

## COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts

Item Type	Journal article
Authors	Kousha, Kayvan;Thelwall, Michael
Citation	Kousha, K. and Thelwall, M. (2020) COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts, Quantitative Science Studies 1 (3), pp. 1068-1091. DOI: 10.1162/qss_a_00066
DOI	<a href="https://doi.org/10.1162/qss_a_00066">10.1162/qss_a_00066</a>
Publisher	MIT Press
Journal	Quantitative Science Studies
Download date	2025-05-01 04:16:17
License	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
Link to Item	<a href="http://hdl.handle.net/2436/623217">http://hdl.handle.net/2436/623217</a>



an open access  journal



Citation: Kousha, K., & Thelwall, M. (2020). COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. *Quantitative Science Studies*, 1(3), 1068–1091. [https://doi.org/10.1162/qss\\_a\\_00066](https://doi.org/10.1162/qss_a_00066)

DOI:  
[https://doi.org/10.1162/qss\\_a\\_00066](https://doi.org/10.1162/qss_a_00066)

Supporting Information:  
[https://doi.org/10.1162/qss\\_a\\_00066](https://doi.org/10.1162/qss_a_00066)

Received: 21 April 2020  
Accepted: 14 May 2020

Corresponding Author:  
Mike Thelwall  
[m.thelwall@wlv.ac.uk](mailto:m.thelwall@wlv.ac.uk)

Handling Editor:  
Ludo Waltman

Copyright: © 2020 Kayvan Kousha and Mike Thelwall. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



## RESEARCH ARTICLE

# COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts

Kayvan Kousha  and Mike Thelwall 

Statistical Cybermetrics Research Group, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

**Keywords:** altmetrics, COVID-19, Dimensions, Mendeley, Scopus, Web of Science

### ABSTRACT

The COVID-19 pandemic requires a fast response from researchers to help address biological, medical, and public health issues to minimize its impact. In this rapidly evolving context, scholars, professionals, and the public may need to identify important new studies quickly. In response, this paper assesses the coverage of scholarly databases and impact indicators during March 21, 2020 to April 18, 2020. The rapidly increasing volume of research is particularly accessible through Dimensions, and less through Scopus, the Web of Science, and PubMed. Google Scholar's results included many false matches. A few COVID-19 papers from the 21,395 in Dimensions were already highly cited, with substantial news and social media attention. For this topic, in contrast to previous studies, there seems to be a high degree of convergence between articles shared in the social web and citation counts, at least in the short term. In particular, articles that are extensively tweeted on the day first indexed are likely to be highly read and relatively highly cited 3 weeks later. Researchers needing wide scope literature searches (rather than health-focused PubMed or medRxiv searches) should start with Dimensions (or Google Scholar) and can use tweet and Mendeley reader counts as indicators of likely importance.

### 1. INTRODUCTION

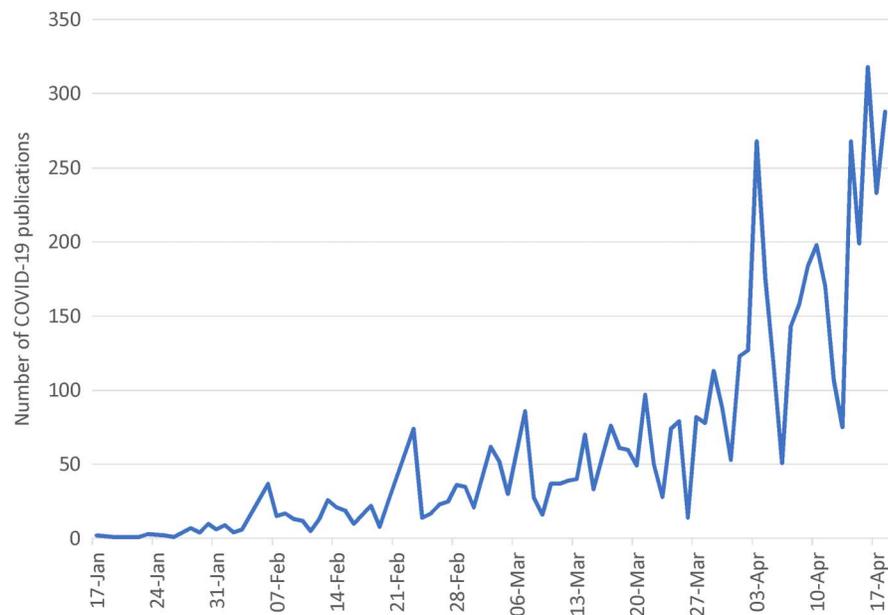
The international scientific effort to mitigate COVID-19 is unprecedented in scale and rapidity. For instance, PubMed added related publications daily between January 17 and April 18, 2020<sup>1</sup> (Figure 1), reaching over 300 in a single day. This effort is in response to the lethality and rapid spread of the disease, as well as the major economic and social consequences of COVID-19 lockdowns. As part of the response, researchers, professionals, and the public may need to consult the scientific literature for the latest findings. Although this is normal for science, standard literature search methods may be ineffective in a rapid publishing environment. Traditional citation indexes may not be fast enough, especially given that they do not index most preprints, and citation counts may not help point to important studies. The more inclusive online citation indexes of sites such as Google Scholar and Dimensions.ai seem like suitable alternatives because they index both the traditional scholarly literature and documents not published in journals, including preprints (Herzog, Hook, & Konkiel, 2020; Kousha & Thelwall, 2019a). There are initiatives to help various communities with curated collections of COVID-19

<sup>1</sup> [https://www.nlm.nih.gov/pubs/techbull/nd08/nd08\\_pm\\_new\\_date\\_field.html](https://www.nlm.nih.gov/pubs/techbull/nd08/nd08_pm_new_date_field.html)

documents, such as published biomedical documents from PubMed Central (PMC, 2020), preprints from medRxiv and bioRxiv (medRxiv, 2020), and a data mining collection (Allen Institute, 2020; Colavizza, Costas, et al., 2020), but none are complete. It is therefore important to assess the COVID-19 coverage and growth of scholarly publication indexes, as well as the value of citation counts for new COVID-19 research.

In parallel with scholarly needs for literature, the public, professionals, and policy-makers also need to access current COVID-19 research to inform their decision-making, such as whether to recommend wearing protective masks. This may be in addition to, or to clarify, World Health Organization guidelines (WHO, 2020). They may therefore share relevant academic research in the social web (e.g., Merchant & Lurie, 2020), generating interest that may be picked up by alternative indicators (altmetrics). Thus, altmetrics may be useful in helping the public to identify the most relevant research or may help point researchers to topics considered important by the public. It would therefore be helpful to assess whether altmetrics can perform this role. In particular, because altmetrics can reflect both academic and nonacademic interests (Mohammadi, Barahmand, & Thelwall, 2019; Mohammadi, Thelwall, et al., 2018), it is not clear whether they will essentially be early indicators of citation impact or whether they reflect societal or other impacts for COVID-19. Altmetrics have already been shown useful to identify the spread of a misleading COVID-19 paper that was subsequently withdrawn (Ioannidis, 2020).

This paper addresses the above issues through a primarily descriptive analysis of the evolution of four online scholarly databases, and associated altmetrics, over 4 weeks in March–April 2020, when many countries were experiencing a lockdown. A previous study of January 20 to April 12, 2020 has shown continually increasing growth in the COVID-19 coverage of scholarly databases, with substantial variations between fields (Torres-Salinas, 2020). Individual highly cited



**Figure 1.** Daily additions of COVID-19 publications to PubMed (January 17 to April 18). Query used (((((((("COVID-19") OR "Novel coronavirus") OR "2019-nCoV") OR "SARS-CoV-2") OR "coronavirus 2") OR "Coronavirus disease 2019") OR "Corona virus disease 2019") AND ("2019/12/01"[Date - Publication] : "3000"[Date - Publication])).

or shared papers are also important to examine for qualitative insights into the types of research that are attracting substantial attention. The following research questions drive this paper:

- Which scholarly databases index the most COVID-19 publications (extending: Torres-Salinas, 2020)?
- Which COVID-19 documents have become highly cited or highly discussed?
- Do altmetrics and early citation counts reflect similar types of COVID-19 impact?
- Can any altmetrics serve as early indicators of future citation impact for COVID-19 documents?

