RoRI RESEARCH ON RESEARCH INSTITUTE

Scholarly Communication in Times of Crisis

The response of the scholarly communication system to the COVID-19 pandemic

Ludo Waltman, Stephen Pinfield, Narmin Rzayeva, Susana Oliveira Henriques, Zhichao Fang, Johanna Brumberg, Sarah Greaves, Phil Hurst, Andy Collings, Arianne Heinrichs, Nick Lindsay, Catriona J. MacCallum, Dan Morgan, Susanna-Assunta Sansone & Sowmya Swaminathan

DECEMBER 2021

About the authors

Ludo Waltman is Associate Director of RoRI and Professor of Quantitative Science Studies at the Centre for Science and Technology Studies (CWTS) at Leiden University.

0000-0001-8249-1752

Stephen Pinfield is Associate Director of RoRI and Professor of Information Services Management at University of Sheffield.

0000-0003-4696-764X

Narmin Rzayeva is PhD candidate at CWTS, Leiden University, working with RoRI, and Head of the Scientific Research Department at Azerbaijan University of Architecture and Construction.

0000-0003-0397-5412

Susana Oliveira Henriques is PhD candidate at CWTS, Leiden University, working with RoRI, and Head of the Central Library at Lisbon University Medical School.

0000-0002-0947-5083

Zhichao Fang is PhD candidate at CWTS, Leiden University, working with RoRI.

0000-0002-3802-2227

Johanna Brumberg was Program Director of RoRI and is Program Director at Volkswagen Foundation.

0000-0001-9403-4591

Sarah Greaves was Publishing Director at Hindawi and is now an independent publishing consultant.

0000-0001-7802-0261

Phil Hurst is Publisher at the Royal Society.

0000-0001-7971-1005

Andy Collings is Executive Editor of *eLife*.

0000-0001-9570-2061

Arianne Heinrichs is Editorial Director for Nature Portfolio and BMC journals at Springer Nature.

0000-0002-0061-9925

Nick Lindsay is Journals and Open Access Director at the MIT Press.

0000-0002-7442-6579

Catriona J. MacCallum is Director of Open Science at Hindawi.

0000-0001-9623-2225

Dan Morgan is Director of Communications and Community Relations at PLOS.

0000-0003-0625-2770

Susanna-Assunta Sansone is Associate Director of the Oxford e-Research Centre, Associate Professor at the Department of Engineering Science at University of Oxford, and Principal Investigator at FAIRsharing.

0000-0001-5306-5690

Sowmya Swaminathan is Head of Editorial Policy for the Nature Portfolio journals at Springer Nature.

0000-0002-0285-4910

Cover image by Vanessa Lovegrove.

DOI: 10.6084/m9.figshare.17125394

Contents

Executive summary

The COVID-19 pandemic has been the greatest global public-health crisis in a century. Scientific research and its effective communication have been at the centre of the worldwide response to the emergency. This report analyses how the scholarly communication system – involving the production, evaluation, and dissemination of research outputs – has responded to this crisis, focusing on the period until mid-2021. It evaluates ways in which the scholarly communication system, including its quality control mechanisms, has operated during the pandemic. It also examines how the global crisis has enabled innovations in scholarly communication, and the effects they have had on the system, or may have in future.

Early in the pandemic, a statement on **'Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak'** was issued by Wellcome and signed by 160 organisations worldwide, including research funders, publishers, infrastructure providers, and research institutions. The statement called on actors in the research system to implement a set of agreed principles. In addition, a group of publishers and related organisations came together and launched the **COVID-19 Rapid Review Initiative** in April 2020, to maximise the efficiency and speed of peer review of COVID-19 research.

Together the commitments made as part of these initiatives involve major change in scholarly communication, including making COVID-19 research outputs openly or freely accessible, preprinting of COVID-19 research, sharing data from COVID-19 research, speeding up publication times of COVID-19 articles, and facilitating peer review of COVID-19 preprints.

This report examines the extent to which these key commitments made at the beginning of the pandemic have been realised. The report has been written by a team comprising researchers, publishers, and other scholarly communication experts, all associated with the COVID-19 Rapid Review Initiative. It presents the results of research undertaken by the team and reviews research conducted by others. The report helps to identify opportunities for scholarly communication stakeholders to effect change that will extend beyond the pandemic and have long-lasting benefits.

The main findings show that **the scholarly communication system has been successful in making research openly or freely accessible**, with about 90% of peer-reviewed COVID-19 outputs accessible in these ways.

Especially in medical fields, the pandemic has led to an increased interest in posting articles on preprint servers before submitting them to a peer-reviewed journal. However, while COVID-19 research has been preprinted relatively often, **the proportion of peer-reviewed COVID-19 outputs with a preprint is still low**. This study identifies a preprint for just 5% of all peer-reviewed COVID-19 outputs.

Early and ongoing data sharing of the SARS-CoV-2 genome sequences has clearly been successful in combating the pandemic, but overall, **sharing of COVID-19 research data has remained relatively low**.

Many journals managed to speed up publication times. Analysis of the response of the journal peer review system to the pandemic yields a generally positive picture. On average, the time from submission to publication has been substantially shorter for COVID-19 articles than for similar non-COVID-19 articles.

With the impetus of the pandemic, **significant innovation has taken place in the area of peer reviewing preprints, but initiatives remain small-scale and experimental**.

The report offers a series of recommendations. It concludes that no one player has the solution to the major challenges faced by the scholarly communication system. **Improving scholarly communication is a joint responsibility that requires collaboration and coordinated action across stakeholders in the research system.**

The pandemic has illustrated the importance of openness – open access, open data, and open science more widely. The rapid and open sharing of the SARS-CoV-2 genomic sequencing data, along with opening up of the majority of the COVID-19 literature, are open science success stories of the pandemic. The experience of COVID-19 has further strengthened the case for more widespread adoption of open practices beyond the immediate crisis. Efforts to promote open science, particularly open access of published outputs and open sharing of data, need to be further intensified.

There is a need for more concerted action in the area of preprinting if larger-scale adoption is to be achieved. All stakeholders in the scholarly communication system can play a role in this, including mandating preprints, at least in the case of emergencies, and possibly more generally.

The prevalence of data sharing can be increased through joint efforts of key players. Across stakeholders, common data policy templates should be developed to require data sets and software to be posted to a trusted repository, and to require formal citations to data sets and software.

Additional investment in preprint peer review platforms are needed to scale up operations and to develop best practices for preprint peer review. Approaches to combining or integrating preprint peer review and journal peer review also need to be considered.

Efforts should be intensified to improve the availability and quality of data and metadata on scholarly publishing, allowing for robust evidence-informed approaches to innovation in scholarly communication.

It is hoped that this report will contribute to the ongoing discussion and debate about the future of scholarly communication, as we emerge from the pandemic.

1 An unprecedented challenge for scholarly communications

The COVID-19 pandemic has been the greatest global public-health crisis in a century. The worldwide medical emergency has put massive pressure on health systems, and necessitated the rapid development of new drugs and treatments. Many countries have responded by introducing national lockdowns, with profound social and economic consequences. Rapid decision-making, often with implications for millions of people, has been essential. Citizens, communities, and organisations at local, national, and international levels, public and private, have been affected, and having access to the best possible evidence in a timely way has been crucial. The role of science and scientists, as providers of such evidence, has been elevated in the public sphere in ways previously unseen.

The communication of scientific research results has, therefore, been at the centre of the global response to the crisis. The sharing of scientific findings has informed clinical decision making, drug development, policy making, and a host of other activities. Getting high-quality scientific work into the public domain as rapidly as possible has never been more important. The pandemic has illustrated the importance of an effective scholarly communication system that works both for the academy and for society more widely.

"The communication of scientific research results has been at the centre of the global response to the crisis"

The pandemic has put a strain on the research system in general, and on the scholarly communication system in particular. For the research system as a whole, challenges have included allocating funding to emergency projects investigating COVID-19 and building research capacity and infrastructure so that high-quality research can be undertaken quickly. From a scholarly communication point of view, the pandemic gave rise to what has been described as a "torrent" of submissions to journals across all subjects [16]. For COVID-19 research, it has been particularly challenging to ensure research outputs are disseminated widely and are peer reviewed rapidly by qualified scientific experts. The pandemic has given rise to innovation in scholarly communication to address some of these challenges, such as new ways to rapidly evaluate outputs prior to peer review by journals, and has raised the profile of existing innovations, including greater adoption of open science practices. In addition to open access to published outputs, open practices also include the sharing of data and software underpinning research, which have become crucial parts of the response to the pandemic.

1.1 Contribution of the report

The scholarly communication system is centred on the production, evaluation, and dissemination of the outputs of research. The system comprises a wide range of activities, undertaken using a complex set of infrastructures, involving different actors. This report analyses how the scholarly communication system has responded to the pandemic, focusing on the period until mid-2021. It also examines how the global crisis has enabled innovations in scholarly communication and the ef-

"This report analyses how the scholarly communication system has responded to the pandemic"

fects they have had on the system, or may have in future.

Peer-reviewed journals have been put under particular pressure to deal with COVID-19 research in a timely way. At the same time, other channels of scholarly communication have received more attention during the pandemic, partly because of the role they can play in easing strain on journals. Preprinting is a particularly relevant channel. Preprints are versions of research articles shared on open access preprint servers, typically before they have been peer reviewed by a journal. Some commentators have claimed that the pandemic has heralded a new era in preprinting [8]. This report examines ways in which preprints have contributed to scholarly communication during the pandemic, and also some of the challenges created by their use.

In addition, this report explores ways in which peer review processes have been expanded and adapted in response to the pandemic. These adaptations include attempts to speed up peer review processes for journals and other innovations in evaluation and quality assurance activities, notably the review of preprints, supplementing journal-based peer review.

All of these developments have implications for the ways in which the scholarly communication system may develop in the future. Based on the insights presented in this report, we draw out lessons about how the system can be improved for the future. A resilient scholarly communication system is needed, not merely to respond to future times of crisis, but also at all other times to serve the research community and society more broadly, in expanding the borders of human knowledge.

1.2 Wellcome-coordinated COVID-19 statement

Policy-based responses began early in the pandemic. These included statements from organisations such as the World Health Organization (WHO) and OECD, as well as national governments, and a wide range of other stakeholders. Perhaps the most prominent example of such a response in the research system was the statement on 'Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak', issued by Wellcome on 31 January 2020 and signed by 160 organisations worldwide, including research funders, publishers, infrastructure providers and research institutions.

The statement called on different actors in the research system to commit to a set of agreed principles:

> "We call on researchers, journals and funders to ensure that research findings and data relevant to this outbreak are shared rapidly and openly to inform the public health response and help save lives."

The key commitments can be summarised under the following three headings:

1. *Open or free access:* All COVID-19 publications should be made openly accessible, or made freely available at least for the duration of the pandemic

2. *Preprinting:* COVID-19 papers should be made available via open access preprint servers prior to publication in journals

3. *Data sharing:* Data from COVID-19 research should be shared openly as early as possible

On the first commitment, some publishers had already started making COVID-19 research freely accessible as early as January 2020, and by mid-March 2020, over 30 publishers announced their commitment to doing so. Often this sharing amounted to 'free access', where outputs are free to read at least for the duration of the pandemic. This is different from full 'open access', which also includes reuse rights in addition to free accessibility, not restricted to the duration of the pandemic.

"Some publishers had already started making COVID-19 research freely accessible as early as January 2020"

The statement described the second commitment, preprinting, as where:

> "research findings are made available via preprint servers before journal publication, or via platforms that make papers openly accessible before peer review."

This was clearly designed to accelerate the scholarly communication process, meaning the sharing of findings is not delayed by the often lengthy process of peer review carried out by journals. It could happen via the pre-existing infrastructure of preprint servers and similar platforms. The statement also committed to ensure "authors are clear that data or preprints shared ahead of submission will not preempt its publication in these journals". This commitment was designed to ensure publishers allow the posting of preprints and sharing of data without those negatively affecting the ability of authors to publish their work in journals subsequently, thus enabling early preprinting and data sharing to work effectively as part of accelerating the dissemination of research findings.

In relation to the third commitment above, data sharing, the statement called for all data and accompanying "protocols and standards used to collect data" to be shared to provide evidence of their credibility and to enable reuse by third parties. The statement also specified that research papers should include clear statements of data availability, enabling shared data to be located.

The statement also included the specific commitment that the outputs of research, including data and published articles, should be shared with "public health and research communities and the WHO", to ensure they could be used in combatting the pandemic in a timely way.

Finally, the statement specified the intention to apply the above commitments not only to the COVID-19 pandemic but also to similar outbreaks in the future.

This report provides an initial analysis of how the commitments made in the Wellcome-coordinated COVID-19 statement have been implemented, covering the period from the outbreak of the crisis to the middle of 2021. In Box 1.1, a number of publishers that have contributed to this report explain the steps they have taken to implement these commitments.

Box 1.1: Case study – How did publishers implement the commitments made in the Wellcome-coordinated COVID-19 statement?

eLife

By Andy Collings

eLife implemented its commitment to the Wellcome-coordinated COVID-19 statement as follows:

- *Ensuring all peer-reviewed research publications relevant to the outbreak are made immediately open access, or freely available at least for the duration of the outbreak:*
  - By default eLife publishes all research articles under a CC-BY licence and deposits published articles to PubMed Central.
- *Ensuring research findings relevant to the outbreak are shared immediately with the WHO upon journal submission, by the journal and with author knowledge:*
  - Early in the pandemic, in the email sent to authors after submission to eLife, information was added to inform authors that eLife supports the Wellcome coordinated COVID-19 statement and that submissions relevant to COVID-19 are shared with the WHO.
  - For an initial period, eLife manually sent COVID-19 submissions to the WHO.
- *Ensuring research findings are made available via preprint servers before journal publication, or via platforms that make papers openly accessible before peer review, with clear statements regarding the availability of underlying data:*
  - Authors can post a preprint to bioRxiv or medRxiv and then transfer for consideration by eLife.
  - In March 2020, eLife made preprints the default for all research submissions [15].
  - From July 2021, eLife only peer reviews research articles posted by the authors on a preprint server, and facilitates deposit of preprints if the authors have not already done so during the full submission.
  - All eLife research articles include a data availability statement; research articles should also adhere to TOP guidelines level 2 for data transparency, requiring data to be posted to a trusted repository and exceptions to be identified at article submission.

Hindawi

By Catriona J. MacCallum and Sarah Theissen

Hindawi implemented its commitment to the Wellcome-coordinated COVID-19 statement as follows:

- *Ensuring all peer-reviewed research publications relevant to the outbreak are made immediately open access, or freely available at least for the duration of the outbreak.* As a fully open access publisher, all articles published in our 200+ journals are already automatically made publicly available under a CC-BY licence. The abstracts and references of every paper are also submitted as metadata to Crossref, and made publicly available via the Crossref API to reduce barriers to machine access and discovery. For the first 12 months of the pandemic, we waived APCs for all COVID-19 research articles.
- *Ensuring research findings relevant to the outbreak are shared immediately with the WHO upon journal submission, by the journal and with author knowledge.* We sent an email to authors of relevant articles informing them that we would share their article with the WHO, in accordance with the Wellcome-coordinated statement. We sent articles to the WHO until they asked us to stop as they were inundated with articles.
- *Ensuring immediate sharing of research findings and data.* Hindawi already requires that authors of all published articles include a data availability statement explaining if and how their data and code are made publicly available. Our data policy covers most of STEM, including mathematics, engineering, chemistry and biomedicine. Because of disciplinary differences in data sharing norms, our policy allowed data to be made available on request. In response to the Wellcome-coordinated statement and our participation in the COVID-19 Rapid Review Initiative, we strengthened this policy for COVID-19 articles specifically to mandate data deposition in a public repository. This required additional changes to our workflow across all journals, which is still ongoing.
- *Ensuring research findings are made available via preprint servers before journal publication.* In our emails to authors of any relevant submitted research article, we strongly encouraged them to make their article immediately available as a preprint.

In addition, Hindawi staff also implemented a fast-track process for identifying and screening COVID-19 submissions and prioritising these articles. This included, for example, rapidly identifying article types not accepted by our journals, such as letters and opinion articles. These were triaged by in-house staff and desk-rejected quickly so that authors could submit them elsewhere without delay. We also created a dedicated editorial workflow,

which included bespoke template letters for editors and reviewers of any relevant COVID-19 article asking them to prioritise the review process where possible, and an additional approval and quality check by in-house staff of all accepted articles.

PLOS

By Dan Morgan

PLOS implemented its commitment to the Wellcome-coordinated COVID-19 statement as follows:

- *Ensuring all peer-reviewed research publications relevant to the outbreak are made immediately open access, or freely available at least for the duration of the outbreak.*

  All PLOS content is always openly available, reusable, ready for text- and data-mining, and pushed to PubMed Central as standard. PLOS content is therefore always ready to support a global crisis, such as the COVID-19 pandemic, among the myriad other important benefits of open access.

- *Ensuring research findings relevant to the outbreak are shared immediately with the WHO upon journal submission, by the journal and with author knowledge.*

  PLOS sent all COVID-19 articles to the WHO and notified authors about this. In the notification email, authors who had not yet preprinted their work were encouraged to post a preprint.

- *Ensuring research findings are made available via preprint servers before journal publication, or via platforms that make papers openly accessible before peer review, with clear statements regarding the availability of underlying data.*

  PLOS journals always require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data. PLOS strongly encourages preprint posting for all articles and directly facilitates preprint posting if authors opt in and if bioRxiv is the target. (Further connections are planned in 2022, for instance for medRxiv.). PLOS cannot directly facilitate the preprint posting if authors did not opt in or if there was no established partnership with the preprint server.

- *Ensuring authors are clear that data or preprints shared ahead of submission will not pre-empt its publication in these journals.*

  PLOS strongly and publicly supports preprint posting. On our website we state that "authors are strongly encouraged to share their research on preprint servers and to provide links to their preprint during submission, and editors are advised to use comments posted on preprints and preprint reviews where applicable".

Early in the pandemic, we published a blog post introducing the above measures, along with others such as fast-tracking COVID-19 research and encouraging reviewers to carefully consider the need for additional experiments or analyses in revision requests.

Royal Society

By Phil Hurst

The Royal Society implemented its commitment to the Wellcome-coordinated COVID-19 statement as follows:

- *Ensuring that primary research articles are freely available.* All COVID-19 articles published since the start of the pandemic across the Royal Society portfolio are made open access under a CC-BY licence. We plan to continue to publish new COVID-19 content in this way at least until the pandemic is over. Our collection of COVID-19 content is made available directly from the publishing homepage. In addition, this and the available data supporting it, is immediately accessible in PubMed Central.
- *Ensuring immediate sharing of research findings and data.* As with all submissions to Royal Society journals, it is a condition of publication that authors make available the data, code and research materials supporting the results in the article. We joined the COVID-19 Rapid Review Initiative committing that all COVID-19 submissions include a mandatory data availability statement.
- *Ensuring research findings are made available via preprint servers before journal publication.* We strongly encouraged authors to post a preprint at submission if they had not done so already. Several of our journals are integrated with bioRxiv allowing direct submission to them at the time of posting the preprint.

Springer Nature

By Sowmya Swaminathan

Building on actions that had already been taken unilaterally in support of the global research efforts addressing the pandemic, Springer Nature was a launch signatory to the Wellcome-coordinated COVID-19 statement [9]. As such we committed to the following actions:

- *Ensuring that primary research articles are freely available.* All COVID-19 articles published across Springer Nature that were not already published open access were systematically identified and made freely available (without a CC-BY licence and without incurring an APC) through journal websites. This policy was limited to primary research journal articles. Although it did not include books, we provided nearly 3,000 book chapters with relevance for COVID-19 to PubMed Central.

We also made COVID-19 content available by depositing the full text in PubMed Central, resulting in a corpus that now includes more than 65,000 published works. Metadata provision via Crossref also raised the visibility of this content and facilitated publisher-agnostic searching of the freely available articles.

- *Ensuring immediate sharing of research findings and data.* Springer Nature authors with COVID-19 research articles were made aware of our commitment to the Wellcome-coordinated statement and our expectation that preprints, data, code, and protocols where available should be shared immediately through repositories. Authors whose articles went out for review were provided information about these requirements and offered support through the Springer Nature Research Data Helpdesk for depositing and curating data, and support for preprint sharing through our *In Review* service, which provides preprint deposition to the Research Square preprint server in parallel with journal peer review. Data availability statements are required in published papers across many Springer Nature journals, including all Nature journals, Communications journals, and BMC journals. Authors whose articles were declined prior to peer review were also encouraged to share research findings, data, and other outputs.

  While all Springer Nature journals publishing research relevant to COVID-19 put in place processes to ensure that these requirements were communicated to authors, systematically monitoring author compliance was more challenging and varied across our 3,000+ journals, as efforts and resources were focused as a priority on providing a rapid and constructive editorial, peer review, and publication process.

- *Ensuring that a copy of all COVID-19 articles was submitted to the WHO.* We put in place a process for ensuring that a copy of every COVID-19 research article received and peer reviewed at a Springer Nature journal was submitted to the WHO. The process for sending articles ended on 26 November 2020, by mutual agreement with the WHO. Authors were notified and given 24 hours to respond if they had concerns about sharing with the WHO.

1.3 COVID-19 Rapid Review Initiative

Partly in response to the Wellcome-coordinated statement, many organisations in the research system took steps in the first half of 2020 to alter their policies and processes. A group of publishers and related organisations came together to set up the COVID-19 Rapid Review Initiative, launched in April 2020, which grew to eventually include 22 members (see Appendix A) [29]. The group consisted of a wide range of publishers, including both large commercial publishers and smaller not-for-profit organisations, fully open access publishers and hybrid publishers, plus other organisations in the scholarly communi-

cation space, such as preprint servers and preprint peer review platforms. These organisations came together partly to agree how policies, such as the Wellcome-coordinated COVID-19 statement, could best be implemented. The Research on Research Institute (RoRI) also joined the COVID-19 Rapid Review Initiative. As scientific partner in the initiative, RoRI played a leading role in the preparation of the present report and in the design and execution of the methods and analyses within it.

The COVID-19 Rapid Review Initiative was launched with a call to action targeted specifically at "reviewers, editors, authors, and publishers in the research community" aiming to "maximize the efficiency and speed of the triage and peer review process of COVID-19 research". It endorsed the Wellcome-coordinated statement, including its general call for open or free access, and proposed a number of specific actions, focused partly on implementing the commitments made in the Wellcome-coordinated statement and partly on accelerating peer review. The aims of the initiative can be summarised under four headings:

"The COVID-19 Rapid Review Initiative was launched with a call to action targeted at reviewers, editors, authors, and publishers in the research community"

1. *Preprinting:* Ensure articles submitted to a journal have a preprint
2. *Data sharing:* Ensure authors of articles share their data
3. *Speeding up publication times:* Speed up the time from submission to publication in a journal by creating a pool of reviewers ready to respond quickly and by enabling transfer of articles and reviews between publishers
4. *Encouraging peer review of preprints:* Enable early evaluation of papers, which could identify important work and inform triage of journal submissions

With regard to the first aim, the members of the initiative promoted the use of preprints in COVID-19 research highlighting a number of established preprint servers.

Sharing data, the second aim, was seen as important in speeding up research, enabling the verification of its robustness, and ensuring the reproducibility of results. Rapid Review Initiative partners worked with FAIRsharing to ensure partner policies were registered, and, at the same time, mandated data availability statements from participating organisations.

Under the third aim, the initiative established a pool of over 2,000 researchers from 110 countries to act as rapid reviewers. Publishers could use these reviewers to accelerate peer review of COVID-19 research. The initiative also set up a process to transfer rejected submissions and the associated reviews between publishers.

Finally, under the fourth aim, peer review of preprints, the initiative supported Outbreak Science PREreview, a platform enabling peer review of COVID-19-related preprints, and *Rapid Reviews: COVID-19*, an 'overlay' journal set up by MIT Press to perform peer review of selected COVID-19 preprints.

Not all scientific publishers or scholarly communication service providers signed up to the Wellcome-coordinated statement or the COVID-19 Rapid Review Initiative. Those that signed up to the former represent

a large proportion of the global scientific literature, but there are still a large number who were not involved. Those who participated in the latter were a smaller subset of publishers, with others undertaking a variety of separate actions. More research is needed, however, on why other organisations did not become involved in these coordinated initiatives. It seems likely that a whole range of possibilities came into play, including lack of awareness of the initiatives, insufficient capacity to engage, and disagreement with the approaches proposed.

1.4 Overview of the report

"This report examines the implementation of the key commitments made in the Wellcome-coordinated COVID-19 statement and the COVID-19 Rapid Review Initiative"

This report examines the implementation of the key commitments made in the Wellcome-coordinated COVID-19 statement and the COVID-19 Rapid Review Initiative combined. These commitments represent a significant set of developments involving major change in the scholarly communication system, either by accelerating pre-existing developments or by initiating new ones:

- Making COVID-19 research outputs open or free access
- Preprinting of COVID-19 research
- Sharing data from COVID-19 research
- Speeding up publication times of COVID-19 articles
- Facilitating peer review of COVID-19 preprints

This report explores the extent to which these commitments have been realised. In Chapter 2, we examine the dissemination of COVID-19 research, addressing the question of how COVID-19-related research outputs have been disseminated in terms of open or free access, preprinting and data sharing. In Chapter 3, we investigate how COVID-19 research has been handled by the journal peer review system, including the time taken to publish articles. Chapter 4 explores initiatives to enable peer review of preprints, how these have worked during the pandemic, and what they might mean for the future of scholarly communication. In Chapter 5, we deal with the level of attention given to COVID-19 research in subsequent scientific research, in news media, and on social media. Based on the findings presented in the report, we conclude by making recommendations to publishers, scholarly communication organisations, funders, and other stakeholders in Chapter 6.

This report has been written by a team comprising researchers, publishers, and other scholarly communication experts, all associated with the COVID-19 Rapid Review Initiative. It presents the results of research undertaken by the team, and it reviews research conducted by others. By drawing on evidence from a variety of sources, we hope to offer a rich picture of the complex system under investigation. As well as carrying out scientometric analyses, drawing on a variety of data sources, we have also compiled case studies of specific developments. We have additionally obtained relevant insights from a survey of authors of COVID-19 preprints. The survey addresses many of the key questions studied in our report, and various chapters of the report will therefore be informed by results obtained from the

survey. The methodology of the survey is discussed in Box 1.2. Similar boxes will be presented in later chapters of the report to discuss other methodological aspects of our work.

Since the pandemic has not yet subsided, and it will take time before the long-term impact of the pandemic is clear, our report will need to be followed by other studies which provide more detail and further longitudinal analysis. In many cases the data we have worked with has limitations, and we explain these as we go along. These limitations highlight the need for additional data gathering and ongoing research and dialogue in this field. We suggest areas for further work as we present our own findings.

