Comparison of COVID-19 and non-COVID-19 papers

Cristina Candal-Pedreira, Alberto Ruano-Ravina, Mónica Pérez-Ríos

A Health Research Institute of Santiago de Compostela, Santiago de Compostela, A Coruña, Spain
b CIBER in Epidemiology and Public Health (CIBERESP), Spain
c Department of Preventive Medicine and Public Health, University of Santiago de Compostela, Santiago de Compostela, A Coruña, Spain

Objective: The need to generate evidence related to COVID-19, the acceleration of publication and peer-review process and the competition between journals may have influenced the quality of COVID-19 papers. Our objective was to compare the characteristics of COVID-19 papers against those of non-COVID-19 papers and identify the variables in which they differ.

Method: We conducted a journal-matched case-control study. Cases were COVID-19 papers and controls were non-COVID-19 papers published between March 2020 and January 2021. Journals belonging to five different Journal Citations Reports categories were selected. Within each selected journal, a COVID-19 paper (where there was one) and another non-COVID-19 paper were selected. Conditional logistic regression models were fitted.

Results: We included 81 COVID-19 and 143 non-COVID-19 papers. Descriptive observational studies and analytical observational studies had, respectively, a 55-fold (odds ratio [OR]: 55.12; 95% confidence interval [IC95%]: 7.41-409.84) and 19-fold (OR: 19.28; IC95%: 3.09-120.31) higher likelihood of being COVID-19 papers, respectively, and also a higher probability of having a smaller sample size (OR: 7.15; IC95%: 2.33-21.94). COVID-19 papers had a higher probability of being cited since their publication (OR: 4.97; IC95%: 1.63-15.10).

Conclusions: The characteristics of COVID-19 papers differed from those of non-COVID-19 papers published in the first months of the pandemic. In order to ensure the publication of good scientific evidence the quality of COVID-19-papers should be preserved.

© 2022 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Comparación de los artículos COVID-19 y no COVID-19

Resumen

Objetivo: La necesidad de generar evidencia relacionada con la COVID-19, la aceleración del proceso de revisión por pares y publicación, y la competencia entre revistas, pueden haber influido en la calidad de los artículos sobre la COVID-19. El objetivo fue comparar las características de los artículos sobre la COVID-19 con las de los artículos no COVID-19 e identificar las variables en las que difieren.


Resultados: Se incluyeron 81 artículos COVID-19 y 143 artículos no COVID-19. Los estudios observacionales descriptivos y los estudios observacionales analíticos tenían, respectivamente, una probabilidad 55 veces mayor (odds ratio [OR]: 55.12; intervalo de confianza del 95% [IC95%]: 7.41-409.84) y 19 veces mayor (OR: 19.28; IC95%: 3.09-120.31) de ser artículos COVID-19 y de tener un tamaño de muestra menor (OR: 7.15; IC95%: 2.33-21.94). Los artículos más citados desde su publicación tenían 5 veces más probabilidades de ser artículos COVID-19 (OR: 4.97; IC95%: 1.63-15.10).

Conclusiones: Los artículos COVID-19 parecen tener mayor impacto bibliométrico, a pesar de tener menor calidad metodológica. Para asegurar la publicación de buena evidencia científica se debe preservar la calidad de los artículos relacionados con la COVID-19.

© 2022 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY (http://creativecommons.org/licenses/by/4.0/).
Introduction

Since January 2020, the disease caused by the new coronavirus (SARS-CoV-2), named COVID-19 by the World Health Organization, has spread rapidly causing a global pandemic, with multifactor consequences and giving rise to an unprecedented worldwide health and economic crisis in times of peace.1,2 The interest aroused by the new coronavirus, added to the urgent need to generate evidence about the disease and its treatment, among other things, has led to an exponential increase in COVID-19-related papers.3 This has in turn produced an avalanche of information difficult to manage and apply, whether by health professionals or by bodies tasked with implementing public health policies.4

