

# **13 weeks in a pandemic: a descriptive study of Imperial College London's COVID-19 publications**

Robyn Price and Yusuf Ozkan

## **Introduction**

In a period of 13 weeks between 6<sup>th</sup> January 2020 to 5<sup>th</sup> April 2020, the coronavirus pandemic spread globally. We identified 41 outputs on the topic published in this period with Imperial authorship. The diversity of outputs found range from reports of statistical modelling, expert editorial commentary, research code, early original research and even applied practical and clinical skills items. The authors of these works chose appropriately diverse range of platforms with which to disseminate; with the first emerging works using the institutional repository Spiral for rapid communications, followed by preprint servers, peer reviewed journals and lastly to software communities. Finally, for research contributing to a question of perhaps unprecedented global need; the open access status varies across these 41 outputs. Crucially, 100% of the outputs are, at time of writing, free to read but for some licensing that ensures this in perpetuity or ensures translation and reuse is not present. The aim of this report is to describe the types and methods of output.

## **Executive Summary**

In a 13 week period, 41 outputs were published from a range of Imperial authors. From the groups of outputs defined: output type (Report, Preprint, Published Article, Code); non-peer reviewed and peer reviewed and open access status (CC-BY Gold, free access) a pattern can be seen.

Between Week 1 – Week 5 the outputs were entirely Reports published CC-BY. As Reports continued to be released, they were joined by Preprints in Week 6 – Week 8, also published either Gold CC-BY or under a specific preprint server licence permitting free access. From Week 9, Published Articles in journals began to appear with a mix of Gold OA outputs and introducing free access to the set. In Week 12, Code emerges for the first time as an output.

From the introduction of Preprints in Week 6 and Published Articles in Week 9 we can also see the types of articles being produced; Research outputs in the form of Preprints being first seen, followed by further Research and Editorial content following in Published Articles. Later in Week 12 we see the first emergence of 'Other' article types – clinical guidelines and clinical education outputs.

The open access status of outputs seems largely linked to their output type. 100% of the outputs were freely available at the time of writing, but 29% (12 outputs) of them were not published with licenses that guarantee this in perpetuity. This 29% are all Published Articles in subscription journals that the publisher has made free to read. There are no reuse rights associated with this. All of the rest of the outputs- Reports, Preprints and Code are open access either by a CC-BY licence or the preprint server's licence to host the article.

Of the platforms disseminating these outputs, 34% are published by Imperial's repository Spiral, with a following 27% published by other repositories and preprint servers. The remaining 39% are published by journal publishers, with Elsevier being the most frequent publisher of journals in this set.

## **Data collection methods**

A search performed on Dimensions for “COVID-19” as a keyword, filtered to “Imperial College London” as an affiliated Research Organisation. Dimensions search results include the term found anywhere in document Full Text. A second search in the Imperial College London CRIS, Symplectic Elements, for the term “COVID-19” in a publication’s keyword record or title. 5 publications found in this search were eliminated due to erroneous name based-matching by Symplectic. Both searches were performed on 2<sup>nd</sup> April and again on 6<sup>th</sup> April to add any new outputs. The results of the Dimensions and the Symplectic searches were then combined and deduplicated for outputs existing as both a Preprint and Published Article. In these cases, the output is represented only once in the data as the Preprint entity. Two Published Articles were omitted for this reason. Where multiple versions of the same Preprint were found against the same DOI, the earliest version’s publication date has been used. Each publication included in the final dataset was manually verified to have relevance to the COVID-19 search term and to have at least one Imperial co-author, including honorary appointments. Other data fields of Article Type, Source, Publisher and Open Access Status were manually added.

In line with Imperial’s commitment to value all types of research outputs, the Imperial College London software repository<sup>1</sup> was also searched for items with a keyword of ‘COVID-19’. This produced one code output. A search for the same term on Zenodo returned a second code output identified to have Imperial authorship. Both code outputs are associated with a publication output in the dataset, e.g. are the shared version of code described in an article. As a proxy for ‘publication date’, the date of public release of the earliest found version of code in the source has been used. We acknowledge that there could be further code or software outputs that could not be identified due to differing authorship identity practices in these creator communities.

Datasets were also searched for using similar techniques described above and none identified.

A total of 41 research outputs published in the period were identified by the search described above.

## **Analysis**

### **Rate of daily publication growth by type, excluding Code**

As demonstrated in Figure 1, Reports were the earliest identified output type, with the first publication date found to be 17<sup>th</sup> January. Followed by the first Preprint identified 26 days later on 12<sup>th</sup> February; and the first Published Article a following 13 days later on 25<sup>th</sup> February.

---

<sup>1</sup> <https://imperialcollegelondon.github.io/research-software-directory/>

Whilst Published Article was the slowest output type to appear, after the first publication it grows at faster rate than the other types, at a daily growth rate for the entire period of 0.21 in comparison to 0.13 for Preprints and 0.17 for Reports<sup>2</sup>.

