

Tracing open data in emergencies: The case of the COVID-19 pandemic

Konstantinos Gkiouras¹  | Meletios P. Nigdelis²  | Maria G. Grammatikopoulou³  |
Dimitrios G. Goulis² 

¹Laboratory of Clinical Pharmacology, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

²Unit of Reproductive Endocrinology, 1st Department of Obstetrics and Gynecology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

³Faculty of Health Sciences, International Hellenic University, Thessaloniki, Greece

Correspondence

Dimitrios G. Goulis, Unit of Reproductive Endocrinology, 1st Department of Obstetrics and Gynecology, Medical School, Aristotle University of Thessaloniki, Thessaloniki GR56429, Greece.
Email: dgg@auth.gr

Abstract

Background: The coronavirus disease 2019 (COVID-19) pandemic constitutes an ongoing, burning Public Health Emergency of International Concern (PHEIC). In 2015, the World Health Organization adopted an open data policy recommendation in such situations.

Objectives: The present cross-sectional meta-research study aimed to assess the availability of open data and metrics of articles pertaining to the COVID-19 outbreak in five high-impact journals.

Methods: All articles regarding the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), published in five high-impact journals (*Ann Intern Med*, *BMJ*, *JAMA*, *NEJM* and *Lancet*) until March 14, 2020 were retrieved. Metadata (namely the type of article, number of authors, number of patients, citations, errata, news and social media mentions) were extracted for each article in each journal in a systematic way. Google Scholar and Scopus were used for citations and author metrics respectively, and Altmetrics and PlumX were used for news and social media mentions retrieval. The degree of adherence to the PHEIC open data call was also evaluated.

Results: A total of 140 articles were published until March 14, 2020, mostly opinion papers. Sixteen errata followed these publications. The number of authors in each article ranged from 1 to 63, whereas the number of patients with a laboratory-confirmed SARS-CoV-2 infection reached 2645. Extensive hyperauthorship was evident among case studies. The impact of these publications reached a total of 4210 cumulative crude citations and 342 790 news and social media mentions. Only one publication (0.7%) provided complete open data, while 32 (22.9%) included patient data.

Conclusions: Even though a large number of manuscripts was produced since the pandemic, availability of open data remains restricted.

KEYWORDS

altmetrics, coronavirus, publication metrics, research waste, SARS-CoV-2, scientometrics

Konstantinos Gkiouras and Meletios P. Nigdelis contributed equally to the manuscript.

© 2020 Stichting European Society for Clinical Investigation Journal Foundation

1 | INTRODUCTION

Forty-six years ago, Monto, a Coronavirus epidemiology expert reported the complete lack of literature suggesting that human Coronaviruses could be involved in lethal infections.¹ Today, according to the World Health Organization (WHO), a pandemic is an appropriate analogy for what the coronavirus disease 2019 (COVID-19) has developed to,² sparking the need for information concerning the characteristics of the virus, the underlying pathology and possible management. Lessons learned from the Ebola outbreak include the importance of prompt and facilitated availability of data in Public Health Emergencies of International Concern (PHEIC), as globally agreed in the 2015 WHO consultation,³ to enhance preparedness. This paradigm shift in the approach to information sharing in emergencies encompassed open data, elimination of embargo policies, fit-for-purpose platforms to present relevant research items, as well as expedited, or post-publication peer review policies.^{3,4}

The proposed framework was adopted immediately in the COVID-19 outbreak, being the second public health emergency to implement this specific WHO consultation after the Zika virus outbreak. In parallel, the Wellcome Trust issued a similar statement, high-lightening the need for rapid dissemination of research data and findings relevant to the virus.⁵ Subsequently, editors of scientific journals adhered to this strategy by immediately issuing calls for research items on COVID-19 to feed the existing demand gap,^{6,7} often expediting the review process to disseminate information as promptly as possible, in an open access manner.

These calls for COVID-19 intel sparked a new era in scientific publishing, with individual journals receiving 20⁸ to 100⁹ articles every day, concerning the virus. On the other hand, scientists have called for taking everything with a pinch of salt as the circulation of fake news was pedalled from non-evidence-based data.¹⁰ Given that we are currently living unprecedented times in both terms of health outbreak and rapid publication analogy, a close look at the metrics behind these rapidly produced publications in five high-impact journals (*BMJ*, *NEJM*, *JAMA*, *Lancet* and *Ann Intern Med*) is of utmost importance to better understand this unique situation we are currently experiencing.

2 | METHODS

2.1 | Search strategy

Three independent researchers accessed the web platforms designed specifically for articles pertaining to COVID-19 pandemic developed by five high-impact journals. The search strategy was designed to locate any COVID-19-related publication, with a publication type normally involving peer

review in each journal, published from inception, until March 14th, 2020. The COVID-19 resource platforms of the *Annals of Internal Medicine* (*Ann Int Med*),¹¹ *The British Medical Journal* (*BMJ*),¹² *The Journal of the American Medical Association* (*JAMA*),¹³ *The New England Journal of Medicine* (*NEJM*),¹⁴ and *The Lancet* (*Lancet*)¹⁵ were searched. Articles were included in this study if their type was eligible for a peer review process in each journal, as ascertained by the information provided on each journal's guidelines for authors.