## 2. BACKGROUND

The novel coronavirus SARS-CoV-2, which causes COVID-19, was first recorded in Wuhan City, China in December 2019. Quickly disseminating scientific results about COVID-19 is vital to allow the rapid exploitation of successful clinical results (Song & Karako, 2020). The importance of scientific publishing to respond to infectious disease outbreaks has been emphasized by many bibliometric studies of previous cases (Rethlefsen & Livinski, 2013), such as SARS (Kostoff & Morse, 2011; Tian & Zheng, 2015), H7N9 influenza (Tian & Zheng, 2015), HIV/AIDS (Pouris & Pouris, 2011), Ebola (Pouris & Ho, 2016), and Zika (Delwiche, 2018).

One recent study using Dimensions, Scopus, Web of Science (WoS), and the LitCovid (Chen, Allot, & Lu, 2020) curated list has investigated the daily growth of COVID-19 related publications in citation databases and digital libraries from January 1 to April 7, finding that Dimensions had the best coverage (9,435 publications) compared to WoS (718) and Scopus (1,568). The weekly growth of PubMed was about 1,000 publications and the PubMed Central (1,398), medRxiv (989), and SSRN (608) repositories had the best coverage of open access COVID-19 publications (Torres-Salinas, 2020). Google Scholar was not assessed, and all evidence was extracted from Dimensions, so the counts for other repositories may not be complete.

### 2.1. Dimensions Citations

Dimensions.ai (Herzog et al., 2020) is an online scholarly database that operates similarly to Google Scholar, in the sense of indexing documents using public information from the Web, but has an Applications Programming Interface (API) that supports automatic downloading for all query matches. It indexes most documents in Scopus (Thelwall, 2018b), although not for all fields (Orduña-Malea & Delgado-López-Cózar, 2018). It seems to have substantial coverage of preprint servers, such as arXiv, and so probably has much larger coverage overall, especially for recently published papers. Its coverage seems to be higher than Scopus and WoS, comparable to CrossRef but lower than Google Scholar and Microsoft Academic (Harzing, 2019). In line with this, citation counts for papers in Dimensions can be expected to be slightly higher than for Scopus and WoS but substantially higher for newer documents.

### 2.2. Altmetrics: Mendeley Readers

Counts of readers from the social reference sharing site Mendeley form the most extensively researched and understood altmetric. A nontrivial minority of researchers (about 5%) used Mendeley by 2014 according to one survey, with disciplinary differences (Van Noorden, 2014). People typically register documents in Mendeley when they have read them or intend to read them (Mohammadi, Thelwall, & Kousha, 2016), so it is reasonable to regard Mendeley counts as an indicator of readership. According to self-reports in the site, users are predominantly academics and postgraduate students, with a few undergraduates, librarians, and people in nonacademic

occupations (Mohammadi, Thelwall, et al., 2015). Thus, Mendeley is an indicator of predominantly academic readership, with an element of student readership. One difference is that nonarticle publications in journals, such as editorials and news items, are relatively more likely to be registered in Mendeley than to be cited (Zahedi & Haustein, 2018). Mendeley reader counts can help with the early identification of highly cited documents (Zahedi, Costas, & Wouters, 2017).

A range of studies have investigated the relationship between Mendeley reader counts and citation counts, finding moderate or strong positive correlations (Costas, Zahedi, & Wouters, 2015). Correlations between mature citation counts and Mendeley reader counts are strong and positive in almost all narrow fields in Scopus (Thelwall, 2017a), supporting their use as a citation impact type of indicator. Although the two types of data seem to be close to interchangeable for sets of mature articles (although they can differ sharply for individual education-oriented papers: Thelwall, 2017c), the advantage of Mendeley reader counts is that they appear and are useful a year before citation counts (Thelwall, 2017b). They may even be common enough to be used for scientometric purposes by the publication month of the publishing journal. Moreover, as early Mendeley reader counts correlate positively with later citation counts (Thelwall, 2018a), Mendeley reader counts are early academic impact indicators. They should therefore be a better academic impact indicator than citation counts for fast-moving issues, such as COVID-19.

### 2.3. Altmetrics: Tweepers, Facebook Walls

Twitter is potentially a source of societal attention evidence (Holmberg & Vainio, 2018; Priem, Taraborelli, et al., 2010). More articles have nonzero tweet counts than nonzero scores on any other altmetric, other than Mendeley (Costas et al., 2015; Thelwall, Haustein, et al., 2013). As Twitter is a news-oriented social media platform, articles can expect to get a substantial proportion of their tweets in the week of publication, so tweets are visible long before citations (Ortega, 2018a,b).

Tweeter counts (counting the number of tweeters rather than the number of tweets) are problematic to interpret. About half of people that tweet academic research are not academics (Mohammadi et al., 2018), and tweets typically contain just article titles or brief summaries (Thelwall, Tsou, et al., 2013; Robinson-García, Costas, et al., 2017), serving as publicity rather than evidence of impact. Many academic tweets are also created by bots (Haustein, Bowman, et al., 2016; Robinson-García, Costas, et al., 2017). Together with often close to zero correlations with citation counts (Costas et al., 2015; Haustein, Larivière, et al., 2014; Thelwall, Haustein, et al., 2013), there is insufficient evidence to claim that tweeter counts are indicators of either academic or societal impact. Nevertheless, they may have some value for health-related research, where there is more public interest in academic research (Haustein, Larivière, et al., 2014; Mohammadi, Gregory, et al., 2020). Editorials and news articles are relatively more likely to be tweeted than cited (Haustein, Costas, & Larivière, 2015), reflecting the news orientation of Twitter.

Facebook wall posts function like tweeter counts except that they are rarer (Costas et al., 2015; Thelwall, Haustein, et al., 2013). As most of Facebook is private and Altmetric.com obtains its Facebook wall counts only from public pages, this altmetric probably reflects a tiny fraction of all Facebook posts and may be oriented to organizational uses of Facebook (including journals) rather than typical users; few posts are directly from academics (Mohammadi et al., 2019).

### 2.4. Altmetrics: News and Reddit

Altmetric.com harvests citations from online free news websites and the news-oriented site Reddit. Altmetrics from both are relatively rare and have very low correlations with citation

counts (Costas et al., 2015; Thelwall, Haustein, et al., 2013). Nevertheless, health-related topics are newsworthy (Clark & Illman, 2006; Kousha & Thelwall, 2019b), including for infectious diseases (e.g., SARS: Lewison, 2008), so they may be useful for COVID-19.

### 3. METHODS

The research design is in three parts. First, to assess the relative coverage of scholarly databases, the main candidates were queried daily from March 21, 2020 to record the number of COVID-19 documents indexed. Second, lists of documents matching a set of COVID-19 queries were downloaded from Dimensions.ai and altmetrics for these were gathered from Mendeley (Gunn, 2014) and Altmetric.com (Adie & Roe, 2013; Robinson-García, Torres-Salinas, et al., 2014) daily and the individual scores and documents compared. Third, a March 24 data set was created to track a set of documents indexed on the same day.

#### 3.1. Scholarly Database Indexing of COVID-19 Publications

To assess the indexing of COVID-19-related publications, the two mainstream scholarly databases, Scopus and WoS, were queried as well as other major academic sources that may index relevant documents. After testing with the original and current names of the virus and disease and “Corona virus disease 2019” and “Coronavirus disease 2019”, the core queries used to identify relevant documents were as shown in Table 1. The queries are designed to be as inclusive as possible for the database in terms of document type and part of the document searched: full text, if available in the database, otherwise all metadata fields (e.g., title, abstract, keywords). The queries are not comprehensive but are high precision, unless stated, and should include the most recent research focusing on the issue, assuming that it includes the current official disease description.

The combined queries did not work in Google Scholar, giving false matches. The results for Google Scholar seemed to be substantially inflated by its web search component indexing advertisements or warnings in webpages alongside articles irrelevant to the disease, so its results are not reported. To illustrate the existence of these false matches, a search for “COVID-19” in Google Scholar with a date range specified as 1990–2000 (i.e., 20 years before the name was coined) on April 21, 2020 returned an estimated 5,010 matches<sup>2</sup>. Each incorrect Google Scholar match reported snippets not from the paper, such as, “PEDIATRICS COVID-19 COLLECTION We are fast-tracking and publishing the latest research and articles related to COVID-19 for free.” The exact COVID-19 coverage of Google Scholar is difficult to assess because it is not possible to download and check all matches in the absence of a Google Scholar API to download large sets of publication records. Because of these issues, no results are reported for Google Scholar.