We also call for further action. Our report is designed to help stakeholders in the scholarly communication system to effect change that will extend beyond the pandemic and have long-lasting benefits, creating a more open, diverse and effective system. We hope this report can make a contribution to the ongoing discussion and debate about the future of scholarly communication, as we emerge from the pandemic.

"Our report is designed to help stakeholders in the scholarly communication system to effect change that will extend beyond the pandemic and have long-lasting benefits"

Box 1.2: Methodological note – Survey of authors of COVID-19 preprints.

Our survey of authors of COVID-19 preprints was designed to enable insight into the experience of authors of COVID-19 preprints in relation to preprinting, journal submission and peer review. It consisted of three main sections:

- Experience of preprinting and the feedback received: including motivations for preprinting, type of feedback received on the preprint, changes made to the paper as a result of feedback etc.
- Experience of journal submission and the peer review process: including motivations for submitting to particular journals, experience of the journal peer review process, changes made to the paper in response to peer review etc.
- Demographic questions: including country of the researcher, type of institution, experience in research etc.

The survey consisted of a total of 36 questions, although the number of questions answered by any individual participant varied depending on which 'pathway' they took through the survey based on a number of factors, such as whether or not their paper had been published in a journal.

Ethical approval to carry out the survey was granted by the Ethics Review Committee of the Social Sciences of the Faculty of Social and Behavioural Sciences of Leiden University. The online survey was created using the Qualtrics software. It was piloted by two journal publishers, two preprint service providers, and six researchers before being finalised. In addition to quantitative data, the survey collected a substantial amount of qualitative responses to open questions and comments boxes.

Most survey questions related to the author's experience associated with a particular COVID-19 preprint. This preprint was identified in a targeted email sent to the corresponding author. We extracted information on COVID-19 preprints and their corresponding authors from arXiv, bioRxiv, chemRxiv, and medRxiv. Invitations were sent by email to a total of 12,230 corresponding authors. The survey was open from 20 May 2021 to 14 July 2021.

A total of 673 responses were received, a response rate of 6%. Of the respondents, 516 (77%) described their gender as "man", 137 (20%) as "woman", 16 (2%) responded "prefer not to say", and 4 (1%) "prefer to self-describe" (free-text responses included "non-binary" and "genderfluid").

Responses were received from researchers with different levels of experience in conducting research. 126 (19%) had experience of up to 5 years, 381 (57%) had 6–25 years of experience, and 162 (24%) had 25 years or more experience (with 4 respondents answering "not applicable"). As we sent the invitation email to addresses of corresponding authors, it is not surprising that most responses were from researchers with at least 6 years of experience.

Responses were received from a total of 78 countries. Unsurprisingly, the country with the highest number of respondents was the USA, with 131 (19%) responses. 93 (14%) responses were received from the UK, 70 (10%) from India, and 33 (5%) from Brazil. 283 (42%) responses were received from Europe, 169 (25%) from North America, 131 (19%) from Asia, 51 (8%) from South America, 21 (3%) from Africa, and 15 (2%) from Australasia (with 3 not disclosing their country).

Most respondents, 424 (63%), were based in universities or colleges, with a further 91 (14%) in hospitals or medical schools. Smaller numbers were based in other organisation types, comprising public research organisations (56, 8%), governments (28, 4%), industrial/commercial organisations (27, 4%), non-governmental organisations (17, 3%), and "other" kinds of organisations (30, 4%).

The survey data is available in figshare [51], except for the free-text responses, which may contain sensitive information.

2 Accessibility, preprinting and data sharing

Highlights

- Our methodology for identifying COVID-19 outputs yields an estimated number of 150,000 peer-reviewed COVID-19 outputs published between January 2020 and April 2021. In addition, an estimated 40,000 COVID-19 preprints were posted in this period.
- The commitment made at the beginning of the pandemic to make COVID-19 research openly or freely accessible has been largely fulfilled. About 90% of all peer-reviewed COVID-19 outputs are openly or freely accessible.
- The commitment made at the beginning of the pandemic to post COVID-19 research on a preprint server before it appears in a journal, or in some other peer-reviewed outlet, has not been fulfilled. Only a small share of all COVID-19 research was posted on a preprint server. We were able to identify a preprint for 5% of all peer-reviewed COVID-19 outputs (but the actual percentage of peer-reviewed COVID-19 outputs that have a preprint is likely to be a little higher).
- The commitment made at the beginning of the pandemic to share COVID-19 data sets has not been fulfilled, although the efforts to share genome sequencing data are a notable exception to this. Only a small share of the COVID-19 preprints and journal articles made data openly available.

As discussed in Chapter 1, a large number of organisations signed the COVID-19 statement coordinated by Wellcome. These organisations made three key commitments: 1. Making COVID-19 outputs open or free access; 2. Preprinting of COVID-19 research; 3. Sharing of data from COVID-19 research. Have these commitments been realised? This chapter addresses this question by presenting an in-depth scientometric analysis, combined with insights from our survey of COVID-19 preprint authors and from other studies reported in the literature.

Box 2.1: Methodological note – Scientometric analysis.

The scientometric analyses presented in this chapter are based on data from the Dimensions database, made available to us by Digital Science in May 2021. This database covers all scientific outputs that have a Digital Object Identifier (DOI) registered at Crossref, as well as various other types of scientific outputs, for instance outputs that have a record in PubMed and outputs posted on arXiv. These scientific outputs include not only peer-

reviewed work such as journal articles, conference papers, and book chapters, but also non-peer-reviewed preprints on a large number of preprint servers. In addition to data from Dimensions, we also use data from Crossref and Unpaywall.

Some outputs do not fall neatly into one specific category. Outputs made available in SSRN are sometimes classified as journal articles, but we classify them as preprints, since SSRN operates in a similar way to a preprint server. *F1000Research* and its various sister platforms offer an approach to publishing that combines features of preprint servers and journals. Outputs published on these platforms cannot easily be classified as either a preprint or a journal article, and we therefore do not include these outputs in our analyses. Box 2.2 offers a more detailed discussion of COVID-19 research published on F1000 platforms and the post-publication peer review process operated by these platforms.

An output is classified as a COVID-19 output if it was published in 2020 or 2021 and if at least one of the following terms is included in its title or abstract:

- coronavirus
- covid-19
- sars-cov
- ncov-2019
- 2019-ncov
- hcov-19
- sars-2

The publication date of an output is defined as the date on which the output was published online. The official publication date of a journal issue is not taken into consideration. If the online publication date of an output is not available in the Dimensions database, the approach of Fraser et al. [19] is followed and the date at which the DOI of the output was created in Crossref is used as a proxy of the online publication date. If this date is not available either, the online publication date of the output is considered unknown. This means that the output is excluded from most of the analyses presented in this chapter.

An alternative data source for analysing COVID-19 outputs is the COVID-19 Open Research Dataset (CORD-19) [52]. CORD-19 was created by the Semantic Scholar team at the Allen Institute for AI in collaboration with a number of other organisations. It provides metadata and full-text data for a very large number of COVID-19 outputs. We decided not to use CORD-19 because we are interested not only in analysing COVID-19 outputs but also in making comparisons with non-COVID-19 outputs.

Box 2.2: Case study – F1000: Post-publication peer review of COVID-19 research.

F1000 operates a number of platforms: *F1000Research*, *Wellcome Open Research*, *Gates Open Research*, and several others. An article submitted to one of these platforms first goes through a number of pre-publication checks performed by F1000's editorial team. If the article passes these checks, it is immediately published on the platform. The article is open access under a CC-BY licence. Data underlying the results are also shared in line with the FAIR data principles and are uploaded to a repository under an open licence. After the article has been published, it goes through an open peer review process. Peer review reports are published on the platform and are linked to the article. The names of the reviewers are published as well. If the article is approved by two reviewers, possibly after one or more rounds of revisions, it is considered to have passed peer review. The article is then indexed in databases such as PubMed and Scopus.

The F1000 platforms perform peer review after publication of an article rather than before. This post-publication peer review approach contrasts with the pre-publication peer review approach taken by most journals. Essentially, before an article has passed peer review on an F1000 platform, its status is similar to a preprint on a preprint server. After an article has passed peer review, its status is similar to an article published in a regular journal.

Like many other publishers, F1000 has signed the Wellcome-coordinated COVID-19 statement, including the commitment to preprinting of COVID-19 research. Since all articles published on an F1000 platform initially have a status similar to a preprint, F1000 by definition fulfills the commitment. As will be shown in Section 2.4, this is quite different for other publishers, for which it has been challenging to fulfill the commitment.

Between January 2020 and April 2021, 364 COVID-19 articles were published on F1000 platforms. 50% of these articles were published on *F1000Research*, 31% on *Wellcome Open Research*, and the rest on other platforms. Of the 364 COVID-19 articles published on F1000 platforms, 178 (49%) had passed peer review at the time of our analysis. According to statistics provided to us by F1000, for COVID-19 articles that have passed peer review, the peer review process typically took between two and three months (excluding the time required for pre-publication checks). 182 COVID-19 articles (50%) had not passed peer review at the time of our analysis. Four COVID-19 articles are editorials and were not subject to peer review.

2.1 Journal articles and other peer-reviewed outputs

We start by providing an overview of the peer-reviewed COVID-19 literature. This literature includes journal articles and other peer-

reviewed outputs, such as conference papers and book chapters. It does not include preprints, which have not been peer reviewed and which we analyse separately in Section 2.3.

Box 2.3: Methodological note – How to distinguish between peer-reviewed and non-peer-reviewed literature.

The Dimensions database covers both peer-reviewed and non-peer-reviewed content. Content on preprint servers has gone through basic quality checks performed by the preprint server, but preprint servers do not perform more in-depth forms of peer review. We therefore consider preprints to be non-peer-reviewed content. Preprints will be considered in more detail in Sections 2.3 and 2.4.

For content in journals, conference proceedings, and books, it is not always clear whether it has been peer reviewed or not. In addition to peer-reviewed research articles, journals may for instance also contain non-peer-reviewed content such as editorials, corrections, and news items. The metadata of content in journals, conference proceedings, and books typically does not indicate whether the content has been peer reviewed, and the Dimensions database therefore does not provide this information. As a proxy for peer-reviewed content, we consider only content that has an abstract in our analysis.

Our methodology identified about 150,000 peer-reviewed COVID-19 outputs that were published between January 2020 and April 2021. Figure 2.1 shows the monthly number of peer-reviewed COVID-19 outputs per broad disciplinary area. Outputs were assigned to these disciplinary areas by clustering them based on citation links [50] and by assigning the resulting clusters to disciplinary areas based on the journals in which outputs were published. As expected, a large majority of the COVID-19 outputs can be found in the biomedical and health sciences. There are also a substantial number of COVID-19 outputs in the social sciences and humanities. The physical sciences and engineering have only a small share of the COVID-19 outputs.

"Our methodology identified about 150,000 peer-reviewed COVID-19 outputs that were published between January 2020 and April 2021"

Figure 2.2 shows the monthly number of peer-reviewed COVID-19 outputs per broad disciplinary area as a percentage of the total monthly number of peer-reviewed outputs in a disciplinary area. In all disciplinary areas, the percentage of COVID-19 outputs has increased over time during the pandemic. In the most recent months, about 8% of the peer-reviewed outputs in the biomedical and health sciences were about COVID-19. This was the case for almost 6% of the peer-reviewed outputs in the social sciences and humanities. In the physical sciences and engineering, only about 1% of the peer-reviewed outputs dealt with COVID-19.

Figure 2.3 presents a VOSviewer term map showing the main topics covered by peer-reviewed COVID-19 outputs. The map includes the most relevant terms extracted from the titles and abstracts of COVID-19 outputs. The size of a term reflects the number of COVID-19 outputs in which the term occurs. The distance between two terms approximately indicates the relatedness of the terms, determined by the number of COVID-19 outputs in which the terms occur together. In general, the larger the number of COVID-19 outputs in which two

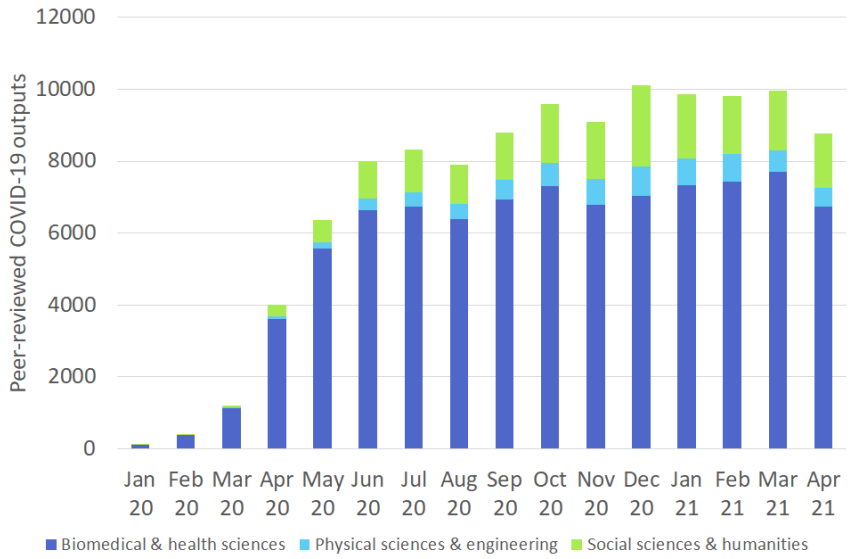


Figure 2.1: Monthly number of peer-reviewed COVID-19 outputs per broad disciplinary area.

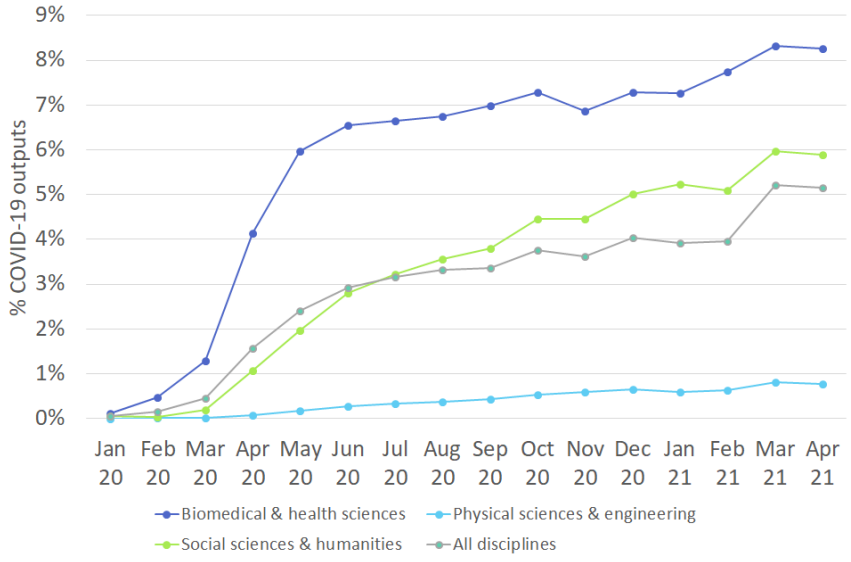


Figure 2.2: Monthly percentage of peer-reviewed COVID-19 outputs per broad disciplinary area.

terms co-occur, the smaller the distance between the terms. Based on their relatedness, the terms have been algorithmically grouped into six clusters, each representing a broad topic studied in the literature. The colour of a term indicates the cluster to which the term belongs.

The left side of the term map in Figure 2.3 shows biomedical research on COVID-19. Life sciences research is located in the top left (i.e., the light green cluster) and medical research in the bottom left (i.e., the light blue cluster). The right side of the map shows COVID-19 research in the social and behavioural sciences (i.e., the dark blue and orange clusters).

Figure 2.4 presents the same term map as Figure 2.3, but colours now indicate how the topical focus of peer-reviewed COVID-19 outputs has shifted over time. Blue terms occur relatively often in older

"In the most recent months, about 8% of the peer-reviewed outputs in the biomedical and health sciences were about COVID-19"

COVID-19 outputs. Red terms occur relatively often in more recent COVID-19 outputs. The term map shows a gradual shift from biomedical topics, which dominated the early literature on COVID-19, towards topics in the social and behavioural sciences, which play a more prominent role in the more recent COVID-19 literature.

"The term map shows a gradual shift from biomedical topics towards topics in the social and behavioural sciences"

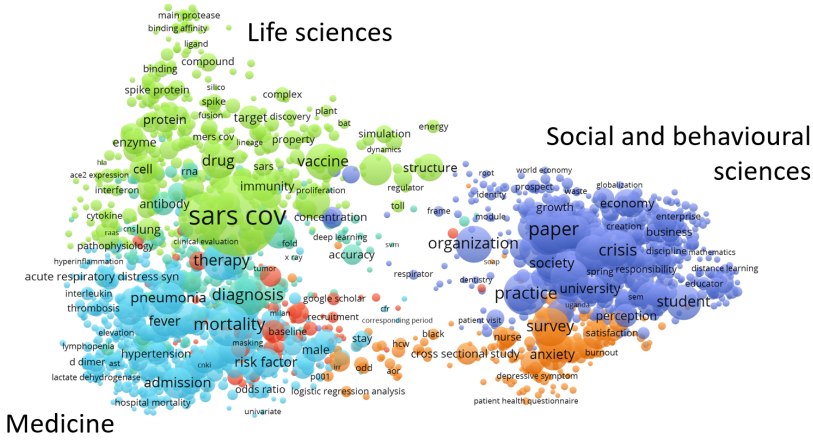


Figure 2.3: Term map showing the main topics covered by peer-reviewed COVID-19 outputs.

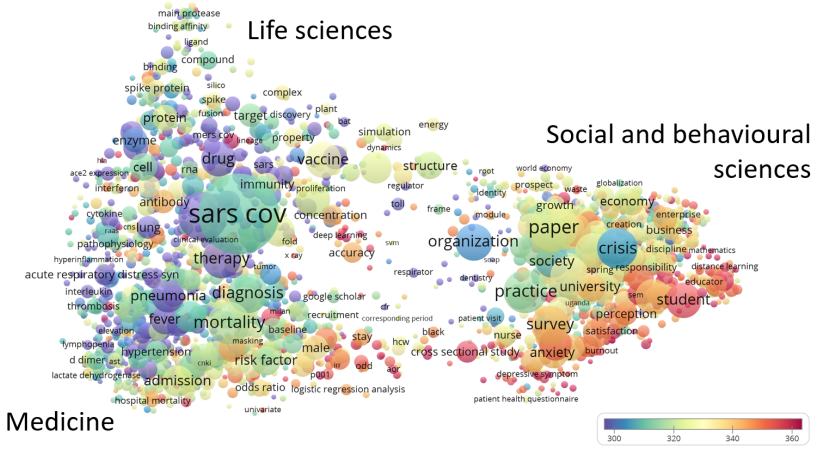


Figure 2.4: Term map showing the development over time of the topical focus of peer-reviewed COVID-19 outputs. The colour of a term reflects the average publication date of the outputs in which the term occurs (expressed in number of days after 1 January 2020).

2.2 Has the commitment to open or free access been fulfilled?

At the beginning of the pandemic, publishers and other organisations made a commitment to make all COVID-19 research outputs available in an open or free access form. Has this commitment been fulfilled? In our survey of authors of COVID-19 preprints, 87% of the respondents whose work had already been published in a journal stated that the final version of their work is openly accessible, either in the

journal or in a repository. This suggests that to a large extent the commitment to open or free access has been fulfilled. However, to obtain a more comprehensive understanding of the open or free accessibility of the COVID-19 literature, we look at some scientometric statistics.

Box 2.4: Methodological note – How to distinguish between different forms of open or free access.

There are different ways in which research can be made openly or freely accessible and different ways in which the various forms of open or free access can be classified. We have followed the classification introduced by Piwowar et al. [41]. This classification distinguishes between four forms of open or free access, which are all labeled 'open access' by Piwowar et al.:

- *Gold open access:* Outputs that have been published in an open access journal.
- *Hybrid open access:* Outputs that have been made openly accessible in a subscription-based journal with a licence that allows reuse (e.g., a Creative Commons licence).
- *Bronze open access:* Outputs that have been made freely accessible in a subscription-based journal without a licence that allows reuse.
- *Green open access:* Outputs that are not openly or freely accessible in a journal, but that are openly accessible on a preprint server or in a repository.

To determine the open access status of a COVID-19 output, we used data from Unpaywall.

These categories are now widely used for analysis, and are helpful in our understanding of the open access landscape. We do, however, recognise that they are contested. In particular, if open access is understood as being not only freely accessible but also reusable, bronze open access outputs should not be called 'open access', but should instead be referred to as 'free access'. Importantly, there is no guarantee that bronze open access outputs will remain freely accessible when the pandemic has ended. Hybrid open access is sometimes defined as a subset of gold open access, where gold is defined broadly as an output published open access in a journal. However, it is pragmatically useful to distinguish between gold and hybrid open access in order to inform a wider understanding of the current shape of the scholarly communication environment.

To avoid complexity, there is no overlap between the categories. The green open access category excludes outputs that are also available in gold, hybrid, or bronze form. So an open access version of an output available in a repository is classified as green open access only if the version in a journal is closed.

Figure 2.5 shows a breakdown by open or free access status for the monthly number of peer-reviewed COVID-19 outputs. Figure 2.6 shows

"Overall, 88% of the peer-reviewed COVID-19 outputs are openly or freely accessible"

the same breakdown for 19 selected publishers (i.e., the 15 publishers with the largest number of COVID-19 outputs and four additional publishers participating in the COVID-19 Rapid Review Initiative). Except for IEEE, these publishers have all signed the Wellcome-coordinated COVID-19 statement.

Overall, 88% of the peer-reviewed COVID-19 outputs are openly or freely accessible. More specifically, 44% of the COVID-19 outputs are gold open access. These outputs have been published in an open access journal. 10% of the COVID-19 outputs are hybrid open access. They have been made openly accessible in a subscription-based journal. Bronze open access outputs, which represent 28% of the COVID-19 outputs, have been made freely accessible in a subscription-based journal. These outputs do not have a licence that allows reuse and there is no guarantee that they will remain freely accessible when the pandemic has ended. 6% of the COVID-19 outputs are green open access. According to our data, these outputs are not openly or freely accessible in a journal, but they are openly accessible on a preprint server or in a repository. The final version of these outputs may for instance be openly accessible in PubMed Central or an earlier version may be openly accessible on a preprint server.

12% of the peer-reviewed COVID-19 outputs are not openly or freely accessible. A manual examination of a small random sample of these outputs showed that some of them have been misclassified. They do not represent peer-reviewed scientific content, or they have been incorrectly classified as closed instead of open or free. Other outputs mention COVID-19 briefly in their abstract but are not really about the pandemic. However, we also found outputs that clearly represent peer-reviewed scientific content dealing with COVID-19 but that have not been made openly or freely accessible. This goes against the commitment made to open or free access to COVID-19 research, although some of these were published by publishers who did not commit to making their outputs accessible in this way.

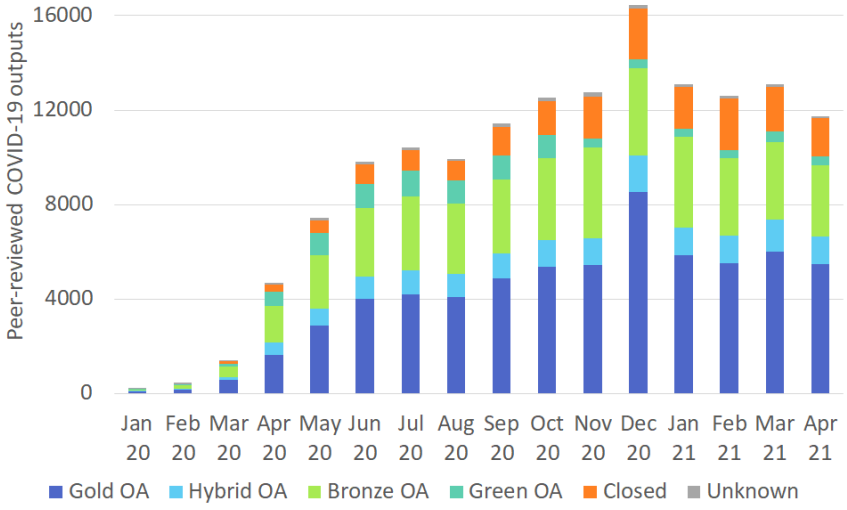
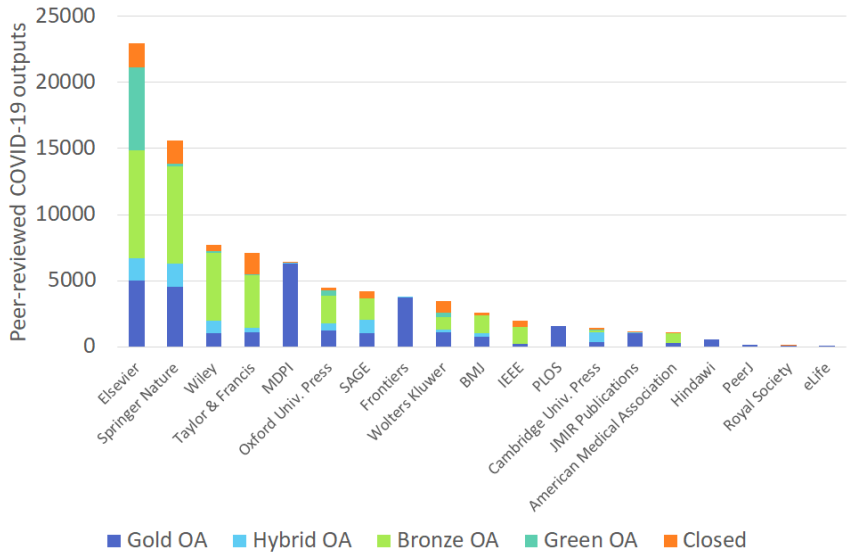


Figure 2.5: Breakdown by open or free access status for monthly number of COVID-19 outputs.

As shown in Figure 2.6, with the exception of the open access publishers (i.e., eLife, Frontiers, Hindawi, JMIR, MDPI, PeerJ, and PLOS), most

publishers have a large share of bronze open access COVID-19 outputs. These outputs have been made freely accessible, but there is no guarantee that they will remain free to access. Elsevier also has a large share of green open access outputs. However, this appears to be due to a mistake in the data. It seems that these outputs should have been classified as bronze open access.

Based on the above findings, our conclusion is that the commitment to open or free access of COVID-19 research made at the beginning of the pandemic has been largely fulfilled.



"The commitment to open or free access of COVID-19 research made at the beginning of the pandemic has been largely fulfilled"

Figure 2.6: Breakdown by open or free access status for COVID-19 outputs of 19 selected publishers.

2.3 Preprints

There has been a significant growth in awareness and posting of preprints during the pandemic [8, 19, 25, 46]. In our survey of authors of COVID-19 preprints, early and rapid dissemination of research results was the most frequently reported motivation for preprinting, mentioned by 86% of the respondents (see Figure 2.7). As one respondent stated:

> "We felt it was important to make our research available in a timely manner as the peer-review process can be time consuming."