2.2 | Data extraction

Metadata characteristics were extracted from each publication by three researchers for each publication in a predefined Microsoft Excel[®] spreadsheet. Extracted information involved (a) characteristics of the article including type, country of origin, number of authors, number of patients (with a laboratory-confirmed SARS-CoV-2 infection); (b) citations based on Google Scholar,¹⁶ (c) errata; (d) data availability; (e) news mentions based on alternative metrics (altmetrics) provided by each journal's website for every article; (f) social media mentions including Facebook,¹⁷ Twitter,¹⁸ and Reddit,¹⁹ according to the altmetrics provided by each journal, as well as (g) first and last author publication metrics including the total number of publications based on PubMed²⁰ and the Hirsch-index (h-index)²¹ according to Scopus.²²

Citations of the published items were retrieved from Google Scholar,¹⁶ using the title of each manuscript as the searched term. This method provided results related to the specific publication only. Then, the total number of citations of each article was normalized by the immediacy index of the journal, provided by the Web of Science.²³ The ratio of the number of authors to patients (confirmed cases of SARS-CoV-2 infection) was also calculated for ascertaining any hyperbole in the number of authors claiming authorship. Articles were categorized in data driven (if they included patients), reviews (when summarizing evidence), or opinion papers.

Concerning altmetrics, the *BMJ*, *JAMA* and *Ann Int Med* used the Altmetric website²⁴ to measure the impact of their publications; *The Lancet* applied the Elsevier relevant metric website (PlumX),²⁵ whereas the *NEJM* used a journal-specific impact metric method, based on the Altmetric.²⁴ First and last author data were retrieved directly for authors with Scopus²² profiles. In those lacking a Scopus²² author profile, the affiliation used in the COVID-19 publication was applied in the search to identify them within the Scopus database.

Data availability was ascertained if articles provided all data in an accessible way (eg either as supplements of their online publications or as stored databases on publicly available depositories/websites). In the case of case reports/series

(describing either only one or more patients), data availability was determined by checking if data was provided for all patient characteristics and measurements mentioned within the publications.

2.3 | Statistical analyses

The present study follows the reporting guidelines for health research.²⁶ Data are presented as means \pm standard deviations (SDs), followed by medians alongside the minimum and maximum values for each variable. Descriptive analysis of the data was conducted on the Statistical Package for Social Sciences version 25.0 (SPSS).

3 | RESULTS

A total of 140 research items were retrieved from the five journals (Table S1). Figure 1 details the number of research items published by researchers in each country, until the search end-date. Most publications originated from China, followed by the UK and the USA.

In Table 1, characteristics and metrics of the retrieved publications are presented. Most of the publications involved opinion papers, with only 25% ($n = 35$) being data driven. The majority of items were published by *The Lancet*, being mainly opinion papers. Mean crude citations reached 34.3 ± 109 , whereas normalized citations were calculated at 2.2 ± 6.8 per research item. New mentions reached 609 for

individual publications, whereas the maximum cumulative social media mentions skyrocketed at 41 150. Many errata ($n = 16$, 11.4%) were noted among the published research. The number of authors ranged between 1 and 63, and the cumulative patient data totalled 2645 COVID-19 cases. The *Ann Int Med* mother journal had published the fewer COVID-19 research items and exhibited null errata concerning the related publications.

Table 2 details the degree of adherence to the WHO call for open data in PHEIC among the five examined journals. The open data suggestion was fulfilled by all journals as far as unrestricted access to the COVID-19-related manuscripts was concerned. A very small percentage of the research, however (0.7%), fulfilled the open database WHO recommendation *per se*. Actual patient data were included in 22.9% of the retrieved items. All five journals adhered to the expedited peer review policy and the development of hubs specifically for COVID-19 research items.

4 | DISCUSSION

The present results indicate that publication norms have changed to facilitate information dissemination concerning the COVID-19 pandemic. In further detail, a surge of manuscripts emerged as soon as the SARS-CoV-2 was identified, with distinct authorship etiquette, citation and dissemination effects, as well as peer review policies, all differing from what was considered as “acceptable” until now. Instead of actual open data and dataset sharing, most of the published



FIGURE 1 Number of research items by country of origin ($N = 140$)

