
Olatz Arrizabalaga1, David Otaegui2, Itziar Vergara3, Julio Arrizabalaga1, Eva Méndez4

1Innovation Group, Biodonostia Health Research Institute, San Sebastian, 20014, Spain
2Multiple Sclerosis Group, Biodonostia Health Research Institute, San Sebastian, 20014, Spain
3Group of Research in Primary Care, Biodonostia Health Research Institute, San Sebastian, 20014, Spain
4Library and Information Science Department, Universidad Carlos III de Madrid, Madrid, 28903, Spain

Abstract

Background: The COVID-19 outbreak has made funders, researchers and publishers agree to have research publications, as well as other research outputs, such as data, become openly available. In this extraordinary research context of the SARS CoV-2 pandemic, publishers are announcing that their coronavirus-related articles will be made immediately accessible in appropriate open repositories, like PubMed Central, agreeing upon funders’ and researchers’ instigation.

Methods: This work uses Unpaywall, OpenRefine and PubMed to analyse the level of openness of articles about COVID-19, published during the first quarter of 2020. It also analyses Open Access (OA) articles published about previous coronavirus (SARS CoV-1 and MERS CoV) as a means of comparison.

Results: A total of 5,611 COVID-19-related articles were analysed from PubMed. This is a much higher amount for a period of 4 months compared to those found for SARS CoV-1 and MERS during the first year of their first outbreaks (335 and 116 articles, respectively). Regarding the levels of openness, 88.8% of the SARS CoV-2 papers are freely available; similar rates were found for the other coronaviruses. Deeper analysis showed that (i) 67.4% of articles belong to an undefined Bronze category; (ii) 76.4% of all OA papers don’t carry any license, followed by 10.4% which display restricted licensing. These patterns were found to be repeated in the three most frequent publishers: Elsevier, Springer and Wiley.

Conclusions: Our results suggest that, although scientific production is much higher than during previous epidemics and is open, there is a caveat to this opening, characterized by the absence of fundamental elements and values on which Open Science is based, such as licensing.

Keywords
This article is included in the Science Policy Research gateway.

This article is included in the Disease Outbreaks gateway.

**Corresponding author:** Olatz Arrizabalaga (olatz.arrizabalaga@biodonostia.org)

**Author roles:** Arrizabalaga O: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Otaegui D: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; Vergara I: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; Arrizabalaga J: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; Méndez E: Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2020 Arrizabalaga O et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


**First published:** 26 Jun 2020, 9:649 https://doi.org/10.12688/f1000research.24136.1
Introduction

In the last four months (January–April 2020), due to the COVID-19 pandemic, funders\textsuperscript{1,2}, researchers and publishers (such as Springer or Wiley) seem to agree upon making research outcomes related to the SARS CoV-2 pandemic openly available, including research papers (from preprints - MedRxiv and bioRxiv - to different mechanisms for waiving Article Processing Charges (APCs) or new specific Open Research platforms, as Elsevier or The Lancet). However, traditional practices for scholarly publishing and regular practices to access scientific content might not be mature enough for this massive open endeavour.

Throughout history, research and innovation have been key in the transformation of our society. It has been observed that, in addition to a direct economic benefit, only those societies with a certain level of scientific culture have the capacity to face new risks and participate in new ethical dilemmas, like the ones that we are currently facing. The more scientifically educated societies are, the freer they become, since answers to big social challenges arise from this interaction\textsuperscript{1}. Open Access (OA)/Open Science has been promoted over the last few decades by different stakeholders of the scientific system to make publications openly accessible, and more recently, also data and other research outcomes, in order to make them FAIR (Findable, Accessible, Interoperable and Reusable). All these initiatives aim to boost a democratic scientific advance in which scientists but also citizens are involved.

In the current situation of a global pandemic, OA becomes urgent. The emergence of the virus that causes the disease known as COVID-19 first reported by the Chinese authorities in late December 2019, has resulted in an unprecedented level of collaboration among researchers around the world\textsuperscript{1-6}. A health crisis, such as the SARS CoV-2 pandemic, requires special effort and collaboration within the scientific community in order to generate and disseminate new results, while trying to avoid duplication of efforts globally.