In the period from the outbreak of the pandemic to May 2021, a total of 419,950 publications on COVID-19 were published, and of these more than 47,000 (11.2% of the total) were pre-prints, i.e., publications published online before peer review (www.dimensions.ai/covid19). Both peer-reviewed papers and preprints have served to generate a great volume of information on COVID-19. Yet it has been observed that a substantial number of peer-reviewed papers have methodological limitations or give rise to erroneous conclusions, with the resulting potential harmful effect for public health.5-6

The need to generate evidence quickly in the case of COVID-19, coupled with the competition between scientific journals to be first to publish what might prove to be the results of greatest relevance, may have undermined the quality of the research papers published. Added to this is the fact that the filters in the editorial process, such as peer review, may be more lax in the case of COVID-19 papers, due to pressure to publish.7-8 Evidence of the change in the editorial process can be seen in the number of retractions observed in COVID-19 papers.7 By May 2021, a total of 108 COVID-19-related papers had been retracted according to the Retraction Watch database (www.retractiondatabase.org). Bearing in mind that the mean time for retracting a paper is long, this may probably be only the tip of the iceberg. What is more, the publication of papers which address other relevant scientific topics may well be affected by the lack of space for publishing them.

Since the outbreak of the pandemic, a number of authors have published bibliometric analyses aimed at describing the characteristics of COVID-19-related studies and analyzing publication trends.3,4,8-11 Only one study has analyzed the quality of COVID-19 papers, comparing it against that of non-COVID-19 papers, using a case-control design covering the initial months of the pandemic.14

A comparative analysis of papers on COVID-19 with respect to those on other topics published in the same journals (therefore controlling the role of editorial teams) could inform us about methodological differences that might influence or determine the quality of such studies. The objective of this study, therefore, was to compare the methodological and bibliometric characteristics of COVID-19 papers against those of non-COVID-19 papers and identify the variables in which the two differed, using a purpose-made case-control design.

Method

Study design

We conducted a case-control study of published papers matched by journal of publication of the included papers. Cases were papers on COVID-19 published from March 2020 through January 2021, and controls were papers published during the same period on other topics in the same journal. A minimum of one control was matched with each case. The number of controls per case varies between one and ten, depending on the number of COVID-19 and non-COVID-19 papers published by the selected journals.

We selected indexed journals profiled in Journal Citation Reports (JCR), belonging to the categories “Respiratory System”, “Infectious Diseases”, “Public, Environmental and Occupational Health”, “General/Internal Medicine” and “Oncology”. We selected the four journals with the higher impact factor of each quartile from each category, yielding a total of 20 scientific journals (5 categories × 4 quartiles). As inclusion criteria, all journals selected were required to have published one issue per month and to publish original papers in English, Spanish or Portuguese. For each journal selected, we reviewed all the monthly issues published from March 2020 through January 2021.

From each issue included, we selected one COVID-19 paper and one non-COVID-19 paper by order of appearance. In any case, should a given number might not contain a COVID-19 paper that fulfilled the inclusion criteria, this was selected from another issue of the same journal. Hence, the first of the COVID-19 papers of that particular issue of the journal was extracted for the month of publication in question, and a second paper was then selected for the issue of journals which did not contain a COVID-19 paper.

By way of inclusion criteria, papers were required to report on original research conducted on human beings and be written in English, Spanish or Portuguese. We excluded modeling studies, case studies, ecological studies and qualitative studies. Retracted papers were not included.

Due to the nature of the study, ethics committee approval was not required.

Data-collection

From each journal selected the following variables were obtained from the JCR database: JCR category; relative position in 2019 (by quartile); journal name and abbreviation; country and impact factor. We also obtained the number of COVID-19-related papers from each of the journal issues included, as well as the number of non-COVID-19 papers, and the total number of papers published in each number, with the aim of calculating the proportion of COVID-19 papers and non-COVID-19 papers in each of the numbers.