TABLE 1 Metadata of the COVID-19 published research items

Publications per type in each journal	Research items (n)	Open data (n)	Citations ^a			
			Crude (Mean ± SD [Median, range])	Normalized citations ^d (Mean ± SD [Median, range])	Cumulative (n)	Errata (n)
<i>Ann Int Med</i>	6	2	2.5 ± 6.1 [0,0-15]	0.4 ± 0.9 [0, 0-2.3]	15	0
Data driven	2	2	0	0	0	0
Reviews	0	NA	—	—	—	—
Opinions	4	NA	3.8 ± 7.5 [0, 0-15]	0.6 ± 1.2 [0, 0-2.3]	15	0
<i>BMJ</i>	36	0	1.3 ± 7.2 [0, 0-43]	0.1 ± 0.6 [0, 0-3.8]	47	2
Data-driven	1	0	43 ^c	3.8 ^c	NA	1
Reviews	1	NA	0 ^c	0 ^c	NA	1
Opinions	34	NA	0.1 ± 0.5 [0, 0-2]	0.01 ± 0.04 [0, 0-0.18]	4	0
<i>JAMA</i>	25	1	18.2 ± 51.6 [1, 0-255]	1.6 ± 4.5 [0.1, 0-22]	456	3
Data driven	8	1	41.3 ± 87.3 [7.5, 0-255]	3.6 ± 7.5 [0.6, 0-22]	330	1 ^f
Reviews	0	NA	—	—	—	—
Opinions	17	NA	7.4 ± 16.3 [0, 0-64]	0.6 ± 1.4 [0, 0-5.5]	126	2
<i>Lancet</i>	56	1	42 ± 126.1 [2, 0-740]	2.8 ± 8.5 [0.1, 0-49.7]	2354	5
Data-driven	13	1	155.9 ± 229.4 [11, 0-740]	10.5 ± 15.4 [0.7, 0-49.7]	2027	4
Reviews	4	NA	7.5 ± 11.7 [2.5, 0-25]	0.5 ± 0.8 [0.2, 0-1.7]	30	0
Opinions	39	NA	7.6 ± 24.5 [1, 0-153]	0.5 ± 1.6 [0.1, 0-10.3]	297	1 ^f
<i>NEJM</i>	17	7	113.3 ± 185.4 [24, 0-573]	6.1 ± 10 [1.3, 0-30.8]	1926	6
Data driven	11	7	164.4 ± 214.8 [55, 0-573]	8.8 ± 11.6 [3, 0-30.8]	1808	6
Reviews	0	NA	—	—	—	—
Opinions	6	NA	19.7 ± 37.8 [3, 0-96]	1.1 ± 2 [0.2, 0-5.2]	118	0
Total	140	11^e	34.3 ± 109 [1, 0-740]	2.2 ± 6.8 [0.1, 0-49.7]	4798	16

Note: Data are presented in the form of means ± SDs [median, minimum-maximum values], or n.

From inception until March 14, 2020, all publications that would originally undergo peer review in each journal based on the instructions for authors;

Abbreviations: *Ann Int Med*, *Annals of Internal Medicine*; *BMJ*, *British Medical Journal*; COVID-19, coronavirus disease 2019; *JAMA*, *Journal of the American Medical Association*; NA, not applicable; *NEJM*, *New England Journal of Medicine*; PCR, polymerase chain reaction; SD, standard deviation; WoS, web of science.

^aBased on Google Scholar¹⁶;

^bFacebook,¹⁷ twitter,¹⁸ and/or reddit¹⁹;

^cRefers to data from one publication only;

^dBased on WoS metrics²³;

^eOut of which, n = 7 consisted of case reports, n = 1 estimated the virus incubation period, n = 1 referred to a retrospective analysis on children, n = 1 tested the presence of SARS-CoV-2 on air and surfaces, and n = 1 presented newspaper data;

^fDuring the revision of the present manuscript, one additional errata was noted;

^gRefers to confirmed COVID-19 positive cases based on the PCR only.