Another survey respondent specified the benefits they aimed to realise in posting a preprint:

> "Wanted to get this work on vaccine design disseminated as early as possible for further work in urgent vaccine development."

The aim of early and rapid dissemination was followed by the motivation to make research openly available (64%). Interestingly, two-third

of the respondents had not posted any preprints before 2020. This is in line with similar findings reported by Fraser et al. [19]. It supports the idea that the pandemic has given a major impetus to preprinting of biomedical research. Moreover, 83% of the respondents stated that they will continue to post preprints, either for all their work (33%) or for at least some of it (50%). 6% of the respondents indicated that they will not post preprints in the future.

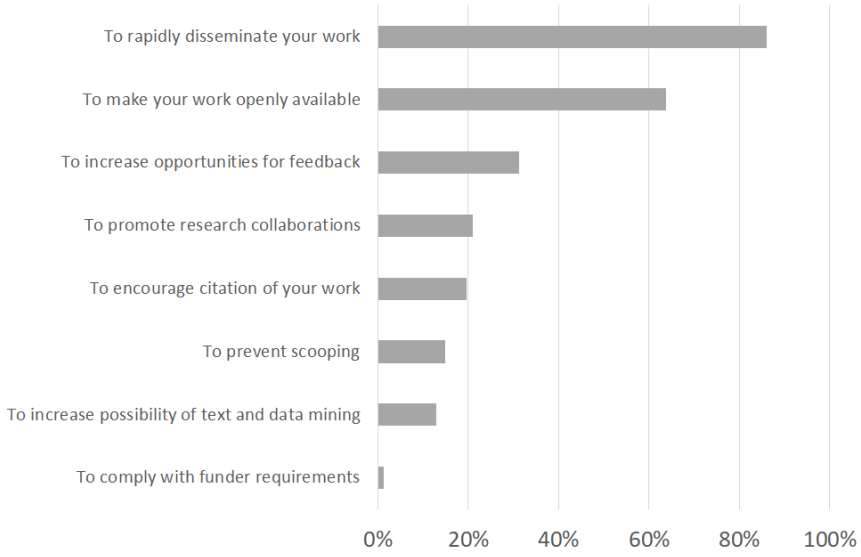


Figure 2.7: Motivations of survey respondents for preprinting their COVID-19 research.

Preprints also play an essential role in the development of COVID-19 guidelines by the WHO. A recent analysis showed that almost half of the references in these guidelines point to preprints, while less than one quarter point to journal articles [48].

"Preprints also play an essential role in the development of COVID-19 guidelines by the WHO"

Box 2.5: Methodological note – How to deal with multiple versions of the same preprint.

Researchers that have posted a preprint on a preprint server may at any time update the preprint by posting a new version. A preprint server may therefore contain multiple versions of a preprint. Our scientometric analyses include only the first version of a preprint. Later versions are not taken into account.

Researchers sometimes post the same preprint on multiple preprint servers, but this is a relatively rare phenomenon [19] and is discouraged by some preprint servers. Our scientometric analyses treat identical preprints posted on different preprint servers as separate preprints. Deduplication would involve some algorithmic challenges and therefore was not performed.

Almost 40,000 COVID-19 preprints were posted between January 2020 and April 2021. Figure 2.8 shows the number of COVID-19 preprints posted per month. For comparison, the monthly number of peer-reviewed COVID-19 outputs is shown as well. The monthly number of COVID-19 preprints reached its maximum in May 2020, when over 5,000 preprints were posted. In February and March 2020, the num-

ber of COVID-19 preprints that were posted exceeded the number of peer-reviewed COVID-19 outputs that were published. The reverse situation can be observed in all other months. Especially in more recent months, the number of peer-reviewed COVID-19 outputs was much larger than the number of COVID-19 preprints.

"Almost 40,000 COVID-19 preprints were posted between January 2020 and April 2021"

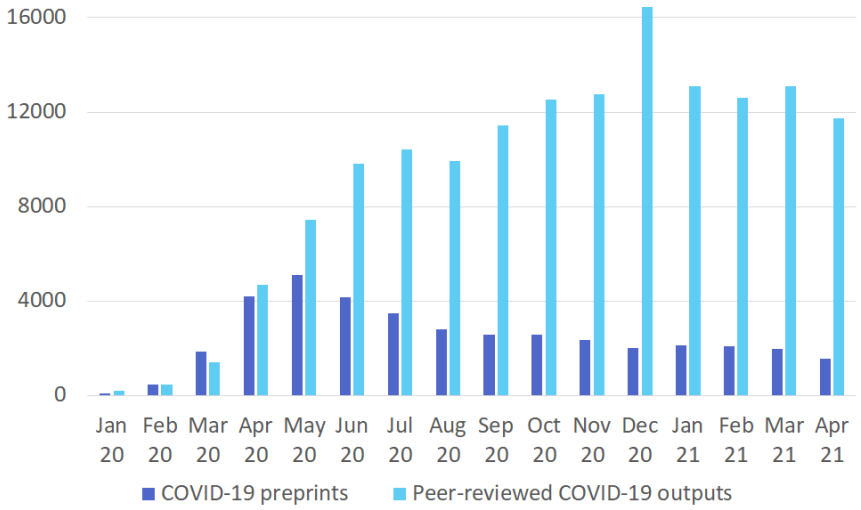


Figure 2.8: Monthly number of COVID-19 preprints and peer-reviewed COVID-19 outputs. Preprint data is incomplete for the most recent months.

Figure 2.9 shows the monthly number of COVID-19 preprints per preprint server. The five preprint servers with the largest number of COVID-19 preprints are medRxiv, SSRN, arXiv, bioRxiv, and Research Square. These preprint servers are presented individually in Figure 2.9. Other preprint servers are grouped together. For the most recent months, preprints posted on Research Square and SSRN are missing in our data. We refer to the work by Fraser and Kramer [18] for more complete and more up-to-date statistics on the number of COVID-19 preprints per preprint server, including a number of preprint servers not covered by our analysis.

Figure 2.10 shows the monthly number of COVID-19 preprints per preprint server as a percentage of the total monthly number of preprints. For most preprint servers, the percentage of COVID-19 preprints was below 20% in all or almost all months. However, for medRxiv, it was above 60% in most months and even reached almost 80% in May 2020. The growth of medRxiv as a result of the pandemic is impressive. medRxiv is the most frequently used preprint server for posting COVID-19 research. This is remarkable especially when taking into account that it was launched in June 2019, just half a year before the first confirmed cases of COVID-19.

Figure 2.11 shows the monthly number of COVID-19 and non-COVID-19 preprints in the period from January 2015 to April 2021, combining all preprint servers covered by our data. The number of preprints in 2020 was about 150% larger than the number of preprints in 2015, confirming the rapid increase in preprinting also reported in a recent analysis by Xie et al. [54]. The contribution of COVID-19 research to the overall growth in preprinting is relatively modest. arXiv is still by far the largest preprint server, and only about 2% of the preprints posted on arXiv during the pandemic were about COVID-19 (see Figure 2.10).

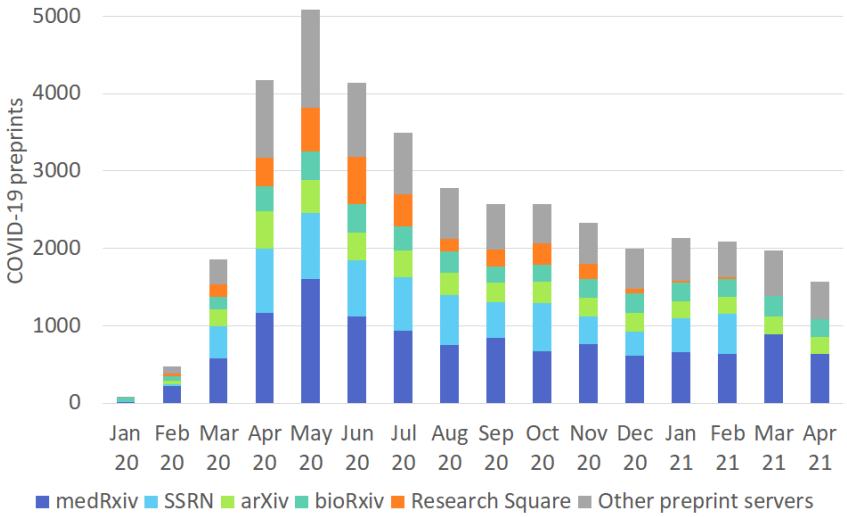


Figure 2.9: Monthly number of COVID-19 preprints per preprint server. Data for Research Square and SSRN is incomplete for the most recent months.

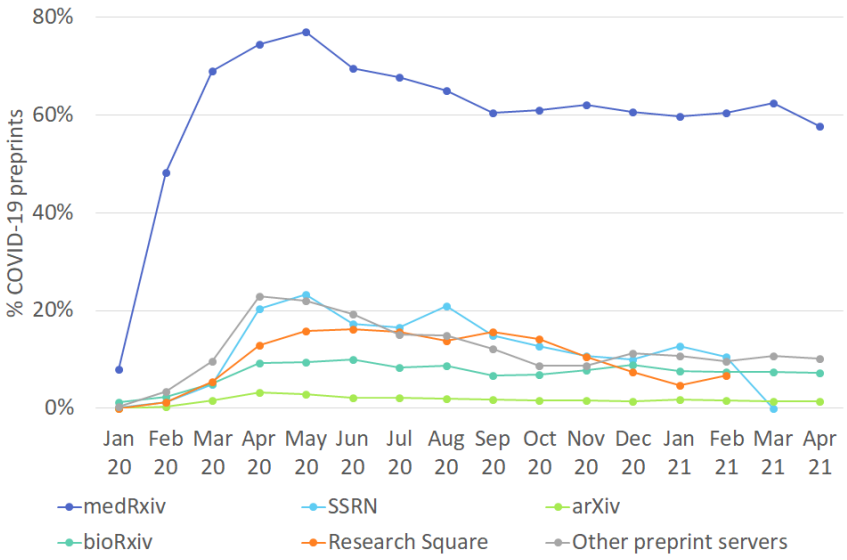


Figure 2.10: Monthly percentage of COVID-19 preprints per preprint server. Data for Research Square and SSRN is incomplete for the most recent months.

2.4 Has the commitment to preprinting been fulfilled?

"The growth of medRxiv as a result of the pandemic is impressive"

Has the commitment to preprinting of COVID-19 research, made at the beginning of the pandemic, been fulfilled? To what extent did researchers post preprints of their COVID-19 research before publishing their work in a peer-reviewed outlet, and conversely, to what extent did researchers manage to turn their preprints into peer-reviewed outputs? Answering these questions requires identifying links between preprints and the corresponding peer-reviewed outputs.

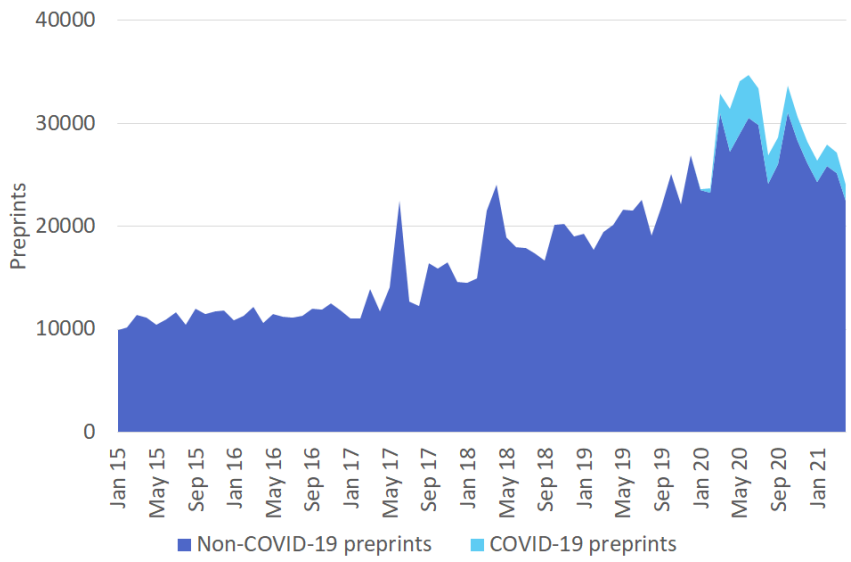


Figure 2.11: Monthly number of COVID-19 and non-COVID-19 preprints (2015–2021).

Box 2.6: Methodological note – How to identify links between preprints and peer-reviewed outputs.

The analyses presented in this section rely on links between preprints and the corresponding peer-reviewed outputs. Unfortunately, it is difficult to obtain accurate and complete data on such links [7, 20, 32, 35]. We combined data obtained from three sources. First, we used data on links between preprints and peer-reviewed outputs obtained from the Dimensions database. Second, we used data from Unpaywall. For many peer-reviewed outputs, Unpaywall provides links to versions of these outputs that are openly accessible on preprint servers and in repositories. Third, for a few preprint servers (i.e., bioRxiv and medRxiv), we used data obtained directly from the preprint server.

For many preprint servers, the above approach enabled us to capture a large share of the links between preprints and peer-reviewed outputs. However, for some preprint servers, our approach did not yield good results. In particular, our approach could not be used to identify links between preprints in SSRN and peer-reviewed outputs. Missing links between preprints and peer-reviewed outputs represent an important limitation of the analyses presented in this section.

Researchers normally post their research on a preprint server before it is published in a peer-reviewed outlet. They may choose to post their work on a preprint server after it has been published in a peer-reviewed outlet, for instance because they prefer not to share a non-final version of their work, but this is uncommon and some preprint servers actively discourage it. Less than 2% of the COVID-19 preprints that have a link to a peer-reviewed output were posted after the publication of the peer-reviewed output.

In our survey of authors of COVID-19 preprints, 53% of the respondents stated that they posted their research on a preprint server

"Researchers normally post their research on a preprint server before it is published in a peer-reviewed outlet"

"Most respondents were using preprint posting as a way of making their work available earlier than would be the case if they waited for publication in a journal"

before submitting it to a journal and a further 35% posted simultaneously with the submission to a journal. Most respondents were, therefore, using preprint posting as a way of making their work available earlier than would be the case if they waited for publication in a journal. 8% of the respondents posted their work on a preprint server while it was under peer review at a journal and 4% after the paper had been accepted or rejected.

Figure 2.12 shows the monthly number of COVID-19 preprints with and without a link to a peer-reviewed output. The figure also shows the monthly percentage of COVID-19 preprints with a link to a peer-reviewed output. About 30% of the preprints posted in the first months of the pandemic have a corresponding peer-reviewed output. The other preprints do not have a corresponding peer-reviewed output (or they do have a corresponding peer-reviewed output, but our methodology was unable to identify a link between the preprint and the peer-reviewed output). The percentage of preprints that have a corresponding peer-reviewed output decreases in more recent months, which is to be expected, given the time it takes to publish in a peer-reviewed outlet.

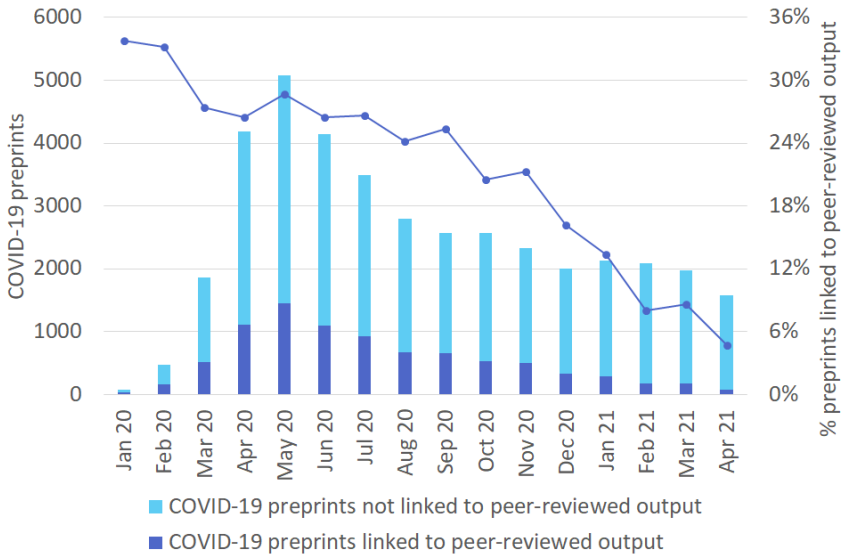


Figure 2.12: Monthly number of COVID-19 preprints with and without a link to a peer-reviewed output and monthly percentage of COVID-19 preprints with a link to a peer-reviewed output.

For COVID-19 and non-COVID-19 preprints posted in April 2020, Figure 2.13 shows the percentage of preprints per preprint server that have a link to a peer-reviewed output. About 40% of the preprints posted on arXiv, bioRxiv, and medRxiv have a corresponding peer-reviewed output. For Research Square, this is the case for 25% of the preprints. Data on links between preprints and peer-reviewed outputs is missing for SSRN. For most preprint servers, COVID-19 preprints are less likely to have a corresponding peer-reviewed output than non-COVID-19 preprints. The difference is especially large for arXiv (37% vs. 60%) and Research Square (25% vs. 41%). Of course, for each of the preprint servers, the percentage of preprints that have a corresponding peer-reviewed output will still increase. Most likely, a substantial share of the preprints posted in April 2020 were still undergoing peer review at the time of our analysis and will be published in a peer-reviewed outlet in the future.

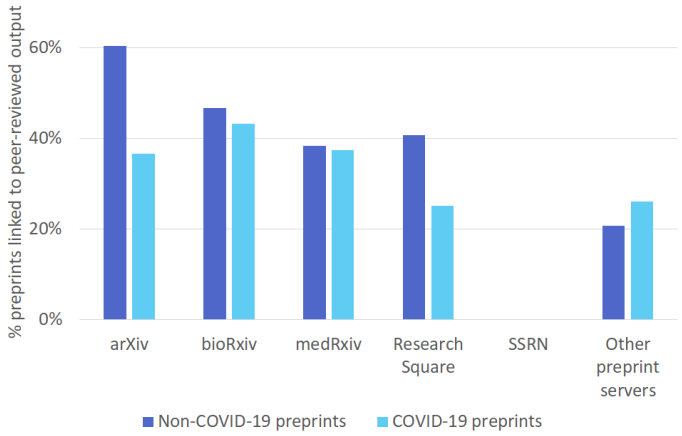


Figure 2.13: Percentage of COVID-19 and non-COVID-19 preprints in April 2020 with a link to a peer-reviewed output per preprint server.

Figure 2.14 shows the monthly number of peer-reviewed COVID-19 outputs with and without a link to a preprint. The figure also shows the monthly percentage of peer-reviewed COVID-19 outputs with a link to a preprint. 5% of all peer-reviewed COVID-19 outputs have a corresponding preprint, and this percentage is fairly stable over time. The other 95% of the peer-reviewed COVID-19 outputs do not have a corresponding preprint (or they do have a corresponding preprint, but our methodology was unable to identify a link between the peer-reviewed output and the preprint).

Based on Figure 2.14, we conclude that the commitment to preprinting made at the beginning of the pandemic has not been fulfilled. Only a small share of all COVID-19 research has been preprinted. Some COVID-19 research could not be posted on a preprint server because preprint servers, in particular bioRxiv and medRxiv, decided not to post certain types of research [34] (see also the case study for *PLOS ONE* presented in Box 2.7). However, in most cases, researchers themselves decided not to post their work on a preprint server.

"The commitment to preprinting made at the beginning of the pandemic has not been fulfilled"

Figure 2.15 presents the same term map as Figures 2.3 and 2.4, but colours now show the level of preprinting for different COVID-19 topics. The level of preprinting is highest for life sciences research on COVID-19, somewhat lower for medical research, and lowest for some of the COVID-19 topics in the social sciences.

Figure 2.16 shows the level of preprinting of COVID-19 outputs for 19 selected publishers (i.e., the 15 publishers with the largest number of COVID-19 outputs and four additional publishers participating in the COVID-19 Rapid Review Initiative). The highest level of preprinting can be observed for some of the smaller publishers, in particular eLife, JMIR, PeerJ, PLOS, and Royal Society. As discussed at the beginning of this chapter, outputs published on F1000 platforms are not included in the analysis, but these outputs can all be considered to have been preprinted.

For peer-reviewed COVID-19 and non-COVID-19 outputs published in January 2021, Figure 2.17 shows the percentage of outputs per broad disciplinary area that have a link to a preprint. In the biomedical and health sciences, the percentage of COVID-19 outputs that

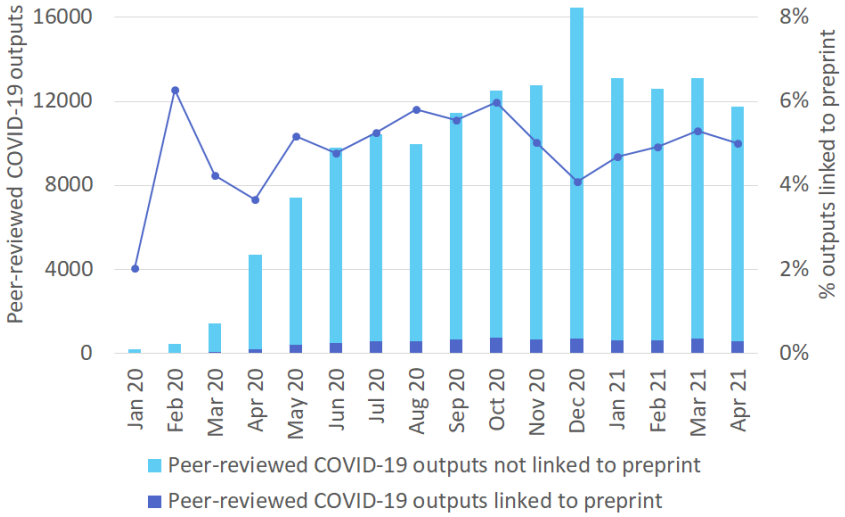


Figure 2.14: Monthly number of peer-reviewed COVID-19 outputs with and without a link to a preprint and monthly percentage of peer-reviewed COVID-19 outputs with a link to a preprint.

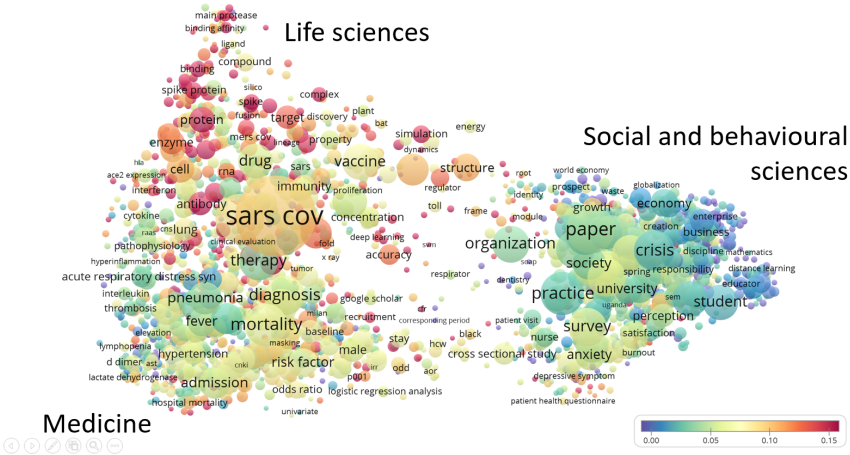


Figure 2.15: Term map showing the level of preprinting for different COVID-19 topics. The colour of a term reflects the proportion of the peer-reviewed COVID-19 outputs in which the term occurs that have been preprinted.

"In the biomedical sciences, the percentage of COVID-19 outputs with a preprint is more than three times higher than in the social sciences and humanities"

have a corresponding preprint is more than three times higher than in the social sciences and humanities (7% vs. 2%). The physical sciences and engineering have an intermediate level of preprinting of COVID-19 outputs (5%). These disciplinary differences in the level of preprinting may to some extent reflect differences between preprint servers in the availability of data on links between preprints and peer-reviewed outputs. They may in particular be affected by the missing data for SSRN.

Figure 2.17 can also be used to compare the level of preprinting for COVID-19 and non-COVID-19 outputs. In the biomedical and health sciences and the social sciences and humanities, the level of preprinting is higher for COVID-19 outputs than for non-COVID-19 outputs. In the physical sciences and engineering, in contrast, COVID-19 outputs have a lower level of preprinting than non-COVID-19 outputs.

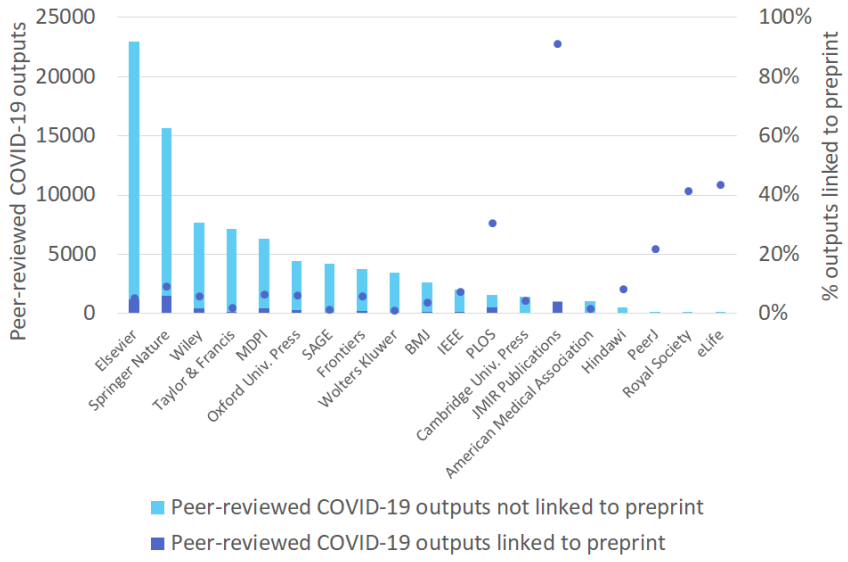


Figure 2.16: Number of peer-reviewed COVID-19 outputs with and without a link to a preprint and percentage of peer-reviewed COVID-19 outputs with a link to a preprint, for 19 selected publishers.

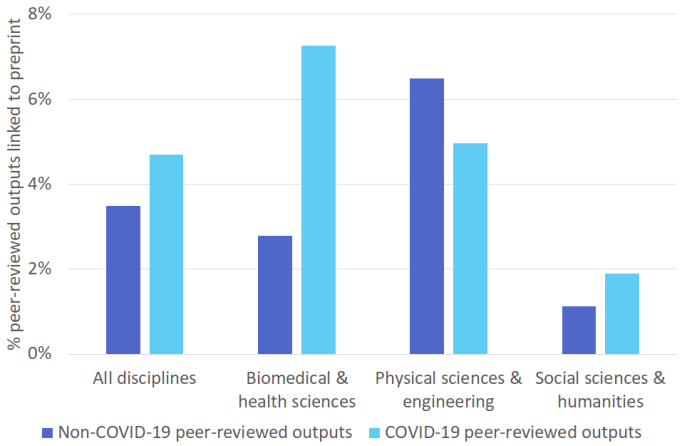


Figure 2.17: Percentage of peer-reviewed COVID-19 and non-COVID-19 outputs in January 2021 with a link to a preprint per broad disciplinary area.