In this unique context of the pandemic, publishers are announcing massive OA changes, primarily by making their coronavirus-related articles freely available through databases, such as PubMed Central, together with other public repositories. SPARC Europe stated that overnight COVID-19 heightens the need for Open Science, and we cannot agree more. But we wonder if this openness might be enough in such a demanding and urgent episode for Science, and coincidently we wonder if the scientific community is ready to share and consume openly such information. This work aims to make an initial analysis of scientific production concerning COVID-19 and its level of openness as a first step to assess the current research publication model and the unpredicted outcome of openness of research in this global health emergency. Thus, this paper analysed all scientific content openly available from PubMed database.

Methods

Publication source

In order to analyse publications concerning COVID-19 and their level of openness, we have chosen PubMed instead of other multidisciplinary databases, like Web of Science (WoS) or Scopus. PubMed is one database developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) in the USA. It is one of the most used databases to find biomedical scientific content. This database gathers over 14 million bibliographic citations and it provides access to MEDLINE articles and PubMed Central (PMC), an extensive digital repository created in 2000 for biomedical and life sciences Open Access publications. Unlike many other research databases, such as WoS, PubMed also includes articles that are “in process”; this means a status prior to being indexed with MeSH terms, and articles submitted by publishers as pre-prints (i.e. articles that haven’t gone through peer review)\textsuperscript{7}. This aspect is crucial for this study since, at this moment, scientific papers are being published very fast and may not have yet undergone peer review\textsuperscript{8}.

Search terms

Since during the global pandemic period, the scientific community is posting articles that are freely accessible through the NCBI, data were collected from the PubMed database in order to analyse every COVID-19-related scientific paper that is currently published (including PMC)\textsuperscript{9}. In an attempt to evaluate the most accurate list of publications, we exported all results obtained from the suggested search queries offered by NLM (NCBI webpage), as follows: “2019-nCoV OR 2019nCoV OR COVID-19 OR SARS-CoV-2 OR (wuhan AND coronavirus)”. Only articles published from January 1\textsuperscript{st} to April 23\textsuperscript{rd} of 2020 were considered. No exclusions were made in the type of article (journal article, books, reviews, clinical trial or meta-analysis) or in the language, choosing in each case every article offered by PubMed.

In line with the objective of analysing published papers during other emergency circumstances, similar search procedures were applied to the SARS CoV-1 pandemic (query: “SARS CoV” OR “Severe Acute Respiratory Syndrome Coronavirus”; period searched: from 2003 to 2006) and MERS CoV epidemic (query: “MERS CoV” OR “Middle East Respiratory Syndrome Coronavirus”; period searched: from 2013 to 2016).

In order to determine the effect that this health emergency is having on the availability of the scientific production, we decided to compare it with the availability in a normalized situation, for which we performed the same analysis using two chronic diseases: low grade glioma (query: “low grade glioma”) and peptic ulcer (query: “peptic ulcer”), which, as seen by our search, have stable publication patterns for the last three years (2017 to 2019).
Data analysis

Obtained results, without exclusion, were exported and uploaded to OpenRefine, a free open source tool that helps exploration of large data sets, and has the capability to link and extend these data sets with different webservices. In this study, OpenRefine was used to manage data but also as the key element in order to link our PubMed data set with Unpaywall, the selected tool for analysing the OA content of all these data. Unpaywall (previously known as oaDOI) is a database introduced in 2016 as a service to check OA availability of journal articles identified by their Digital Object Identifier (DOI). Unpaywall is currently used more than 50,000 times a day and is maintained by Our Research, a non-profit company previously called Impactstory. It offers access to the OA status of scientific journals, through an open application programming interface (API). Unpaywall also shows license information and variable version availability from different repositories.