#### 3.2. Document and Altmetric Comparison Data Sets

Initial testing suggested that Dimensions and Google Scholar had the largest coverage of COVID-19 documents. As Google Scholar does not have an API and the number of matches exceeds its 1,000 limit per query, it was not possible to extract Google Scholar’s set of matching documents. In contrast, Dimensions.ai has an API allowing complete sets of matching document records to be downloaded and does not seem to include false matches, so it was chosen as the base source of COVID-19 documents. It was checked daily with the following set of queries in the Dimensions API, designed to match publications about COVID-19 using various related

<sup>2</sup> [https://scholar.google.co.uk/scholar?q=covid-19&hl=en&as\\_sdt=0%2C5&as\\_ylo=1990&as\\_yhi=2000](https://scholar.google.co.uk/scholar?q=covid-19&hl=en&as_sdt=0%2C5&as_ylo=1990&as_yhi=2000)

**Table 1.** COVID-19 queries for a range of scholarly sources

Source	Query	Scope/ Year	Comments
Google Scholar	"COVID-19"	Full text 2019–2020	OR does not work False matches.
Dimensions	"COVID-19" OR "Novel coronavirus" OR "2019-nCoV" OR "SARS-CoV-2" OR "coronavirus 2" OR "Coronavirus disease 2019" OR "Corona virus disease 2019"	Full text 2019–2020	
PubMed	(((((("COVID-19") OR "Novel coronavirus") OR "2019-nCoV") OR "SARS-CoV-2") OR "coronavirus 2") OR "Coronavirus disease 2019") OR "Corona virus disease 2019") AND ("2019/12/01"[Date - Publication] : "3000"[Date - Publication]))	All metadata from Dec 2019	
Mendeley	"COVID-19"	Probably metadata <sup>3</sup>	OR does not work.
medRxiv and bioRxiv	Self-reported repository statistics for self-curated collection.	Full text	Repository statistics for COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv <sup>4</sup> .
Scopus	(ALL ("COVID-19") OR ALL ("Novel coronavirus") OR ALL ("2019-nCoV") OR ALL ("SARS-CoV-2") OR ALL ("coronavirus 2") OR ALL ("Coronavirus disease 2019") OR ALL ("Corona virus disease 2019")) AND PUBYEAR = 2020 OR PUBDATETXT (december 2019)	All metadata 2019–2020	
WoS Core Collection	TOPIC=("COVID-19" OR "Novel coronavirus" OR "2019-nCoV" OR "SARS-CoV-2" OR "coronavirus 2" OR "Coronavirus disease 2019" OR "Corona virus disease 2019")	All metadata 2019–2020	Including Conference Proceedings Citation Index.
PMC	(((((("COVID-19") OR "Novel coronavirus") OR "Novel coronavirus") OR "2019-nCoV") OR "SARS-CoV-2") OR "coronavirus 2") OR "Coronavirus disease 2019") OR "Corona virus disease 2019") AND ("2019/12/01"[Publication Date] : "3000"[Publication Date]))	Full text from Dec 2019	
ClinicalTrials.gov	COVID OR "SARS-CoV-2" OR "2019-nCoV"		Query predefined statistics <sup>5</sup> .

names. These queries are all designed to be precise, but there were still a few false matches. All queries ended in “return publications [basics + extras]”.

- search publications for "COVID-19" where year >= 2019
- search publications for "Novel coronavirus" where year >= 2019
- search publications for "2019-nCoV" where year >= 2019
- search publications for "SARS-CoV-2" where year >= 2019
- search publications for "coronavirus 2" where year >= 2019

<sup>3</sup> Mendeley does not report the scope of its search feature <https://www.mendeley.com/guides/web/02-paper-search>

<sup>4</sup> <https://connect.biorxiv.org/relate/content/181>

<sup>5</sup> <https://clinicaltrials.gov/ct2/results?cond=COVID-19>

**Table 2.** The number of articles for the top 10 Dimensions subject codes for the complete and March 24 data sets

FOR code	All	All (frac)*	%	Mar-24	Mar-24 (frac)*	%
1117 Public Health and Health Services	3,072	2,762.9	13%	78	73.3	21%
1108 Medical Microbiology	2,773	2,240.8	10%	32	27.2	8%
1103 Clinical Sciences	2,159	1,860.9	9%	32	28.8	8%
0601 Biochemistry and Cell Biology	1,192	946.3	4%	14	11.3	3%
1107 Immunology	1,096	873.4	4%	5	2.8	1%
0604 Genetics	803	642.7	3%	4	2.8	1%
1102 Cardiorespiratory Medicine & Haematology	459	384.4	2%	7	6.5	2%
0801 Artificial Intelligence and Image Processing	383	336.0	2%	6	6.0	2%
1109 Neurosciences	316	257.9	1%	2	1.3	0%
0605 Microbiology	364	224.3	1%	0	0.0	0%

\* Counting 1/n for a paper with n subject codes.

- search publications for "Coronavirus disease 2019" where year >= 2019
- search publications for "Corona virus disease 2019" where year >= 2019

The resulting 21,395 publications were mainly open access (53%; 75% for the March 24 set—see later) and predominantly from health-related specialties (Table 2).

The data sets analyzed include substantial numbers of papers from preprint planforms, including medRxiv, SSRN, arXiv, bioRxiv, ChemRxiv, and Research Square (Table 3, as in Torres-Salinas, 2020), as well as books and more traditional journals (Table 3).

Although most documents were classified as Articles by Dimensions, this type includes medRxiv preprints and diverse types of document published in journals, such as notes, short

**Table 3.** The top 10 journals, as recorded in Dimensions, for the complete and March 24 data sets

Journal	All	%	Mar-24	%	Comment
[None]	2,932	14%	13	4%	Books, book chapters, theses
medRxiv	1,234	6%	30	9%	Health sciences preprints
SSRN Electronic Journal	855	4%	0	0%	Social science preprints
arXiv	389	2%	16	5%	Physics/computing preprints
bioRxiv	358	2%	1	0%	Biological sciences preprints
Research Square	341	2%	13	4%	Preprint platform
BMJ	262	1%	9	3%	Core medical journal
ChemRxiv	210	1%	8	2%	Chemistry preprints
Viruses	196	1%	1	0%	MDPI open access journal
Journal of Medical Virology	176	1%	4	1%	Wiley journal

**Table 4.** The top 10 document types, as recorded in Dimensions, for the complete and March 24 data sets

Type	All	%	Mar-24	%	Comments
Article	16,330	76%	295	85%	Includes preprints from medRxiv, editorials, commentaries
Book	832	4%	4	1%	Matches more general “Coronavirus 2” research
Chapter	1,645	8%	5	1%	Matches more general “Coronavirus 2” research
Preprint	2,236	10%	43	12%	Includes arXiv, Research Square, chemRxiv, JMIR Preprints, SSRN
Monograph	166	1%	2	1%	Matches more general “Coronavirus 2” research
Proceeding	186	1%	0	0%	Conference proceedings
<b>Total</b>	<b>21,392</b>	<b>100%</b>	<b>349</b>	<b>100%</b>	

communications, editorials, and commentaries (Table 4). As many editorials seemed to discuss the impact of COVID-19 on the journal or field, this added fewer citable documents to the Article class. The surprising number of books and book chapters (13% overall) seems to be primarily due to pre-COVID-19 discussions about coronaviruses, matching the query “Coronavirus 2”. The low number of conference proceedings may be due to conference cancellations, or the inability of most conferences to respond to the COVID-19 timescale.

Because the Dimensions type Article includes documents that would not be classed as standard journal articles in scientometric analyses, the 295 Dimensions “Articles” from March 24 were visited to classify them by type. Only 106 of these seemed to be standard journal articles. The rest were mainly editorials, letters (called *letters*, *letters to the editor*, or *correspondence*; one detailed letter was classed as an article), or news stories. In some cases, documents were called “article” by the publishing journal but were clearly news stories published in a news-focused magazine/journal. The reduced set of 106 journal articles from March 24, 2020 was used for follow-up correlation tests.