News mentions		Social media mentions ^b		Number of authors		Cumulative patient sample size ^g (n)
Mean \pm SD [Median, range]	Cumulative (n)	Mean \pm SD [Median, range]	Cumulative (n)	Mean \pm SD [Median, range]	Cumulative (n)	
98 \pm 208.7 [14.5, 1-523]	588	549.2 \pm 1213.8 [60, 7-3026]	3295	6.2 \pm 6.9 [3, 1-19]	37	182
262 \pm 369.1 [262, 1-523]	524	1516.5 \pm 2134.8 [1516.5, 7-3026]	3033	14 \pm 7.1 [14, 9-19]	28	182
—	—	—	—	—	—	NA
16 \pm 17 [14.5, 1-34]	64	65.5 \pm 34.6 [60, 33-109]	262	2.3 \pm 1.3 [2, 1-4]	9	NA
5 ^c	NA	648.2 \pm 1239.8 [179.5, 0-5450]	23 336	2.6 \pm 2.5 [1.5, 1-14]	95	62 ^c
5 ^c	NA	389 ^c	NA	14 ^c	NA	62 ^c
NA	NA	1681 ^c	NA	4 ^c	NA	NA
NA	NA	625.5 \pm 1263 [146, 0-5450]	21 266	2.3 \pm 1.6 [1, 1-6]	77	NA
78.4 \pm 133.2 [30, 0-609]	1960	2552.7 \pm 3973.2 [783, 37-16 306]	63 817	5 \pm 5.4 [3, 2-27]	125	396
102.1 \pm 102.8 [67, 6-309]	817	3729.4 \pm 4106.8 [2800.5, 426-12 885]	29 835	10.3 \pm 7.2 [7, 6-27]	82	396
—	—	—	—	—	—	NA
67.2 \pm 146.9 [10, 0-609]	1143	1998.9 \pm 3909.3 [563, 37-16 306]	33 982	2.5 \pm 0.5 [3, 2-3]	43	NA
29.7 \pm 57.5 [4, 0-285]	1217	3490.6 \pm 7543.5 [452.5, 6-41 150]	174 530	7.1 \pm 8.5 [4, 1-38]	400	429
92.6 \pm 87.7 [74, 1-285]	926	11 506 \pm 12 877 [9223, 114-41 150]	126 566	15.8 \pm 12.1 [14, 3-38]	205	429
14 \pm 18.2 [5, 2-35]	42	1617 \pm 1621 [1244, 215-3392]	4851	4 \pm 2.2 [3.5, 2-7]	16	NA
8.9 \pm 19.3 [2, 0-98]	249	1197.6 \pm 2497.6 [369, 6-12 906]	43113	4.6 \pm 5 [3, 1-27]	179	NA
25.9 \pm 38.6 [3, 0-114]	440	4577.2 \pm 5764.3 [1471, 122-21 996]	77 812	16.1 \pm 17.7 [9, 1-63]	273	1576
36.4 \pm 44 [18, 0-114]	400	6199.7 \pm 6572.6 [5657, 122-21 996]	68 197	23.5 \pm 18.1 [19, 4-63]	258	1576
—	—	—	—	—	—	NA
6.7 \pm 14.4 [0.5, 0-36]	40	1602.5 \pm 1875.6 [887, 482-5355]	9615	2.5 \pm 1.5 [2.5, 1-5]	15	NA
46.8 \pm 98.4 [6, 0-609]	4210	2558.1 \pm 5503.4 [497, 0-41 150]	342 790	6.6 \pm 9.4 [3, 1-63]	930	2645

Open data (shared datasets)	Fulfilled by 0.7% of the research items (n = 1 out of 140) ^a
Patient data	Included in 22.9% of the research items (n = 32 out of 140), with the remaining being mainly opinion papers ^c
Expedited peer review policy	Fulfilled by all five ^b journals
Elimination of embargo policy (open access for COVID-19-related manuscripts)	Fulfilled by all five ^b journals (open access to all items)
Fit-for-purpose platforms to present COVID-19 research	Fulfilled by all five ^b journals

Abbreviations: *BMJ*, *British Medical Journal*; COVID-19, coronavirus disease 2019; PHEIC, Public Health Emergencies of International Concern; WHO, World Health Organization.

^aSince the April 20, 2020 (after the present search was conducted), the *BMJ* has requested for open datasets concerning COVID-19 related-research.

^bSearched journals included the *Annals of Internal Medicine*, *The British Medical Journal*, *The Journal of the American Medical Association*, *The Lancet* and *The New England Journal of Medicine*.

^cAn additional n = 3 publications were also data driven without, however, any patient data (involved samples from surfaces, air or incubation period data).

items were in fact opinion papers, case reports and reviews, indicating that the PHEIC call was not adhered.

4.1 | “Publish or perish” in the times of COVID-19

Carefully poring over these five journals since the COVID-19 outbreak indicates that the majority of publications were produced by Chinese researchers, followed by UK and US scientists (Figure 1). Although China is known to exhibit a “publish or perish” culture,²⁷ in the case of COVID-19, Chinese research appears highly justified given that the country was the initial epicentre of the virus. In parallel, many studies originating from China included a large number of patients (>100), presenting most of the COVID-19 data synthesis, with scientists from other countries either presenting case studies, case series or opinion and narrative review papers. Following the COVID-19 call for data, opinion and reviews lacking patient data consisted of 75% of the published research in these five journals.

Interestingly, many research items were characterized by extensive authorship lists (Table 1). Hyperauthorship is common in multicentre studies, with an increased collaboration size being needed to orchestrate data collection. In China, the need to reform the quality of submitted manuscripts²⁸ and reduce the number of hyperprolific authors²⁷ stemming from a local cash-for-publication policy²⁹⁻³² has been high-lightened. On these premises, fairly recently, Chinese authorities opted for tackling these incentives, by limiting publication rewards and promoting better research practices.³³ On the other hand, the cash-for-publication policy has been adopted by institutions from many other countries, including the USA.³⁴

TABLE 2 Degree of adherence to the WHO call for open data in PHEIC in five milestone Medical journals^b

However, in the call for COVID-19 open data, the authorship norm appears to be stretched out. Although some of the research was submitted by prominent figures in medicine or the political scene (eg Bill Gates), many items were authored by more “rookie” scientists. In further detail, the COVID-19 era was the perfect opportunity for many researchers to publish, with some identified as publishing their first manuscript following the COVID-19 call, and others having an h-index of 0. Often, when authors with Chinese/Korean names were encountered, it was not possible to calculate the exact number of publications or h-index, as Scopus has been reported to distinguish inadequately between Chinese/Korean names,²⁷ and the same problem has also been reported in other databases, including Google Scholar³⁵ resulting in confusing and erroneous metadata records.³⁶