WoS, which includes OA information from Unpaywall, classifies OA papers in five-categories that we consider in this work: Gold, OA journal indexed by the Directory of Open Access Journals (DOAJ); Hybrid, subscription-based journals including some OA articles; Green, toll-access on the publisher page, but there is a free copy in an OA repository; and Bronze, articles freely available on websites hosted by their publisher, either immediately or following an embargo, but are not formally licensed for reuse. Unpaywall also provides information about Creative Commons (CC) licensing of each document (commonly Gold OA or Hybrid journals). Copyright licenses, released by Creative Commons, are variable and range from more open permissions (CC or CC-BY) to more restrictive ones (CC-BY-ND, CC-BY-NC, CC-BY-NC-ND or CC-BY-NC-SA).

Scope of the analysis and limitations

Articles from dates other than the ones specified were not considered (even if PubMed includes some out-of-date articles in its results). Only articles with a DOI were considered, and among them, there was a proportion not recognized by Unpaywall and thus, also not considered. Hence, the exclusion criteria after Unpaywall analysis includes out-of-date and those not scanned by Unpaywall (including papers without DOI).

Also, the Unpaywall system indexes thousands of institutional and subject repositories, but there are some still missing, and the database updates periodically, so some data might have changed.

Results

COVID-19 and SARS CoV-2 pandemic publications

The data obtained about SARS CoV-2 from January 1st to April 23rd 2020 are shown in Figure 1. In total, 6,223 articles were retrieved from PubMed. Of these 10 were from 2019, 182 did

Figure 1. PubMed-hosted SARS CoV-2 related papers published in Q1 of 2020 and their Open Access (OA) information. (a) Number of total and OA papers published during SARS CoV-2 pandemic. (b) Percentage of publications divided by their OA publishing mode. (c) Unpaywall-used source to obtain OA papers. (Data extracted from PubMed: 23rd April 2020).
not have a DOI assigned and 485 were not recognized by Unpaywall, and so were excluded from analysis; therefore, analysis was performed on a total of 5,611 articles.

From the data, it can be seen that the number of articles published during the selected period increases daily. Figure 1a shows that 88.8% (n=4,986) of articles were published as OA. Regarding the type of OA, 67.4% (n=3,359) are classified as Bronze OA, followed by Gold OA (21.5%), Hybrid journals (7.8%), and Green OA (3.3%) (Figure 1b). All these OA articles (n=4,986) were found by Unpaywall through different sources of information (Figure 1c), mostly (73.8%) as free articles (PDF or HTML). It is worth mentioning that 43% of the OA papers (n=2,414) have a copy in a repository, even if they are Gold, Hybrid or Bronze, which is known as shadowed Green documents.

In order to deeply analyse the OA situation, we also reviewed license information of all the OA papers. Figure 2 shows that most of these articles lack a license (76.4%). Most open licenses (CC, CC-BY and Public Domain (PD)) are present in 13% of the papers, while the most restrictive ones (CC-BY-NC, CC-BY-NC-ND, CC-BY-NC-SA, CC-BY-SA and CC-BY-ND) are represented by more than 10% of all the considered papers (Figure 2b). Publisher implied licenses (implied OA) are included as the more restrictive ones. From all licensed papers (n=1,175), 44.3% bear a restricted one. It is remarkable that 258 of the articles classified as Gold OA (24%) don’t bear any license.

Furthermore, the most frequent publishers and journals during this period in relation to SARS CoV-2 were studied. The most frequent publisher is Elsevier, who published ~30% of papers, followed by Wiley (13.6%) and Springer (10.7%) (Figure 3a). In terms of journals, The British Medical Journal (The BMJ), Journal of Medical Virology and The Lancet are those with the largest number of papers: 4.2, 3.1 and 2.2% of all analysed papers, respectively (Figure 3b).

Based on these results, we specifically studied the COVID-19-related articles published by Elsevier, Wiley and Springer (Figure 4). While Elsevier and Springer release almost all SARS CoV-based articles as OA (96.3%), Wiley retains 28.3% as closed access (Figure 4a). All three publishers publish the majority of their papers as Bronze OA (Figure 4b). Note that Elsevier is the only one (out of these three) that classifies more than 2% of its articles as Green OA (n=130; 8.1% of all OA papers). Elsevier has also published approximately 17% (n=274) of these documents as Gold OA, 1.25% and 12.1% more than Springer and Wiley, respectively. Looking at licensing, most of the OA publications from these publishers lack a license, being Springer the one with highest license number (24.3%) (Figure 4c). Regarding specific OA licensing, Springer publishes 89.9% of its licensed articles under CC-BY, Wiley does the same but with less than the half of its collection (44.4%) and Elsevier has the most restrictive conditions: 89.5% of the licensed papers carry CC-BY-NC-ND licenses (Figure 4d).