After Webometric Analyst had downloaded a complete set of records each day, the Mendeley API was used to identify the number of Mendeley readers for each document, again using Webometric Analyst. It queries by DOI and by title/author/year and combines nonoverlapping results for the most complete reader count. This follows best practice (Zahedi, Haustein, & Bowman, 2014). Webometric Analyst was also used to identify counts of citations in Twitter, Facebook, Reddit, and online news outlets to these documents, as identified by DOI queries to Altmetric.com. This data provider seems to have the most comprehensive coverage of Twitter, the largest of the sources (Ortega, 2018a). Twitter and Facebook are logical choices to investigate because they seem to be the social media sources that most cite academic research (Costas et al., 2015; Thelwall, Haustein, et al., 2013). Reddit and news may give a news perspective, although Reddit is a multipurpose site (Ovadia, 2015; Stoddard, 2015) and the news sources harvested by Altmetric.com presumably exclude some major paywalled press sources.

There were some gaps in the data collection due to documents not being returned by a query on one day when they had been returned on a previous day. This produced missing citation and altmetric scores, affecting the analysis. To avoid this issue, these missing values were replaced with approximate values by linear interpolation (when scores were available for previous and subsequent dates), linear extrapolation (when at least two previous but no subsequent scores were available), or constant values (when only one previous value was available).

### 3.3. Analysis

The coverage of the different sources was evaluated by comparing (on a graph) the number of query matches over time. This is not a fair comparison because the queries are not equivalent, a researcher may use other queries, and the sources index with different levels of comprehensiveness. For example, a source that indexed the full text of documents would get more and probably less relevant hits than a source indexing the title and abstract, even if they had the same coverage.

To assess the types of document generating the most impact for each source, the top 5 for each indicator was extracted to give a manageable set. A comparison of the relative ranks of these documents for the different indicators was used to guide the evaluation of the relative importance of the document characteristics, along with the document age (younger documents would tend to have lower scores in less rapidly evolving indicators). This focus on the highest scoring documents seems reasonable because they are likely to be the most influential or important, even though different trends may apply to more average documents.

To compare the average accumulation speed and scores of COVID-19 documents, a base set was chosen, consisting of documents first indexed in Dimensions on March 24, 2020. This was the date from the first week with the most new documents (excluding the first day). These documents form a set that are likely to have been published on or shortly before March 24, 2020. The altmetric and citation scores for this set were compared over time to assess their evolution and relative magnitude. Averages were calculated with geometric means (with a +1 offset: Fairclough & Thelwall, 2015) rather than arithmetic means due to the highly skewed nature of citations (de Solla Price, 1976; Wallace, Larivière, & Gingras, 2009) and altmetrics (Thelwall & Wilson, 2016; Yu, Xu, et al., 2017). The scores of this set were then compared using Spearman correlations to assess the extent to which they may reflect similar types of impact (Sud & Thelwall, 2014). Because altmetrics other than Mendeley tend to have very weak correlations with citation counts (Costas et al., 2015; Haustein et al., 2014; Thelwall, Haustein, et al., 2013), high correlations are not expected. Correlations were used rather than regression because this is the standard technique for altmetrics and in this case matches the hypothesis' use case: sorting documents matching COVID-19 queries using altmetrics or citations.

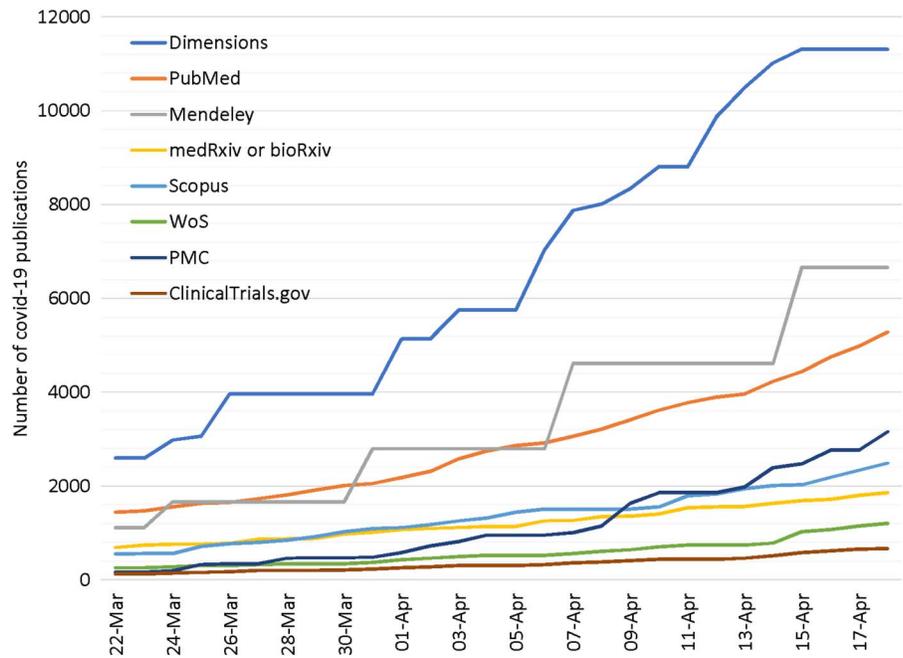
Field normalization was not used for either analysis because (a) the papers cover a relatively narrow topic (COVID-19) even though they span many subject areas and (b) it is impractical to field normalize the values because this would require daily updates of the whole of Dimensions, Altmetric.com, and Mendeley for the world reference sets.

## 4. RESULTS

### 4.1. Coverage of Scholarly Databases

Based on the estimated number of manual search results returned by the sources queried, it seems that Dimensions has substantially wider coverage of COVID-19 publications than all other sources or finds more because of its full text indexing rather than just searching metadata (Figure 2). Google Scholar probably indexes at least as many documents as Dimensions, although this could not be checked because of the number of false matches it returned (a graph including Google Scholar is in version 1 of this paper at <https://arxiv.org/abs/2004.10400v1>).

Google Scholar and Dimensions index both publisher records and other online publications (preprint archives for Dimensions, wider web sources for Google Scholar). As Dimensions seems able to identify COVID-19 publications more quickly or more widely than WoS and Scopus, academics studying the area should consider Dimensions (or Google Scholar, if the false matches are not a concern) if more specialist databases, such as PubMed, are not adequate.



**Figure 2.** The daily number of hits for COVID-19 queries (see Table 1) from a range of scholarly sources (March 22 to April 18).

This argument does not take into account the importance of the documents, however, and it is possible that the key publications are quickly peer reviewed, published, and indexed by Scopus and WoS. The Dimensions results include editorials, news, and letters, and may include recent documents not about the disease but that mention it for background information in their full text.

**Overlaps Between Dimension, Scopus and WoS**

The extent of overlaps between the COVID-19 query results for Dimensions, Scopus, and WoS were estimated on April 19, 2020 to assess whether they were indexing the same publications. To obtain a relevant set of COVID-19 publications, only publications from 2019–2020 with the terms "COVID" OR "coronavirus" OR "2019-nCoV" OR "SARS-CoV-2" OR "Corona" in their titles were selected. Publications with DOIs were matched between the three databases to assess the percentage overlap between them (Table 5). Few of the Dimensions publications were also in Scopus (23.3%) or WoS (11.8%). Two-fifths (40.4%) of the Scopus publications were in WoS and four-fifths (81.9%) of WoS publications were in Scopus. Google Scholar could not be compared without a comprehensive list of search matches.

**Table 5.** Overlaps for COVID-19 publications with DOIs between Dimension, Scopus, and Web of Science (April 18, 2020)

	Total publications	Overlap % (No.)		Nonoverlapping % (No.)	
		Scopus	WoS	Scopus	WoS
<b>Dimensions</b>	8,642	23.3% (2,010)	11.8% (1,017)	76.7% (6,632)	88.2% (7,624)
<b>Scopus</b>	2,166	–	40.4% (874)	–	59.6% (1,292)
<b>WoS</b>	1,067	81.9% (874)	–	18.1% (193)	–

Publications in Dimensions but neither WoS nor Scopus were investigated to identify the document types uniquely found by Dimensions. The Dimensions-only publications were almost all from 2020 (95%), as were the publications that were also in Scopus or WoS (99%). The biggest single source was preprint archives (39% of the documents unique to Dimensions): medRxiv, SSRN, Research Square, bioRxiv, chemRxiv, and JMIR Preprints. An additional 2% were other nonjournal publications (e.g., book chapters). The remainder were either in journals not indexed by WoS and Scopus or in journals indexed more slowly by WoS and Scopus. For example, Dimensions had indexed 58 *Nature* articles that Scopus and WoS had not yet recorded (although they included 19 others in *Nature*), and neither Scopus nor WoS had indexed *Medical Gas Research* or *Chinese Journal of Internal Medicine*.

