic, both public and professional interest in scientific literature has grown as people around the world attempt to educate themselves on the virus. According to the World Health Organization (WHO) COVID research database, in 2020 and 2021 over 500,000 pandemic-related journal articles had appeared, representing a paper surge never before seen in history (WHO COVID-19 Research Database, 2022). This has contributed to what some scientists refer to as the "COVID-ization" of research, which includes the fear of high paper volume resulting in lower quality scientific studies (Dinis-Oliveira, 2020). As submission volumes to journals grow at an unprecedented rate, so grows the burden of journal editors tasked with policing misinformation, plagiarism, medical factual error, and data discrepancies within those papers.

As the COVID-19 pandemic has progressed, research and scientific literature have played a major role in the development of vaccines, treatments, public health guidance, and healthcare policy. However, the number of papers pertaining to COVID-19 which have been retracted continues to rise to over 200, based on the Retraction Watch Database (RWD) COVID-19 blog (Marcus & Oransky, 2022). As we continue to learn about COVID-19 from scientific journals, there is reason for concern that expedited publication or incomplete medical and statistical review resulting in retraction could cause future errors in our fight against the pandemic (Dinis-Oliveira, 2020). Those future medical errors may occur through citation and dissemination of information from a retracted publication spreading incorrect or incomplete conclusions about disease course, vaccination, or symptomatology.

The study of retractions as they relate to the COVID pandemic is not novel, the first calls for concern appear within the literature as early as June 2020 (Yeo-Teh & Tang, 2020). Several "high profile" cases of retraction have been studied, including the now infamous

"Surgisphere" publications (Ledford & van Noorden, 2020). Other analyses have mainly examined origin of the paper, prestige of journal, and reason for retraction (Moradi & Abdi, 2021; Shimray, 2022; Soltani & Patini, 2020). Quantitative analysis of citations for COVID related retractions has been sparse, no peer-reviewed article to date has compared the citation count of retracted papers to the metrics of the journals they were submitted in (Peterson et al., 2022).

We hypothesized that retracted COVID-19 articles received more attention, measured in citations, than would be expected for the average article in the journal they were published in over the same time period. In presenting a quantitative analysis of retracted COVID-19 articles published in medical and life science journals, we hope to shine light on the attention these articles garnered and provide possible rationale for this occurrence. The inclusion of SCImago Journal Ranking (SJR) is used to demonstrate that this phenomenon was not isolated to smaller journals but included journals of all sizes and prestige levels. It is our hope to add to the conversation regarding COVID-19 retractions and the state of scientific literature during the pandemic.

Methods

The Retraction Watch Database COVID-19 blog was accessed in June and then November of 2022, at which time there were a total of 270 articles. Articles were excluded from analysis if they were in pre-print at time of retraction, were retracted because they were duplicates, were not published in medical or life science journals, were published in journals without an SJR or CiteScore, or were unable to be found on Google Scholar or Scopus. Of the original 270 articles, 90 articles met inclusion criteria and were analyzed (Fig. 1).

Pre-print articles were not examined given the lack of a complete peer review process. Articles not published in medical or life science journals were excluded in keeping with our hypothesis and desire to elucidate trends in this specific field. Duplicate articles published in other journals were not thought to constitute true retractions because they represented clerical error in the publishing process and were not related to the content of the article
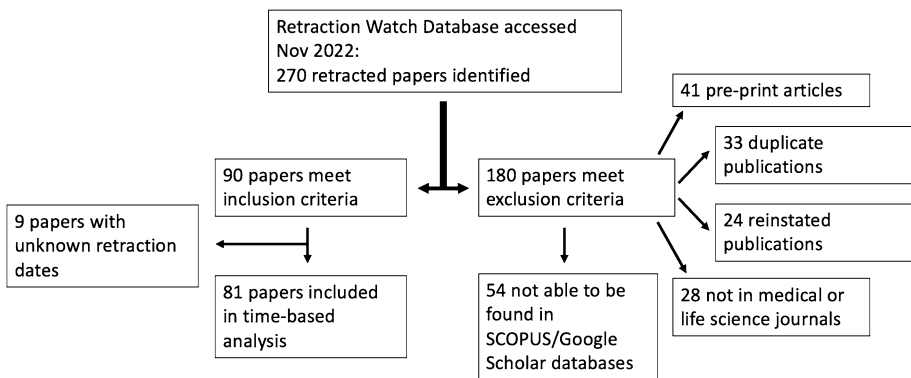

Fig. 1 Flowchart showing the inclusion/exclusion of articles based on categorization of criteria. All 90 articles meeting inclusion criteria were analyzed for citation analysis, however only 81 were able to be used for time-based analysis

itself. Articles published in journals without SJR or CiteScore were excluded because they couldn't be analyzed using our chosen standardized metrics. Articles published in 2022 were not included for two reasons. First, it would not be possible to compare their citation number to the final CiteScore of the publishing journal as the 2022 score will not be reported until 2023. Second, these articles, having been published for less than one year, have not had time to accrue citations and might skew results.