Publications about other coronaviruses and epidemics: SARS CoV-1 and MERS CoV
In order to compare the scientific production and OA publication during global health emergencies, both SARS CoV-1 and MERS CoV-related publications were studied using the PubMed database, taking into account the times for the beginning of each outbreak.

In the case of the SARS CoV-1 (Severe Acute Respiratory Syndrome CoronaVirus-1) epidemic, the first case was discovered in China during November 2002. We therefore analysed...
Figure 3. Publishers and journals that published the highest number of COVID-19-related papers hosted in PubMed in Q1 of 2020. Number and percentage of total publications distributed by most frequent publishers (a) and journals (b). (Data extracted from PubMed: 23rd April 2020).

publications published in 2003, 2004, 2005 and 2006 (Figure 5). For the period from 2003 to 2006, PubMed returned a total of 2,396 articles, of which, after exclusion criteria, 1,858 were considered (476 lacked DOI, 58 were out-of-date and 4 were not recognized by Unpaywall). There was an increase in the number of publications from 2003 to 2004, with a decline onwards. The percentage of OA publications increased from 80 to 87% in the first year, maintaining a stable average of 84% throughout the analysed period (Figure 5a). Among these open articles, 63.1% were published as Bronze OA, 19.6% as Green OA, 13.9% as Gold OA, and 3.3% as Hybrid journals (Figure 5b). From all the OA papers, almost 88.8% (1,389) lacked a license, including a high proportion (44.5%) of Gold OA papers.

Next we performed the searches for the MERS CoV (Middle East Respiratory Syndrome Coronavirus) epidemic, whose outbreak began in September 2012 in Saudi Arabia. A total of 1,129 papers were obtained for the specified period (2013 to 2016), of those 78 don’t have any DOI and Unpaywall did not recognize 8, giving as a result a total of 1,043 analysed articles. In this case, this number is significantly lower than the one found for SARS CoV-1 over time. In 2016, the year in which most papers are registered (n=345), the percentage of these published as OA remains constant and is very high, with an average of 93.5% (Figure 6a). Unlike SARS CoV-2 and SARS CoV-1, 44.3% of MERS-related OA publications were published as Gold OA (Figure 6b). From all the OA papers, 61.3% (n=598) lack a license, an important proportion corresponding to Gold OA papers (29.4% of Gold).

In order to determine if these results are a consequence of the current extraordinary circumstances, a control of the research was established through the analysis of open content of chronic diseases considered constant over time. We performed searches
Figure 4. Analysis of the three most frequent publishers with more Open Access (OA) COVID-19 papers hosted in PubMed in Q1 of 2020: Elsevier, Wiley and Springer. (a) Percentage of OA publications of the most relevant publishers: Elsevier, Wiley and Springer. (b) Distribution of their open content by Gold, Hybrid, Green or Bronze status. (c) Distribution of the licensed and non-licensed articles of the three publishers. (d) Distribution of the most frequent licensing type by each publisher: from the most restrictive licenses to the more open ones. (Data extracted from PubMed: 23rd April 2020).

Figure 5. Publications related to SARS CoV-1 epidemic hosted in PubMed from 2003 to 2006 and their Open Access (OA) indicators. (a) Number of total and OA publications about SARS CoV-1 epidemic during the first 4 years from the start of the epidemic. (b) OA category of the OA published articles. (Data extracted from PubMed: 19th April 2020).
for “low grade glioma” and “peptic ulcer”, which harbour similar output levels compared to SARS CoV-1 and MERS, obtaining a constant OA proportion for each case over the last 3 years (Figure 7). This rate is low for all cases, with an average of 55.1% and 51.5% for low grade glioma (Figure 7a) and peptic ulcer (Figure 7b), respectively. In addition, articles concerning both diseases were mostly published as Gold OA (Figure 7a and 7b). In these two cases, the number of OA articles without a license represents around 40%.