Below we present two case studies in which we examine in more detail how publishers handle preprinting. The first case study considers *PLOS ONE* (Box 2.7). The second case study focuses on Springer Nature journals (Box 2.8).

Box 2.7: Case study – How did *PLOS ONE* stimulate preprinting?

By Dan Morgan and Emily Chenette (PLOS)

PLOS strongly encourages all authors to preprint their work. PLOS has a direct connection with the bioRxiv preprint server, and authors can opt in to facilitated bioRxiv posting when they submit an article to PLOS. PLOS does not yet have such a direct connection with medRxiv, although one is planned.

PLOS sent all COVID-19 articles to the WHO and notified au-

thors about this. In the notification email, authors who had not yet preprinted their work were encouraged to post a preprint. However, most COVID-19 articles needed to go to medRxiv, and without a direct connection with medRxiv, the ability to facilitate posting of preprints was limited.

In early 2020, PLOS was notified by Cold Spring Harbor Laboratory (CSHL), which operates bioRxiv and medRxiv, that not all COVID-19 articles should be sent to bioRxiv for immediate posting. There were significant concerns around the danger of unvalidated drug predictions. As a result, PLOS staff needed to bear in mind the following additional policies when conducting preprint checks and screenings:

- *Predictions of drug efficacy/potential for treatment of COVID-19 based entirely on in silico work.* PLOS was informed that articles describing predictions of inhibitors/-compounds/drug efficacy/potential for treatment of COVID-19 based entirely on in silico work would not be posted by bioRxiv and medRxiv. These include articles reporting in silico structural docking studies with any type of OTC, natural compounds, metabolic analysis, etc., any in silico prediction for COVID-19 treatment. As a result, *PLOS ONE* did not send these articles to bioRxiv.
- *Predictions of drug efficacy/potential for treatment of COVID-19 that include experiments in vitro or using in vivo model systems.* PLOS was informed that articles reporting predictions of drug efficacy/potential for treatment of COVID-19 that include experiments in vitro or using in vivo model systems would undergo special screening by several Outbreak Affiliates at bioRxiv to determine their suitability. Per CSHL's request, *PLOS ONE* did not send these articles to bioRxiv.
- *Articles examining drug treatment for COVID-19 in humans.* Articles examining drug treatment for COVID-19 in humans are outside of bioRxiv's scope, but could be considered for medRxiv. Per PLOS' standard workflow for articles that are outside of bioRxiv's scope, *PLOS ONE* declined to send to bioRxiv articles reporting treatment studies in humans.

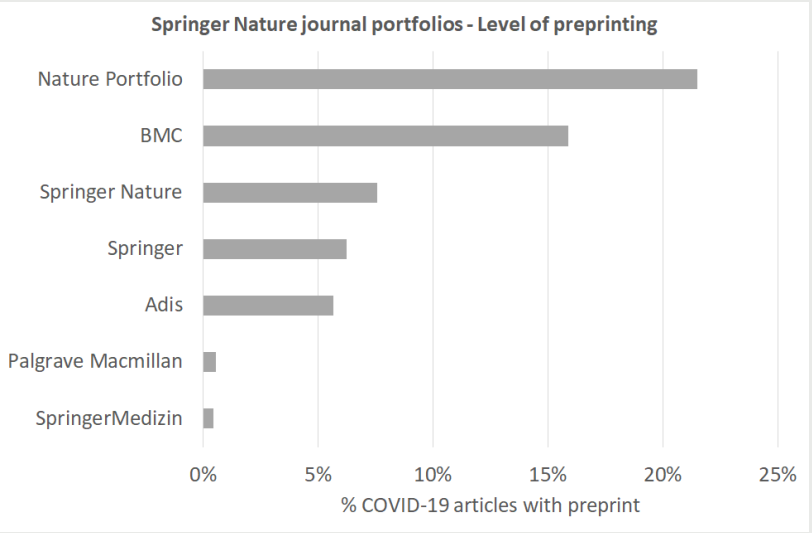
In summary:

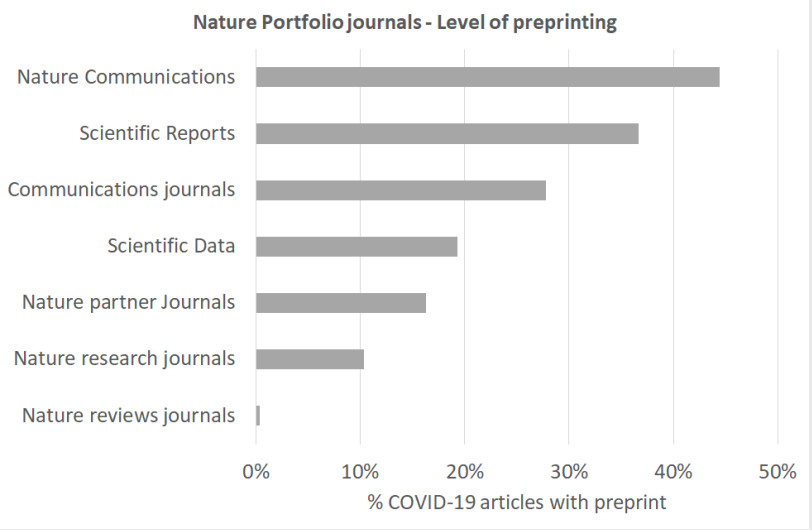
- PLOS encourages preprint posting for all articles and can directly facilitate preprint posting if authors opt in and if bioRxiv is the target. (Further connections are planned.)
- PLOS cannot facilitate preprint posting if authors do not opt in or if there is no established partnership with the preprint server
- PLOS screens articles prior to sending them to bioRxiv.
- Guidelines from bioRxiv emerged in 2020 that prevented certain articles being directly sent to bioRxiv.

Box 2.8: Case study – How did Springer Nature authors share preprints?

By Andy Needham and Sowmya Swaminathan (Springer Nature)

Only 10% of the authors who published COVID-19 research across all Springer Nature journals between January 2020 and April 2021 posted a preprint of their work. Nevertheless, there are notable differences in author behaviour across imprints and journals, as shown in the figures presented below. 21% of the authors publishing in Nature Portfolio journals posted a preprint. A significantly higher percentage of articles in the open access journals within the Nature Portfolio have an associated preprint – – 44% at *Nature Communications* and 36% at *Scientific Reports*. Across the BMC portfolio of journals, which are all open access, 15% of the COVID-19 articles have an associated preprint, increasing to 24% across the BMC Series journals. The correlation between posting of preprints and open access status of journals may reflect a predisposition of authors selecting to publish in open access journals toward other open research practices.





The top preprint servers of choice for Springer Nature authors of COVID-19 research were bioRxiv/medRxiv (for technical reasons, we are unable to distinguish between the two), Research Square and arXiv. Of the COVID-19 articles published in Springer Nature journals that have an associated preprint, 45%, 38%, and 13% respectively have their preprint on bioRxiv/medRxiv, Research Square, and arXiv. There are interesting differences in author behaviour across Springer Nature imprints and journals with respect to the choice of preprint servers. Whereas a greater percentage of authors publishing in BMC journals posted their article on the Research Square platform (via Springer Nature's *In Review* service, which allows authors to easily and seamlessly post preprints on Research Square), the trend is reversed for authors publishing in Nature Portfolio journals, with a much higher percentage of preprints on bioRxiv/medRxiv.

Of authors publishing in BMC journals that posted their article as a preprint, a majority (64%) posted their work on the Research Square platform, with three quarters of these making use of Springer Nature's *In Review* service. *In Review* was made available for these journals in Spring 2019 (BMC Series journals) and Winter 2019/2020 (BMC Academic Journals). In contrast, for Nature Portfolio journals, where *In Review* was made available later in 2020, a smaller proportion of authors posted on Research Square (34%, 8% of these via *In Review*), with a much higher proportion using bioRxiv/medRxiv (63%). This may reflect inherent author preferences for certain preprint servers, guided in part by emerging community norms and usage trends establishing some platforms as the venue of choice for COVID-19 preprints. The uptake through *In Review* underscores a role for publisher-supported open research infrastructure in making a transition to open research practices easy and beneficial for researchers.

2.5 Has the commitment to data sharing been fulfilled?

"Even in COVID-19 research, data sharing has not always been practised"

The pandemic has illustrated the power of rapid and open data sharing. The sharing of the genome sequencing of the SARS-CoV-2 virus in early January 2020 has been described as "one the most important acts of data-sharing ever undertaken" [43]. It allowed work to take place immediately on developing treatments and vaccines globally. Such data sharing has continued to be a feature of the crisis, with over one million Sars-CoV-2 sequences shared by scientists from around the world via the Wellcome Sanger Institute since March 2020 [2]. The power of data sharing was seen again recently in the sharing of the details of the Omicron variant (Barrett, 2021). Sharing is enabled by a developing infrastructure, including the Nextstrain tool, "an online resource that uses genome data to monitor the evolution of disease-causing organisms such as viruses in real time" [23].

However, even in COVID-19 research, data sharing has not always been practised. On 22 May 2020, *the Lancet* published an article evaluating the effectiveness of hydroxychloroquine or chloroquine

for treatment of COVID-19. The article, which prompted the WHO to halt a hydroxychloroquine trial, was heavily criticised in an open letter signed by almost 200 clinicians, medical researchers, statisticians, and ethicists. The letter questioned the accuracy of the data underlying the article and criticised the authors for not sharing the data. On 5 June 2020, the article was retracted by three of the four authors, who stated that they "can no longer vouch for the veracity of the primary data sources". As shown in Section 5.4, this is one of the most high-profile retractions of COVID-19 research. As a result of the retraction, *the Lancet* changed its author and reviewer guidelines to include an explicit check for the data underlying an article.

This latter example shows the challenges faced by publishers in fulfilling the commitments to data sharing made at the beginning of the pandemic. To illustrate how publishers dealt with these challenges, Box 2.9 present a case study discussing the steps taken by the publishers in the COVID-19 Rapid Review Initiative to reinforce their data sharing policies.

"The publishers participating in the COVID-19 Rapid Review Initiative anticipated that clarifying and strengthening their data sharing policies would be important"

Box 2.9: Case study – Data sharing and the COVID-19 Rapid Review Initiative.

By Catriona J. MacCallum (Hindawi) and Susanna-Assunta Sansone (University of Oxford and FAIRsharing)

The publishers participating in the COVID-19 Rapid Review Initiative anticipated that clarifying and strengthening their data sharing policies would be important and formed a working group dedicated to this issue. This was motivated not only by the commitment to data sharing made in the Wellcome-coordinated COVID-19 statement, but also by a commitment to the FAIR principles [53], which emphasise the need for data to be findable, accessible, interoperable, and reusable by humans and machines. At the core of the FAIR principles are (meta)data standards and identifiers, which are essential technical infrastructure to enable unambiguous reporting and structured descriptions of data. Implemented by data repositories, this technical infrastructure, along with the appropriate policies, training, and social aspects of incentives and rewards, are crucial to turn the FAIR principles into reality (e.g., [13, 12]).

FAIRsharing [44] joined the COVID-19 Rapid Review Initiative with the intent to help shape and coordinate the work of the data sharing working group. FAIRsharing is an informative and educational resource that registers and interlinks data policies (from publishers, funders, and other organisations) to the databases and (meta)data standards that these policies recommend. Policy registration ensures that policies are discoverable by humans and machines, are citable, are transparent as to what they recommend, and are comparable to each other.

The data sharing working group had two aims. The first aim was to have minimum requirements for publishers to be part of the COVID-19 Rapid Review Initiative. It was decided that any publisher participating in the initiative must ensure that authors include a data availability statement in articles related to COVID-19. In the first instance, stating that "data is available on request"

was allowed. Vines et al. [49], for example, had shown that just the inclusion of a data availability statement may increase the extent to which authors make data publicly available.

The second aim was to harmonise the data policies and guidelines for authors across the publishers and journals participating in the COVID-19 Rapid Review Initiative, or to at least make these policies and guidelines more transparent. Often publisher policies are not clear on how data should be made available, or what appropriate data repositories and standards are recommended to researchers.

The COVID-19 Rapid Review Initiative called on publishers to facilitate the stewardship and sharing of FAIR data underlying COVID-19 articles. Registration of data sharing policies in FAIRsharing was also explicitly solicited.

To monitor and track the outcomes of the working group's data policy requirements, information was collected on the policies of all publishers that participated in the COVID-19 Rapid Review Initiative and that agreed to ensure that COVID-19 articles have a data availability statement. BMC, eLife, F1000, GigaScience, Hindawi, PLOS, and Royal Society already had their policies registered with FAIRsharing before the initiative launched. Cambridge University Press, Oxford University Press, PeerJ, and Ubiquity Press registered their policies in FAIRsharing as a result of joining the COVID-19 Rapid Review Initiative. The data policies of the members of the initiative were grouped and made visible under a dedicated FAIRsharing Collection, which includes 14 policies that recommend a total of 137 data repositories and that refer to 27 (meta)data standards. The figure below illustrates the network of the databases and standards recommended by the data policies of the members of the COVID-19 Rapid Review Initiative.

In January 2021, eLife, Hindawi, PeerJ, PLOS, and Royal Society, the publishers that had originally launched the COVID-19 Rapid Review Initiative, agreed to align their data sharing policies to

meet the TOP Data Transparency Level II, which requires that "data must be posted to a trusted repository. Exceptions must be identified at article submission" (see press release). This involved mandating data sharing in a public repository rather than just ensuring that an article includes a data availability statement. Making data available on request is inefficient [47] and was therefore not acceptable, except for legal or ethical reasons. A data availability statement needs to explicitly mention the repository in which the data is publicly available. In addition, providing formal citations to data and software was also strongly encouraged, following the Joint Declaration of Data Citation Principles and the recent Software Citation Guide [31]. Note that enforcing data sharing policies poses particular challenges to publishers of clinical articles, to ensure that publication does not jeopardise patient confidentiality.

Two valuable lessons were learned from the work of the data sharing working group of the COVID-19 Rapid Review Initiative. First, there is a need to collect more data about data sharing, for example about the extent to which articles actually link to the relevant data sets and the extent to which data sets and software are formally cited. Second, additional work is needed to harmonise data guidelines across publishers and journals, and ideally align them with funders, to ensure that requirements on data availability are clear and comparable across policies. The aim is to address this challenge by taking a collaborative approach jointly through the RDA/FORCE11 FAIRsharing Working Group, the Funders and Policy Standardisation groups in the Research Data Alliance (RDA).

In our survey of authors of COVID-19 preprints, 47% of the respondents stated that they had made the data underlying their research publicly available. Another 22% of the respondents stated that they had indicated in their preprint that data is available on request. Only 13% of the respondents told us that they did not share their data. 12% of the respondents had not used new data in their research, and 6% provided "other" miscellaneous explanations.

Our survey results offer a relatively positive picture of the data sharing practices of COVID-19 researchers. However, we suspect that this may be partly due to a self-selection bias, causing the survey respondents not to be representative of the population of all COVID-19 researchers. Other studies of COVID-19 data sharing practices offer a less positive picture.

Larregue et al. [36] and Collins and Alexander [11] used text mining approaches to study data sharing in COVID-19 preprints. Larregue et al. found that 11% of the COVID-19 preprints posted in medRxiv between January and early November 2020 made data openly available. A larger share of the preprints (43%) made data available on request. Based on a different text mining methodology, Collins and Alexander reported that 15% of the COVID-19 preprints posted in medRxiv in 2020 and in the first half of 2021 made data openly available. For bioRxiv this is the case for 28% of the COVID-19 preprints, while for arXiv and SocArXiv this is the case only for, respectively 13% and 12% of the preprints. For arXiv and SocArXiv, Collins and Alexander found that the percentage of preprints with open data was sub-

"Additional work is needed to harmonise data guidelines across publishers and journals, and ideally align them with funders"

stantially higher than before the pandemic, but for medRxiv and bioRxiv this percentage was somewhat lower than before the pandemic.

Similar results have been found for COVID-19 journal articles. Gkiouras et al. [22] investigated data sharing in COVID-19 articles published in five prominent medical journals (i.e., *Annals of Internal Medicine*, *BMJ*, *JAMA*, *the Lancet*, and *New England Journal of Medicine*) in the first months of the pandemic (i.e., the period until March 14, 2020). Of the 32 COVID-19 articles that include patient data, they found that one shared this data in a fully open manner. Lucas-Dominguez et al. [37] studied data sharing in COVID-19 articles published between January and April 2020 and indexed in PubMed Central. Of the almost 6,000 articles, they found that about 800 made data available, either as supplementary material or in a repository. However, in most cases data was made available in PDF files or Word documents containing textual or graphic material. In only 1.2% of the 6,000 articles, data was made available in a way that, according to Lucas-Dominguez et al., allowed for reuse of the data.

"The level of data sharing is low, and the different stakeholders in the research system have not managed to fulfill the data sharing commitments made at the beginning of the pandemic"

Recent work by Maxwell et al. [38] looked at population-specific and discipline-specific resources for collecting and disseminating participant-level data from COVID-19 research. This work highlights two additional challenges for data sharing. First, these resources are concentrated in high-income countries and siloed by comorbidity, body system, and data type. Second, resources for sharing clinical data were less FAIR than those for sharing molecular data.

Although the above-mentioned studies use different definitions of data sharing and different methods to operationalise these definitions, they give a fairly consistent picture. The level of data sharing is low, and the different stakeholders in the research system have not managed to fulfill the data sharing commitments made at the beginning of the pandemic. Even during a global crisis, the barriers to data sharing seem to be high from the perspective of most researchers, and at the same time, the incentives are weak. There is still a great deal to do to create an environment where the sharing of research data becomes part of business as usual for science.

3 Journal peer review

Highlights

- Compared to non-COVID-19 articles, COVID-19 articles were more likely to be rejected for publication in a journal.
- On average, the time from submission to a journal to final decision (acceptance or rejection) and subsequent publication (in case of acceptance) was substantially shorter for COVID-19 articles than for non-COVID-19 articles, especially in the beginning of the pandemic.
- More research is needed on the quality of COVID-19 research outputs following some concerns raised in the early stages of the response to the pandemic. For some journals, peer review may have been more constructive for COVID-19 articles, but this does not seem to apply across the system as a whole.
- For most COVID-19 articles posted on a preprint server and published in a journal, the differences between the preprint version and the journal version were minor. Survey respondents reported major changes were made to the discussion or conclusion in about one-fifth of the COVID-19 articles.

The journal peer review system has been under strong pressure during the pandemic. Large numbers of COVID-19 articles were submitted to journals, and these submissions needed to be peer reviewed as rapidly as possible, ideally without compromising on the rigour of peer review. At the same time, non-COVID-19 submissions in all subjects were also reported by many publishers to have risen rapidly, creating system-wide congestion [16]. Publishers took different approaches to speed up peer review of COVID-19 research. A number of publishers started working together in the COVID-19 Rapid Review Initiative. These publishers established a pool of rapid reviewers. They also developed a process to transfer rejected submissions and the associated reviews between publishers.

To what extent did publishers indeed manage to accelerate peer review? And more generally, how did the journal peer review system respond to the pandemic? To address these questions, we analyse the extent to which COVID-19 articles have been peer reviewed and published more rapidly than non-COVID-19 articles. In addition to general statistics on submission-to-publication times, we also present more detailed statistics for journals published by eLife, Hindawi, PLOS, and Royal Society, four publishers participating in the COVID-19 Rapid Review Initiative. We also discuss possible differences in the way in which peer review was carried out for COVID-19 and non-COVID-19 articles. Finally, we discuss the extent to which peer review has led to changes to COVID-19 articles.

We note that we use the term 'peer review' in a broad sense. It refers not only to the peer review activities of researchers invited by a jour-

nal to assess the quality of an article, but also to quality assurance activities carried out by the editorial staff of a journal.

3.1 Submission-to-publication times

"The speed of peer review was a major priority for COVID-19 research"

For obvious reasons, the speed of peer review was a major priority for COVID-19 research. This was also confirmed by the respondents to our survey of COVID-19 preprint authors. 64% of the respondents stated that the speed of peer review was extremely important or very important in the choice of the journal to which they submitted their COVID-19 research.

We now present an analysis of the time it took to publish COVID-19 research, from the submission of an article to a journal to the publication of the article on the journal website. This includes not only the time required for peer review. It also includes the time needed by authors to revise an article based on the comments provided by editors and reviewers as well as the time needed by a journal to publish an article after it has been accepted for publication.

Box 3.1: Methodological note – How to determine the time from submission to publication.

There is no database that provides complete data on the time between submission of an article to a journal and publication of the article in the journal. We collected this data from PubMed and also directly from publishers. Five publishers provided data: eLife, Hindawi, PLOS, Royal Society, and Springer Nature. These publishers are all associated with the COVID-19 Rapid Review Initiative. In the case of Springer Nature, two BMC journals took part in the COVID-19 Rapid Review Initiative, but data was provided for all Springer Nature journals. The data provided by eLife, Hindawi, PLOS, and Royal Society is available in figshare [51].

After combining and deduplicating the data, we had data on the time from submission to publication for about half a million journal articles submitted in 2020. Our data is incomplete and skewed towards biomedical research and towards specific publishers. There may also be differences between publishers in the way in which the date of submission and the date of publication of an article are defined. Such differences may lead to small inconsistencies in our analysis.

To analyse differences between COVID-19 and non-COVID-19 articles in the time from submission to publication, we focused on a snapshot: articles that were submitted to a journal in April 2020 and that in the meantime have been published in the journal. There are 100 journals to which at least 10 COVID-19 articles and at least 10 non-COVID-19 articles were submitted in April 2020. For each of these journals, we calculated the average time from submission to publication, both for the COVID-19 articles submitted in April 2020 and for the non-COVID-19 articles submitted in the same month.

For these 100 journals, Figure 3.1 shows the average time from submission to publication (in days) for non-COVID-19 articles (horizontal

axis) and for COVID-19 articles (vertical axis). For almost all journals, the time from submission to publicaton was shorter for COVID-19 articles than for non-COVID-19 articles. For some journals the difference was small, while for other journals it was very large. Elsevier journals in particular show a large decrease in the submission-to-publication time. For journals published by Frontiers, Springer Nature, and Wiley, the average decrease is smaller, but it is still highly significant for many of these journals. The decrease in submission-to-publication time is very modest for MDPI journals. These journals have a rapid turnaround time both for COVID-19 and for non-COVID-19 articles. The full data, including the names of the individual journals, is available online. A more detailed discussion for Springer Nature journals is provided in Box 3.2.

"For almost all journals, the time from submission to publication was shorter for COVID-19 articles than for non-COVID-19 articles"

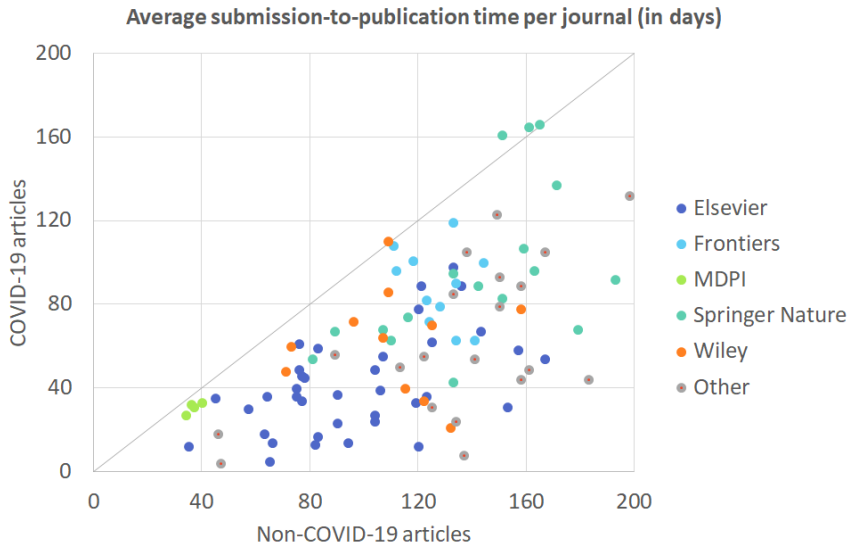


Figure 3.1: Average submission-to-publication time (in days) for non-COVID-19 articles (horizontal axis) and COVID-19 articles (vertical axis) submitted in April 2020. Each dot represents a journal to which at least 10 COVID-19 articles and at least 10 non-COVID-19 articles were submitted.

For each journal in Figure 3.1, we calculated the mean normalised submission-to-publication time for COVID-19 articles, defined as the ratio of the average submission-to-publication time for COVID-19 articles and the average submission-to-publication time for non-COVID-19 articles. We then averaged the mean normalised submission-to-publication times over the 100 journals in Figure 3.1, weighting each journal by its number of COVID-19 articles. In this way, we obtained an overall mean normalised submission-to-publication time for April 2020 of 0.56. This means that on average COVID-19 articles submitted to a journal in April 2020 were published 44% more quickly than non-COVID-19 articles submitted to the same journal in the same month.

We performed the same calculation for each month between February 2020 and November 2020. Figure 3.2 shows the monthly mean normalised submission-to-publication time for COVID-19 articles. In the first months of the pandemic, COVID-19 articles were published about 50% more quickly than non-COVID-19 articles. In later months, the difference was much smaller, and COVID-19 articles were published only about 10% more quickly than non-COVID-19 articles. The statistics for these later months may still change, because many ar-

ticles submitted in these months presumably had not yet been published at the time of analysis.

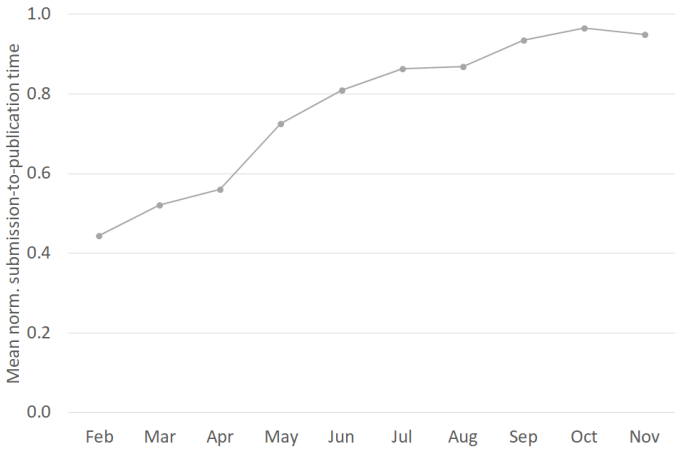


Figure 3.2: Monthly mean submission-to-publication time for COVID-19 articles normalised relative to submission-to-publication times for non-COVID-19 articles.

The rapid way in which many COVID-19 articles were peer reviewed and published in the first months of the pandemic was also observed in other analyses of submission-to-acceptance and submission-to-publication times for COVID-19 research [1, 24, 26, 28, 33, 40].