If possible, articles were accessed on PubMed where there was frequently a retraction notification added to the original publication. Care was taken to determine the rationale behind the retraction by thorough inspection of the available correspondence. Articles were searched by title and author then accessed on Google Scholar and the Scopus database to find number of citations and SJR/CiteScore. As stated previously, 90 articles were able to be found on Scopus search, these articles were used for our primary analyses. 93 articles were able to be found on Google Scholar search, these articles were assessed twice for citation count, once in June and once in November of 2022. The 93 articles found on Google Scholar were only compared to themselves, so as to not over-report citations.

SJR is a metric designed to quantify the prestige of journals using weighted citations. 1.0 is considered average, anything below that is less prestigious than average and anything above it more. CiteScore is an Elsevier metric which shows the average number of citations per paper in a given journal within its first 3 years of publication. The results were analyzed using student's t-test and an alpha of 0.05 was used to determine statistical significance.

Results

At the time of writing, there were 270 retracted articles listed. After eliminating articles based on our exclusion criteria, we were left with 90 articles published in 2020 or 2021 and retracted before November of 2022. Those 90 articles were accessed and broadly divided into categories based on content and reason for retraction. 28 of the articles focused on epidemiology, 20 on disease course, and 22 on treatment (Fig. 2). In 53 of the cases, the publisher initiated the retraction proceedings. In 18 instances the request was initiated by

Fig. 2 Content area of retracted COVID-19 articles fitting inclusion criteria. Epidemiology, treatment, and disease course made up the majority of the retracted articles analyzed
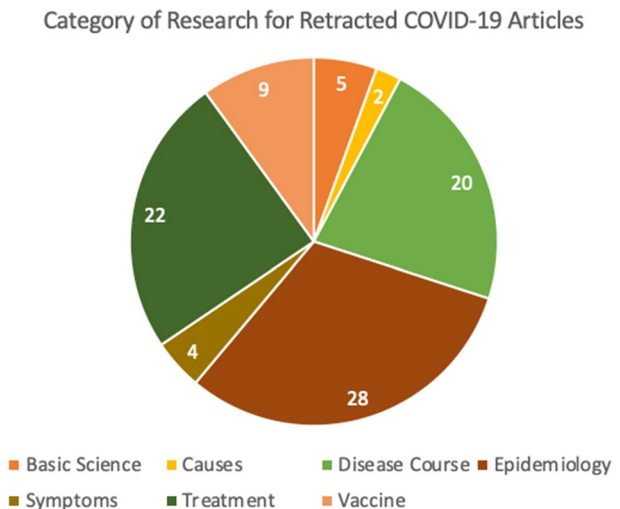
the author. The requesting party was unknown in the remaining 19 instances. There were a wide variety of reasons for retraction with little standardization, at least 7 articles were retracted due to concerns for plagiarism and at least 5 for IRB violations.

The median number of days examined articles remained published for was 175 days (Fig. 3). The average SJR and CiteScore for a journal that published one of the articles was 1.531 (SD 3.019) and 7.3 (SD 13.1) respectively. The retracted articles we examined were cited an average of 44.8 times (SD 138.9). Retracted articles accrued a median number of 10.0 (IQR 26) citations while CiteScore accrued a median of 4.2 (IQR 3.7) citations (Fig. 4). The difference between the average CiteScore and average citation number for a retracted article was 37.5, a statistically significant difference ($p = 0.01$).

93 articles were analyzed via Google Scholar in both June and November 2022, within that time period these articles gained a total of 728 new citations. Of articles receiving new citations, 70.5% had a heading of "retracted:" or "withdrawn:" proceeding their first Google Scholar search result while 75.5% did not. This was not a statistically significant difference. Articles without a heading had an average of 8.55 citations (SD 19.29) while articles with a heading had an average of 7.02 citations (SD 18.18). This difference also proved to be insignificant.

The 53 articles with publisher-initiated retraction garnered an average of 30.9 citations (SD 62.4) while their author-initiated counterparts accumulated an average of 52.8 citations (SD 171.4). This difference was not statistically significant.