**Discussion and conclusion**

Compared to other emergency crises such as, SARS CoV-1 or MERS CoV epidemics, the number of published papers during the current COVID-19 pandemic is huge. Our study (based only on the PubMed database) reveals that in only four months, the number of these articles is 17-times more than the number of documents available in the first year in the case of SARS CoV-1, and 48-times in the case of MERS CoV. Shortening of acceptance rates by journals is giving rise to information overload both for the scientific community but also for society, making it difficult to ascertain what really has a significant scientific value and as a consequence may affect decision-making.

In addition to the massive scientific production, after the pandemic declaration, publishers have made, not only COVID-19 but also previous SARS CoV-1 and MERS CoV related papers, openly available. From our study, both SARS-like viruses share the same limited conditions, i.e. are non-licensed Bronze OA articles. On the contrary, a large number of MERS CoV-related papers present as Gold OA, suggesting high public funding from funders with OA policies during this period. In this context, it is surprising that there is a large number of Gold OA articles without licenses for all three diseases, which raises some uncertainties about whether some journals should still be listed in the DOAJ.

While Gold OA makes papers available immediately by the publishing journal itself, the predominant Bronze OA category, found by the present study, means that papers are freely hosted on publisher websites, without a license at all. Little is discussed in the OA literature about this category, but what is clear is that articles under this group without a categorised license do not allow extended reuse rights beyond reading. Thus, this “open” label removes rights to share or redistribute and, moreover, the publisher can revoke this access at any time. For instance, publishers’ announcements about their temporary fee drop on coronavirus-related research is limited only to the duration of the crisis (Springer Nature or Elsevier).

In line with this, this study found that PubMed-hosted COVID-19 papers that have a copy included in a repository almost reach 50% of OA papers; however, only 3% are assigned under Green OA status. This implies that many of these Bronze OA articles - around 60% - have a copy in the repositories searched by Unpaywall, which can be removed upon publisher request.

Another point to highlight, as defined by Piwowar et al., is the fact that many of these Bronze OA publications have been published in Hybrid journals. These papers, due to their accessibility, benefit from greater citation. It is not surprising that during this emergency situation, they are attracting the attention and curiosity of the entire world, including not only the scientific community but also non-scientific, increasing the citations and so the journals’ reputation. After publishers decide to
reinstate paywalls, as the majority of the documentation is not free all the time, the number of subscriptions might be affected, since it is possible that new non-subscribed readers obtained during this pandemic period have read articles from these journals and want to continue doing it.

What is most interesting about the effect of the COVID-19 emergency on scientific research disclosure is what it says about the current publication model: it fails when a critical need arises for fast data dissemination. Our analysis demonstrates that the current alternative that is in use falls short of expectations of being the best model, since this fast opening lacks basic OA principles, which are required in order to be transparent, reusable and good for the society. This could also have an important impact on a possible scenario where new outbreaks occur in the coming months or years.

We finally conclude that it seems clear that all stakeholders agree that Science only works when knowledge is shared. This unique and exceptional pandemic situation gives the opportunity to analyse the current publishing system in order to start doing things in a way that benefits the whole community, both researchers and society at large. This study has presented a part of Open Science-related issues and hopefully stimulates further research from the OA community regarding the use of Bronze OA and Hybrid journals.

Figure 7. Analysis of the number and OA properties of papers about two chronic diseases: low grade glioma and peptic ulcer. Number of publications, OA percentage and category of articles related to low grade glioma (a) and peptic ulcer (b) during 2017, 2018 and 2019. (Data extracted from PubMed: 20th April 2020).
Data availability

Underlying data


This project contains the following underlying data:
- Excel datafile with Unpaywall analysis of each research query.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements

Dimity Flanagan (Manager, Scholarly Communications, University of Melbourne) for her review and valuable suggestions.

References

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com