"The respondents to the survey had mixed experiences of the time taken for the peer review of their COVID-19 research"

Our preprint author survey partly confirms the above findings. The respondents to the survey had mixed experiences of the time taken for the peer review of their COVID-19 research. A substantial number of respondents whose papers had been published in a journal reported that peer review took less time than it normally does, but an equally large number of respondents reported that it took more time. 40% of the respondents stated that the time from submission of a revised article to acceptance of the article was shorter than normal, while 29% stated that this was not the case. The clearest outcome was obtained for the time from acceptance of an article to publication of the article. 47% of the respondents agreed that this time was shorter than normal; only 22% disagreed.

Box 3.2: Case study – How have Springer Nature journals tried to speed up the publication process?

By Arianne Heinrichs and Sowmya Swaminathan (Springer Nature)

A reduction in turnaround times can be observed for COVID-19 submissions across Springer, Nature Portfolio and BMC Series journals. The greatest reduction in turnaround times was achieved by Nature Portfolio journals, in particular *Nature* and *Nature Medicine*, with submission-to-acceptance times reduced by more than half for COVID-19 articles. At both journals, attention to COVID-19 articles was prioritised at every stage of the editorial, peer review, production and publication process, with accelerated decision times at each stage. In some cases a submission was already discussed with an editor before the complete article was formally submitted. This enabled the editor

to enlist reviewers ahead of receipt of the formal submission. Editors also developed a pool of experts committed to rapid turnaround times. In addition, editors worked with authors and reviewers to carefully prioritise the most fundamental revisions needed to support the claims made in an article. Articles were accelerated through the production process using the Accelerated Article Preview format, which releases the peer-reviewed version of an article before the article has gone through a complete production process.

BMC Medicine nearly halved the turnaround times for COVID-19 articles, thanks largely to the rapid revision of articles after peer review and the accelerated production process in which COVID-19 articles were prioritised. Journals with external academic editors also put in place creative strategies to prioritise COVID-19 articles and highlighted the challenges therein, particularly for journals where authors, editors and reviewers were on the frontlines in hospitals and other health care settings (e.g., [3]). For example, at one journal the society partner appointed a guest editor to lead a team of 14 scientists to prioritise editorial decisions and fast-track peer review for COVID-19 articles.

3.2 Rejection rates: An analysis of four publishers

The analysis reported in Section 3.1 covers only articles that were accepted for publication in a journal. Articles that were rejected are not included. Data for these articles typically is not available. For journals published by eLife, Hindawi, PLOS, and Royal Society, we now present an analysis in which we include not only articles that were accepted for publication, but also articles that were rejected. Considering both accepted and rejected articles offers a more detailed understanding of the role played by journals in the evaluation and dissemination of COVID-19 research.

"Considering both accepted and rejected articles offers a more detailed understanding of the role played by journals in the evaluation and dissemination of COVID-19 research"

Figure 3.3 shows the monthly number of COVID-19 and non-COVID-19 articles submitted to eLife, Hindawi, PLOS, and Royal Society journals in 2020. Between 4 and 5% of the articles submitted to eLife, Hindawi, and Royal Society journals in 2020 dealt with COVID-19. For PLOS journals this was the case for almost 12% of the submitted articles.

The rest of our analysis focuses on articles submitted to eLife, Hindawi, PLOS, and Royal Society journals between April and June 2020, taking into account only articles for which a final decision (i.e., acceptance or rejection) had been made at the time of our data collection. Moreover, we consider only journals for which our analysis covers more than ten COVID-19 articles. This is the case for eLife and for 34 Hindawi, seven PLOS, and three Royal Society journals. For these 45 journals, our analysis covers a total of 3,743 COVID-19 articles, all submitted between April and June 2020. Each article has either been accepted for publication or been rejected. We do not distinguish between different points in the editorial process of a journal at which an article can be rejected (e.g., before or after peer review).

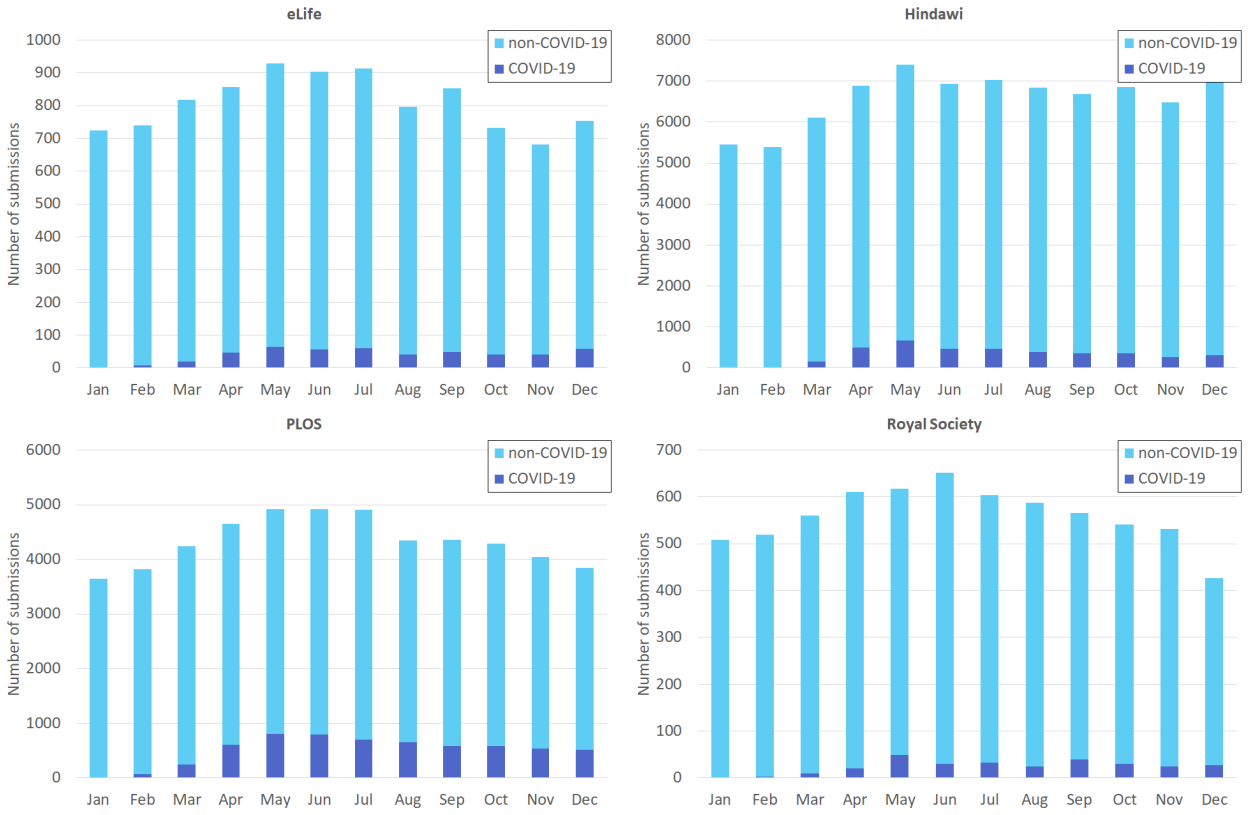


Figure 3.3: Monthly number of COVID-19 and non-COVID-19 articles submitted to eLife, Hindawi, PLOS, and Royal Society journals in 2020.

"For most journals, the rejection rate was substantially higher for COVID-19 articles than for non-COVID-19 articles"

Figure 3.4 shows the rejection rates for COVID-19 and non-COVID-19 articles submitted to the 45 selected journals. For most journals, the rejection rate was substantially higher for COVID-19 articles than for non-COVID-19 articles. The overall rejection rate for COVID-19 articles was 75%, which is much higher than the rejection rate of 56% for non-COVID-19 articles. The high rejection rate for COVID-19 articles suggests that journals played an important role as a quality filter for COVID-19 research. The data here are consistent with evidence presented elsewhere that the pandemic gave rise to large amounts of lower-quality submissions that had to be dealt with by publishers, causing additional pressures on the system [5, 14].

Figures 3.5 and 3.6 show the average time from submission to final decision for COVID-19 and non-COVID-19 articles submitted to the 45 selected journals, distinguishing between rejected articles (Figure 3.5) and accepted articles (Figure 3.6). Most journals rejected COVID-19 articles much more quickly than non-COVID-19 articles. The difference in the time from submission to rejection was especially large for many Hindawi journals, which might be due to additional quality checks that Hindawi journals performed for COVID-19 articles, leading to rapid desk rejection of many of these articles. The high rejection rates for COVID-19 articles (Figure 3.4) combined with the short submission-to-rejection times (Figure 3.5) suggests that many COVID-19 articles may have been of low quality, so that journals could quickly reject them. This seems to fit with a great deal of anecdotal evidence suggesting that the pandemic gave rise to a large amount of low-quality research submitted to journals in the early stages of the pandemic, which put a considerable strain on the system.

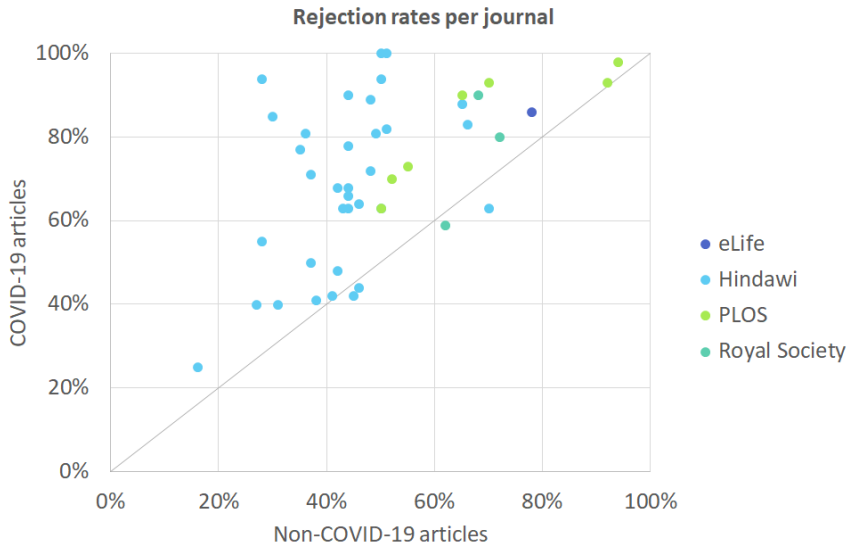


Figure 3.4: Rejection rates for COVID-19 and non-COVID-19 articles submitted to 45 selected eLife, Hindawi, PLOS, and Royal Society journals.

For accepted articles, our findings are somewhat mixed (Figure 3.6). eLife and the PLOS journals accepted COVID-19 articles more rapidly than non-COVID19 articles. For Hindawi and Royal Society journals, on the other hand, there is no clear pattern. Some Hindawi and Royal Society journals accepted COVID-19 articles more rapidly than non-COVID-19 articles, but other journals were slower in accepting COVID-19 articles. These differences can apparently be explained by different approaches taken by different publishers. eLife and PLOS report putting in place special measures at company level to accelerate COVID-19 submissions through their editorial processes. In contrast, Hindawi and Royal Society allowed decisions to made at individual journal levels, which led to different patterns of responses across their portfolios.

"Different approaches were taken by different publishers"

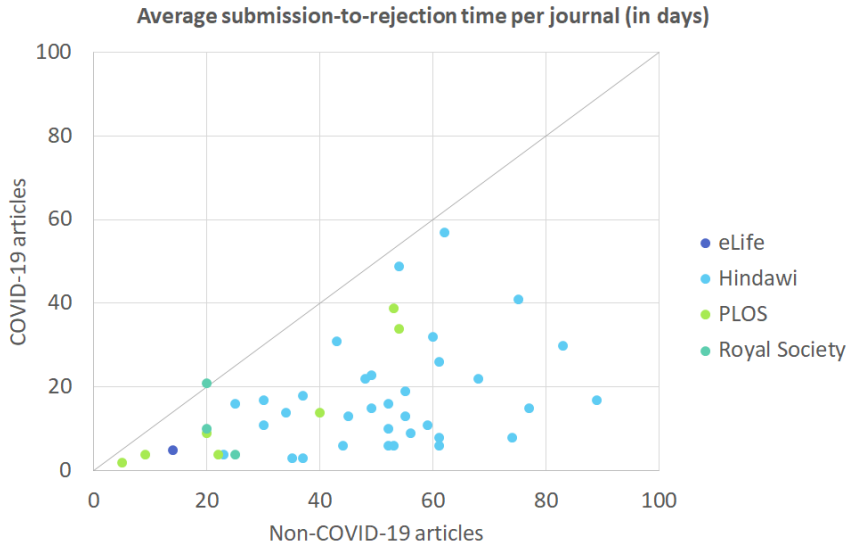


Figure 3.5: Average time (in days) from submission to rejection for COVID-19 and non-COVID-19 articles submitted to 45 selected eLife, Hindawi, PLOS, and Royal Society journals.

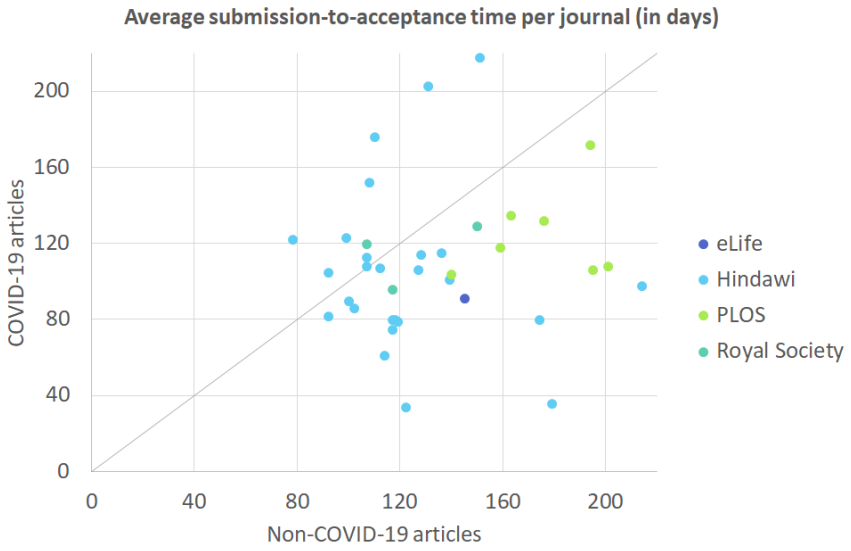


Figure 3.6: Average time (in days) from submission to acceptance for COVID-19 and non-COVID-19 articles submitted to 45 selected eLife, Hindawi, PLOS, and Royal Society journals.

Box 3.3: Case study – Transferring rejected COVID-19 articles between publishers.

One of the innovations introduced by the publishers in the COVID-19 Rapid Review Initiative was the possibility of authors allowing their papers to be automatically transferred to other journals, possibly published by different publishers, following rejection by the first journal. This involved seeking permission from reviewers to transfer their reviews, including revealing their names, to other publishers. It also involved authors opting to take advantage of the service. However, in the event, no authors opted to do so.

A survey of authors who submitted articles to journals part of the COVID-19 Rapid Review Initiative found that at least some authors preferred to submit to another journal of their own choice rather than this being done on their behalf [29]. However, since we know that journal transfer is a popular and well-established practice within publisher portfolios, it remains an open question whether this has potential across publisher portfolios as well.

3.3 Nature of peer review

"It remains an open question whether journal transfer has potential across publisher portfolios"

As shown in the previous sections, many journals managed to reduce the time from submission to final decision for COVID-19 articles, especially in the beginning of the pandemic. How did this affect the nature of the peer review performed for COVID-19 articles?

We asked the respondents to our COVID-19 preprint author survey whether they found peer review to be more constructive for their COVID-19 research than for other work, for instance in terms of the

tone of the reviews and requests for additional experiments. Figure 3.7 shows that this is not the case. On average, our respondents experienced peer review of their COVID-19 research to be as constructive as peer review of other work. Likewise, our respondents found peer review of their COVID-19 work to be of similar quality as peer review of other research.

"Our respondents experienced peer review of their COVID-19 research to be as constructive as peer review of other work"

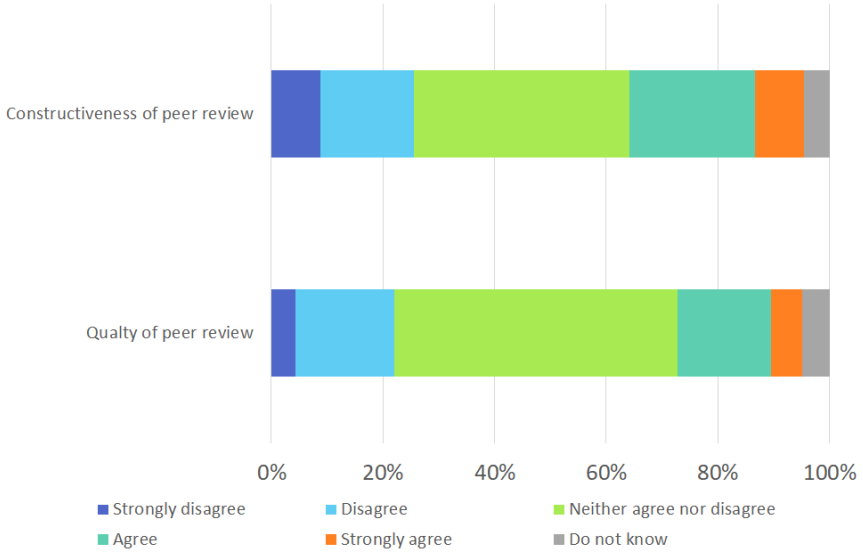


Figure 3.7: The extent to which respondents to the COVID-19 preprint author survey agreed that peer review of their COVID-19 research was more constructive and of higher quality than it normally is.

The latter finding seems to be in agreement with the results of a study by Horbach [27]. In this study, a qualitative comparison was made of review reports and editorial decision letters for COVID-19 and non-COVID-19 articles published in *BMJ* and *eLife*. It was found that review reports for COVID-19 articles were as thorough as review reports for non-COVID-19 articles. However, it was also observed that for COVID-19 articles reviewers and editors were less likely to ask for additional experiments than for non-COVID-19 articles. Reviewers and editors instead asked authors to tone down their conclusions and to acknowledge the limitations of their research. The observation that authors of COVID-19 articles are less likely to be requested to perform additional experiments is not confirmed by the findings of our survey. This observation may be specific to *BMJ* and *eLife* and may not generalise to other journals.

This analysis does raise a question, however, about whether there were compromises on quality involved in pushing through COVID-19 outputs as quickly as possible, even if they were subject to higher rejection rates. Early contributions to this debate have already appeared in the literature raising quality concerns [5, 14]. For example, Quinn et al. [42] used a variety of quality reporting measures assessed by pairs of expert reviewers on a sample of articles from key journals (i.e., *BMJ*, *Journal of the American Medical Association*, *the Lancet*, and *New England Journal of Medicine*) between February and May 2020, finding that "covid-19 research from the first wave of the pandemic was potentially of lower quality than contemporaneous non-covid research". Jung et al. [30], report similar findings from a sample of 686 COVID-19 articles compared with an historical control group of non-COVID-19 articles against various quality measures.

They report, "the accelerated publication of COVID-19 research was associated with lower study quality scores compared to previously published historical control studies". Further work is needed on this topic, especially as studies now available focus on the early stages of the pandemic.

3.4 Changes as a result of peer review

Our findings presented in Section 3.2 show that peer review played an important role as a filter to decide which COVID-19 articles were accepted for publication in a journal and which ones were rejected. In fact, rejection rates were substantially higher for COVID-19 articles than for non-COVID-19 articles.

"For COVID-19 articles that were accepted for publication, peer review may also have played an important role in improving the articles"

For COVID-19 articles that were accepted for publication, peer review may also have played an important role in improving the articles. If a COVID-19 article had been posted as a preprint before it was published in a journal, the journal version of the article may include important improvements compared to the preprint version. In this section, we analyse the extent to which peer review of COVID-19 articles did indeed lead to such improvements – another key dimension to the quality debate.

As Figure 3.8 shows, 21% of the respondents to our COVID-19 preprint author survey reported that they had made major changes to the discussion/conclusion section of their COVID-19 article as a result of comments provided by reviewers and editors. Almost 19% of the respondents had made major changes to the results section. For other sections, making major changes as a result of peer review was less common. In 12% of the cases, respondents had made no changes at all to the discussion/conclusion section. For other sections, this percentage is substantially higher.

Figure 3.9 shows that the most common reason for making major changes to a COVID-19 article was a request by reviewers and editors to include additional analyses from data already collected. 15% of the respondents reported that such a request had led to major changes to their article, and for another 36% of the respondents this had led to minor changes. 43% of the respondents had made changes to the conclusions drawn in their article. Likewise, 44% of the respondents had made changes to the discussion of the limitations of their research. However, in both cases, major changes were quite uncommon. Major changes to the conclusions had been made by only 3% of the respondents, and major changes to the limitations by only 4% of the respondents.

A study by Brierley et al. [6] found a somewhat higher percentage of COVID-19 articles for which peer review had resulted in major changes to the conclusions drawn by the authors. In this study, a manual comparison was made between the abstracts of COVID-19 preprints and the abstracts of the corresponding journal articles. In 15% of the cases, there were major differences between the conclusions drawn in the abstract of the preprint and the conclusions drawn in the abstract of the journal article. There were minor differences in 50% of the cases. The study also analysed non-COVID-19 articles, for which it was found that there were major differences in the conclusions of the preprint version and the journal version in only 6% of the cases.

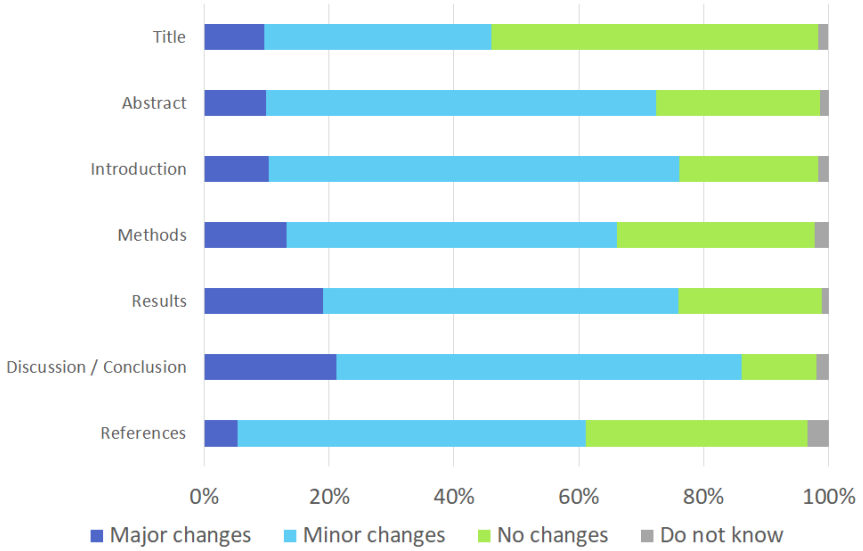


Figure 3.8: Percentage of respondents to the COVID-19 preprint author survey that had made no changes, minor changes, or major changes to different parts of their article in response to comments provided by reviewers and editors.

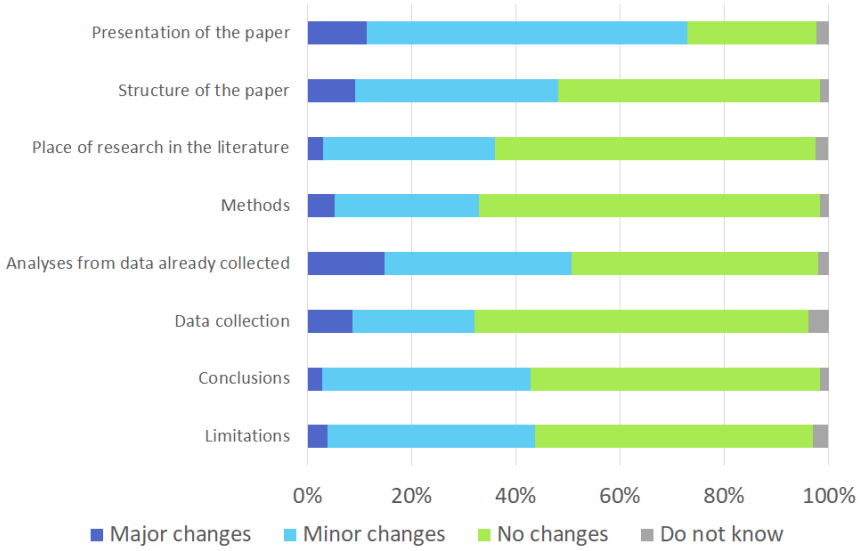


Figure 3.9: Percentage of respondents to the COVID-19 preprint author survey that had made no changes, minor changes, or major changes to different aspects of their article in response to comments provided by reviewers and editors.

In addition to their analysis of abstracts, Brierley et al. also presented an analysis of the number of figures and tables included in the preprint version and the journal version of COVID-19 articles. Based on these two analyses, Brierley et al. concluded that "preprints were most often passing into the 'permanent' literature with only minor changes to their conclusions, suggesting that the entire publication pipeline is having a minimal but beneficial effect".

4 Preprint peer review

Highlights

- The majority of COVID-19 preprint authors report a key motivation for posting preprints is wanting to receive feedback on their work.
- Over half of survey respondents reported receiving feedback on their preprint, but most feedback was 'closed' – in the form of emails or in meetings. Fewer authors received open feedback on the preprint server itself or on social media.
- A number of initiatives have been launched in recent years, some in response to the coronavirus pandemic, aiming to encourage and facilitate review of preprints.
- New services have developed novel approaches to organising peer review and acted as useful foci of debate and discussion of these approaches, but are still operating at a small scale.
- Faculty Opinions and Publons, two prominent services which enable commenting and rating of papers in various forms, also show low levels of commenting on COVID-19 outputs.
- The potential of new forms of reviewing, including reviewing preprints, is yet to be realised. Even during a public health crisis, usage of these services remained at low levels.

One of the objectives of the COVID-19 Rapid Review Initiative was to speed up dissemination and quality assurance of COVID-19 research by encouraging and facilitating peer review of preprints. Since preprint peer review is still in an early stage of development, this was perhaps one of the most ambitious goals of the Rapid Review Initiative. To what extent has the Rapid Review Initiative, and the scholarly communication system more generally, been successful in promoting preprint peer review?

To address this question, this chapter identifies a number of platforms and initiatives for preprint peer review. We then present two case studies which illustrate a number of important issues arising from early experimentation in this area, Outbreak Science PREreview and the MIT Press *Rapid Reviews COVID-19* 'overlay' journal. We go on to present data on other forms of 'informal peer review' carried out using the Faculty Opinions and Publons platforms, which allow commenting on both preprints and articles published in journals.