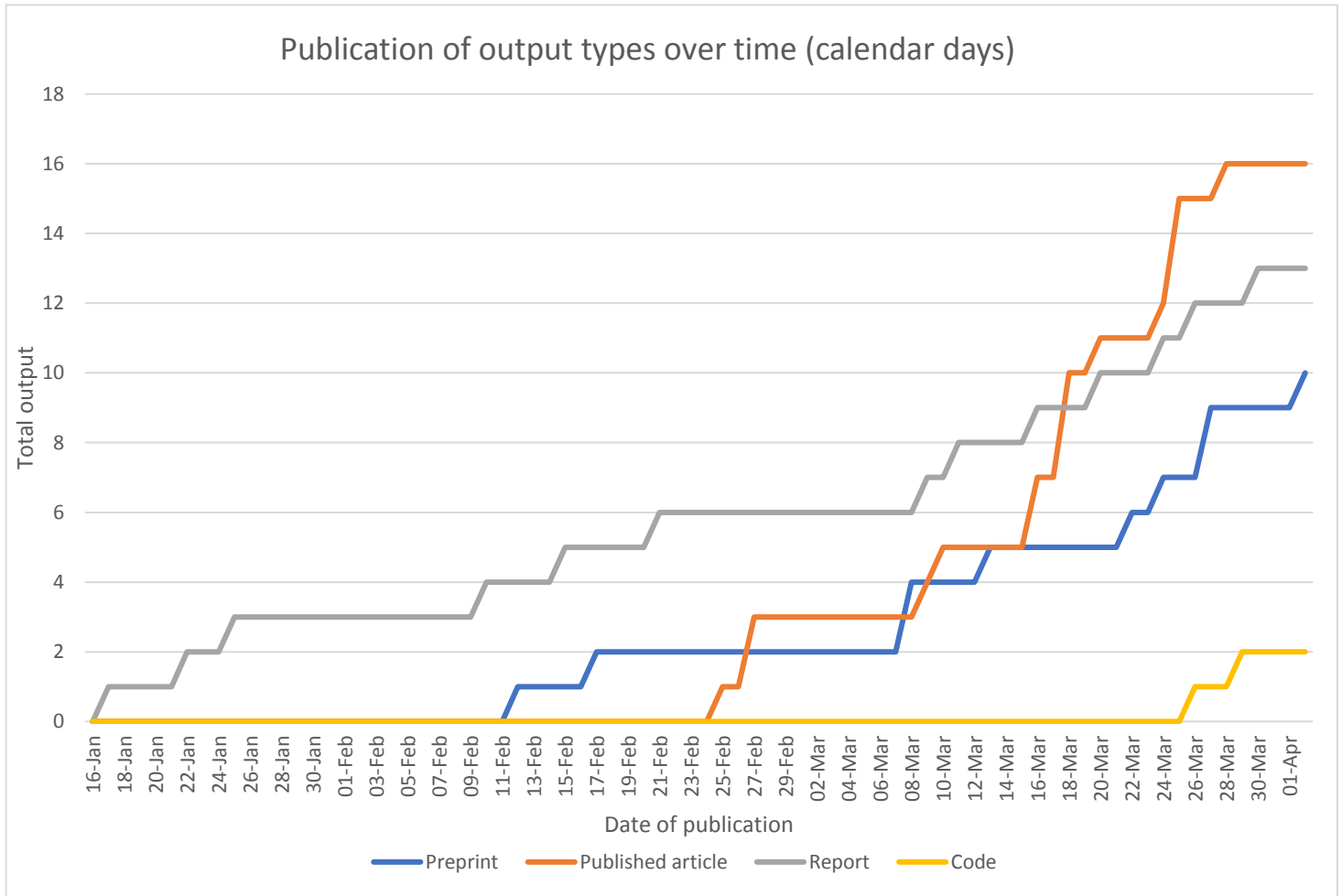


Figure 1: Publication of output types over time (calendar days)

### Publication of Code

The creation and deployment of software/code would have been integral to the research carried out in many of the Reports and Published Articles included in this dataset. That the first identified code in the dataset is found on 26<sup>th</sup> March (Week 12) suggests a slower rate of sharing code than their associated publication based outputs.

<sup>2</sup> Calculated as total output / total days

## Publication growth by week, excluding Code

The week time interval has been determined by calendar Monday to Sunday weeks, assuming Week 1 to be the first week of zero activity starting Monday 6<sup>th</sup> January and ending the time period with Week 13, covering the last recorded activity in the set on Thursday 2<sup>nd</sup> April.