The following bibliometric variables were obtained from the papers selected: paper’s title; month of publication; type of paper (COVID-19 or non-COVID-19); number of authors; name of first author; country of first author; first author’s institutional affiliation; date of submission of manuscript and date of acceptance (where available). We calculated the number of days elapsed between the manuscript’s date of submission and the date of publication of the articles included. Using the Web of Science database, we kept a manual record of the citations received by the papers included, from the date of their publication until 31st January 2021.

Furthermore, the following design variables were sourced from each paper: study design, categorized as descriptive observational (cross-sectional and cases series studies), analytical observational (case-control and cohort studies) or experimental (quasi-experimental and clinical trials); sample size (1-500, >500); study scope (local, national or international); number of centers at which the study was undertaken (uncenter or multicenter); and number of adjustment variables.

All the above variables were entered into a database purpose-designed by the authors.

Statistical analysis

We performed a descriptive statistical analysis to identify the characteristics of the papers included, by reference to the variables of interest collected and case or control status. Quantitative
variables were expressed in terms of median and interquartile range, while qualitative variables were expressed in absolute and relative frequencies.

A multivariate conditional logistic regression model was fitted to identify the bibliometric and study-design variables associated with COVID-19 or non-COVID-19 papers. In the final model the following variables were included: study design (descriptive observational, analytical observational, experimental) and sample size (1-500, >500). In addition, the model was adjusted by the journal in which the articles were published. Adjusted odds ratios (OR) were calculated along with their 95% confidence intervals (95%CI).

Values were deemed statistically significant at \( p < 0.05 \). All statistical analyses were performed using the Stata v17 computer software program.

Results

Data were collected on 301 papers published in 20 scientific journals across the period from March 2020 to January 2021, both inclusive. In March, no COVID-19 paper was published in the selected journals. In the following months, the proportion of COVID-19 papers varied, reaching 80% in the number of the Lancet Respiratory Medicine journal published in December 2020. Seven of the selected journals published no COVID-19 papers across the study period.

For analysis purposes, we excluded non-COVID-19 papers published in the seven journals which had not published any COVID-19 paper, due to the impossibility of matching these controls with cases by reference to the journal. Finally, the analysis included a total of 224 papers published in 13 journals; of these, 81 were COVID-19 papers and 143 were non-COVID-19 papers (Fig. 1).

Table 1 describes the main bibliometric and design characteristics of the papers included, differentiating between COVID-19 and non-COVID-19 articles. Among the COVID-19 papers, 18.5% had more than 20 authors, whereas among the non-COVID-19 papers this percentage was 12.8%; for COVID-19 papers, the country with most publications was China, whereas in 23.1% of the non-COVID-19 papers the first author’s country was the USA. While 63% of the COVID-19 papers had a descriptive design and 3 (3.7%) were experimental studies, among the non-COVID-19 papers, the percentage of papers included with an experimental design was 24.9%.

Table 2 and Figure 2 show that the time to publication of COVID-19 papers was shorter than that of non-COVID-19 papers (\( p < 0.001 \)). In addition, it was observed that COVID-19 articles received a higher number of citations compared to non-COVID-19 articles (\( p < 0.001 \)). COVID-19 articles received a median of 3 citations (interquartile range: 0–25), while non-COVID-19 articles received a median of 1 citation (interquartile range: 0–3).

As can be seen in Table 3, descriptive observational and analytical observational studies had a 56-fold (OR: 56.43; 95%CI: 9.40–338.67; \( p < 0.001 \)) and 20-fold (OR: 19.94; 95%CI: 3.77–105.58; \( p < 0.001 \)) higher likelihood of being COVID-19 papers, respectively, taking experimental design as the reference category. The studies with the smallest sample size (<500 individuals) had a 6-fold higher likelihood of being COVID-19 papers (OR: 5.96; 95%CI: 2.19–16.28; \( p < 0.001 \)) as compared to studies with the largest sample size (>500 individuals included in the study).