The literature on preprints often identifies the possibility of receiving feedback as one of the perceived benefits of preprinting [10]. This is usually assumed to be a kind of "informal peer review", or "'crowdsourced' review" as Paul Ginsparg, one of the founders of arXiv, has called it, which he sees as additional but complementary to "journal-

sourced peer-review" [21]. In this chapter, we draw on various strands of evidence which cast light on how these issues have been played out in the context of the pandemic.

4.1 Feedback on preprints

31% of the respondents to our survey of authors of COVID-19 preprints said they were motivated to preprint their work by wanting to receive feedback. A small majority of respondents, 53%, reported that they had received feedback in some form, a surprisingly high proportion. The most common channel for receiving feedback was a 'closed' one – by email. Of the respondents that had received feedback, 68% said the feedback had been given by email. This is consistent with previous studies (e.g., [45]). A further 21% reported receiving feedback in meetings with colleagues. More open channels for feedback were also reported. Of the respondents that had received feedback, 32% had received it on the preprint platform itself and 30% on social media. 5% reported receiving feedback posted on platforms for peer reviewing preprints. Publicly available comments on preprints have previously been observed to be relatively low on the preprint servers themselves [10]. Whilst our data tells a similar story, it also shows a much higher level of feedback being received by authors in more closed ways – still useful but not part of the public scholarly discourse.

"31% of the respondents to our survey of authors of COVID-19 preprints said they were motivated to preprint their work by wanting to receive feedback"

The nature of the feedback received varied considerably. As Figure 4.1 shows, the most common type/category of feedback was "comments which do not resemble peer reviewer comments e.g. thanks for the paper, retweets etc" (44%). 29% of the respondents had received feedback suggesting areas for further research, and a further 26% had received brief feedback (e.g., correcting a mistake). A smaller number of respondents (23%) had received more detailed feedback on the research presented in the preprint.

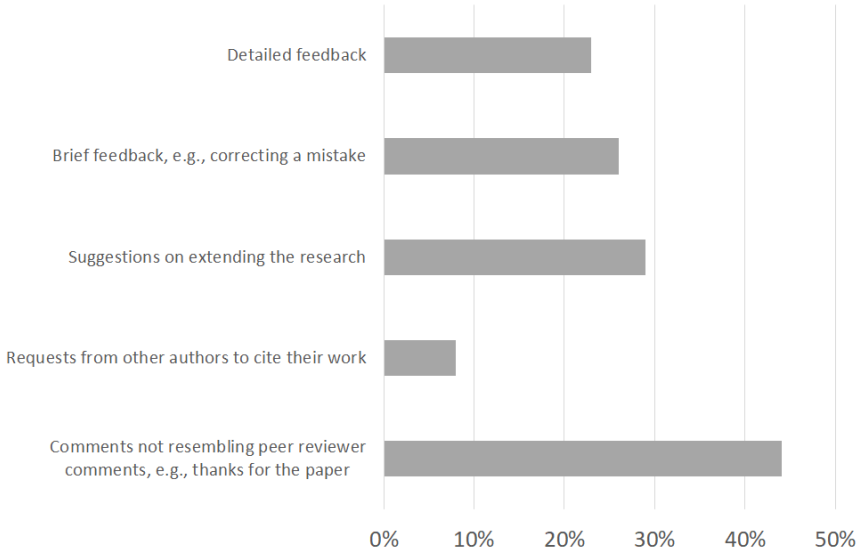


Figure 4.1: Feedback received by preprint authors.

The experiences of these authors show that preprints can be a useful way of prompting feedback. The channels through which feedback was received varied considerably and the nature of that feed-

back also varied. One way to assess the usefulness of feedback is the extent to which it has prompted changes to the preprint. Major changes to preprints were uncommon, although 8% of those who received feedback and answered this question said they had made major changes to the discussion/conclusion section of their paper. 28% reported making minor changes to this section, the most common section changed. 5% reported making major changes to the results section of their paper, and 25% reported making minor changes to that section. Figure 4.2 shows the extent of other changes.

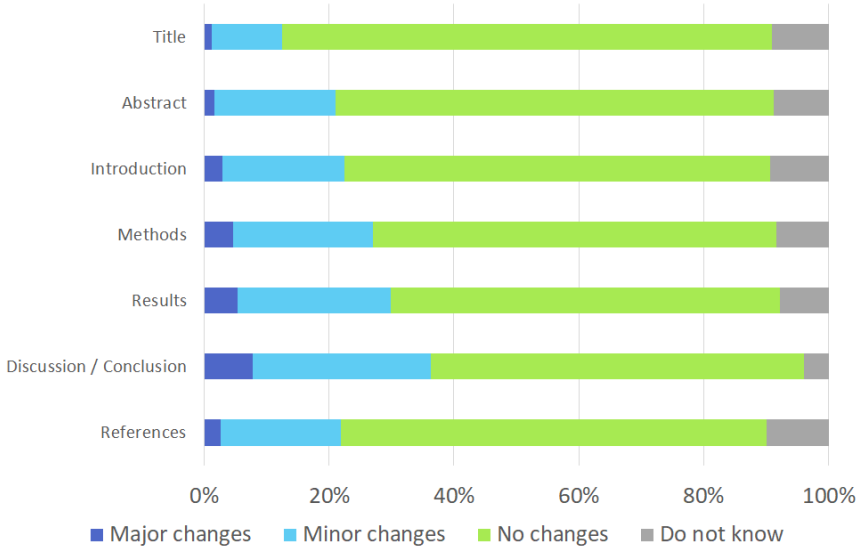


Figure 4.2: The extent to which feedback on a preprint prompted changes to different parts of the paper.

Because of the apparent usefulness of feedback to preprints, a number of innovative services have been created, some of them during the pandemic, to provide venues and standardised approaches to reviewing preprints.

4.2 Preprint peer review initiatives

"Services and initiatives set up to facilitate reviewing preprints have emerged in recent years"

Services and initiatives set up to facilitate reviewing preprints have emerged in recent years, in a range of different disciplinary fields and with various aims and approaches. Some were launched prior to the pandemic, others have been specifically developed to support the sharing and use of COVID-19 literature. We will discuss two of these, Outbreak Science PREreview and the MIT Press *Rapid Reviews: COVID-19* overlay journal, in more detail here, including the way in which the services have been set up, the key decisions they have made in approaching their work, and some of the challenges they have faced.

An overview of such initiatives can be found on the ReimagineReview website, a registry of innovative platforms and experiments around peer review of scientific outputs, including peer review of preprints. It is clear from the information provided by RemagineReview that such platforms and experiments aim to improve peer review in various areas, such as its speed, quality, transparency, incentivisation, and fairness. Services employ different types of quality assessment (e.g.,

free-form commenting, badges, or quantitative scores), new patterns of communication (e.g., interaction between authors and reviewers, public commenting, or review of code or small parts of a manuscript). They are operated in some cases by newly established groups, and in other cases by pre-existing organisations, such as publishers or professional societies.

Some initiatives have attempted to address the challenge of providing access to credible preprints relating to COVID-19. In addition to Outbreak Science PREreview and the *Rapid Reviews COVID-19* overlay journal, projects like the 2019 Novel Coronavirus Research Compendium (NCRC) and Sciety have been developed to summarise evidence and aggregate scientific content on COVID-19.

The NCRC, from Johns Hopkins University and other institutions globally, was developed to "rapidly curate and assess emerging research" in order "to provide accurate, relevant information for global public health action by clinicians, public health practitioners, and policy makers". Describing this process, they state, "experts work in eight teams to summarize the papers selected into the compendium, describe its value added based on what is already known about SARS-CoV-2 and COVID-19, and write a summary of the key findings relevant for action or practice". The papers covered may be published in journals or available in preprint servers. The compendium is designed to provide added value based on what is already known about SARS-CoV-2 and COVID-19, including summaries of the key findings that are relevant for action or clinical practice.

Sciety aggregates evaluations from across the web and brings together evaluation of preprints and curation in one place. Its scope includes COVID-19 content but extends beyond that. Sciety's mission is "to grow a network of researchers who evaluate, curate and consume scientific content in the open." Sciety currently aggregates the work of 16 preprint evaluation groups, from different scientific fields (i.e., NCRC, ScreenIT, PREreview, PeerJ, Review Commons, eLife, preLights, Rapid Reviews: COVID-19, ASAPbio Crowd Review, and Biophysics Colab) and six peer communities (i.e., PCI Zoology, PCI Evolutionary Biology, PCI Ecology, PCI Animal Science, PCI Archaeology, and PCI Paleo).

Infrastructure for preprint peer review is still very much in development. For example, recently the bioRxiv and medRxiv preprint servers also started to perform an aggregative function, providing links to third-party sites reviewing preprints. bioRxiv's "dashboard", introduced in 2021, provides links to reviews and discussion on the preprint from a wide variety of sources.

"Infrastructure for preprint peer review is still very much in development"

More time will be needed to track the various innovations, and to assess their sustainability and impact. They have created services which have potential value in the scholarly communication system, but their value still needs to be tested through further use and evaluation. Moreover, by creating new forms of reviews, these services also raise questions about how such reviews should be produced and used. We can see some of these issues, and how they have been addressed, in the case studies of two initiatives presented here: Outbreak Science PREreview (Box 4.1) and the MIT Press *Rapid Reviews: COVID-19* overlay journal (Box 4.2).

Box 4.1: Case study – Outbreak Science PREreview.

PREreview is an innovative platform which aims to facilitate peer reviewing of preprints independently of reviews associated with journals. It is funded by a number of organisations, including Alfred P. Sloan Foundation, eLife, Mozilla Foundation and Wellcome. COVID-19 preprints can be reviewed on PREreview in two main ways: 'rapid' and 'full' reviews. The rapid review process uses a 12-point review form, in which reviewers are asked to assess a paper against set questions, covering areas such as the preprint's novelty, reproducibility, and rigour. It also covers areas such as data and code availability. All of these are assessed against a simple 'yes/no/unsure' set of options. Full reviews, on the other hand, are more like conventional journal peer reviews, with detailed comments about the paper. These reviews stand alone and are independent of any journal publication venue, although journal editors are invited to make use of material on PREreview to facilitate their journal's peer review process. Multiple reviews can be contributed for single preprints. Members of the platform can request reviews of preprints.

One interesting feature of PREreview is the way it handles reviewer transparency and accountability. Authors of preprints are named, but reviewers are able to maintain a level of anonymity, if they wish. Reviewers are asked to create two "personas" when they register, one with their full details, which they have the option to make public, and another which is anonymous. Reviewers can switch between personas at any time. This means that reviewers are able to review a preprint anonymously, something which is designed to protect particularly early career researchers against the possibility of negative consequences of being critical of the work of others. "Unconstructive and disrespectful" reviews are, however, removed and individual's accounts deactivated where the code of conduct is violated. Anonymous reviewers may still request confirmation of their contributions to the service for various credit-related purposes, and so do not necessarily forgo recognition for their work entirely. As of 28 November 2021, the platform contained records for 360 COVID-19 preprints, with 182 full reviews, 172 rapid reviews, and 310 requests for reviews.

PREreview has created the 'COVID-19 dashboard' as part of the COVID-19 Rapid Review Initiative. This lists COVID-19 preprints available on different platforms, whilst at the same time highlighting them for review. It also identifies data availability for preprints. Reviewers are able to register so they will receive notification of new preprints which are added. The dashboard was also designed to inform journal editors about characteristics of COVID-19 preprints, which may help to filter papers during the journal submission process. The PREreview process provides indications of the quality of a paper and also whether the paper has openly available data and code associated with it.

The COVID-19 Rapid Review Initiative carried out an informal survey of editors of journals published by organisations participating in the initiative, results of which were published in a blog in July 2021. There were 31 responses, mainly from Hindawi and Royal

Society editors. 42% of them said that they would consider using reviews of preprints to help in their triage of manuscripts, with a further 36% replying "maybe". 23% responded in the negative. However, worries were expressed by respondents about the reliability of such reviews by self-selected reviewers. Trust was a major issue, with respondents wishing to know more about the reviewers in order to trust their review. Respondents expressed a preference for knowing information such as the discipline of the reviewer (65%), career level (55%), affiliation and publications record (both with 45%). These responses demonstrate a certain openness of journal editors to the innovative PREreview system, but also considerable caution about using it, views which are in many ways typical of those relating to innovation in the peer review area.

Box 4.2: Case study – MIT Press *Rapid Reviews: COVID-19* overlay journal.

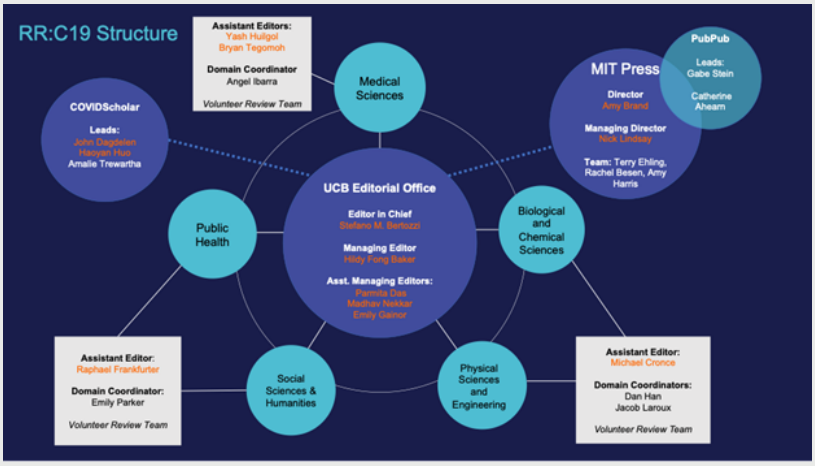
By Nick Lindsay (MIT Press)

In spring 2020, with generous funding from the McGovern Foundation, MIT Press conceived and developed *Rapid Reviews: COVID-19* (*RR:C19*), an open access overlay journal that publishes peer reviews of COVID-19 preprints. Senior leadership at the Press felt there was a clear need to combat misinformation in the public and academic spheres that was leaking in from unreviewed preprints and that changes in how preprints were being used by journalists and policymakers were creating new challenges for publishers.

The Knowledge Futures Group's PubPub was selected as the publishing platform for *RR:C19*. Stefano Bertozzi, dean emeritus of the School of Public Health at the University of California, Berkeley, was recruited to be the editor-in-chief and Hildy Fong Baker, executive director of the UC Berkeley Center for Global Public Health, agreed to become the managing editor. Bertozzi and Fong Baker quickly worked to pull together a novel publishing apparatus to handle a broad range of COVID-19 preprints across five domains: public health, medical sciences, biological and chemical sciences, physical sciences and engineering, and social sciences and humanities. Employing at times up to 70 graduate students and postdocs, the team managed to efficiently and effectively elevate preprints for review out of the tens of thousands available on bioRxiv and elsewhere. COVID-Scholar, an artificial intelligence developed by the Lawrence Berkeley National Laboratory, assisted with the effort and allowed for more fine-grained sifting of the literature to spot preprints for review.

RR:C19 was launched at the end of June 2020 and the first reviews were posted on 10 August 2020. As of August 2021, *RR:C19* has published nearly 500 peer reviews of COVID-19 preprints. The journal has received extensive media coverage in major outlets and contributed to the debunking of high-profile claims related to the origins of the coronavirus. New York Times

science reporter Carl Zimmer met with MIT Press and UC Berkeley at the end of April 2021 to further discuss how we can better work with the science journalism community and support their efforts at combating misinformation. The journal also presented a unique opportunity to educate graduate students and postdocs on issues around scholarly publishing and also draw them into the day-to-day work of scholarly communication. UC Berkeley established two new courses at the School of Public Health (PB HLTH 298 and a directed-group study course entitled "Academic Research & Publishing in the Age of COVID-19") that have provided training and feedback for the student volunteers.



The editorial office operations for *RR:C19* are substantially more labour intensive than a typical journal and this is largely driven by the need to search for preprints rather than process submissions as they come in. *RR:C19* volunteers identify preprints with the potential to advance new and important findings related to COVID-19. *RR:C19* relies on a ground-up effort through which students and early-career researchers lead a critical "pre" peer review process focused on the scraping of preprint servers weekly, and advance important or controversial papers for urgent peer review. Volunteers attend a weekly domain meeting with their assistant editor and domain coordinators, where they discuss the week's important/relevant preprints and provide input on which should potentially be sent for peer review.

From the publisher perspective, the development of new funding models is the most pressing issue facing the ongoing development and flourishing of publish/review/curate models such as *RR:C19*'s. The nature of the outputs from overlay journals presents few opportunities for developing a business model that can successfully offset costs associated with editorial office operations, marketing, and technology investment and maintenance. A multi-stream funding effort that incorporates funders, universities, government agencies, and other institutions may be well positioned to build upon and extend what has started with *RR:C19*.

The *RR:C19* model, which does not request permission to review from authors prior to peer review, also has presented some challenges with regard to norms around scholarly communica-

tion. A small number of authors expressed concern that the process may result in their paper being disqualified for publication at traditional journals. We have seen no evidence so far that this is the case and are optimistic that this is evidence of overlay journals being accepted as a welcome addition by the broader scholarly communications community. We are hopeful that other journals can use *RR:C19* peer reviews to enhance their own peer review processes and perhaps suggest alternate avenues for improvement of articles.

4.3 Commenting: Faculty Opinions and Publons

The potential usefulness of reviewing preprints is apparent from the case studies presented, but other forms of 'informal peer review' have also developed in recent years, allowing, for example, commenting and rating of papers in various forms, including both preprints and papers published in journals. To assess the use of such platforms in relation to COVID-19 research, we chose Faculty Opinions and Publons, two representatives of scholarly recommendation sites and peer review platforms. Faculty Opinions, formerly known as F1000Prime, is a scholarly recommendation site publishing recommendations of articles in the fields of biology and medicine made by over 8,000 'faculty members'. The recommendations act as a kind of post-publication peer review from the peer-nominated researchers.Publons is a peer review platform for both pre-publication and post-publication peer review. The former refers to peer reviews "commissioned by a journal or conference during a manuscript's path to publication (or not)", while the latter refers to peer reviews written by the readers of an article, which are not considered for a journal.

"The recommendations act as a kind of post-publication peer review from the peer-nominated researchers"

For the same dataset we studied in Chapter 2, which consists of 148,072 peer-reviewed COVID-19 outputs and 39,329 COVID-19 preprints, we collected the peer review comments received on Faculty Opinions and Publons and recorded by Altmetric. Table 4.1 presents the number of outputs with peer review comments.

Table 4.1: Number of COVID-19 outputs with Faculty Opinions recommendations and Publons peer review comments.

	Outputs	Faculty Opinions	Publons
Peer-reviewed outputs	148,072	603 (0.4%)	715 (0.5%)
Preprints	39,329	38 (0.0%)	473 (1.2%)
All outputs	187,401	641 (0.3%)	1,188 (0.6%)

Overall, the number of COVID-19 outputs with Faculty Opinions recommendations and Publons peer review comments is rather low, with only 0.3% of the COVID-19 outputs being recommended by Faculty Opinions and only 0.6% having peer review comments on Publons. In comparison with preprints, peer-reviewed outputs are more likely to receive Faculty Opinions recommendations, but less likely to receive Publons peer review comments.

To conclude, although some COVID-19 research was highlighted on Faculty Opinions and Publons, the overall share is very limited. It seems that the potential of these services is yet to be realised.

5 Academic and social attention

Highlights

- Peer-reviewed COVID-19 outputs published in the beginning of the pandemic have received a lot of attention in terms of citations, Twitter mentions, and mentions in news media. For more recent COVID-19 research, the level of attention on Twitter and in news media is lower but still substantially above the level for non-COVID-19 research.
- COVID-19 preprints have received a lot of attention on Twitter and in news media. The academic attention that preprints received in terms of citations is substantial, but much lower than that of peer-reviewed outputs.
- The higher the level of attention received by a peer-reviewed COVID-19 output, in terms of citations, Twitter mentions, or mentions in news media, the more likely the output is to have a preprint. Nevertheless, many peer-reviewed outputs with a high level of attention do not have a preprint.
- Some retracted COVID-19 outputs have received a lot of attention in terms of citations, Twitter mentions, or mentions in news media, both before and after the retraction. Researchers keep citing retracted COVID-19 outputs.

In the preceding chapters, we have analysed the production and peer review of COVID-19 outputs. To develop a deeper understanding of the dissemination of COVID-19 research, we now turn to the attention given to this research. What has been the 'academic attention' and 'social attention' paid to COVID-19 outputs, both peer-reviewed and non-peer-reviewed outputs? We address this question by analysing the number of times COVID-19 outputs have been cited by other scientific outputs and the number of times they have been mentioned on Twitter or in news media. We pay special attention to COVID-19 preprints and COVID-19 outputs that have been retracted.

Box 5.1: Methodological note – How to capture academic and social attention.

The analyses presented in this chapter are based on the same COVID-19 outputs also studied in Chapter 2. These outputs were published between January 2020 and April 2021. Data on these outputs was obtained from the Dimensions database, made available to us by Digital Science in May 2021. The citation statistics reported in this chapter are also based on Dimensions data. Statistics on the number of mentions of COVID-19 outputs on Twitter or in news media, as well as a number of other 'altmetric' statistics, are based on Altmetric data made available to us by Digital Science in June 2021. More detailed information on the Altmetric data is available on the Altmetric website.

5.1 Indicators of academic and social attention

Table 5.1 provides an overall picture of the extent to which peer-reviewed COVID-19 outputs have attracted attention from the scientific community and society at large. For ten different attention indicators, the table shows the percentage of COVID-19 outputs with at least one 'attention event', for instance at least one citation or at least one Twitter mention.

The three attention indicators with the highest percentage of peer-reviewed COVID-19 outputs with at least one event are mentions on Twitter (51.3%), citations in other scientific outputs (48.8%), and mentions in news media (14.9%). For each of the other attention indicators listed in Table 5.1, fewer than 10% of the COVID-19 outputs have at least one event.

The level of attention given to COVID-19 outputs is substantially higher than the typical attention given to global scientific outputs. For instance, Fang et al. [17] found that about one-third of the recent scientific outputs indexed in the Web of Science database had at least one Twitter mention, which is substantially below the 51.3% reported in Table 5.1. Likewise, Fang et al. found that 4% of the recent scientific outputs had been mentioned at least once in news media, which is much lower than the 14.9% reported in Table 5.1.

"The level of attention given to COVID-19 outputs is substantially higher than the typical attention given to global scientific outputs"

In the rest of this chapter, our focus will be on the academic attention given to COVID-19 outputs in terms of citations, and the social attention given to COVID-19 outputs in terms of Twitter mentions and news mentions. The other attention indicators listed in Table 5.1 capture a smaller number of attention events and therefore will not be considered in more detail in this chapter.

5.2 Peer-reviewed outputs

The peer-reviewed COVID-19 outputs that received the most attention are different when viewed through the lens of citations, Twitter mentions, and news mentions (see Tables B.1–B.3 in Appendix B). In terms of the academic attention represented by citations, the COVID-19 outputs that received the most attention from the scientific community were all published at the beginning of the pandemic. Most of them are related to the clinical characteristics of infected patients during the initial COVID-19 outbreak in Wuhan, China. In contrast, the most tweeted COVID-19 outputs include both outputs published in the beginning of the pandemic and more recent outputs. Discussions on Twitter focus on outputs that study the effectiveness of countermeasures to the pandemic, such as lockdowns, facemasks, and vaccines. Similarly, the COVID-19 outputs that attracted the most attention in news media also include both older and more recent outputs. Compared to citations and Twitter mentions, news mentions focus more strongly on outputs that study the transmission means of the virus as well as the safety and efficacy of vaccines. Notably, one out of the top ten outputs with the most news attention has been retracted.

In addition to differences in the research topics with the most attention, the three categories of attention events also exhibit different

Table 5.1: Percentage of peer-reviewed COVID-19 outputs with at least one attention event, for ten different attention indicators.

Indicator	Definition	Outputs with at least one event
Citations	Citations received by a scholarly output recorded by Dimensions	48.8%
Twitter mentions	Tweets that contain a direct link to a scholarly output, including original tweets, reply tweets, quote tweets, and retweets	51.3%
News mentions	News mentions of a scholarly output in a curated list of news media tracked by Altmetric	14.9%
Blog citations	Blog citations to a scholarly output on a curated list of blogs tracked by Altmetric	9.2%
Facebook mentions	Facebook posts that contain a direct link to a scholarly output, including only posts in the Facebook public pages tracked by Altmetric	8.7%
Reddit mentions	Original posts that contain a direct link to a scholarly output in all subreddits	3.2%
Policy document citations	Policy document citations to a scholarly output in a curated list of policy sources tracked by Altmetric	2.4%
Video mentions	Video mentions of a scholarly output in the description section of videos posted by a curated list of YouTube channels tracked by Altmetric	1.5%
Wikipedia citations	Wikipedia citations to a scholarly output in Wikipedia pages written in certain languages (i.e., English, Spanish, Portuguese, Dutch, Finnish, Swedish)	1.4%
Q&A mentions	Posts on StackExchange that contain a link to a scholarly output	0.1%

levels of speed in response to newly published peer-reviewed COVID-19 outputs. As shown in Figure 5.1, discussions of COVID-19 outputs on Twitter typically took place in the first two months after the publication of an output. Likewise, attention in news media was received mostly in the first few months after the publication of an output. For citations the pattern is very different. It was relatively uncommon for COVID-19 outputs to be cited in the first one or two months after their publication. However, while after a few months these outputs received almost no further attention on Twitter and in news media, they kept receiving citations, suggesting a fundamental difference between the way in which scientific knowledge on COVID-19 was absorbed by the scientific community and the way in which it influenced society at large.

"On average, outputs published in the first months of the pandemic received a higher level of attention than more recent outputs"

For each month between January 2020 and April 2021, Figure 5.2 shows the percentage of peer-reviewed COVID-19 outputs published in that month that have received at least 1, 10, or 100 citations, Twitter mentions, or news mentions. On average, outputs published in the first months of the pandemic received a higher level of attention than more recent outputs. The level of attention was highest for outputs published in February and March 2020, just after COVID-19 had been declared a Public Health Emergency of International Concern by the

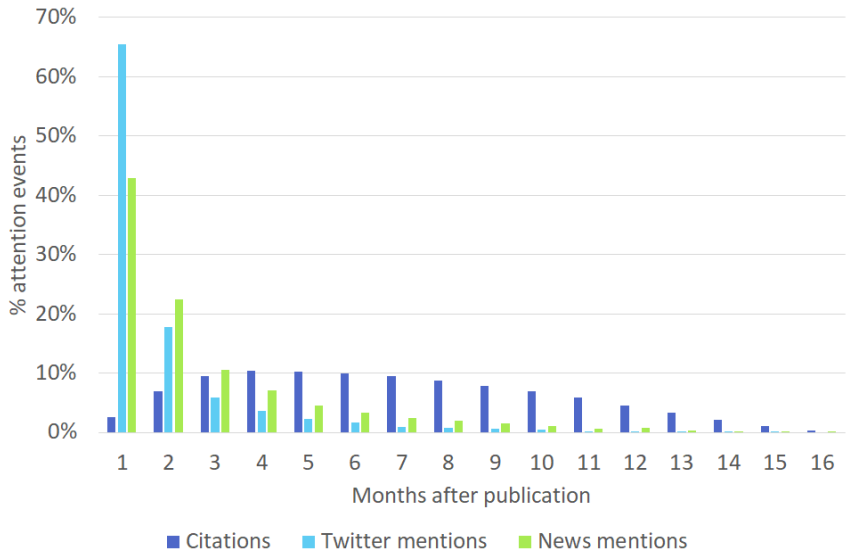


Figure 5.1: Distribution of the time between the publication of a peer-reviewed COVID-19 output and subsequent attention events.