Discussion

There is no doubt that throughout the COVID-19 pandemic, the scientific community has come together to characterize the virus and innovate new diagnostic tests and treatments (Mishra & Tripathi, 2021). As is to be expected given the sheer volume produced, that immense effort included many scientific publications on the COVID-19
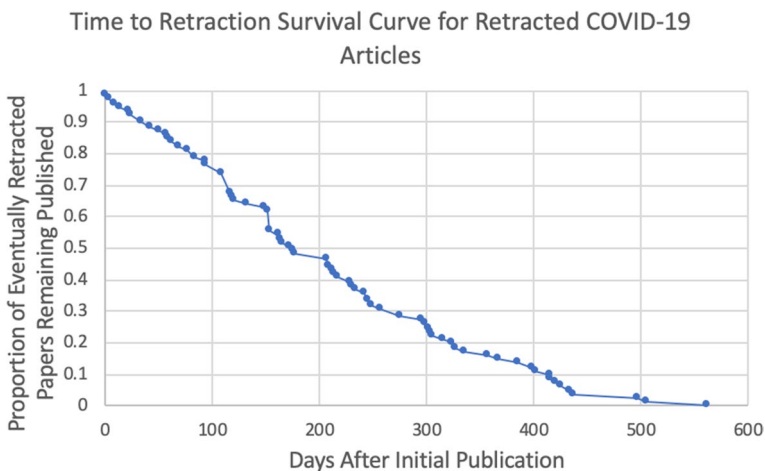

Fig. 3 Survival curve analysis shows the percentage of studies that remained published for a given number of days before being retracted. The median number of days a study took to be retracted was 175 days
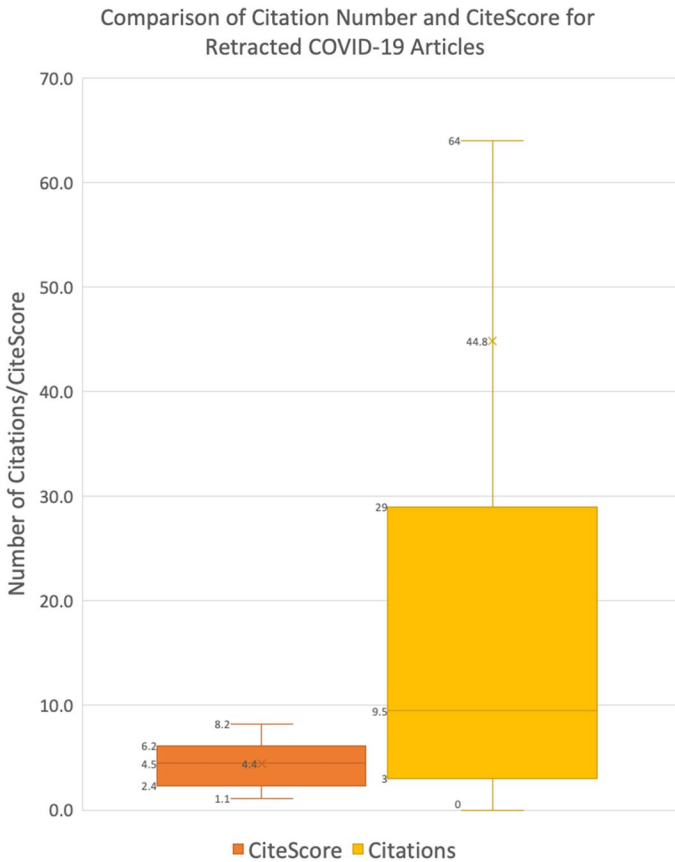
Fig. 4 Box and whisker plot shows difference between the citations of retracted COVID-19 articles and the CiteScore of the journals in which they were published. Means are visualized as "x" marks, outlier points are not visualized

virus that were later disproven or adjusted. These articles were published on average for almost 6 months, enough time to influence both lay people and the scientific community alike.

As of November 2020, the retraction rate for COVID-19 related papers was five times higher than that of general life science literature (Nagella & Madhugiri, 2020). Clear instances of intentional malpractice and fraudulent publications have been identified in some of the world's most noteworthy journals. However, the reasons behind the disproportionately high retraction rate for COVID-19 related publications will require further analysis. It is worth noting that with an average SJR of 1.509, the studies in this analysis included those from prestigious and reputable journals.

Further analysis of the retracted articles reveals that these papers tended to have a disproportionately high impact. On average, a retracted COVID-19 publication had tenfold the number of citations compared to other articles in the journals in which they were published, many of which were presumably also related to the pandemic. Retracted COVID-19 publications may have been more likely to include bold or novel claims that garnered a disproportionately high amount of attention from others in the scientific

community. The phenomenon of "clickbait" titles could explain how certain publications gained more attention and led to further writing on a potentially misleading claim.