In terms of citations found by the three databases for the matching publications, Dimensions citation counts for all its matching COVID-19 publications were 4.9 and 2.8 times as numerous as WoS and Scopus, suggesting that for recently published or in-press articles, Dimensions had faster citation indexing than WoS and Scopus or from faster sources, such as preprint archives. This could be important when scholars want to consider early citation impact evidence for identifying relevant COVID-19 publications or for the impact assessment of published articles.

#### 4.2. Most Cited Papers

Other factors being equal, the most cited papers are likely to be at the core of humanity's early response to COVID-19 and the most mentioned papers illustrate the public perception of the most relevant research. The age, type, publication venue, and titles of these documents may therefore give insights into important early scientific contributions to the disease. Lower ranked documents are likely to have a different character, however, so the results should not be used as proxies for all COVID-19 research.

The documents with the most Mendeley readers and Dimensions citations tended to be similar and to provide primary clinical and epidemiological evidence about COVID-19 (Table 6). Shorter publication formats and analyses are more evident in the social web and news sources, representing a partially different type of document. The social web and news articles also seemed to give information that might be particularly useful, as public health information for the vast majority of the planet's population that had not yet caught COVID-19 by 18 April 2020. These include studies on facemasks, the stability of the virus on surfaces, and pregnancy risks.

None of the five documents most cited on Reddit were also in the top five for the other sources, although they seem to cover similar topics (Table 7). The paper about Malayan pangolins is the exception for not covering the primary characteristics of the disease or public health issues. This may be an artifact of the relatively low numbers of Reddit citations.

Although the top five articles for Dimensions were published in 2020, by March 21 they had all been cited at least 200 times in Dimensions (Figure 3), perhaps mainly by preprints, letters, and short-form fast-publishing formats, such as brief communications, (academic) news, and case reports. All five documents exhibit a reasonably steady rate of increase. The simultaneous jumps in the lines presumably reflect weekly large-scale database refreshing for Dimensions, although there were also smaller daily changes.

The top five Mendeley documents also started March 21 with a high number of readers, but almost five times more than the number of Dimensions citations (Figure 4). There was a similar pattern of steadily increasing numbers of Mendeley readers with periodic interruptions. In this case the interruptions resulted in temporary decreases in the numbers of Mendeley readers. This could be due to two factors. Either the database consolidates weekly, such as by merging duplicates, or

**Table 6.** Characteristics and ranks of COVID-19 papers in the top five for Dimensions (D), Mendeley (M), Twitter (T), Facebook (F), and News (N), and their ranks in these sites on April 18, 2020. Citation and altmetric counts are in the figures below.

Title	Journal*	Date	Type	D	M	T	F	N
Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China	<i>Lancet</i>	January 24	Article	1	1			
A novel coronavirus from patients with pneumonia in China, 2019	<i>NEJM</i>	February 20	Brief report	2	4		3	
Early transmission dynamics in Wuhan, China, of Novel coronavirus-Infected pneumonia	<i>NEJM</i>	March 26	Article	3	3			
Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study	<i>Lancet</i>	January 30	Article	4				
Clinical characteristics of 138 hospitalized patients with 2019 novel Coronavirus-Infected pneumonia in Wuhan, China	<i>JAMA</i>	February 7	Original Investigation	5				
Clinical characteristics of coronavirus disease 2019 in China	<i>NEJM</i>	February 28	Article		2	4		
Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study	<i>Lancet</i>	March 11	Article		5			3
The proximal origin of SARS-CoV-2	<i>Nature Medicine</i>	March 17	Correspondence			1	4	
Treatment of 5 critically ill patients with COVID-19 with convalescent plasma	<i>JAMA</i>	March 27	Preliminary Comm.			2		
Respiratory virus shedding in exhaled breath and efficacy of face masks	<i>Nature Medicine</i>	April 3	Brief Comm.			3		
Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1	<i>NEJM</i>	March 17	Correspondence			5		1
Coronavirus latest: CERN scientists join the COVID-19 fight	<i>Nature</i>	April 8	News					1
Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records	<i>Lancet</i>	February 12	Article				2	
Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China	<i>JAMA</i>	February 24	View-point				5	4
The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application	<i>Annals of Internal Medicine</i>	March 10	Original research					2
Severe outcomes among patients with coronavirus disease 2019 (COVID-19) – United States, February 12–March 16, 2020	<i>Morbidity Mortality Weekly Report</i>	March 18	Report					5

\* *NEJM*: New England Journal of Medicine; *JAMA*: Journal of the American Medical Association.

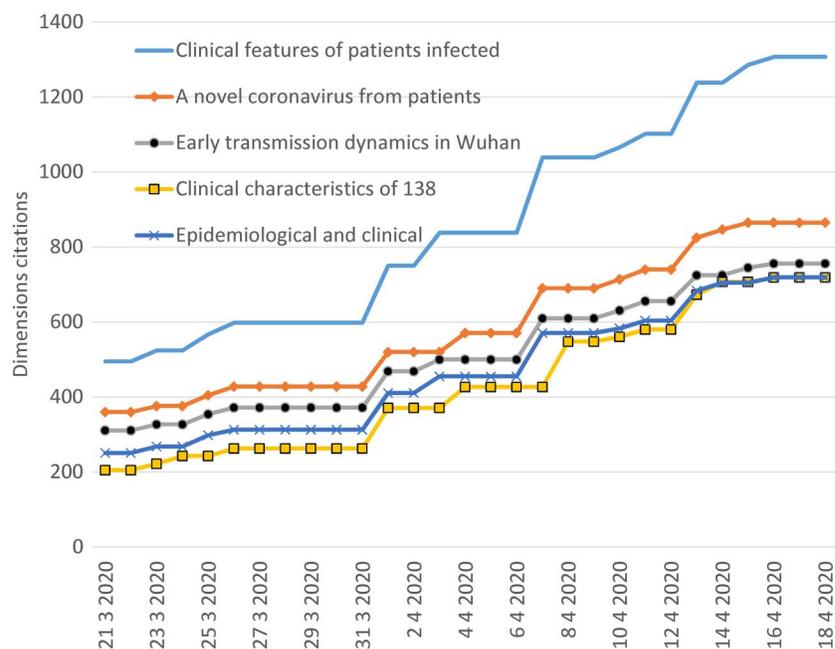
**Table 7.** Characteristics and ranks of COVID-19 papers in the top five for Reddit, and their ranks in these sites on April 18, 2020. There is no overlap with Table 1. Citation and altmetric counts are in the figures below.