WHO on 30 January 2020. The percentage of more recent outputs that have received a substantial number of citations is low, but this percentage can still be expected to increase, as these outputs are still receiving additional citations from newly published research. While the broader societal interest in COVID-19 research seems to have declined somewhat over time, the level of attention on Twitter and in news media remains relatively high.

5.3 Preprints

A number of COVID-19 preprints have also generated a great deal of academic and social interest. The high level of attention received by COVID-19 preprints on Twitter and in news media is particularly striking (see Tables B.5 and B.6 in Appendix B), even though the level of attention is not as high as for peer-reviewed outputs. While the most cited COVID-19 preprints have received substantial numbers of citations (see Table B.4), their citation impact is an order of magnitude lower than for peer-reviewed outputs.

"The high level of attention received by COVID-19 preprints on Twitter and in news media is particularly striking"

How much attention did COVID-19 preprints receive compared to peer-reviewed outputs? To address this question, we consider pairs of a COVID-19 preprint and the corresponding peer-reviewed output, taking into account only those pairs for which the preprint and the peer-reviewed output together have received at least 100 citations, Twitter mentions, or news mentions. We classify each pair of a preprint and a peer-reviewed output into one of five categories based on the percentage of citations, Twitter mentions, or news mentions received by the preprint (i.e., 0–20%, 20–40%, 40–60%, 60–80%, and 80–100%) of attention events received by the preprint version). Figure 5.3 presents the distribution over the five categories.

Take the pairs in which the preprint contributed less than 20% (i.e., the category 0–20%) to the total attention events as an example. Figure 5.3 shows that 69% of the pairs fall into this category in terms of

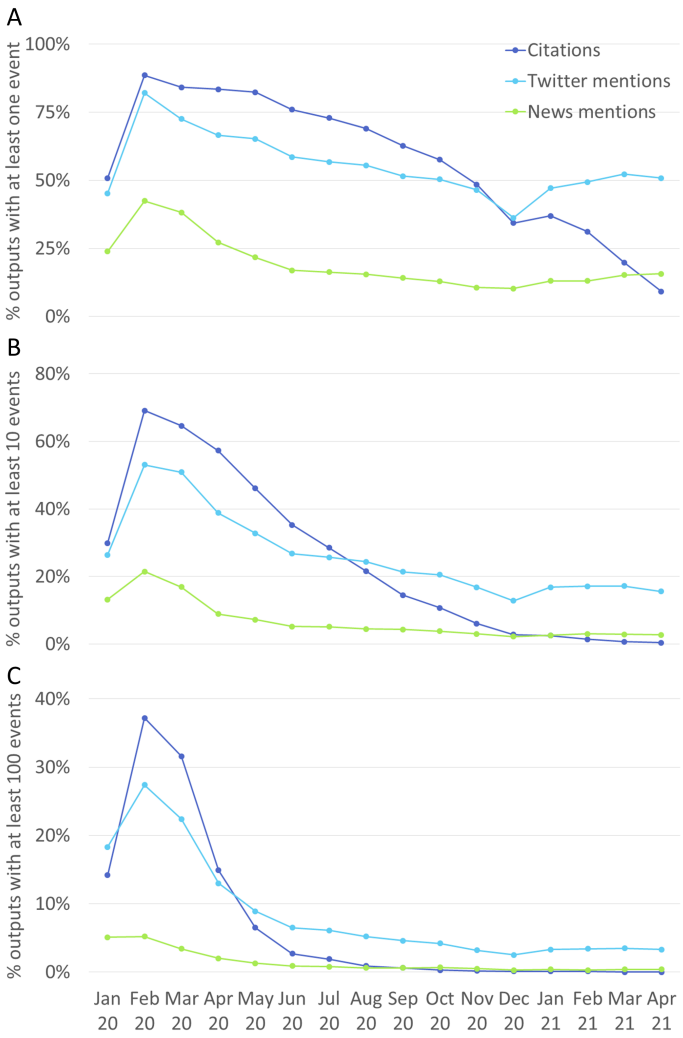


Figure 5.2: Monthly percentage of peer-reviewed COVID-19 outputs with at least 1 (panel A), 10 (panel B), or 100 (panel C) attention events.

citations, indicating that the majority of citations were contributed by peer-reviewed outputs instead of their preprints. Mentions in news media exhibited a similar preference for peer-reviewed literature. In 55% of the cases, the preprint has received less than 20% of the total number of news mentions received by the preprint and the corresponding peer-reviewed output. In contrast, mentions on Twitter were more strongly focused on preprints. Only one-third of the studied pairs have the preprints contributing to less than 20% of Twitter mentions, whereas nearly half of the pairs have at least 40% of Twitter mentions received by the preprint versions, suggesting the more visible role that preprints played on Twitter compared to citations and news media.

"Mentions in news media exhibited a preference for peer-reviewed literature"

As discussed in Section 2.4, only a small share of all peer-reviewed COVID-19 outputs have a corresponding preprint. In our data, a link to a preprint is available for only 5% of the peer-reviewed COVID-19 outputs. While it is disappointing that the commitment to preprinting made at the beginning of the pandemic has not been fulfilled, it might be that it is more common for outputs with more attention received to have a preprint. If this is indeed the case, this could be considered to alleviate the problem of the low level of preprinting.

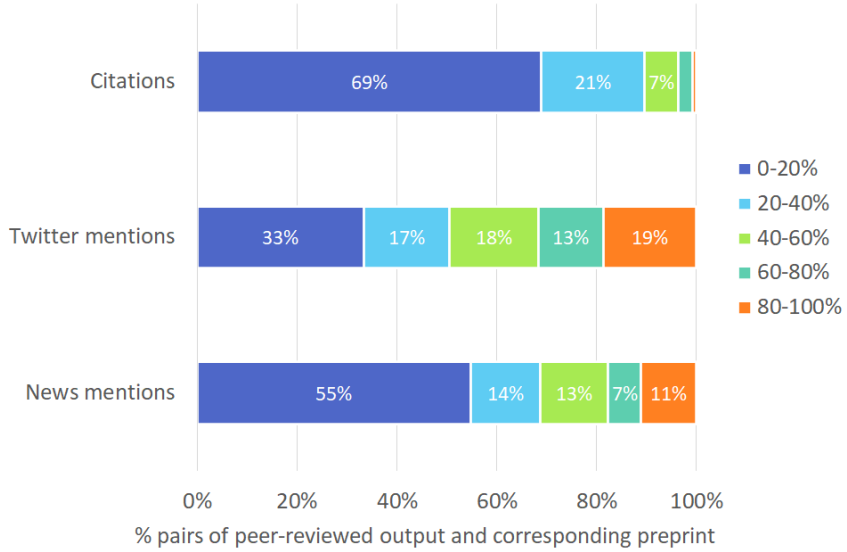


Figure 5.3: Distribution of citations, Twitter mentions, and news mentions between COVID-19 preprints and the corresponding peer-reviewed outputs. Each pair of a preprint and a peer-reviewed output belongs to one of five categories based on the percentage of citations, Twitter mentions, or news mentions received by the preprint. The chart shows the distribution over the five categories.

To determine whether peer-reviewed COVID-19 outputs with higher levels of attention are more likely to have a preprint, we focus on outputs published between April and June 2020. For x ranging from 0 to 100%, Figure 5.4 shows the percentage of the top x% outputs with the most attention that have a preprint. The higher the level of attention received by an output, in terms of citations, Twitter mentions, or mentions in news media, the more likely the output is to have a preprint. This effect is strongest for citations. For instance, while only 5% of all outputs have a preprint, this is the case for 12% of the 20% most cited outputs and for 22% of the 1% most cited outputs.

Figure 5.4 shows a correlation between the attention received by a peer-reviewed COVID-19 output and the availability of a preprint. However, the correlation is not very strong. Even among the outputs reaching the highest level of attention, a large majority do not have a preprint.

"The higher the level of attention received by an output, in terms of citations, Twitter mentions, or mentions in news media, the more likely the output is to have a preprint"

5.4 Retractions

Concerns have been raised about the quality of some COVID-19 research, both non-peer-reviewed research posted on preprint servers and peer-reviewed research published in journals [5, 14, 30, 42]. To provide some insight into problems related to low-quality COVID-19 research, this section presents an analysis of retracted COVID-19 outputs and the attention given to these outputs. According to Retraction Watch there is no evidence that the level of retractions for COVID-19 research is higher than average. However, there is a danger that retracted COVID-19 research may have received more attention than other retracted outputs because of the circumstances of the pandemic, although this in turn may have led to their more rapid

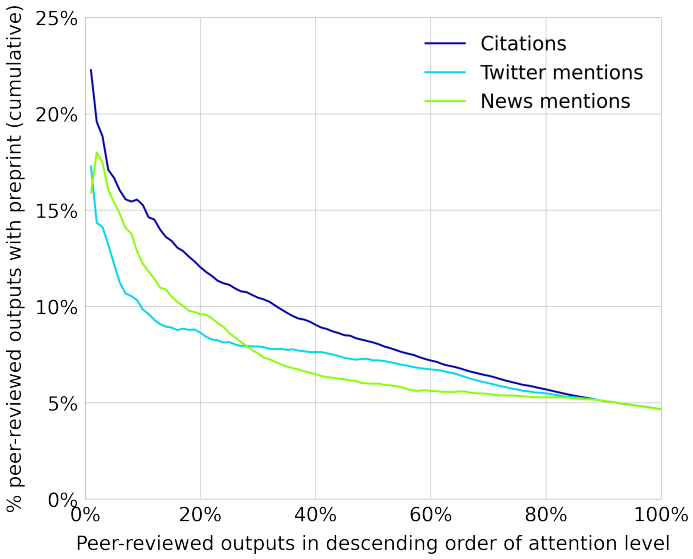


Figure 5.4: Percentage of peer-reviewed COVID-19 outputs with a preprint (vertical axis) among the top x% outputs with the most attention (horizontal axis).

retraction than for other literature [16, 39].

Box 5.2: Methodological note – How to identify retracted COVID-19 outputs.

Retraction Watch maintains a list of COVID-19 outputs, both peer-reviewed outputs and preprints, that have been retracted (or withdrawn). On 5 August 2021, Retraction Watch kindly shared with us DOIs and other metadata for 143 retracted COVID-19 outputs. In 16 cases, the retraction had resulted from a mistake made by the publisher, such as the accidental duplicate publication of an article. We disregarded these 16 cases. Our analysis focuses on the remaining 127 retracted COVID-19 outputs.

Some of the retracted COVID-19 outputs, including peer-reviewed outputs and preprints, have received considerable attention in both the scientific community and in society at large (see Tables B.7–B.9 in Appendix B). One of the retracted outputs topped the list in terms of citations, Twitter mentions, and news mentions. This is an article in *the Lancet* evaluating the effectiveness of hydroxychloroquine or chloroquine for treatment of COVID-19 (see also Section 2.5). Of all retracted COVID-19 outputs, this one has the largest number of mentions in news media, the second largest number of mentions on Twitter, and the second largest number of citations.

"Some of the retracted COVID-19 outputs, including peer-reviewed outputs and preprints, have received considerable attention"

For the 127 retracted COVID-19 outputs included in our data, Figure 5.5 shows the distribution of the time between publication and retraction of an output. Approximately 40% of the retractions took place within 50 days after publication, but there are also retractions that took place more than 200 days after publication. Of course, it is likely that some COVID-19 outputs that have not yet been retracted will be retracted in the future.

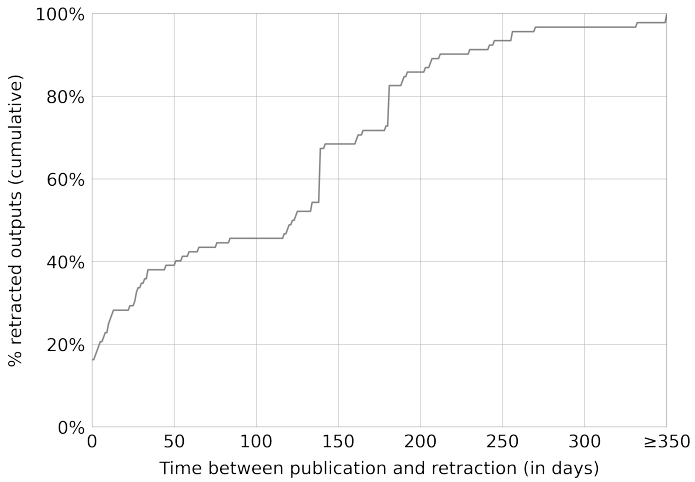


Figure 5.5: Distribution (cumulative) of the time between the publication of a COVID-19 output and the subsequent retraction of the output.

The COVID-19 outputs listed in Tables B.7–B.9 have received a lot of attention in terms of citations, Twitter mentions, and mentions in news media. In some cases most attention was received before the retraction. In other cases attention was mostly received after the retraction.

Figure 5.6 presents the overall picture for all 127 retracted COVID-19 outputs. It shows the temporal distribution of citations, Twitter mentions, and mentions in news media of retracted COVID-19 outputs, with time defined relative to the date of retraction. Twitter mentions mostly took place before the retraction of an output. Presumably outputs were being debated and criticised on Twitter before being retracted. Mentions in news media took place both shortly before and shortly after the retraction of an output, possibly because news media first covered the debate about controversial COVID-19 research and then covered the retraction resulting from the debate. Citations show a completely different pattern. Some citations were received before an output was retracted, but most citations were actually received after the retraction of an output. Moreover, these citations were received not only shortly after the retraction, but also many months later, when one may have expected the citing researchers to be aware of the retraction.

"Outputs were being debated and criticised on Twitter before being retracted"

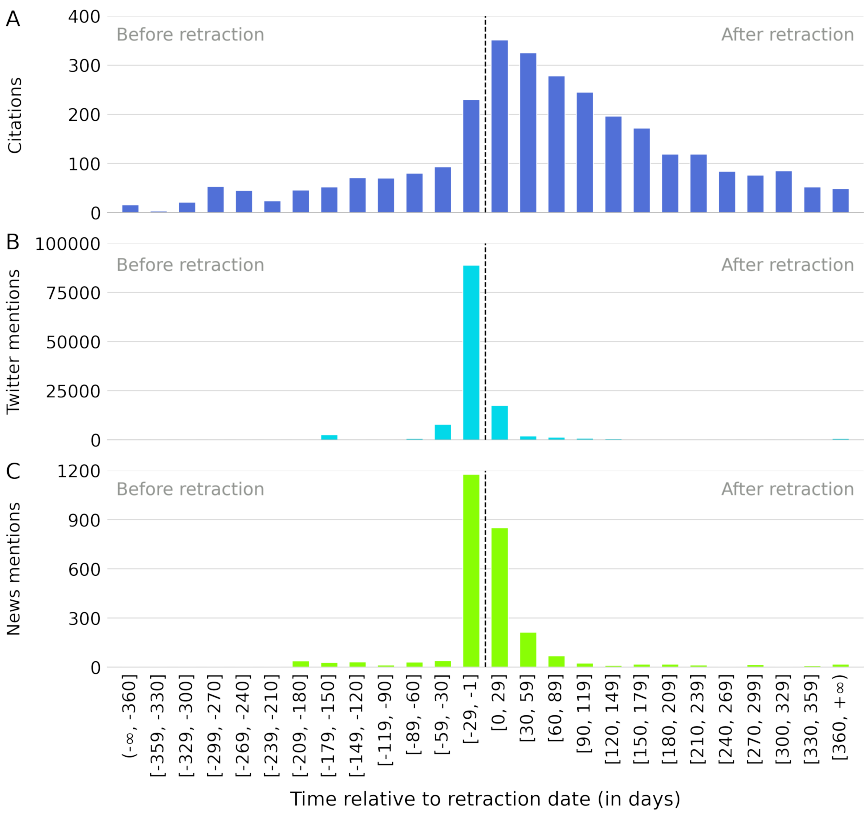


Figure 5.6: Temporal distribution of citations (panel A), Twitter mentions (panel B), and mentions in news media (panel C) of retracted COVID-19 outputs, with time defined relative to the date of retraction.

6 Conclusions and recommendations

This report has attempted to pick out some of the main contours of the scholarly communication landscape during the coronavirus pandemic so far. In this final chapter we present a number of recommendations which arise from our analysis. We have grouped these around the commitments made by publishers and other organisations in the early stages of the pandemic: making COVID-19 research outputs open or free access, preprinting of COVID-19 research, sharing of data from COVID-19 research, speeding up publication times of COVID-19 articles, and facilitating peer review of COVID-19 preprints. We discuss the extent to which these commitments have been fulfilled and we recommend actions that need to be taken to implement them in future. We also present a number of more general recommendations stemming from our work. We believe that the developments of the last two years, as the research community has responded to the global crisis, have created opportunities which now need to be grasped in order to effect ongoing positive change in the scholarly communication system.

One key point has clearly emerged from the pandemic: the importance of openness – open access, open data, and open science more widely. The pandemic has illustrated the power of openness in enabling a concerted global response to the emergency. The rapid and open sharing of the SARS-CoV-2 genomic sequencing data, along with opening up of the majority of the COVID-19 literature, are open science success stories of the pandemic. Open science is not new, of course, but the experience of COVID-19 has further strengthened the case for more systematic and more widespread adoption of open practices beyond the immediate crisis (see also [4]). Numerous other contemporary challenges are all just as likely to benefit from greater openness: the global climate emergency, the fight against cardiovascular diseases and cancers, and combating world poverty, to name just a few. The recently adopted UNESCO Recommendation on Open Science make this point very clearly. It is important not to lose sight of this big picture. The recommendations offered below therefore are not restricted to COVID-19 research but apply to research in general.

6.1 Recommendations on open or free access

The commitment made at the beginning of the pandemic to ensure that COVID-19 outputs are available in an open-access or free-to-read form has largely been realised, at least for the duration of the pandemic so far. Some of the material made freely accessible during the pandemic may be withdrawn from free circulation by publishers, but the extent and timing of this remains to be seen. There are also publishers who did not commit to make COVID-19 outputs openly or freely accessible. We argue, however, that an emergency of the sort we have experienced should mean all relevant material should be made openly or at least freely accessible.

- **Recommendation 1: Publishers** should make all research related to major public health crises openly or freely accessible. Ultimately, **publishers and other stakeholders in the scholarly communication system** should find ways to ensure that all publicly-funded research is made open access.

Enacting this recommendation will help to ensure that in case of emergencies anyone in the world has access to all relevant scientific knowledge. The open availability of research apart from emergencies will also help to address major global challenges and achieve public good.

6.2 Recommendations on preprinting

COVID-19 research has been preprinted more often than non-COVID-19 research, especially in medical fields, but even for COVID-19 research the proportion of peer-reviewed outputs with a preprint is low. **The commitment to make COVID-19 research available in preprint form has not, therefore, been realised.** The need for more concerted action in this area is clear if large-scale adoption of preprinting is going to be achieved.

- **Recommendation 2: Funders, governmental organisations, research institutions, and publishers** should review their policies in relation to preprinting, with a view to mandating, at least in the case of emergencies and possibly more generally.
- **Recommendation 3: Funders, governmental organisations, and research institutions** should reward researchers for posting preprints of their work, for instance in evaluation procedures.
- **Recommendation 4: Publishers, preprint servers, and workflow providers** should facilitate posting of preprints by incorporating it in the submission workflows of journals.
- **Recommendation 5: Leaders of disciplinary communities**, such as scholarly societies, journal editors and others, should advocate and practice preprinting.

These recommendations illustrate the need for coordinated action across the scholarly communication system, an important theme of this report. Between them, these recommendations are likely to increase the incentives for authors to preprint their work, through a combination of mandates and rewards, and make it easier for them to do so as part of their workflows.

6.3 Recommendations on data sharing

Early and ongoing data sharing of the SARS-CoV-2 genome sequences has clearly been successful in combating the pandemic, but overall, evidence available suggests that sharing of COVID-19 research data has remained relatively low. **The data sharing commitment made in the early days of the pandemic has not, therefore, been realised.** The drive towards achieving more widespread data sharing is still in

its early stages and requires ongoing focus amongst various stakeholders. Whilst it is recognised that data sharing may not always be possible, the principle that data should be "as open as possible and as closed as necessary" should guide future policy development, with this being clearly communicated across the research system. Data policies and guidelines for authors need further harmonisation across publishers and other stakeholders.

- **Recommendation 6: Funders, governmental organisations, research institutions, and publishers** should require journal articles to include a data availability statement.
- **Recommendation 7: Funders, governmental organisations, research institutions, and publishers** should collaborate to align their data sharing policies around the principle of "as open as possible and as closed as necessary", to use common data policy templates, to require data sets and software to be posted to a trusted FAIR-enabling repository, and to require formal citations to data sets and software.
- **Recommendation 8: Funders, governmental organisations, research institutions, and publishers** should ensure data policies are registered, discoverable by humans and machines, citable via a DOI, transparent as to what they recommend, and comparable one to another.

These recommendations are designed to make data sharing easier for researchers, with clear and consistent guidance on how to do so, and why. Enacting these recommendations would also enhance the standards, tools and infrastructures around data sharing, developments that are essential if the value of open data is to become apparent in the way science is conducted. These recommendations can contribute to wider developments to make the case for open data and to demonstrate its benefits.

6.4 Recommendations on speeding up publication times

Analysis of the response of the journal peer review system to the COVID-19 pandemic yields a generally positive picture, albeit with caveats. On average, the time from submission to publication has been substantially shorter for COVID-19 articles than for similar non-COVID-19 articles. Many journals managed to speed up publication times. Transferable peer review across publishers, set up in the COVID-19 Rapid Review Initiative, was not taken up by authors [29], suggesting that other innovations or solutions may need to be considered. The current peer review system remains under strong pressure in particular in recruiting peer reviewers, and solving this will require innovation, rather than expecting the existing systems, processes and actors to simply work harder. The system has also seen a number of high-profile retractions. Whilst retractions might be portrayed positively as science self-correcting, in a more open environment publishing defective work can do damage, even if retracted later. Openness can, on the other hand, help by making peer review reports publicly available so that the level of quality assurance that an article has undergone is visible.

- **Recommendation 9: Publishers** should increase the speed and efficiency of peer review by experimenting with new collaborative approaches to peer review, for instance inspired by initiatives such as Review Commons.
- **Recommendation 10: Publishers** should give serious consideration to making the peer review reports of the articles they publish openly available in order to be transparent about the level of quality assurance an article has undergone.

It is important to build on the initiatives taken during the pandemic to address the pressures on the scholarly communication system as it responds to ever-increasing demands. These involve working in more innovative ways, not just expecting the system to work harder, whilst at the same time aiming to make the system more transparent.

6.5 Recommendations on peer review of preprints

Significant innovation has taken place in the area of peer reviewing preprints, but this remains experimental and small-scale. Preprint peer review platforms have very limited resources. Additional investment in these platforms are needed to scale up their operations and to develop best practices for preprint peer review. Approaches to combining or integrating preprint peer review and journal peer review also need to be considered.

- **Recommendation 11: Funders and other stakeholders in the scholarly communication system** should make additional investments in platforms and workflows for peer reviewing preprints in order to explore their value and develop an evidence base around their potential.
- **Recommendation 12: Preprint peer review platform providers, preprint servers, and publishers** should work together to maximise the value of peer reviews of preprints, for instance by developing new approaches to combining or integrating preprint peer review and journal peer review.

There are significant opportunities here, as has been illustrated during the pandemic, but in order for them to be further explored there needs to be commitment to the further development of these new innovative platforms and services. Integrating these new ways of working with existing processes and systems will also help to encourage their use. These recommendations will enable new ways of working to become embedded in disciplinary communities and their role more widely discussed and tested.

6.6 General recommendations

In exploring the response of the scholarly communication system to the pandemic, we have identified a number of more general issues which require further action by stakeholders in the system. These relate to the evidence base, coordination of action, and the widening of

COVID-19-specific initiatives to other parts of the scholarly communication system. We are conscious that implementing these recommendations is likely to be challenging, not least because they require cross-sectoral and international collaborations. However, we believe they are of crucial importance, and we encourage organisations and groups across the scholarly communication system to engage in dialogue and partnership working at different levels in taking these recommendations forward.

Obtaining robust evidence on the response of the scholarly communication system to the COVID-19 pandemic for this report has been challenging. Relevant data is often not available or is of limited quality. This, for instance, applies to data on the peer review status and the open access status of an article, and also to data on links between different versions of an article, and between articles and data sets. To enable robust evidence-informed approaches to innovation in scholarly communication, stakeholders in the scholarly communication system need to intensify their joint efforts to improve the availability and quality of this type of data.

- **Recommendation 13: Publishers and other scholarly communication organisations** should intensify their joint efforts to improve the availability and quality of data and metadata on scholarly publishing, compiled and made available in a standards-based way, allowing for robust evidence-informed approaches to innovation in scholarly communication.

Addressing the challenges faced by the scholarly communication system is a responsibility not only of publishers and other scholarly communication organisations, but also of other stakeholders in the research system, including funders and research organisations. In the early stages of the pandemic, many different stakeholders expressed their support for the Wellcome-coordinated COVID-19 statement, but it seems there was limited concerted action across stakeholders to implement the commitments made. For example, whilst different stakeholders jointly made a commitment to preprint all COVID-19 research, they did not make a collaborative effort to implement this commitment in a systematic way. To address the challenges related to, for instance, preprinting and data sharing, there is a need for more sustained collaboration across stakeholders.

- **Recommendation 14: All stakeholders in the research system** should recognise that improving scholarly communication is a joint responsibility that requires collaboration and coordinated action across stakeholders, including the development of policies with accompanying monitoring and accountability mechanisms.

Many of the initiatives that have emerged or been strengthened in the scholarly communication system during the COVID-19 pandemic are likely to be of general value in improving scholarly communication beyond the pandemic. In particular, the pandemic has highlighted the importance of open science, especially in providing access to publications, data sets and other research outputs as early and as widely as possible. We recommend wherever possible that approaches and initiatives demonstrated to be beneficial during the pandemic are continued and their application broadened beyond the pandemic. At the

same time, it is important to learn the lessons of the pandemic in how the scholarly communication system can best respond to emergencies. This will help stakeholders to mobilise more quickly in similar circumstances in the future.