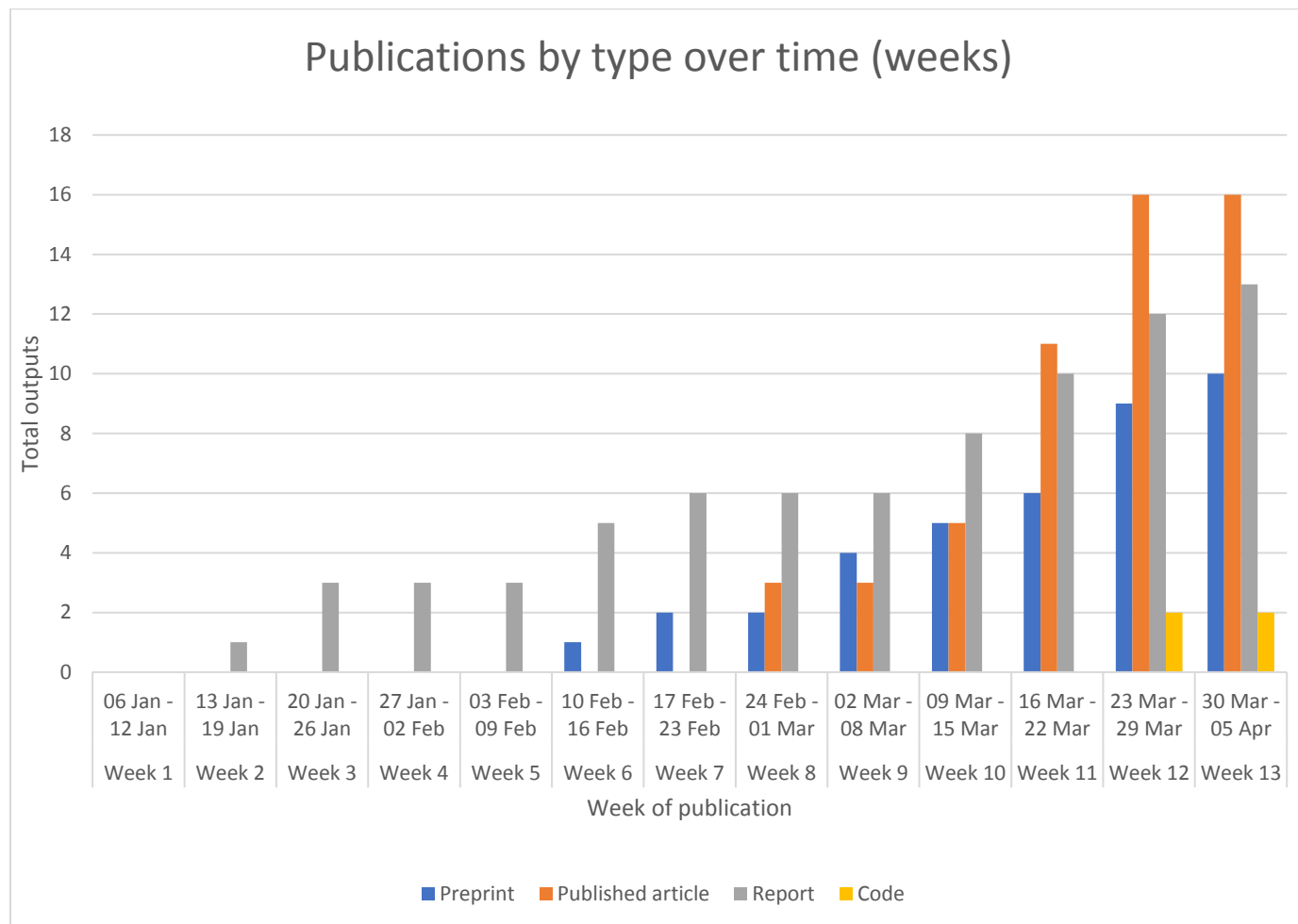


Figure 2 Publication by type over time (weeks)

Until the first release of a Published Article in Week 8, Reports were released at an average of 2 per week. The dominance of Reports as an output type in the first half of the response period is likely a result of the rapidity of the report publishing workflow developed at Imperial. With this workflow a document can be made available in the institutional repository and assigned a DOI on the same day.

This is naturally in contrast to the longer workflow associated with publication of peer reviewed Published Articles in journals. However, Preprints are also non-peer reviewed content and the first Preprint is not found until Week 6. The platforms that support Preprints offer rapid publication but with a slightly longer

expected time between submission and publication, e.g. medRxiv advertises a 4-5 day waiting period<sup>3</sup>. This slightly longer period would account for some moderation, which is not used by the Imperial repository<sup>4</sup>.

Examining the items classified as Reports, all are authored by members of the Imperial College COVID-19 Response Team whilst the Preprint items derive from a variety of Imperial author groups. Since the rapid Report workflow is offered to all Imperial authors, there is no obvious reason why this correlation exists.

### Peer Reviewed versus Non-Peer Reviewed

To avoid the conflation of Reports and one single author group described above, the outputs can also be grouped as non-peer review and peer reviewed. In the graph below, Reports and Preprints have been grouped together as non-peer review outputs, and so represent the outputs of the Imperial COVID-19 Response Team as well as other Imperial researchers publishing in this area but outside of this organizational group. Peer reviewed refers to Published Articles<sup>5</sup>. Code has been excluded.