Even among case studies, although data is being reported for one patient only, there were cases when the number of authors exceeded 60.³⁷ The patient-to-author ratio in case reports ranged between 0.016-0.25, indicating that a minimum of four to a maximum of 63 authors were employed to report a single case study. Therefore, the number of authors might not always reflect the needs of the presented work, but could well be the epiphenomenon of a “publish or perish” trend,^{27,38,39} authorship abuse/manipulation, or even a “you scratch my back and I will scratch yours” strategy among colleagues.^{40,41} When case studies are proliferating excessively, this might entail a higher risk for misconduct,⁴² although it has been suggested that sometimes a collectivistic perspective is exhibited by scientists stemming from more individualistic countries.^{40,43} In an extended version of this multi-author trend in the COVID-19 era, the opportunity to publish in flagging academic journals may exaggerate this effect even further, increasing the number of researchers with authorship credit, even among case reports.

Shared first authorship position was also very prominent among published research items with ten papers having shared first authorship at *The Lancet*, one of which divided the first author position among six authors.⁴⁴

4.2 | Virus dissemination goes viral

The majority of research concerning the SARS-CoV-2 received an astonishing number of citations in less than a month's time post-publication. The cumulative citations of all manuscripts published in the five journals reached 4798, corresponding to a mean of 34.3 citations per research item. Although Google Scholar has been recently portrayed as “the most comprehensive academic search engine” to date,⁴⁵ it also entails several limitations.⁴⁶ For instance, it appears that it is more efficient in detecting grey literature,⁴⁶ and this might produce a greater number of citations results for each research item. Nevertheless, searching specific article titles, as performed in the present analysis, appears to be a valid method for Google Scholar use.⁴⁷ Considering that in the five journals assessed herein, a high-impact manuscript with a “sexy” subject might acquire 0 to a maximum of 2 citations during the first publication month (with one being a possible editorial referring to the data and the other being either a super-fast response in the form of letter, or an UpToDate⁴⁸ citation), it becomes clear that citation farming spirals in PHEIC.

In several cases, altmetrics have been shown to correlate with citations⁴⁹⁻⁵¹ and can even predict highly cited articles.⁵² In the case of COVID-19 intel, the news and social media impact was extremely powerful, reflecting the need for immediate information dissemination in emergencies. The most eye-catching part of the present study is that selected articles received more than 20 000 social media mentions in less than a month. One of the early manuscripts, published in late January 2020⁵³ has been mentioned more than 41 000 in social media, while yielding more than 740 citations in less than 2 months. This indicates that in the COVID-19 era, the use of social media excelled beyond the medical profession lobby and research promotion networks, blending with crisis informatics⁵⁴ to feed the societal information gap concerning the pandemic.

4.3 | Chop chop!

With the need for immediately available information on the course and management of COVID-19, some of the published research items consisted of the first cases detected in selected countries. Interestingly, the SARS-CoV-2 patient 0, suggested to have originated from the Hubei province,^{55,56} was never published as a case report. On the contrary, after

the Wuhan lockdown, many first cases from other countries were published in high-impact journals promptly. For instance, the first US case was confirmed by the Centers for Disease Control and Prevention on January 21st,⁵⁷ and published as a case study as soon as 10 days later.³⁷ Other countries following this first country case publication regime included the UK⁵⁸ and Canada.⁵⁹ It appears that publication of the first country case studies is irrelevant to the COVID-19 burden suffered in each region, as nor the Chinese, neither the Italians, ever reported the characteristics of their country's first patient.

When adhering to the “publish or perish” culture, the likelihood to produce a subsequent erratum of the manuscript is increased, in parallel to the chance for retraction.^{60,61} Among 140 manuscripts, 16 had subsequent corrections revealing pitfalls in the fast-track publication regime. In a follow-up search of our data conducted in late May 2020, 2 additional errata were identified. A quick search in PubMed for retracted articles on COVID-19 showed that three original articles and one editorial (based on one of these manuscripts) to date, were retracted from journals, indicating that editors must be prudent with all submitted research in PHEIC until March 14, 2020. As already noted by Ioannidis,¹⁰ one specific retracted publication⁶² has received an astonishing Altmetric score of 254 647 points. On the other hand, questions have been raised concerning the quality of published information⁶³ with many noting that in the times of COVID-19, the spread of misinformation, in parallel to valid data, is in fact, inevitable.^{64,65} Examples include possible therapies examined via clinical trials published in flagging journals without a comparator, lacking randomization and with authors having clear conflicts of interest (ie funding sources).^{9,66} Nevertheless, on the aftermath of this pandemic, the experience we will have gained can also be used to reduce publication redundancy⁶⁷ and sloppiness, and focus on the quality of the produced research.