It is true that COVID-19 publications accumulated a disproportionate amount of attention regardless of retraction status (Ioannidis et al., 2022). However, the large discrepancy between CiteScore and number of citations is not likely due to an increased amount of interest in COVID literature given the magnitude of the difference between CiteScore and citation number in retracted articles. A direct comparison between retracted and non-retracted COVID-19 publications for each respective journal was beyond the scope of this analysis.

Unfortunately, retracted COVID-19 papers continue to accumulate citations well past their date of retraction. Analysis at two time points, June and November of 2022, showed 728 total additional citations across all 93 articles analyzed across this time frame. This equates to just shy of 5 citations of retracted COVID-19 papers per day. It does not appear that measures taken to curb their citation have much effect either. For example, there was no significant difference in additional citations between those articles bearing "withdrawn" or "retracted" headings and those without them. Ultimately, this places the onus to limit the perpetuation of inaccurate, retracted information squarely on the scientific community itself. It is up to researchers to analyze their citations as closely as they analyze their conclusions.

It is well known that the number of citations per published article is used to quantify the prestige of scientific journals. This also applies to individual authors; citation counts are used to calculate metrics such as the H-index that evaluate the impact and quality of a scholar's research (Nigam & Nigam, 2012). H-index in particular has been shown to be susceptible to falsified elevation via retracted articles, however it has also been identified as a measure that could be used for tenure (Saraykar et al., 2017). With so much at stake, it is easy to see the possible conscious or unconscious motive behind research practices that maximize the attention they garner from the scientific community.

In only 71 cases was it possible to discern which party requested the retraction. Far fewer had clear explanations for why articles were retracted. In our analysis, 32% of retracted articles did not meet the COPE guidelines for retraction statements (Barbour et al., 2009). This is because either the retracting party or retraction reason was not included within the retraction notice. Obscuring the reason for retraction obfuscates the bravery and integrity needed to willingly retract one's own work (Vuong, 2020). Additionally, it prevents researchers who have cited works that have since been retracted from evaluating the legitimacy of their own claims. In the cases that were able to be analyzed, there was no significant difference between citations or time to retraction between those articles whose retraction was initiated by author or publisher. This implies that neither route is "better" than the other and should encourage journals to be forthright about the retraction reason. Retractions could be a tool used to add to the scientific discourse, but as it stands, we are only getting half the data, the what and not the why.

The current study did not examine the overall number of COVID-19 publications during disease spikes, and whether the retractions increased and decreased proportionately. The heavy burden on the healthcare system combined with the increase in available data likely pressured scholars and journals to prioritize speed when publishing on COVID-19 (Twohig et al., 2022). It has been shown that during the pandemic, the time from receiving a COVID related manuscript to publishing it was much less than expected based on previous data (Palayew et al., 2020). This emphasis on speed is a compelling explanation for the disproportionately high retraction rate for COVID-19 publications. Epidemiology and disease course were the main topics of over half of the retractions, which would

further support the notion that a rapid spike in available data may have led to an increased likelihood of attempting to publish. Further research may look at comparing retraction rates during spikes compared to plateaus to test whether the course of the pandemic may have been influential.

Our study had several other limitations, some of which are inherent to the retrospective design. A specific reason for retraction was not available in each case making it difficult to fully assess the trends behind these retractions. Retracted articles were identified using the Retraction Watch blog which may not represent all retracted articles on the COVID-19 pandemic especially those not published in English. Only articles published in journals with both an SJR and Citescore were analyzed. For these reasons, some retracted articles may not have been identified and analyzed by our analysis.

Articles published on the novel Coronavirus have been retracted at a much higher rate than other articles published in similar journals. The reason for the retraction was not clear in all cases, highlighting the benefit of adherence to standardized guidelines for retraction documentation. In addition, these articles tended to be cited much more frequently than would be expected based on data for the journals they were published in. The reasons behind the surge in citations for these articles requires further investigation. Future research is needed to parse the specific content of these publications and reasons for their retraction. Close examination of the titles of the publications, reasons for their retraction, and timing in relation to the pandemic course may help elucidate the trend that has been identified in this study. The results will carry implications about the integrity of the scientific publishing process and its ability to respond effectively during a time of global health crisis.