In Figure 3 the early dominance of non-peer reviewed outputs over peer reviewed matches what was observed at the output type level with Reports appearing 5 weeks ahead of any other type. In Figure 3

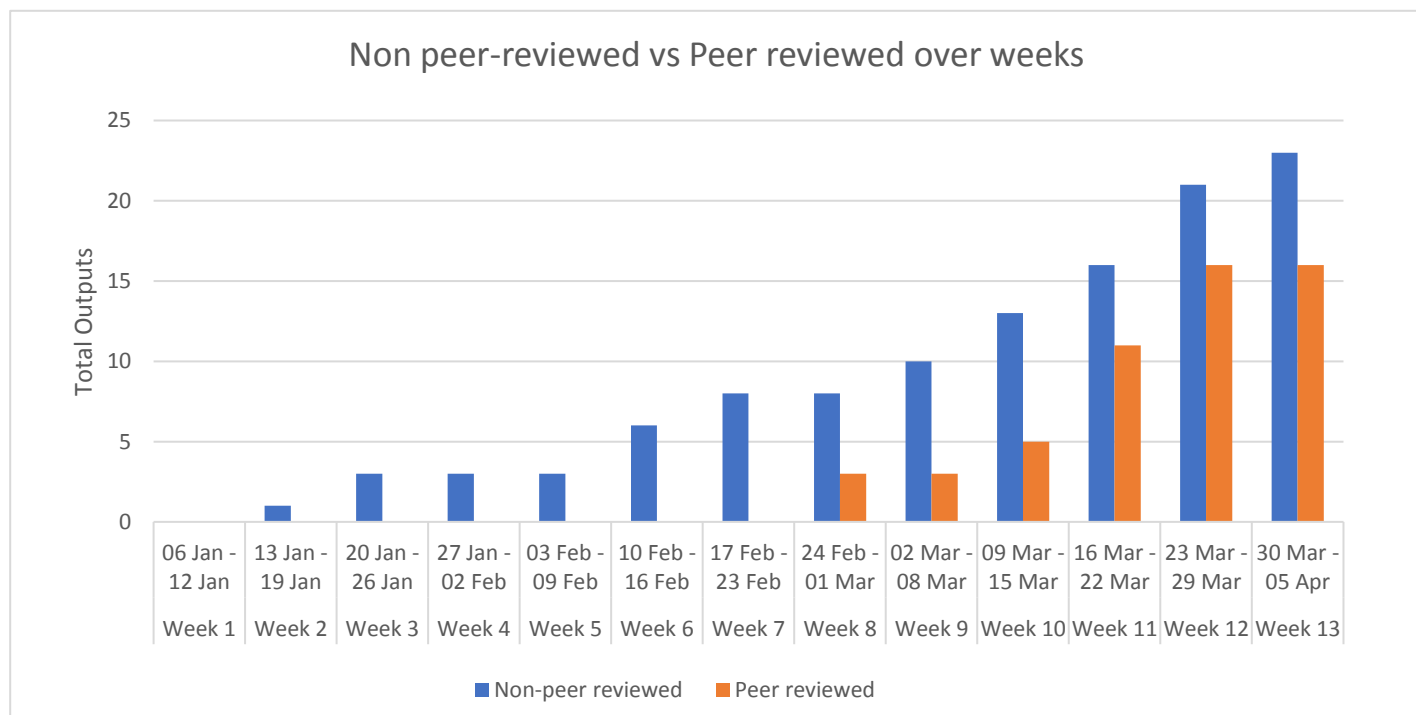


Figure 3 Non peer reviewed vs Peer reviewed outputs over weeks

<sup>3</sup> <https://www.medrxiv.org/about/FAQ>

<sup>4</sup> Submissions to Imperial’s Spiral repository undergo some editorial and administrative checking

<sup>5</sup> As detailed in the Data Collection, Published Articles that also existed as Preprints were eliminated from the data, so the true number of Published Articles would be greater than displayed here

non peer review outputs grew at a rate of 1.77 per week over the entire period, in contrast to peer reviewed outputs at 1.23 per week<sup>6</sup>.

### Article Type

The types of articles within Published Article and Preprint can also be identified. Reports and Code have been excluded from the below analysis.

The most commonly occurring article type is Comment (9 occurrences) followed by Research (7 occurrences).

<b>Article type</b>	<b>Total Outputs</b>
Comment	9
Editorial	2
Guidelines	1
Letter	2
News	1
Practice: 10 Minute Consultation	1
Research	7
Review	3
<b>Grand Total</b>	<b>26</b>

The article types can then be assigned to the following groups: Editorial (Comment, Editorial, Letter, News); Research (Research); Review (Review); and Other (Practice: 10 Minute Consultation, Guidelines). The rate of publication of each of these types is displayed in Figure 4. Time period commences 11<sup>th</sup> February, the first found date of a Preprint or Published Article.