Title	Journal	Date	Type	R
The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients	<i>Journal of Medical Virology</i>	February 27	Review	1
Persistence of coronaviruses on inanimate surfaces and its inactivation with biocidal agents	<i>Journal of Hospital Infection</i>	February 6	Review	2
High temperature and high humidity reduce the transmission of COVID-19	SSRN	March 10	Preprint	3
Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins	<i>Nature</i>	March 26	Article	4
Early release – high contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2 – Volume 26, Number 7-July 2020	<i>Emerging Infectious Diseases</i>	April 7	Research	5

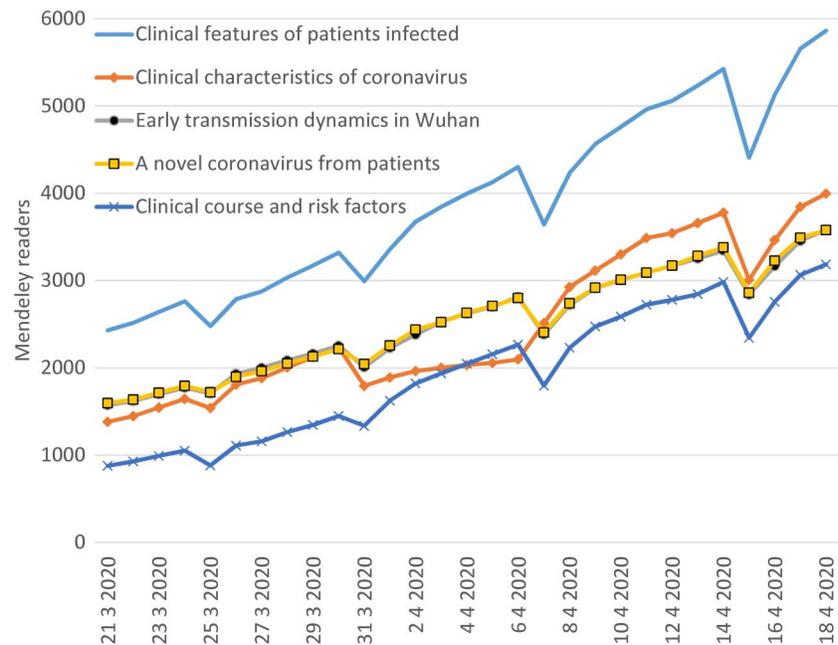
its search is somehow weakened periodically so that the free text search (which is submitted in parallel with the DOI search) matches fewer documents. It is not possible to check which is correct from the data because Mendeley reports reader counts, not the identities of these readers.

Twitter (Figure 5) has a very different pattern to Dimensions and Mendeley. First, some of the documents are much younger, published during the date range analyzed. Second, the number of tweeters achieves close to its maximum when first found by Dimensions, although this is not necessarily the original publication date.

Facebook has a similar growth pattern to Twitter, except that there is a period of increasing interest for the proximal origin paper (Figure 6), which has a more moderate growth on Twitter. An (apparently speculative) news story about CERN scientists that was popular on Facebook did not get traction on Twitter and seems unlikely to become highly cited or read.

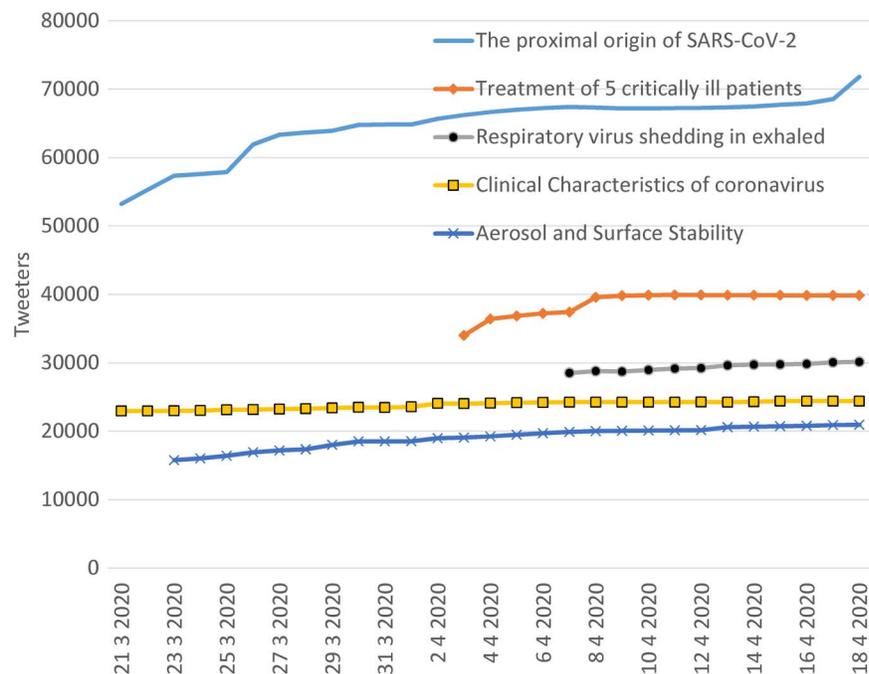


**Figure 3.** The cumulative number of Dimensions citations for the five most cited COVID-19 documents.

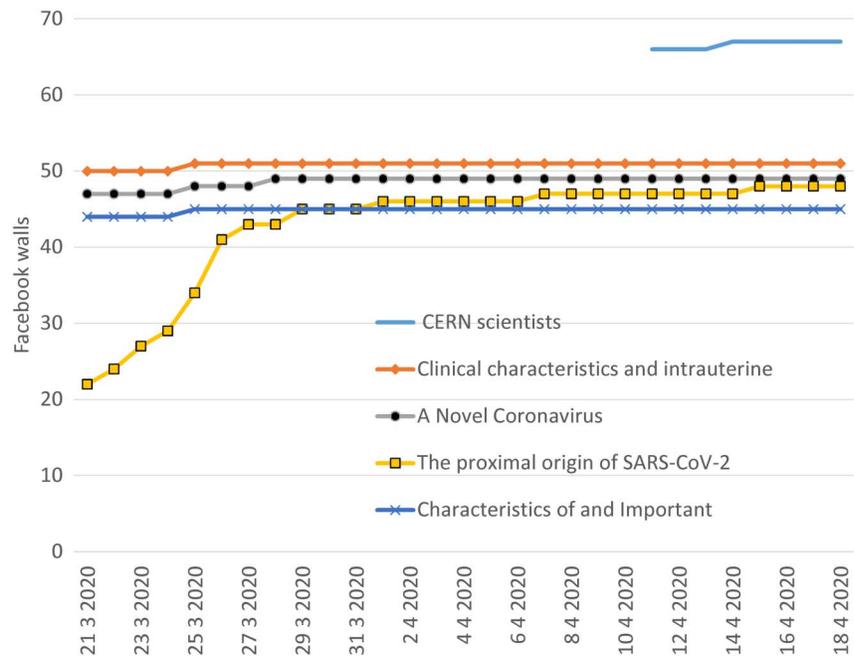


**Figure 4.** The cumulative number of Mendeley readers for the five most read COVID-19 documents.

The top news-cited articles were all covered by at least 400 news sources by the end of the period (Figure 7). Perhaps surprisingly, given that news is very time-dependant, all the sources experienced significant increases in the number of citing sources. *Altmetric.com* is constantly expanding its coverage of news sources (which is possible, but seems unlikely), it is slow to



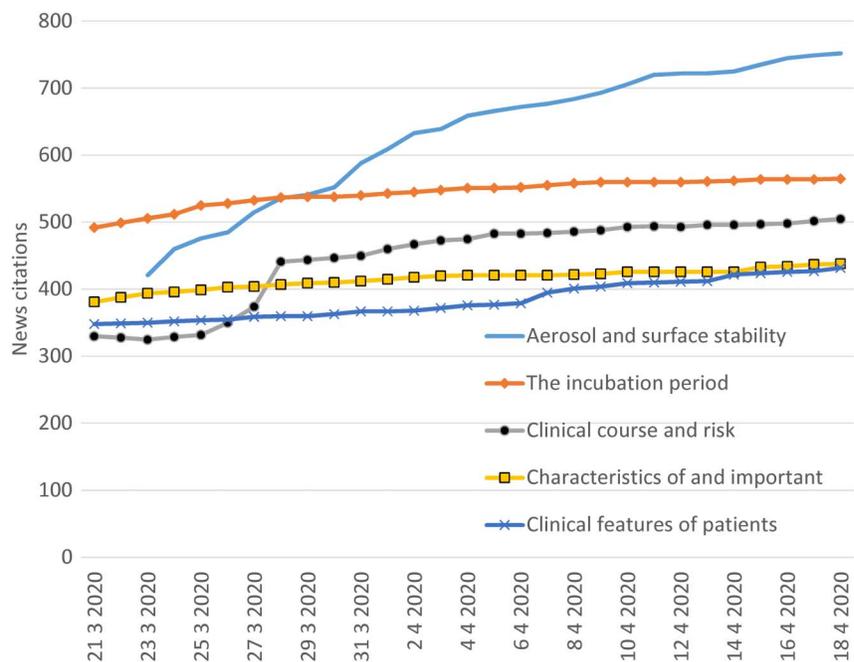
**Figure 5.** The cumulative number of Tweeters for the five most tweeted COVID-19 documents.