4.4 | Under the “open data” umbrella: mind the gap

Although the term “open data” appears to be self-explanatory, researchers and institutions tend to have a different understanding of the concept.⁶⁸ According to the WHO consultation,³ sharing of both data and results should be the default practice; however, the number of COVID-19 research items actually sharing their data was minimal, as 11 out of 140 manuscripts in total, shared actual open patient data. Of note, of these 11 with available patient data (Table 1), only one (0.7%) had a linked dataset file and the used R code, two presented all data within the manuscript tables, one presented newspaper data, and the remaining seven involved either case studies, or case series, where all data were inevitably

presented within the manuscript text. Thus, only one actually fulfilled the call for open datasets (Table 2). This indicates that some researchers make use of the crisis opportunity to submit their manuscripts in an open access manner, without, however, adhering to the WHO recommendations for actual data sharing.³ Luckily, during a revision of the current paper which underwent normal review process, the *BMJ* issued a revision of the call for COVID-19 research (April 20, 2020) asking for a parallel submission of data alongside each SARS-CoV-2-related article.

Inevitably, the lack of actual open data dissemination does not allow for the synthesis of similar data from distinct research items and might allow for data slicing,⁶⁹ as identified in one case herein. Similar concerns were raised by the *JAMA* editorial board⁷⁰ suggesting that duplicate patient records might pollute accuracy of the COVID-19 epidemiology. Nevertheless, it appears that readers must exercise due diligence concerning all published items.¹⁰

4.5 | What's in the pipeline?

At the moment, a total of 101 clinical trials concerning the virus are registered in clinicaltrials.gov, three of which, all China-based, have reported a “completed” status (cumulative patient sample $N = 716$) and many are expected to finish on time (late April 2020). Hopefully, by design, these are bound to produce evidence-based data concerning the management of COVID-19 infection and aid in tampering down the epidemic. The proportion of these that will eventually share their data apart from presenting their results is yet to be discovered, and will help to assess whether the call for open data in PHEIC needs to be re-evaluated in a novel and stricter framework.

5 | CONCLUSIONS

Collectively, the calls for open data in COVID-19 produced a large number of manuscripts in a very short time,⁷¹ although the majority (77.1%) failed to include actual patient data but consisted mainly of opinion papers or reviews. In fact, the surplus of published research was as unanticipated as the spread of COVID-19. In the hypothetical question if the scientists are getting what they expect from this literature burst, the answer is that they are certainly getting what is available, although the call for open data share does not appear to be fully adhered. As a result, at the moment, the signal-to-noise ratio is inflated, with opinions and personal views creating buzz; what is needed are actual data that could be synthesized to aid evidence-based practice and mitigate the pandemic. Therefore, the most appropriate question to consider is whether the call for open data

in PHEIC actually serves its purpose or if it compromises research integrity and quality.⁷²

Undoubtedly, publication norms differ in emergencies, and the rigging review process entails a series of bottlenecks, allowing for a few scientists to abuse relevant calls for open data and use them as a means to publish through a loophole. Nevertheless, as indicated by the social media impact, in PHEIC, the importance of data dissemination is pivotal, and the flow of research should not be ceased until the pandemic has been tackled.

CONFLICTS OF INTEREST

None.

AUTHOR CONTRIBUTION

KG, MPN and MGG extracted data from published research. KG and MPN performed the descriptive analysis. DGG and MGG conceived the study, contributed to the first draft of the manuscript and provided critical revisions. All authors gave intellectual input to improve the manuscript and have read and approved the final version. DGG is the guarantor.

ORCID

Konstantinos Gkiouras  <https://orcid.org/0000-0002-6407-8494>

Meletios P. Nigdelis  <https://orcid.org/0000-0003-3865-6833>

Maria G. Grammatikopoulou  <https://orcid.org/0000-0003-4167-6595>

Dimitrios G. Goulis  <https://orcid.org/0000-0002-5005-1995>

REFERENCES

1. Monto AS. Medical reviews. Coronaviruses. *Yale J Biol Med*. 1974;47(4):234-251.
2. Mahase E. Covid-19: WHO declares pandemic because of “alarming levels” of spread, severity, and inaction. *BMJ*. 2020;12(368):m1036.
3. World Health Organization [WHO]. *Developing global norms for sharing data and results during public health emergencies*. Geneva: World Health Organization [WHO]; 2015.
4. Dye C, Bartolomeos K, Moorthy V, Kieny MP. Data sharing in public health emergencies: a call to researchers. *Bull World Health Organ*. 2016;94(3):158.
5. Wellcome Trust. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Wellcome [Internet]. 2020 [cited March 8, 2020]. <https://wellcome.ac.uk/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-covid-19-outbreak>. Accessed June 22, 2020
6. The Lancet. Emerging understandings of 2019-nCoV. *Lancet (London, England)*. 2020;395(10221):311.
7. Kickbusch I, Leung G. Response to the emerging novel coronavirus outbreak. *BMJ*. 2020;31(368):m406.
8. Yasinski E. Journals open access to coronavirus resources. *The Scientist*. 2020. <https://www.the-scientist.com/news-opinion/journ>