---

<sup>6</sup> Calculated as total output / number of weeks

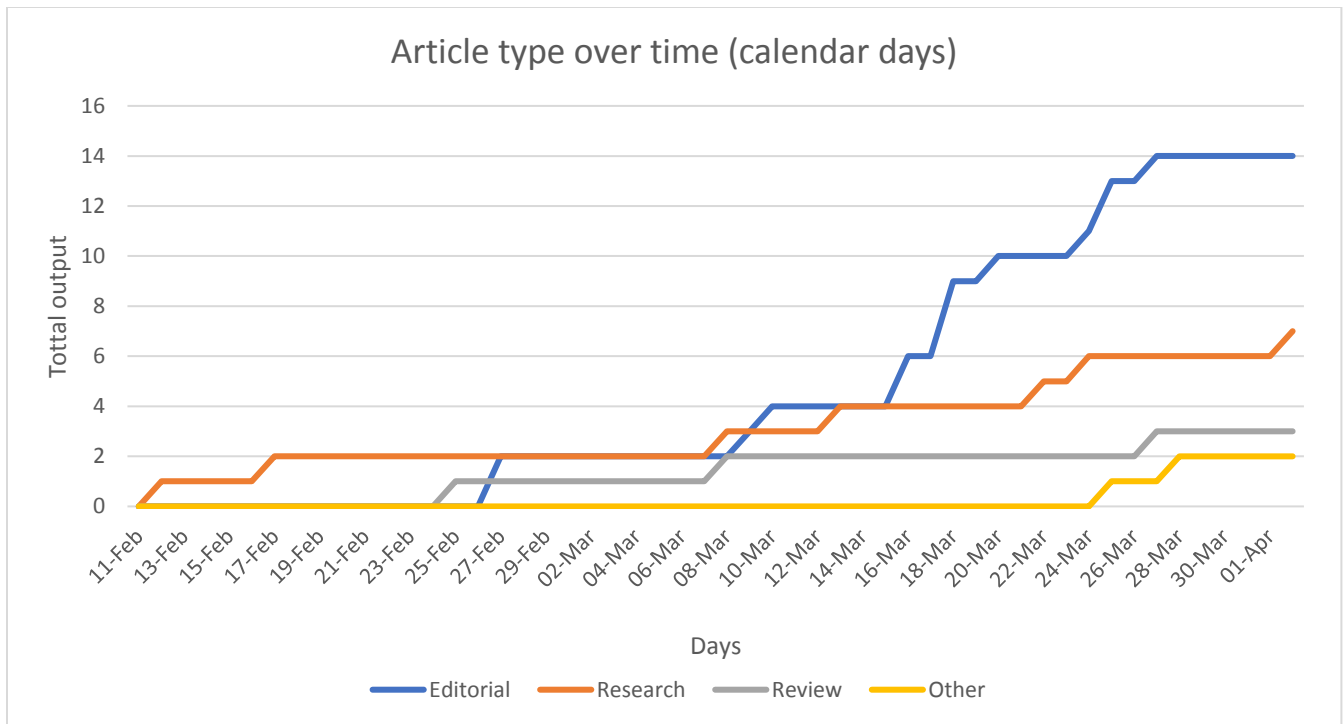


Figure 4 Article type over time (calendar days)

Research is the earliest article types found, but surpassed in quantity on 10<sup>th</sup> March by Editorial content. The Other group represents Practice: 10 minute consultation publication found in the BMJ and a set of clinical guidelines. The emergence of these Other types at a slower rate than Research or Editorial is not surprising. To produce a total of 10 different types of articles (including Reports and Code) in this period demonstrates rapid production of diverse outputs.

#### Open access by licence status

At the time of writing, 100% of the outputs of all types are free to read. However, only 60% of these are Gold open access<sup>1</sup> (see Figure 5) by licence.

37% are free access, either through the publisher’s actions or the licence of the preprint server it has been licenced to. ‘Free access: publisher’ is the removal of a paywall by publishers to content that either previously were, or at some point in the future could return, behind a subscription or pay-to-view paywall and has no guaranteed reuse rights.