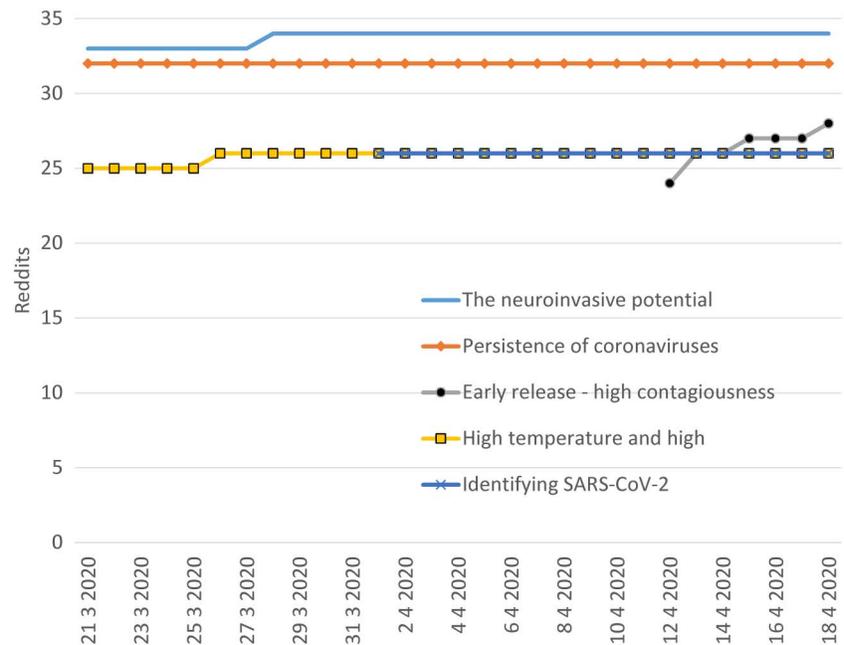
**Figure 6.** The cumulative number of Facebook wall posts for the five most walled COVID-19 documents.

update its news coverage, or news stories about COVID-19 are prepared to cite old articles, perhaps for a more in-depth commentary or as background context for new articles.

There were relatively few citations from Reddit, despite its use as a news source and many academic themes (subreddits) within the site (Figure 8). Perhaps reflecting its news status, older articles do not seem to increase their Reddit citation counts.



**Figure 7.** The number of news citations for the five most News-cited COVID-19 documents.



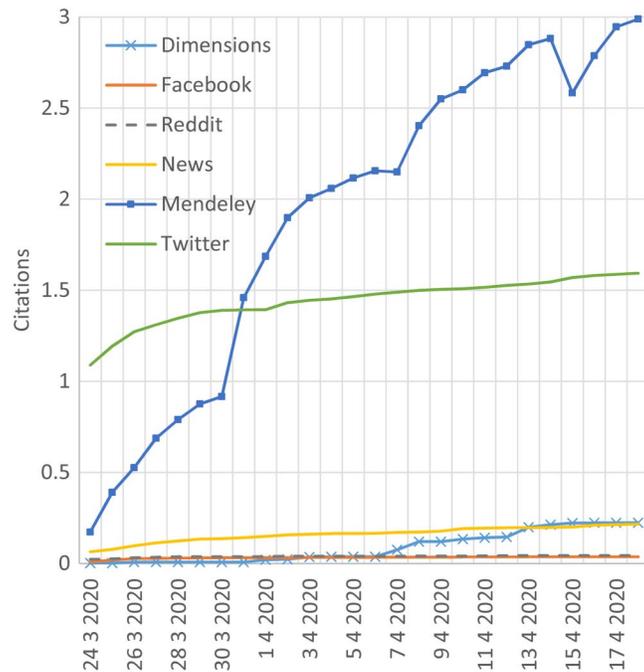
**Figure 8.** The cumulative number of Reddit posts for the five most posted COVID-19 documents.

#### 4.3. A Comparison Between Average Scores for Different Sources

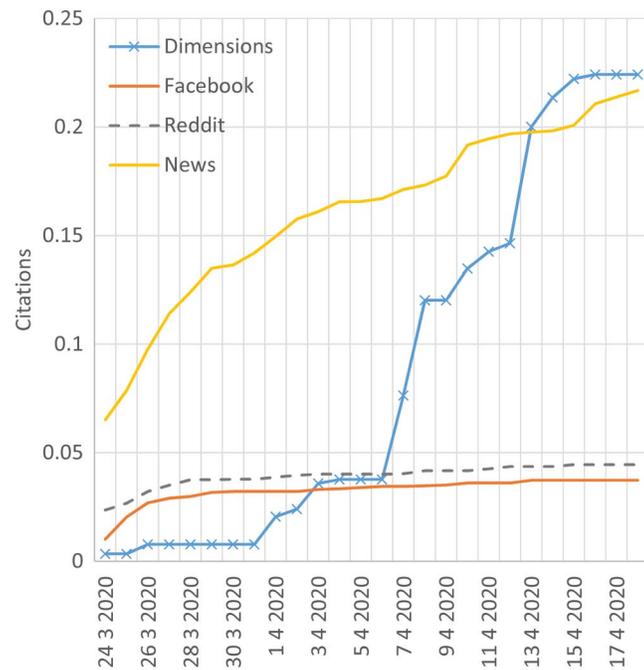
March 24, 2020 was selected for a time series analysis because this date in the first week had the most new articles (349) found by Dimensions. For documents first found by Dimensions on March 24, 2020 and matching the COVID-19 queries, the average score was highest for Twitter and already above 1 on the start day (Figure 9). Average tweeter counts then increased slowly after the first few days. In contrast, average Mendeley reader counts for these 349 articles started close to zero and increased rapidly, except for weekly Mendeley indexing adjustments. Mendeley overtook Twitter after a week.

The low initial value for Dimensions citations and the high initial average number of tweets are unsurprising, given that citations take time to accrue due to publication delays (even for preprints), but articles can be tweeted as soon as they are published. As Twitter is a news source and authors/editors/publishers/current awareness browsers might tweet to announce a publication, high initial tweet counts are to be expected. Mendeley users can also add papers to their libraries as soon as they are published, but the slow growth might represent researchers and students saving the articles to read on the day of publication and then adding them to Mendeley after reading them. The figures for Mendeley are likely to also include people that found the articles through literature searches rather than current awareness, adding them to Mendeley when read, found, or cited in a paper (Mendeley automatically builds reference lists).

The average citation counts for the remaining three sources were all much lower than for Mendeley and Twitter (Figure 10). Although Facebook and Reddit both displayed a similar growth pattern to Twitter (rapid initially, then slow), both News citations and Dimensions citations increased steadily. The low number of Dimensions citations is unsurprising, given publication delays, and the small but nontrivial increase for Dimensions suggests that many authors of COVID-19 papers quickly found current research and added it to their papers, then published them as preprints. The constant growth for news sources is unexpected, given that news is supposed to be current. Possible explanations for this (in addition to those discussed



**Figure 9.** Daily average (geometric mean) citations by source for documents first found by Dimensions on March 24, 2020 ( $n = 349$ ).



**Figure 10.** Daily average (geometric mean) citations by source for documents first found by Dimensions on March 24, 2020, excluding Mendeley and Twitter ( $n = 349$ ).

above) are delays in the production of slower news sources (e.g., magazine-type articles), delays in writing university press releases, and articles being discussed in the press after passing peer review or after being formally published in early view or an online volume.

**4.4. Overlaps in Citation Counts Between Sources**

Spearman correlation tests reveal the extent to which the same documents that are cited by one source are also cited by another source, together with the extent that they are cited. By April 18, 2020, correlations between Dimensions citations and altmetrics for documents first found by Dimensions on March 24 were strong, except for Reddit (Table 7). Because most (229; 66%) documents were uncited by April 18, the correlation mainly confirms that, except for Reddit, news stories, publishing authors, and users of the different platforms tended to select the same documents for attention. The altmetrics also correlated moderately or strongly with each other, except for Reddit, in agreement with this conclusion. Thus, for the narrow topic of COVID-19, there seems to be a researcher-news-social media consensus about the most important topics, at least in the (very) short term.

The correlations (Table 8) do not take into account field differences or document type differences. The relatively high correlations could be at least partially due to ignoring contributions of low relevance to COVID-19, such as book chapters mentioning the possibility of a coronavirus 2, editorials, letters, and subject areas making relatively peripheral contributions to immediate needs.