Data availability The data that support the findings of this study are openly available in Mendeley Data at https://doi.org/10.17632/yts7nxtyx4.1

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Barbour, V., Kleinert, S., Wager, E., & Yentis, S. (2009). *Guidelines for retracting articles*. https://doi.org/10.24318/cope.2019.1.4

Brand, R. A., Heckman, J. D., & Scott, J. (2004). Changing ethical standards in scientific publication. *Journal of Bone and Joint Surgery - Series B*. https://doi.org/10.1302/0301-620X.86B7.15819

Budd, J. M., Sievert, M. E., & Schultz, T. R. (1998). Phenomena of retraction. Reasons for retraction and citations to the publications. *Journal of the American Medical Association*. https://doi.org/10.1001/jama.280.3.296

Coats, A. J. S. (2009). Ethical authorship and publishing. *International Journal of Cardiology*. https://doi.org/10.1016/j.ijcard.2008.11.048

Dinh, L., Sarol, J., Cheng, Y. Y., Hsiao, T. K., Parulian, N., & Schneider, J. (2019). Systematic examination of pre- and post-retraction citations. *Proceedings of the Association for Information Science and Technology, 56*(1), 390–394. https://doi.org/10.1002/pra2.35

Dinis-Oliveira, R. J. (2020). COVID-19 research: pandemic versus "paperdemic", integrity, values and risks of the "speed science." *Forensic Sciences Research*. https://doi.org/10.1080/20961790.2020.1767754

Ioannidis, J. P. A., 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*. https://doi.org/10.1073/pnas

Ledford, H., & van Noorden, R. (2020). High-profile coronavirus retractions raise concerns about data oversight. *Nature, 582*(7811), 160–161. https://doi.org/10.1038/D41586-020-01695-W

Marcus, A., & Oransky, I. (2022). *Retracted coronavirus (COVID-19) papers*. Retraction Watch Blog

Mishra, S. K., & Tripathi, T. (2021). One year update on the COVID-19 pandemic: Where are we now? *Acta Tropica*. https://doi.org/10.1016/j.actatropica.2020.105778

Moradi, S., & Abdi, S. (2021). Pandemic publication: correction and erratum in COVID-19 publications. *Scientometrics*. https://doi.org/10.1007/s11192-020-03787-w

Nagella, A. B., & Madhugiri, V. S. (2020). Journal retraction rates and citation metrics: An ouroboric association? *Cureus*. https://doi.org/10.7759/cureus.11542

National Science Board, National Science Foundation. (2021). Publications Output: U.S. and International Comparisons. Science and Engineering Indicators 2022. NSB-2021-4. Alexandria, VA. https://ncses.nsf.gov/pubs/nsb20214/

Nigam, A., & Nigam, P. K. (2012). citation index and impact factor. *Indian Journal of Dermatology, Venereology and Leprology*. https://doi.org/10.4103/0378-6323.98093

Palayew, A., Norgaard, O., Safreed-Harmon, K., Andersen, T. H., Rasmussen, L. N., Lazarus, J., & v. (2020). Pandemic publishing poses a new COVID-19 challenge. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-020-0911-0

Peterson, C. J., Alexander, R., & Nugent, K. (2022). COVID-19 article retractions in journals indexed in PubMed. *American Journal of the Medical Sciences*. https://doi.org/10.1016/j.amjms.2022.01.014

Saraykar, S., Saleh, A., & Selek, S. (2017). The association between NIMH funding and h-index in Psychiatry. *Academic Psychiatry*. https://doi.org/10.1007/s40596-016-0654-4

Shimray, S. R. (2022). Research done wrong: A comprehensive investigation of retracted publications in COVID-19. *Accountability in Research*. https://doi.org/10.1080/08989621.2021.2014327

Soltani, P., & Patini, R. (2020). Retracted COVID-19 articles: a side-effect of the hot race to publication. *Scientometrics*. https://doi.org/10.1007/s11192-020-03661-9

Steen, R. G. (2011). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics*. https://doi.org/10.1136/jme.2010.040923

Twohig, K. A., Nyberg, T., Zaidi, A., Thelwall, S., Sinnathamby, M. A., Aliabadi, S., Seaman, S. R., Harris, R. J., Hope, R., Lopez-Bernal, J., Gallagher, E., Charlett, A., de Angelis, D., Presanis, A. M., Dabrera, G., Koshy, C., Ash, A., Wise, E., Moore, N., … Gunson, R. (2022). Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious Diseases*, https://doi.org/10.1016/S1473-3099(21)00475-8

Vuong, Q. H. (2020). Reform retractions to make them more transparent. *Nature*. https://doi.org/10.1038/d41586-020-01694-x

WHO COVID-19 Research Database. (2022). World Health Organization. Retrieved June 13, 2022, from https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/

Yeo-Teh, N. S. L., & Tang, B. L. (2020). An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). *Accountability in Research*. https://doi.org/10.1080/08989621.2020.1782203