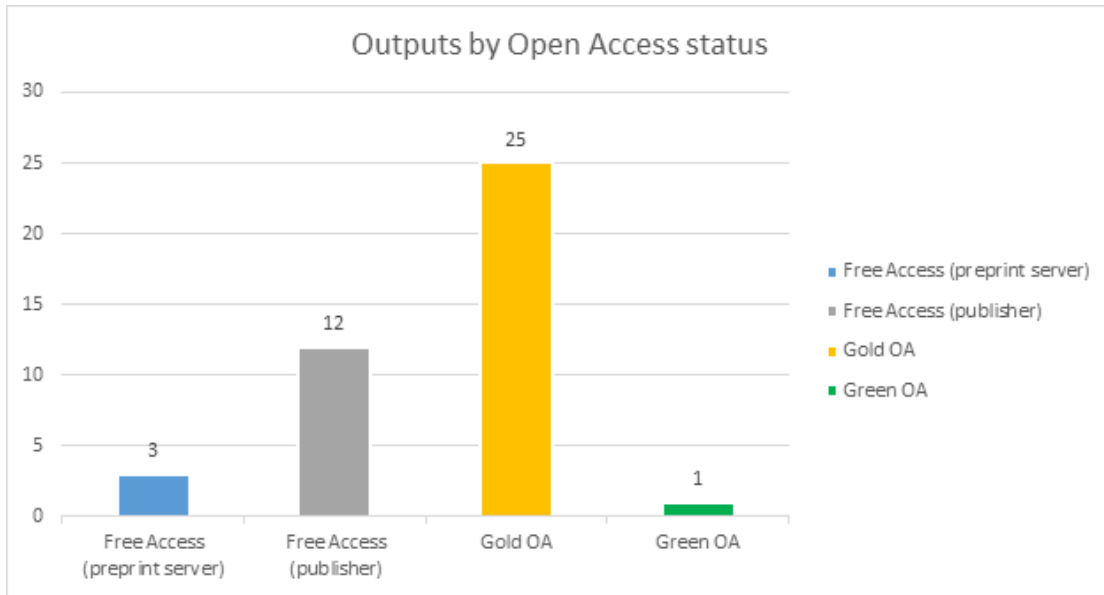


Figure 5 Outputs by Open Access status

As well as a matter of paying or not paying to read the research, the CC BY licence also ensures that research findings and results can be more widely reused by the public, news outlets industry and translations into another language. As demonstrated in Figure 6, 61% of the outputs are published under a CC BY licence that will make clear the permissions for readers on these matters. In contrast, permission is actively required for non-CC BY licensed outputs, relying on communications and decision making for requests that could slow the sharing, reuse and application of this research. However, it should be acknowledged that it is legitimate practice to withhold reuse rights in some cases to help prevent non-peer reviewed medical research being deployed in a clinical setting.

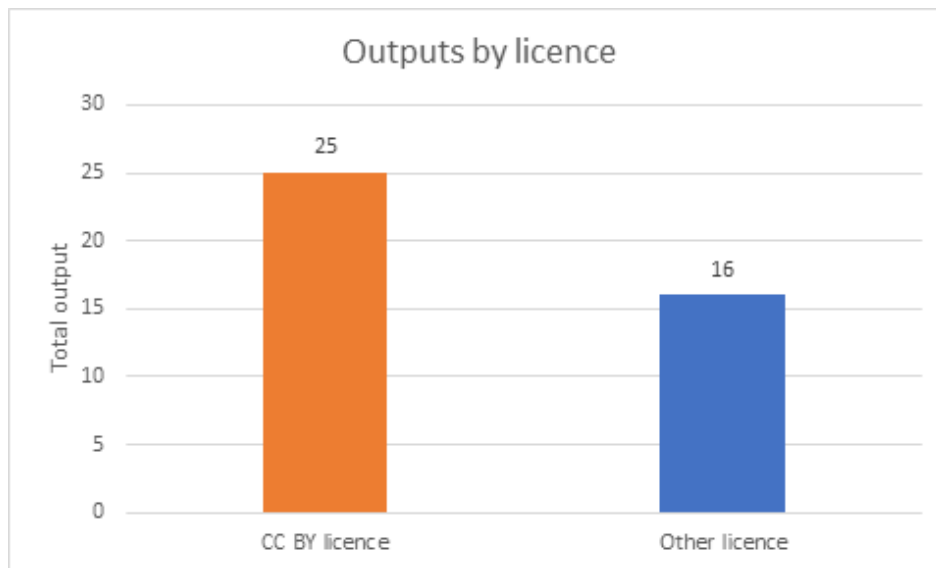


Figure 6 Outputs by licence



### Daily publication rate by open access status

The simplest definitions of open access and free access have been used to show rate of publication of each in the below Figures. Open access refers to those published under Gold, CC BY licences and the 3 Preprint outputs with their preprint server specific licence. Free access refers to the remaining publications (Published Articles in subscription journals where the publisher had at time of writing, removed the paywall).