- als-open-access-to-coronavirus-resources--67105. Accessed June 22, 2020
9. Mogensen JF. Science has an ugly, complicated dark side. And the coronavirus is bringing it out. [Internet]. *Mother Jones*. 2020. <https://www.motherjones.com/politics/2020/04/coronavirus-science-rush-to-publish-retractions/>. Accessed June 22, 2020
 10. Ioannidis JPA. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. *Eur J Clin Invest*. 2020;19:e13222.
 11. American College of Physicians. Annals of internal medicine coronavirus disease 2019 (COVID-19) [Internet]. *Ann Intern Med*. 2020 [cited March 14, 2020]. <https://www.acpjournals.org/topic/category/coronavirus>. Accessed June 22, 2020
 12. BMJ. BMJ's coronavirus (covid-19) hub [Internet]. *Br Med J*. 2020 [cited May 23, 2020]. <https://www.bmj.com/coronavirus>. Accessed June 22, 2020
 13. JAMA Network. JAMA coronavirus disease 2019 (COVID-19) [Internet]. *J Am Med Assoc*. 2020 [cited May 23, 2020]. <https://jamanetwork.com/journals/jama/pages/coronavirus-alert#clinical-information>. Accessed June 22, 2020
 14. NEJM. NEJM coronavirus (Covid-19) [Internet]. *N Engl J Med*. 2020 [cited March 14, 2020]. <https://www.nejm.org/coronavirus>. Accessed June 22, 2020
 15. The Lancet. COVID-19 Resource Centre [Internet]. *Lancet*. 2020 [cited March 14, 2020]. <https://www.thelancet.com/coronavirus>. Accessed June 22, 2020
 16. Google. Google Scholar [Internet]. [cited March 14, 2020]. <https://scholar.google.com>. Accessed June 22, 2020
 17. ©2020 Facebook. Facebook [Internet]. 2020. www.facebook.com. Accessed June 22, 2020
 18. Twitter I. Twitter [Internet]. <https://twitter.com/home>. Accessed June 22, 2020
 19. Advance Publications. Reddit [Internet]. 2020. <https://www.reddit.com>. Accessed June 22, 2020
 20. National Library of Medicine, National Centre for Biotechnology Information. PubMed [Internet]. 2020 [cited May 23, 2020]. <https://pubmed.ncbi.nlm.nih.gov/>. Accessed June 22, 2020
 21. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*. 2005;102(46):16569-16572.
 22. Elsevier. Scopus [Internet] [cited March 14, 2020]. <https://www.scopus.com/>
 23. Clarivate Analytics Company. Web of science [Internet]. 2020 [cited May 23, 2020]. <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>. Accessed June 22, 2020
 24. Digital Science. Altmetric [Internet]. Digital Science. 2020 [cited February 23, 2020]. <https://www.altmetric.com/>. Accessed June 22, 2020
 25. Elsevier Inc. Plumx [Internet]. Elsevier. 2020 [cited May 23, 2020]. https://plu.mx/sign_in. Accessed June 22, 2020
 26. Simeria I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *Eur J Clin Invest*. 2010;40(1):35-53.
 27. Ioannidis JPA, Klavans R, Boyack KW. The scientists who publish a paper every five days. *Nature*. 2018;561:167-169.
 28. Publish or perish. *Nature*. 2010;467(7313):252.
 29. Hvistendahl M. China's publication bazaar. *Science*. 2013;342(6162):1035-1039.
 30. Hvistendahl M. The numbers game. *Science*. 2013;340(6136):1037-1039.
 31. Byrne JA, Grima N, Capes-Davis A, Labbé C. The possibility of systematic research fraud targeting under-studied human genes: causes, consequences, and potential solutions. *Biomark Insights*. 2019;5(14):117727191982916.
 32. Chen X. High monetary rewards and high academic article outputs: are China's research publications policy driven? *Ser Libr*. 2019;77(1-2):49-59.
 33. Mallapaty S. China bans cash rewards for publishing papers. *Nature*. 2020;579(7797):18.
 34. Abritis A. Cash bonuses for peer-reviewed papers go global. *Science*. 2017. <https://doi.org/10.1126/science.aan7214>
 35. Gasparyan AY, Yessirkepov M, Gerasimov AN, Kostyukova EI, Kitas GD. Scientific author names: errors, corrections, and identity profiles. *Biochem Medica*. 2016;26(2):169-173.
 36. Cheng TO. The continuing confusion in figuring out the surname of a Chinese author—a proposed solution. *Chin J Integr Med*. 2012;18(4):243-244.
 37. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382(10):929-936.
 38. Papatheodorou SI, Trikalinos TA, Ioannidis JPA. Inflated numbers of authors over time have not been just due to increasing research complexity. *J Clin Epidemiol*. 2008;61(6):546-551.
 39. Cosgriff CV, Ebner DK, Celi LA. Data sharing in the era of COVID-19. *Lancet Digit Heal*. 2020;2(5):e224.
 40. Chapman CA, Bicca-Marques JC, Calvignac-Spencer S, et al. Games academics play and their consequences: how authorship, h-index and journal impact factors are shaping the future of academia. *Proc R Soc B Biol Sci*. 2019;286(1916):20192047.
 41. Fong EA, Wilhite AW. Authorship and citation manipulation in academic research. *PLoS One*. 2017;12(12):e0187394.
 42. Fanelli D. Pressures to publish: What effects do we see? In: Biagioli M, Lippman A, editors. *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. 1st edn. Cambridge, MA: The MIT Press;2020. p. 111-122.
 43. Ren X, Su H, Lu K, Dong X, Ouyang Z, Talhelm T. Culture and unmerited authorship credit: who wants it and why? *Front Psychol*. 2016;27(7):2017.
 44. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet (London, England)*. 2020;395(10229):1054-1062.
 45. Gusenbauer M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*. 2019;118(1):177-214.
 46. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS One*. 2015;10(9):e0138237.
 47. Giustini D, Boulos MNK. Google Scholar is not enough to be used alone for systematic reviews. *Online J Public Health Inform*. 2013;5(2):214.
 48. Wolters Kluwer Health. UpToDate. 2020. <https://www.uptodate.com/login>. Accessed June 22, 2020
 49. Hayon S, Tripathi H, Stormont IM, Dunne MM, Naslund MJ, Siddiqui MM. Twitter mentions and academic citations in the urologic literature. *Urology*. 2019;123:28-33.
 50. Chang J, Desai N, Gosain A. Correlation between altmetric score and citations in pediatric surgery core journals. *J Surg Res*. 2019;243:52-58.