Discussion

The results observed in this study suggest that there are differences among original research papers published in the first months of the pandemic, according to whether they report COVID-19 or non-COVID-19 studies. Despite receiving more citations, COVID-19 studies are more likely to have certain characteristics (study design, sample size) which are generally associated with low methodological quality than are non-COVID-19 studies, according to some quality assessment tools, such as GRADE.15

The process of scientific publication has been influenced by the COVID-19 pandemic. On the one hand, in line with the results of this study, a number of authors had already found evidence to show that time to publication is significantly shortened in the case of COVID-19 papers, in comparison with papers on other scientific topics.2,16,17 This phenomenon is the result of scientific journals expediting the publication process and peer review of COVID-19 papers, with the aim of disseminating results more quickly. The results of the study conducted by Horbach indicate that COVID-19 papers are accepted after a single peer-review round, whereas non-COVID-19 papers require more rounds before being accepted by the journal, thus increasing the time between manuscript submission and publication.18 In this context, doubts arise as to whether

---

editorial teams apply the same quality criteria when it comes to evaluating COVID-19 and non-COVID-19 articles.

Added to this is the fact that, according to our results, COVID-19 studies published in the first months of the pandemic have a greater likelihood of having a descriptive design and smaller sample size than do non-COVID-19 papers. Different studies have analyzed the quality of COVID-19 papers using different methods and all agree on the fact that, generally speaking, their quality is low.14-18 Accorsi et al.19 have identified many biases in COVID-19-related observational studies in aspects such as interpretation of results or enrollment of a non-representative sample of the study population, among others. This poses a problem, since papers with a low methodological quality and a greater presence of biases can induce readers to draw erroneous conclusions, which is, in itself, a potential health threat. Associated with this is the fact that health professionals may base themselves on erroneous or invalid conclusions to guide their clinical practice, with the potential harm that this entails for patients; similarly, the health authorities could take decisions which might endanger population health.20,21

Mention should be made of the great popularity enjoyed by preprint servers during the pandemic. The use of preprint servers had already been on the rise before the pandemic, due to the criticisms leveled at scientific journals’ current publication system, and peer review in particular, for slowing down the dissemination of results. In the context of the pandemic, a considerable number of COVID-19-related manuscripts were published as preprints.7,22 It should be born in mind that there is no mechanism which would verify that information published as a preprint is in fact genuine and reliable.8 During the pandemic, results published in the form of preprints have led some political and medical personalities to promote certain treatments for COVID-19, without these having been validated in any way, with the potential harm that this could entail for the population.23,22 A formal approach to the role of preprints in COVID-19 pandemic is clearly needed.

It should be noted that only 3.5% of COVID-19 studies included in the analysis had an experimental design versus 29.4% in the case of non-COVID-19 studies. During the first months of the pandemic, it was common for studies to have an observational design and small sample size, thus making them methodologically inferior.23 With the passage of time, however, it is to be hoped that COVID-19-related studies may improve their methodological quality, since higher-quality studies require better planning and execution and, thus, more time.4 Even so, the results of our study highlight the fact that almost one year after the outbreak of the pandemic, the number of experimental studies published on COVID-19 is still low in comparison with non-COVID-19 studies.

Another worrying phenomenon is the impact and visibility of COVID-19 studies. The results indicate that COVID-19 papers accumulate a higher number of citations than do non-COVID-19 papers. Our results are in line with the study conducted by Yang et al.18 and by Zdravkovic et al.14, which not only indicate that COVID-19 papers receive a higher number of citations, but also show that the citations received by papers on other scientific topics have decreased with respect to 2019.

The information which has been published on COVID-19 does not always make for any real advance in scientific knowledge. Encouragement should be given to the publication of systematic literature reviews which compile and summarize the information generated on different aspects of COVID-19, with the aim of making it more manageable.4 It should likewise be borne in mind that there is an unnecessary overlapping of COVID-19-related scientific evidence, with the ensuing loss of time and resources (human and economic) that this entails.5,21,24 However, there are COVID-19-related fields that are hardly being studied, such as the effects of the pandemic on other diseases or the use of new technologies.11 One solution to this problem could be the pre-registration of studies on purpose-designed platforms.5 It should also be noted that by prioritizing the publication of COVID-19 studies, relevant higher-quality studies addressing other scientific topics might possibly be overlooked or even ruled out.