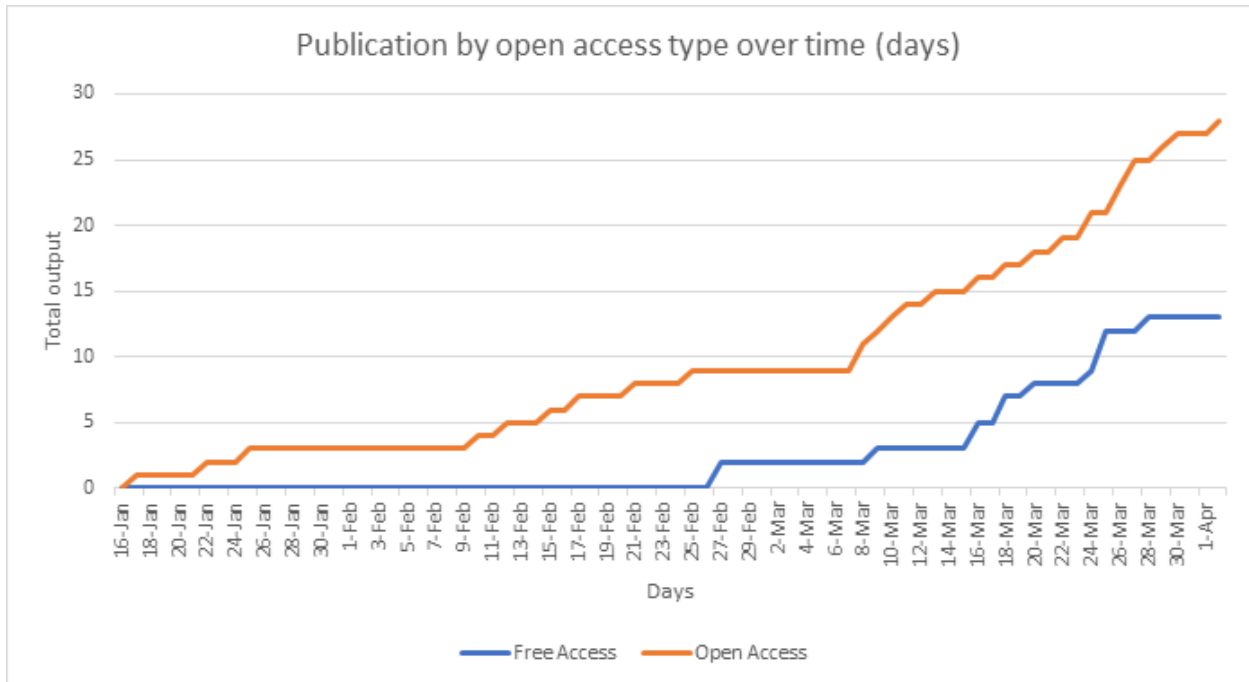


Figure 7 Outputs by Open Access type across days

The daily growth rate of open access publications (0.36) is higher than free access research outputs (0.17) for this period<sup>7</sup>. 9 open access research outputs were published within the first 41 days, after which the first free access output was found. This correlates with the article type rate explained above, that the earliest identified outputs in the series, Reports, are all published (Gold) open access.

<sup>7</sup> Calculated as total output / total days

## Weekly publication rate by open access status

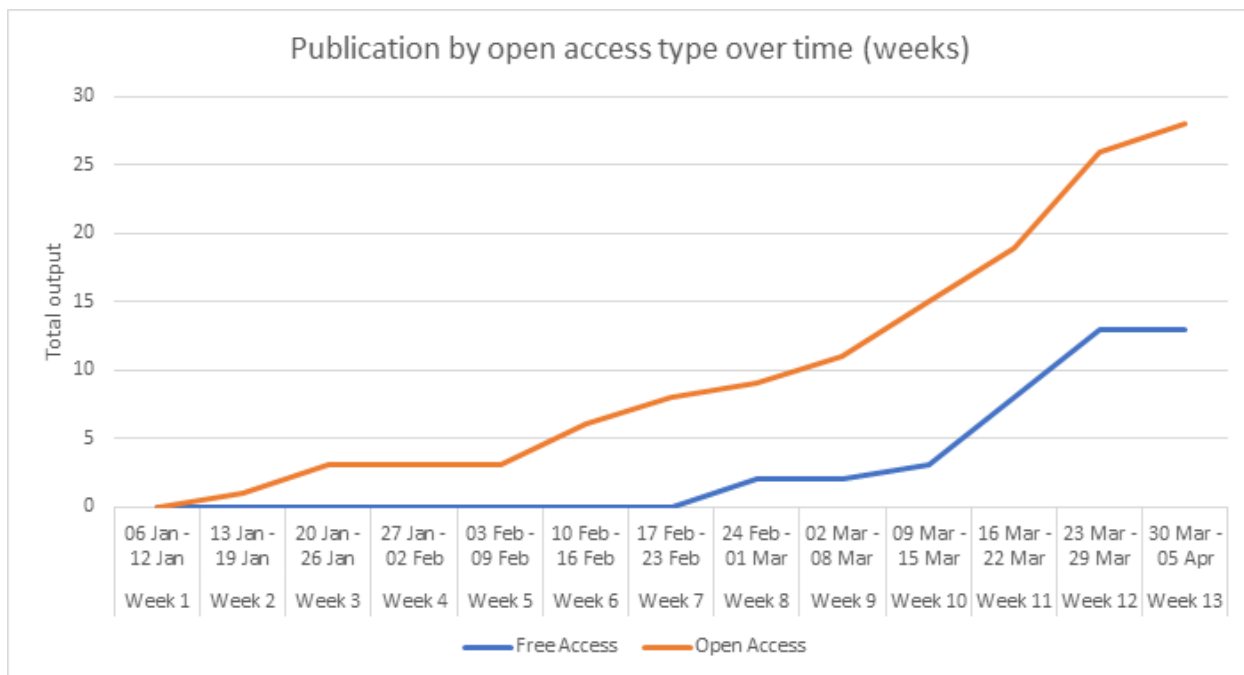


Figure 8 Publication by open access type over time (weeks)

The daily and weekly growth rate figures support Open Access as a fast solution in scientific responses to national and global crises. The first open access report was produced and posted on Spiral on 16<sup>th</sup> January 2020, ahead of Europe experiencing the COVID-19 outbreak. The first free access research article was published on 27<sup>th</sup> February 42 days after the first Open Access article and one week before the first recorded death in the UK<sup>8</sup>.