51. Amath A, Ambacher K, Leddy JJ, Wood TJ, Ramnanan CJ. Comparing alternative and traditional dissemination metrics in medical education. *Med Educ*. 2017;51(9):935-941.
52. Eysenbach G. Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res*. 2011;13(4):e123.
53. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)*. 2020;395(10223):497-506.
54. Reuter C, Kaufhold M-A. Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics. *J Contingencies Cris Manag*. 2018;26(1):41-57.
55. Bassetti M, Vena A, Giacobbe DR. The novel Chinese coronavirus (2019-nCoV) infections: challenges for fighting the storm. *Eur J Clin Invest*. 2020;50(3):e13209.
56. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China. *JAMA*. 2020;323(13):1239.
57. Centers for Disease Control and Prevention. First travel-related case of 2019 novel coronavirus detected in United States. CDC online newsroom. CDC [Internet]. 2020 [cited March 8, 2020]. <https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html>. Accessed June 22, 2020
58. Moss P, Barlow G, Easom N, Lillie P, Samson A. Lessons for managing high-consequence infections from first COVID-19 cases in the UK. *Lancet (London, England)*. 2020;395(10227):e46.
59. Silverstein WK, Stroud L, Cleghorn GE, Leis JA. First imported case of 2019 novel coronavirus in Canada, presenting as mild pneumonia. *Lancet (London, England)*. 2020;395(10225):734.
60. Fanelli D, Ioannidis JPA. US studies may overestimate effect sizes in softer research. *Proc Natl Acad Sci USA*. 2013;110(37):15031-15036.
61. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci USA*. 2012;109(42):17028-17033.
62. Pradhan P, Pandey AK, Mishra A, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.01.30.927871>
63. Ornell F, Schuch JB, Sordi AO, et al. "Pandemic fear" and COVID-19: mental health burden and strategies. *Br J Psychiatry*. 2020;42(3):232-235.
64. Jacobsen KH, Vraga EK. Improving communication about COVID-19 and emerging infectious diseases. *Eur J Clin Invest*. 2020;50(5):e13225.
65. Chrousos GP, Mentis AA. Medical misinformation in mass and social media: an urgent call for action, especially during epidemics. *Eur J Clin Invest*. 2020;50(5):e13227.
66. Grein J, Ohmagari N, Shin D, et al. Compassionate use of remdesivir for patients with severe Covid-19. *N Engl J Med*. 2020;382(24):2327-2336.
67. Papes D, Ozimec E. Redundancy in reporting on COVID-19. *Eur J Clin Invest*. 2020;50(6):e13257.
68. D'Agostino M, Samuel NO, Sarol MJ, et al. Open data and public health. *Rev Panam Salud Publica*. 2018;42:e66.
69. Ding D, Nguyen B, Gebel K, Bauman A, Bero L. Duplicate and salami publication: a prevalence study of journal policies. *Int J Epidemiol*. 2020;49(1):281-288.
70. Bauchner H, Golub RM, Zylke J. Editorial concern—possible reporting of the same patients with COVID-19 in different reports. *JAMA*. 2020;323(13):1256.
71. Ghebreyesus TA, Swaminathan S. Scientists are sprinting to outpace the novel coronavirus. *Lancet (London, England)*. 2020;395(10226):762-764.
72. Cobey KD, Moher D. Publishing in 2020: a checklist to support a shift in behaviour to achieve best practice. *Eur J Clin Invest*. 2020;50(1):e13186.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Gkiouras K, Nigdelis MP, Grammatikopoulou MG, Goulis DG. Tracing open data in emergencies: The case of the COVID-19 pandemic. *Eur J Clin Invest*. 2020;50:e13323. <https://doi.org/10.1111/eci.13323>