It is important that editors be more alert to the quality of COVID-19 papers submitted for publication by using of checklists, such as STROBE or CONSORT25 or by enhancing open peer reviews, making it known what reviews have been conducted and by whom these have been conducted. Open peer reviews could have some benefits as it might increase the transparency of the peer review process by shifting the responsibility to reviewers to perform more thoughtful reviews. Also, open peer review enables greater visibility for peer review activities.26 It is worth mentioned that most authors, reviewers, and editors are in favor of this type of peer review, especially among younger generations. However, a large number of authors also believe that open identities could lead to worse and less critical peer reviews, mostly because of the potential consequences from the aggrieved authors.27 Another strategy could be the promoting of unsolicited reviews carried out spontaneously after the publication of the paper in a peer reviewed journal as it might favor early detection of errors or scientific misconduct in all scientific papers, and not only in those pertaining to COVID-19.5

This paper has some limitations. On the one hand, the sample size is small for the design used. This is due to the inclusion criterion which determines the selection of original research studies. Rather than being original research studies, a good part of the COVID-19 articles published to date are in the form of letters or editorials.2 We, however, chose to include original papers only, whether COVID-19 or non-COVID-19, since these must meet a single common standard of quality. On the other hand, this study included papers published in scientific journals which were subjected to peer review, and as a consequence other type of publication, such as preprints, are not taken into account. A further limitation is the number of missing values in some of the variables collected, due to lack of availability of data, especially on aspects such as time to publication and enrollment centers. Another limitation is that the selection of the papers where made based on the order of appearance of the articles in the selected issues. A further limitation is that we have not used any tool or scale to formally measure the methodological quality of the studies included and compare it between COVID and non-COVID studies. It would have been interesting to include the authors gender as a variable, however, it is difficult to determine it through the name and affiliation of the authors. Because of that, the collection of the variable gender could be prone to errors, especially in the case of Asiatic authors, and this is especially relevant in COVID-19 papers.

One of the strengths of this study is its design, established to favor comparison of the quality of studies in a single journal across the period analyzed, as well as between quartiles of different areas of JCR knowledge. This design facilitates comparison between COVID-19 and non-COVID-19 papers.

In conclusion, characteristics of COVID-19 papers differed from those of non-COVID-19 papers published in the first months of
the pandemic. Many of the aspects relating to COVID-19 remain unknown, despite the effort of the scientific community to generate and disseminate information about the disease. Hence, more evidence of better quality must be generated with the aim of guiding decision-making in the context of the pandemic. That said, however, the urgency with which information is needed is no excuse for conducting low-quality research, which is ultimately of little use, leads to a waste of research resources, and may even pose a health threat and/or undermine scientific credibility. For all the above-mentioned reasons, measures are required that guarantee and improve the quality of COVID-19-related research studies.

Availability of databases and material for replication

These data are not public since they are being used for the PhD work of the first author. Data are available upon reasonable request (contact: A. Ruano-Ravina, alberto.ruano@usc.es).

What is known about the topic?

The COVID-19 pandemic has impacted in biomedical publishing. The rapid increase in COVID-19-related articles, as well as their rapid publication in scientific journals, may have led to a large number of COVID-19-related studies being of low methodological quality.

What does this study add to the literature?

The results show that COVID-19 studies published in 2020 were of lower quality than those published during the initial 3-month period of the pandemic. A preliminary bibliometric analysis during the early stage of the outbreak of the pandemic showed that research on COVID-19-related topics has been rapid, but of lower quality than that of non-COVID-19 topics.

What are the implications of the results?

It is necessary to ensure the quality of published studies in order to have the best possible evidence and to avoid the circulation of unreliable results and conclusions, which may have consequences for the population and the scientific community.

Funding

This work is part of the research conducted to the PhD degree of C. Candal-Pedreire, who has received a FPI fellowship (reference number FI21/00149) from the Health Institute Carlos III (ISICIII).

Conflicts of interest

None.

References