- **Recommendation 15: All stakeholders in the scholarly communication system** should review activities initiated in response to the pandemic and explore how those with positive outcomes can become more widely embedded across the system as a whole. In particular, efforts to promote open science, like open access of published outputs and open sharing of data, need to be further intensified.
- **Recommendation 16: All stakeholders in the scholarly communication system** should capture the lessons of the pandemic and identify ways in which the system can mobilise more quickly and effectively in the event of future global emergencies.

6.7 Conclusion

The coronavirus pandemic has created an unprecedented global crisis. The research system in general, and the scholarly communication system in particular, have been put under enormous strain. This has, however, created opportunities – opportunities to initiate and accelerate innovations in producing and disseminating research. The pandemic has shown that rapid change is possible, and that collaboration across the scholarly communication system is key to enable that change. Learning the lessons of the COVID-19 emergency for the scholarly communication system, we now have the opportunity to create long-lasting positive change beyond the pandemic. Responding during the time of crisis has been of crucial importance, and it is now equally important to ensure we create and sustain a better system for the future.

Acknowledgements

We are grateful to everyone who has made a contribution to this report. The following colleagues, grouped by organisation, have offered valuable input for the report, for instance by sharing data, by answering questions and providing information, or by contributing to case studies:

- Digital Science: Ian Calvert, Mario Diwersy, Daniel Hook, and Brian Kirkegaard Lunn
- eLife: Fred Atherden
- Elsevier: Andrew Plume, Nicoline van der Linden, and Etienne Vignola-Gagné
- F1000: Liz Allen and James Barker
- Hindawi: Alessandra Auddino and Sarah Theissen
- Our Research: Richard Orr
- PLOS: Emily Chenette and Geno Urbano
- Research Square: Michele Avissar-Whiting
- Retraction Watch: Ivan Oransky
- Springer Nature: Can Becleyici, Nick Campbell, Ed Gerstner, Andrew Needham, Magdalena Skipper, and Rembert Wohlers
- SSRN: Shirley Decker-Lucke

We also would like to thank the following colleagues for the specific contributions they have made:

- Contribution to the COVID-19 Rapid Review Initiative, in particular to the reporting group: Peter Binfield, Ian Caswell, Joerg Heber, Michael Markie, Daniela Saderi, and Jon Treadway
- Testing of the preprint author survey: Giovanni Colavizza, João Eurico da Fonseca, Serge Horbach, John Inglis, Jessica Polka, Ruy Ribeiro, and Nicolas Robinson-Garcia
- Feedback on draft version of the report: Michele Avissar-Whiting, Daniel Hook, Hannah Hope, Rob Johnson, Wolfgang Kaltenbrunner, Robert Kiley, Jessica Polka, Richard Sever, Robert Terry, and James Wilsdon
- Communication: Ruth Francis and Philippa Shelton
- Cover illustration: Anthony Lewis and Vanessa Lovegrove
- Layout of the report: Daniel Hook

Author contributions

Ludo Waltman: Co-led the conceptualisation and organisation of the research, led the scientometric analysis, and contributed to other analysis; co-led the drafting, reviewing, and finalisation the report.

Stephen Pinfield: Co-led the conceptualisation and organisation of the research, led the survey analysis, and contributed to other analysis; co-led the drafting, reviewing, and finalisation the report.

Narmin Rzayeva: Carried out the scientometric analysis on scientific outputs, and quantitative analysis of survey data; contributed advice, writing, and editing of drafts of the report; contributed to data curation.

Susana Oliveira Henriques: Carried out qualitative analysis of survey data, drafted some case studies; contributed advice, writing, and editing of drafts of the report; contributed to data curation.

Zhichao Fang: Carried out the scientometric analysis on attention, and original drafting in that area; contributed advice, writing, and editing of drafts of the report.

Johanna Brumberg: Contributed to the conceptualisation and organisation of the research.

Sarah Greaves: Advised on framework for collection of publication and peer review data; contributed advice, writing, and editing of drafts of the report.

Phil Hurst: Coordinated contribution of data from the Royal Society; advised on framework for collection of publication and peer review data; contributed advice, writing, and editing of drafts of the report.

Andy Collings: Coordinated contribution of data from eLife; advised on framework for collection of publication and peer review data; contributed advice, writing, and editing of drafts of the report.

Arianne Heinrichs: Coordinated contribution of data from Springer Nature; advised on framework for collection of publication and peer review data; contributed advice, writing, and editing of drafts of the report.

Nick Lindsay: Contributed advice, writing, and editing of drafts of the report.

Catriona J. MacCallum: Coordinated contribution of data from Hindawi; advised on framework for collection of publication and peer review data; co-led the the data policy work; contributed advice, writing, and editing of drafts of the report.

Dan Morgan: Coordinated contribution of data from PLOS; advised on framework for collection of publication and peer review data; contributed advice, writing, and editing of drafts of the report.

Susanna-Assunta Sansone: Co-led the data policy work; contributed advice, writing, and editing of drafts of the report.

Sowmya Swaminathan: Coordinated contribution of data from Springer Nature; advised on framework for collection of publication and peer

review data; contributed advice, writing, and editing of drafts of the report.

Competing interests

Ludo Waltman, Stephen Pinfield, Andy Collings, and Susanna-Assunta Sansone are affiliated with organisations that receive funding from Wellcome.

Phil Hurst, Andy Collings, Arianne Heinrichs, Nick Lindsay, Catriona J. MacCallum, Dan Morgan, Susanna-Assunta Sansone, and Sowmya Swaminathan are affiliated with organisations whose response to the COVID-19 pandemic is analysed in this report.

Catriona J. MacCallum is on the advisory board for the Experiments in Peer Review project of the Research on Research Institute (RoRI).

Sowmya Swaminathan is on the advisory board for Research Square and serves as editorial advisor for *In Review*.

Funding information

Ludo Waltman and Stephen Pinfield were supported by Wellcome [221297/Z/20/Z] as part of its core funding of the Research on Research Institute (RoRI).

Susanna-Assunta Sansone was supported by Wellcome [212930/Z/18/Z] as part of its funding of the FAIRsharing service.

Data availability

The data presented in the figures in this report is available in figshare [51].

The data of the survey of authors of COVID-19 preprints is available in figshare [51], except for the free-text responses, which may contain sensitive information.

The Dimensions and Altmetric data used in this report cannot be shared publicly, but Digital Science welcomes requests for free access to Dimensions data for research purposes. Crossref, PubMed, and Unpaywall data is publicly available.

The data made available by eLife, Hindawi, PLOS, and Royal Society for the analysis of submission-to-publication times (see Section 3.1) is available in figshare [51]. The data made available by Springer Nature cannot be shared publicly.

To protect the confidentiality of peer review, the data made available by eLife, Hindawi, PLOS, and Royal Society for the analysis of rejection rates and submission-to-final-decision times (see Section 3.2) cannot be shared publicly.

All data made available by eLife, Hindawi, PLOS, Royal Society, and Springer Nature for the analyses presented in this report is archived at Leiden University subject to institutional requirements. It can be made available to third parties requesting access for verification or replication purposes.

A Members of the COVID-19 Rapid Review Initiative

AfricArXiv
Bill & Melinda Gates Foundation
BMC
Cambridge University Press
Copyright Clearance Center
eLife
F1000
FAIRsharing
GigaScience
Hindawi
Life Science Alliance
The MIT Press
Open Access Scholarly Publishers Association (OASPA)
Oxford University Press
PeerJ
PLOS
PREreview
Research on Research Institute (RoRI)
The Royal Society
SSRN
Ubiquity Press
UCL Press

B Detailed attention statistics

Table B.1: Top 10 peer-reviewed COVID-19 outputs with the largest number of citations.

Title	Journal	Pub. date	Citations
Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China	The Lancet	24/01/20	19,967
Clinical characteristics of coronavirus disease 2019 in China	New England Journal of Medicine	28/02/20	13,096
A novel coronavirus from patients with pneumonia in China, 2019	New England Journal of Medicine	24/01/20	11,561
Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study	The Lancet	11/03/20	11,462
Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China	JAMA	17/03/20	11,178
Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study	The Lancet	30/01/20	9,426
A pneumonia outbreak associated with a new coronavirus of probable bat origin	Nature	03/02/20	8,588
Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China	JAMA	07/04/20	7,798
Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia	New England Journal of Medicine	29/01/20	7,634
SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor	Cell	05/03/20	7,045

Table B.2: Top 10 peer-reviewed COVID-19 outputs with the largest number of Twitter mentions.

Title	Journal	Pub. date	Twitter mentions
In pursuit of PPE	New England Journal of Medicine	17/04/20	77,261
Effectiveness of adding a mask recommendation to other public health measures to prevent SARS-CoV-2 infection in Danish mask wearers: A randomized controlled trial	Annals of Internal Medicine	18/11/20	73,374
Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: An interim analysis of a randomised controlled phase 3 trial in Russia	The Lancet	02/02/21	56,878
Assessing mandatory stay at home and business closure effects on the spread of COVID 19	European Journal of Clinical Investigation	06/01/21	55,146
Post-lockdown SARS-CoV-2 nucleic acid screening in nearly ten million residents of Wuhan, China	Nature Communications	20/11/20	49,696
Tripartite combination of candidate pandemic mitigation agents: Vitamin D, quercetin, and estradiol manifest properties of medicinal agents for targeted mitigation of the COVID-19 pandemic defined by genomics-guided tracing of SARS-CoV-2 targets in human cells	Biomedicines	21/05/20	47,121
COVID-19 outpatients: Early risk-stratified treatment with zinc plus low-dose hydroxychloroquine and azithromycin: A retrospective case series study	International Journal of Antimicrobial Agents	26/10/20	46,966
Covid-19: Politicisation, "corruption," and suppression of science	BMJ	13/11/20	43,464
Respiratory virus shedding in exhaled breath and efficacy of face masks	Nature Medicine	03/04/20	40,108
Treatment of 5 critically ill patients with COVID-19 with convalescent plasma	JAMA	28/04/20	39,502

Table B.3: Top 10 peer-reviewed COVID-19 outputs with the largest number of mentions in news media.

Title	Journal	Pub. date	News mentions
Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1	New England Journal of Medicine	17/03/20	3,549
Mental health, substance use, and suicidal ideation during the COVID-19 pandemic —- United States, June 24–30, 2020	Morbidity and Mortality Weekly Report	14/08/20	2,446
Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: A preliminary report of a phase 1/2, single-blind, randomised controlled trial	The Lancet	20/07/20	2,300
High SARS-CoV-2 attack rate following exposure at a choir practice – Skagit County, Washington, March 2020	Morbidity and Mortality Weekly Report	15/05/20	1,912
Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine	New England Journal of Medicine	17/02/21	1,892
Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis	The Lancet	01/06/20	1,775
Temporal dynamics in viral shedding and transmissibility of COVID-19	Nature Medicine	15/04/20	1,772
COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020	Emerging Infectious Diseases	02/04/20	1,733
Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine	New England Journal of Medicine	10/12/20	1,646
Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis	The Lancet	22/05/20	1,630

Table B.4: Top 5 COVID-19 preprints with the largest number of citations.

Title	Preprint server	Pub. date	Citations
Clinical characteristics of 2019 novel coronavirus infection in China	medRxiv	09/02/20	998
Severe acute respiratory syndrome-related coronavirus: The species and its viruses -- A statement of the Coronavirus Study Group	bioRxiv	11/02/20	730
Exposure to air pollution and COVID-19 mortality in the United States	medRxiv	07/04/20	566
Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions	medRxiv	24/01/20	552
Efficacy of hydroxychloroquine in patients with COVID-19: Results of a randomized clinical trial	medRxiv	30/03/20	548

Table B.5: Top 5 COVID-19 preprints with the largest number of Twitter mentions.

Title	Preprint server	Pub. date	Twitter mentions
COVID-19 antibody seroprevalence in Santa Clara County, California	medRxiv	17/04/20	28,670
Indoor transmission of SARS-CoV-2	medRxiv	07/04/20	24,021
Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag	bioRxiv	31/01/20	19,269
Effect of hydroxychloroquine in hospitalized patients with COVID-19: Preliminary results from a multi-centre, randomized, controlled trial	medRxiv	15/07/20	19,148
Efficacy of hydroxychloroquine in patients with COVID-19: Results of a randomized clinical trial	medRxiv	30/03/20	17,529

Table B.6: Top 5 COVID-19 preprints with the largest number of mentions in news media.

Title	Preprint server	Pub. date	News mentions
Outcomes of hydroxychloroquine usage in United States veterans hospitalized with Covid-19	medRxiv	21/04/20	750
Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera	bioRxiv	07/01/21	740
Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70	bioRxiv	14/12/20	740
COVID-19 antibody seroprevalence in Santa Clara County, California	medRxiv	17/04/20	723
Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2	bioRxiv	30/04/20	697

Table B.7: Top 5 retracted COVID-19 outputs with the largest number of citations.

Title	Journal or preprint server	Pub. date	Ret. date	Citations
Cardiovascular disease, drug therapy, and mortality in Covid-19	New England Journal of Medicine	01/05/20	04/06/20	668
Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis	The Lancet	22/05/20	04/06/20	571
Clinical manifestations and outcome of SARS-CoV-2 infection during pregnancy	Journal of Infection	05/03/20	31/01/21	314
Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China	medRxiv	11/02/20	21/02/20	313
Obesity and mortality of COVID-19. Meta-analysis	Obesity Research & Clinical Practice	09/07/20	06/02/21	101

Table B.8: Top 5 retracted COVID-19 outputs with the largest number of Twitter mentions.

Title	Journal or preprint server	Pub. date	Ret. date	Twitter mentions
Facemasks in the COVID-19 era: A health hypothesis	Medical Hypotheses	22/11/20	03/05/21	35,891
Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis	The Lancet	22/05/20	04/06/20	33,854
Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag	bioRxiv	31/01/20	02/02/20	16,481
Hydroxychloroquine plus azithromycin: A potential interest in reducing in-hospital morbidity due to COVID-19 pneumonia (HI-ZY-COVID)?	medRxiv	11/05/20	19/05/20	9,039
Potential false-positive rate among the 'asymptomatic infected individuals' in close contacts of COVID-19 patients	Chinese Journal of Epidemiology	05/03/20	09/03/20	4,142

Table B.9: Top 5 retracted COVID-19 outputs with the largest number of mentions in news media.

Title	Journal or preprint server	Pub. date	Ret. date	News mentions
Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis	The Lancet	22/05/20	04/06/20	1,630
Cardiovascular disease, drug therapy, and mortality in Covid-19	New England Journal of Medicine	01/05/20	04/06/20	308
Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag	bioRxiv	31/01/20	02/02/20	150
mRNA vaccines to prevent COVID-19 disease and reported allergic reactions: Current evidence and approach	Journal of Allergy and Clinical Immunology In Practice	31/12/20	31/12/20	127
Seeding of outbreaks of COVID-19 by contaminated fresh and frozen food	bioRxiv	18/08/20	13/03/21	60

References

[1] Shir Aviv-Reuven and Ariel Rosenfeld. "Publication patterns' changes due to the COVID-19 pandemic: A longitudinal and short-term scientometric analysis". In: *Scientometrics* 126.8 (2021), pp. 6761–6784. DOI: 10.1007/s11192-021-04059-x.

[2] Jeffrey Barrett. *Scientists sharing Omicron data were heroic. Let's ensure they don't regret it*. The Guardian. 2021. URL: https://www.theguardian.com/commentisfree/2021/nov/28/scientists-sharing-omicron-date-were-heroic-lets-ensure-they-dont-regret-it.

[3] Carol K. Bates, Jeff Jackson, and Steven Asch. "From the editor's desk: JGIM and COVID-19". In: *Journal of General Internal Medicine* 35.6 (2020), pp. 1627–1628. DOI: 10.1007/s11606-020-05852-9.

[4] Lonni Besançon et al. "Open science saves lives: Lessons from the COVID-19 pandemic". In: *BMC Medical Research Methodology* 21 (2021), p. 117. DOI: 10.1186/s12874-021-01304-y.

[5] Katrina A. Bramstedt. "The carnage of substandard research during the COVID-19 pandemic: A call for quality". In: *Journal of Medical Ethics* 46.12 (2020), pp. 803–807. DOI: 10.1136/medethics-2020-106494.

[6] Liam Brierley et al. *Preprints in motion: Tracking changes between preprint posting and journal publication during a pandemic*. bioRxiv. 2021. DOI: 10.1101/2021.02.20.432090.

[7] Guillaume Cabanac, Theodora Oikonomidi, and Isabelle Boutron. "Day-to-day discovery of preprint–publication links". In: *Scientometrics* 126.6 (2021), pp. 5285–5304. DOI: 10.1007/s11192-021-03900-7.

[8] Ewen Callaway. "Will the pandemic permanently alter scientific publishing?" In: *Nature* 582 (2020), pp. 167–168. DOI: 10.1038/d41586-020-01520-4.

[9] "Calling all coronavirus researchers: Keep sharing, stay open". In: *Nature* 578 (2020), p. 7. DOI: 10.1038/d41586-020-00307-x.

[10] Andrea Chiarelli et al. "Preprints and scholarly communication: An exploratory qualitative study of adoption, practices, drivers and barriers". In: *F1000Research* 8 (2019), p. 971. DOI: 10.12688/f1000research.19619.2.

[11] Annie Collins and Rohan Alexander. *Reproducibility of COVID-19 pre-prints*. 2021. arXiv: 2107.10724.

[12] European Commission. *Six recommendations for implementation of FAIR practice by the FAIR in practice task force of the European open science cloud FAIR working group*. 2020. URL: https://op.europa.eu/s/u6VD.

[13] European Commission. *Turning FAIR into reality*. 2018. URL: https://op.europa.eu/s/u6VB.

[14] Claudia C. Dobler. "Poor quality research and clinical practice during COVID-19". In: *Breathe* 16.2 (2020), p. 200112. DOI: 10.1183/20734735.0112-2020.

[15] Michael B. Eisen et al. "Publishing in the time of COVID-19". In: *eLife* 9 (2020), e57162. DOI: 10.7554/elife.57162.

[16] Holly Else. "How a torrent of COVID science changed research publishing – in seven charts". In: *Nature* 588 (2020), p. 553. DOI: 10.1038/d41586-020-03564-y.

[17] Zhichao Fang et al. "An extensive analysis of the presence of altmetric data for Web of science publications across subject fields and research topics". In: *Scientometrics* 124.3 (2020), pp. 2519–2549. DOI: 10.1007/s11192-020-03564-9.

[18] Nicholas Fraser and Bianca Kramer. *COVID-19 preprints*. GitHub. 2021. URL: https://github.com/nicholasmfraser/covid19_preprints.

[19] Nicholas Fraser et al. "The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape". In: *PLOS Biology* 19.4 (2021), e3000959. DOI: 10.1371/journal.pbio.3000959.

[20] Nicholas Fraser et al. "The relationship between bioRxiv preprints, citations and altmetrics". In: *Quantitative Science Studies* 1.2 (2020), pp. 618–638. DOI: 10.1162/qss_a_00043.

[21] Paul Ginsparg. "Preprint déjà vu". In: *EMBO Journal* 35.24 (2016), pp. 2620–2625. DOI: 10.15252/embj.201695531.

[22] Konstantinos Gkiouras et al. "Tracing open data in emergencies: The case of the COVID-19 pandemic". In: *European Journal of Clinical Investigation* 50.9 (2020), e13323. DOI: 10.1111/eci.13323.

[23] Ian Le Guillou. *Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research*. Horizon: The EU Research & Innovation Magazine. 2020. URL: https://ec.europa.eu/research-and-innovation/en/horizon-magazine/covid-19-how-unprecedented-data-sharing-has-led-faster-ever-outbreak-research.

[24] J. Homolak, I. Kodvanj, and D. Virag. "Preliminary analysis of COVID-19 academic information patterns: A call for open science in the times of closed borders". In: *Scientometrics* 124.3 (2020), pp. 2687–2701. DOI: 10.1007/s11192-020-03587-2.

[25] Daniel Hook and Simon Porter. *How Covid-19 is changing research culture*. Research Professional News. 2020. URL: https://www.researchprofessionalnews.com/rr-news-uk-views-of-the-uk-2020-5-how-covid-19-is-changing-research-culture/.

[26] Daniel W. Hook et al. "Real-time bibliometrics: Dimensions as a resource for analyzing aspects of COVID-19". In: *Frontiers in Research Metrics and Analytics* 5 (2021). DOI: 10.3389/frma.2020.595299.

[27] Serge P. J. M. Horbach. "No time for that now! Qualitative changes in manuscript peer review during the COVID-19 pandemic". In: *Research Evaluation* 30.3 (2021), pp. 231–239. DOI: 10.1093/reseval/rvaa037.

[28] Serge P. J. M. Horbach. "Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19". In: *Quantitative Science Studies* 1.3 (2020), pp. 1056–1067. DOI: 10.1162/qss_a_00076.

[29] Phil Hurst and Sarah Greaves. "COVID-19 Rapid Review cross-publisher initiative: What we have learned and what we are going to do next". In: *Learned Publishing* 34.3 (2021), pp. 450–453. DOI: 10.1002/leap.1375.

[30] Richard G. Jung et al. "Methodological quality of COVID-19 clinical research". In: *Nature Communications* 12 (2021), p. 943. DOI: 10.1038/s41467-021-21220-5.

[31] Daniel S. Katz et al. "Recognizing the value of software: A software citation guide". In: *F1000Research* 9 (2021), p. 1257. DOI: 10.12688/f1000research.26932.2.

[32] Bianca Kramer. *COVID-19 preprints linked to published journal articles in Crossref metadata*. GitHub. 2021. URL: https://github.com/bmkramer/covid19_preprints_published.

[33] Ádám Kun. "Time to acceptance of 3 days for papers about COVID-19". In: *Publications* 8.2 (2020), p. 30. DOI: 10.3390/publications8020030.

[34] Diana Kwon. "How swamped preprint servers are blocking bad coronavirus research". In: *Nature* 581 (2020), pp. 130–131. DOI: 10.1038/d41586-020-01394-6.

[35] Francois Lachapelle. *COVID-19 preprints and their publishing rate: An improved method*. medRxiv. 2020. DOI: 10.1101/2020.09.04.20188771.

[36] Julien Larregue et al. *COVID-19: Where is the data?* LSE Impact Blog. 2020. URL: https://blogs.lse.ac.uk/impactofsocialsciences/2020/11/30/covid-19-where-is-the-data/.

[37] Rut Lucas-Dominguez et al. "The sharing of research data facing the COVID-19 pandemic". In: *Scientometrics* 126.6 (2021), pp. 4975–4990. DOI: 10.1007/s11192-021-03971-6.

[38] Lauren Maxwell et al. *FAIR, ethical, and coordinated data sharing for COVID-19 response: A review of COVID-19 data sharing platforms and registries*. Zenodo. 2021. DOI: 10.5281/zenodo.5642635.

[39] Ivan Oransky. *List of retracted COVID-19 papers grows past 70*. Retraction Watch. 2020. URL: https://retractionwatch.com/2020/12/30/list-of-retracted-covid-19-papers-grows-past-70/.

[40] Adam Palayew et al. "Pandemic publishing poses a new COVID-19 challenge". In: *Nature Human Behaviour* 4.7 (2020), pp. 666–669. DOI: 10.1038/s41562-020-0911-0.

[41] Heather Piwowar et al. "The state of OA: A large-scale analysis of the prevalence and impact of open access articles". In: *PeerJ* 6 (2018), e4375. DOI: 10.7717/peerj.4375.

[42] Terence J. Quinn et al. "Following the science? Comparison of methodological and reporting quality of COVID-19 and other research from the first wave of the pandemic". In: *BMC Medicine* 19 (2021), p. 46. DOI: 10.1186/s12916-021-01920-x.

[43] Anna Salleh. *Eddie Holmes wins Prime Minister's Prize for Science for work on COVID-19 and viral evolution*. ABC News. 2021. URL: https://www.abc.net.au/news/science/2021-11-03/prime-ministers-science-prize-eddie-edward-holmes-covid-19-virus/100552752.

[44] Susanna-Assunta Sansone et al. "FAIRsharing as a community approach to standards, repositories and policies". In: *Nature Biotechnology* 37.4 (2019), pp. 358–367. DOI: 10.1038/s41587-019-0080-8.

[45] Richard Sever et al. *bioRxiv: The preprint server for biology*. bioRxiv. 2019. DOI: 10.1101/833400.

[46] Dario Taraborelli. *How the COVID-19 crisis has prompted a revolution in scientific publishing*. Fast Company. 2020. URL: https://www.fastcompany.com/90537072/how-the-covid-19-crisis-has-prompted-a-revolution-in-scientific-publishing.

[47] Leho Tedersoo et al. "Data sharing practices and data availability upon request differ across scientific disciplines". In: *Scientific Data* 8 (2021), p. 192. DOI: 10.1038/s41597-021-00981-0.

[48] Robert Terry. *Open science – Time to unlock the potential of the digital age*. YouTube. 2021. URL: https://youtu.be/TrZrRcCoQSo.

[49] Timothy H. Vines et al. "Mandated data archiving greatly improves access to research data". In: *FASEB Journal* 27.4 (2013), pp. 1304–1308. DOI: 10.1096/fj.12-218164.

[50] Ludo Waltman and Nees Jan van Eck. "A new methodology for constructing a publication-level classification system of science". In: *Journal of the American Society for Information Science and Technology* 63.12 (2012), pp. 2378–2392. DOI: 10.1002/asi.22748.

[51] Ludo Waltman et al. *Scholarly communication in times of crisis: Supplementary data*. figshare. 2021. DOI: 10.6084/m9.figshare.17125529.

[52] Lucy Lu Wang et al. *CORD-19: The COVID-19 Open Research Dataset*. 2020. arXiv: 2004.10706.

[53] Mark D. Wilkinson et al. "The FAIR guiding principles for scientific data management and stewardship". In: *Scientific Data* 3 (2016), p. 160018. DOI: 10.1038/sdata.2016.18.

[54] Boya Xie, Zhihong Shen, and Kuansan Wang. *Is preprint the future of science? A thirty year journey of online preprint services*. 2021. arXiv: 2102.09066.

About RoRI

Worldwide, interest is intensifying in how research is funded, practiced and evaluated, and in how research cultures and systems can be made more efficient, open, inclusive and impactful. But research on research (RoR) – also called science of science, or meta-research – is often poorly joined-up.

Over the next decade, policymakers, funders, universities, publishers and researchers themselves are likely to require greater RoR capacity to navigate continued expansion of the research enterprise, accompanied by greater emphasis on interdisciplinarity, mission and challenge-directed research; open scholarship and open data; collaboration and team science; and creative solutions to the linked imperatives of diversity, inclusion, integrity and reproducibility.

To strengthen this field, Wellcome Trust, Digital Science and the Universities of Sheffield and Leiden have joined forces to establish the Research on Research Institute (RoRI), and to build an international consortium of funders, academics and technologists committed to transformative & translational RoR.

Find out more at www.researchonresearch.org

RoRi is a partnerhip of

Report partners

www.researchonresearch.org