## Publishers

The outputs can also be seen by publisher, see below Figure 9. In the Figure; repositories and preprints servers are highlighted in blue, and journal publishers in orange.

In total, Imperial College London's repository, Spiral is the most frequent publisher (14 outputs), followed by the commercial journal publisher Elsevier is the most frequent publisher (8 outputs).

<sup>8</sup> <https://www.bbc.co.uk/news/uk-51759602>

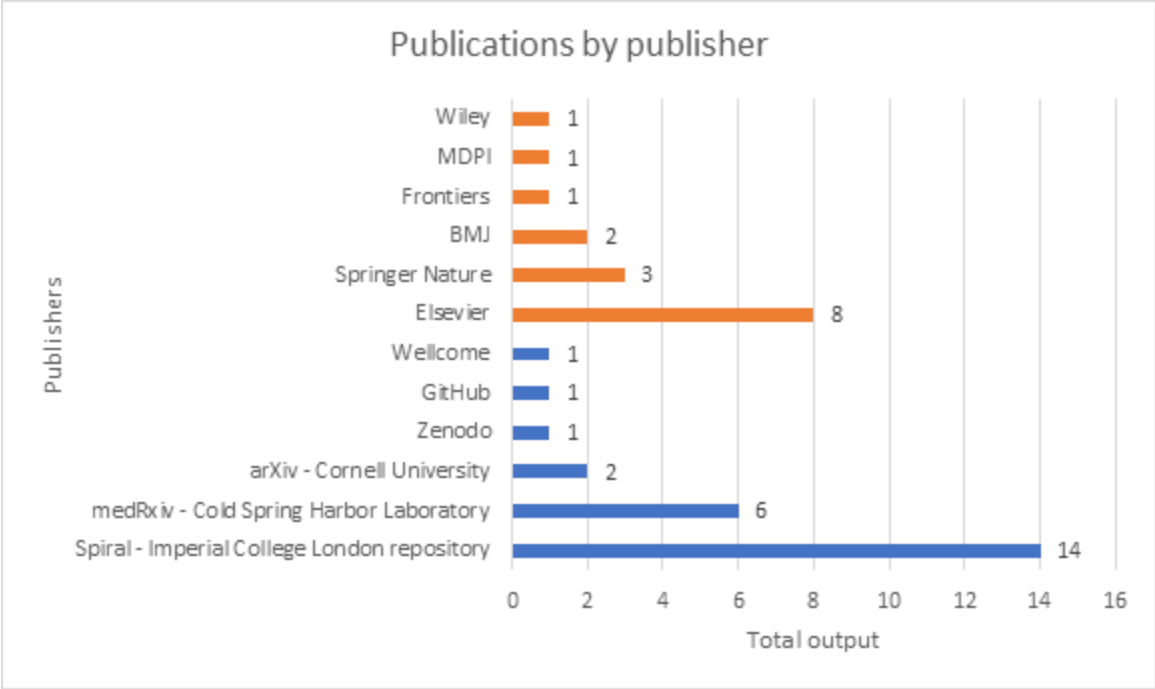


Figure 9 Outputs by publisher

Publisher and Journal	Total output
<b>Elsevier</b>	<b>8</b>
Clinical Radiology	1
The Lancet	3
The Lancet Gastroenterology & Hepatology	1
The Lancet Infectious Diseases	1
The Lancet Respiratory Medicine	2
<b>Springer Nature</b>	<b>3</b>
BMC Medicine	1
Intensive Care Medicine	1
Nature Reviews Materials	1
<b>BMJ</b>	<b>2</b>
The BMJ	2
<b>Frontiers</b>	<b>1</b>
Frontiers in Cell and Developmental Biology	1
<b>Wiley</b>	<b>1</b>
European Journal of Immunology	1
<b>MDPI</b>	<b>1</b>
International Journal of Environmental Research and Public Health	1
<b>Grand Total</b>	<b>16</b>

Only 3 out of 17 journals associated with the above Published Articles are Gold Open Access titles (Figure 10). The remainder are subscription or hybrid titles applying free access to COVID-19 articles. The Gold open access journals are indicated below.

Publisher and Journal	Total Gold OA output of set
<b>Frontiers</b>	<b>1</b>
Frontiers in Cell and Developmental Biology	1
<b>MDPI</b>	<b>1</b>
International Journal of Environmental Research and Public Health	1
<b>Springer Nature</b>	<b>1</b>
BMC Medicine	1
<b>Grand Total</b>	<b>3</b>

Data is available at: <https://doi.org/10.5281/zenodo.3754126>

**Copyright:** © 2020 Imperial College London. This is an open access report distributed under the terms of [Creative Commons Attribution-Non Commercial License](https://creativecommons.org/licenses/by-nc/4.0/) (CC BY-NC 4.0).



**Suggested Citation:** Price RC and Ozkan YA. 13 weeks in a pandemic: a descriptive study of Imperial College London's COVID-19 publications. Imperial College London (April 2020), doi: <https://doi.org/10.25561/77970>