The positive correlations might be influenced by a mix of publication venues and document types. The 349 documents included 239 (68.5%) papers in journals, 67 (19.2%) papers in preprint archives, and 27 (7.7%) magazine articles, with 16 (4.6%) not assigned to a publication venue by Dimensions (e.g., book chapters, reports). In terms of rank order, for all five sources, on average, journal articles were more highly ranked than the other types and preprints were more highly, or equally ranked with, magazine articles. The average ranks were Journals (D: 158; M: 136; T: 154; F: 168; N: 165; R: 167); preprints (D: 198; M: 258; T: 195; F: 191; N: 186; R: 191.5); magazines (D: 235; M: 258; T: 280; F: 191; N: 221; R: 191.5). The magazines (*Alcoholism & Drug Abuse Weekly*; *Focus on Catalysts*) included news stories about the societal side-effects of the disease rather than research about the disease (e.g., “China refineries reduce operating rates”). Preprints presumably attract less attention because they have not been peer reviewed. In addition, the documents in journals included letters and news stories, which may also have lower relevance to COVID-19 research and many received little attention from any

**Table 8.** Spearman correlations between citation counts and altmetrics from April 18, 2020 for COVID-19 documents first found by Dimensions on March 24, 2020. All are statistically significant at  $p = .001$  ( $n = 349$ )

	Mendeley	Twitter	Facebook	News	Reddit
Dimensions	.653*	.659*	.453*	.529*	.249*
Mendeley	1	.689*	.375*	.473*	.354*
Twitter		1	.411*	.626*	.363*
Facebook			1	.376*	.251*
News				1	.335*

\* Statistically significant at  $p = 0.001$ .

source (e.g., the uncited news article “Seven days in medicine: 11–17 March 2020”, in the *BMJ*, with 27 readers and two tweets). Thus, both altmetrics and citations seem to focus on contributions of *types* that are more core to COVID-19 as a medical and public health research issue.

The influence of nonarticle document types on the correlations were tested by filtering out all nonarticles. After manually removing documents that were not journal articles (mainly editorials, news, and letters), there were 106 standard journal articles (including reviews). Nevertheless, the correlations did not substantially change (Table 9). Some of the removed editorials had been cited, read, and shared, explaining the similar positive correlations.

The two most common Dimensions subject codes for the March 24 set were 1117 Public Health and Health Services ( $n = 78$ ) and 1103 Clinical Sciences ( $n = 32$ ). Except for Reddit (correlations close to 0), the pairwise correlations change little if the set is restricted to only subject categories 1117 or 1103, with or without excluding nonarticle types. For example, the lowest correlation between Twitter and Dimensions for any of these four restricted sets is .638 (category 1103 with all document types,  $n = 32$ ). Thus, except for Reddit, the strong positive correlations between indicators do not seem to be due to field differences in the data set.

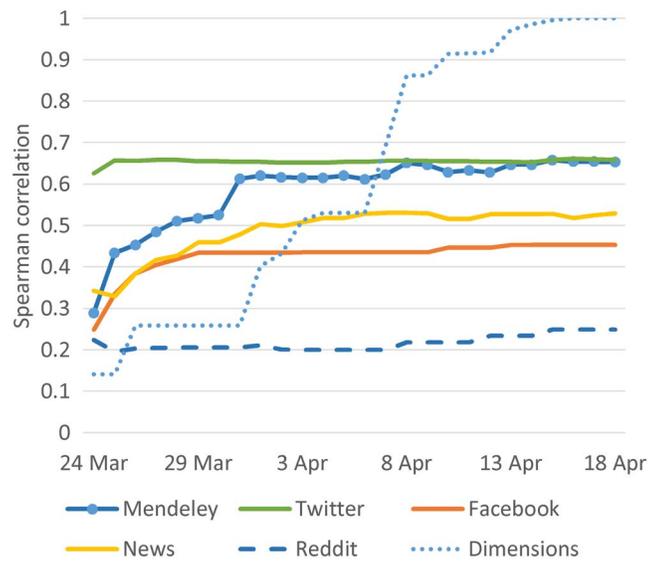
Focusing on the smallest data set mentioned above, 1103 Clinical Sciences, journal articles only ( $n = 19$ ), one of the reasons for the strong positive correlations is that three of the articles were in national journals (*Korean Journal of Radiology*, *Chung-Hua Wai Ko Tsa Chih*, *Chinese Journal of Gastrointestinal Surgery*) and the rest were in international journals or a prestigious national journal that is effectively international (*JAMA*). The three national articles collectively had one citation, three tweets, no mentions in the other altmetrics, and two of the three lowest Mendeley reader counts. This is consistent with lower quality or impact national research being less cited and less read, which is unsurprising. It is more surprising that national research is less tweeted than international research, which has not previously been found by altmetrics studies. In this case, two articles were not in English and this, combined with Twitter not being used in China, might be the explanation.

The top article for all metrics in the 1103 Clinical Sciences journal articles only set (32 citations, 604 readers, 1504 tweeters, three Facebook walls, 31 news stories, one Reddit) was the research letter (classified here as an article) “Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State” from *JAMA* (published March 19, 2020, but picked up by Dimensions on March 23/24). This seems similar to the *Annals of Palliative Medicine* article, “Risk factors associated with disease progression in a cohort of [17] patients infected with the 2019 novel coronavirus” from March 22, 2020, which had low scores on all metrics (zero citations, 92 (third fewest) Mendeley readers, one tweet, zero on the

**Table 9.** Spearman correlations between citation counts and altmetrics from April 18, 2020 for COVID-19 journal articles first found by Dimensions on March 24, 2020 ( $n = 106$ )

	Mendeley	Twitter	Facebook	News	Reddit
Dimensions	.693***	.734***	.589***	.585***	.250**
Mendeley	1	.687***	.401***	.473***	.316***
Twitter		1	.562***	.719***	.382***
Facebook			1	.440***	.215*
News				1	.334***

\*Statistically significant at  $p = 0.05$ ; \*\*Statistically significant at  $p = 0.01$ ; \*\*\*Statistically significant at  $p = 0.001$ .



**Figure 11.** Spearman correlations between altmetrics and Dimensions citation counts from April 18 for COVID-19 documents first found by Dimensions on March 24, 2020 ( $n = 349$ ). Correlations with Dimensions citation counts on the same dates are also reported for context.

others). The first article was in a more prestigious journal and concerned patients from the United States, whereas the second article was more detailed (e.g., pictures, full article, more words, statistical analysis, more references) and was about patients from Nanchang, China. This, combined with the previous three cases, suggests that the regional bias of Twitter (a natural side-effect of news focusing on local issues) coincides with the US/UK or Western domination of more prestigious medical journals. This might not have been visible previously in altmetric studies for the medical domain because the current data set presumably has a higher proportion of Chinese articles than normal, given China’s earlier research into the disease. This is a speculative conclusion, however, and may not be correct. Not all research from China or in nonprestigious sources was ignored in academia. The second most cited article (in an apparently nonprestigious international journal) was, “Clinical features of severe pediatric patients with coronavirus disease 2019 in Wuhan: a single center’s observational study” from the *World Journal of Pediatrics* (seven citations, 347 readers (second highest), only 32 tweets, zero others). This article’s focus on children may have been relatively unique, and therefore particularly valuable for researchers.

Also for the same set of 19 Clinical Sciences articles from March 24, there seemed to be a tendency for articles attracting more attention to be more central to COVID-19. Ignoring the four articles discussed above, the remaining uncited article, also with low social media scores, was “Coronavirus Disease (COVID-19): Spectrum of CT Findings and Temporal Progression of the Disease,” from *Academic Radiology*, which focuses on the radiology dimension. Another relatively specialist and universally low scoring article was, “COVID-19 – what should anaesthesiologists and intensivists know about it?” from *Anesthesiology Intensive Therapy* (1 citation, 218 readers, 18 tweets, 0 others).

#### 4.5. Early Altmetrics and Later Citation Counts

Ideally, an indicator would help researchers and policy-makers to identify important articles when they are first published, without having to wait for enough citations. To check for early evidence of later citation impact, the indicators were correlated with Dimensions citation

counts on April 18, representing longer term citation counts (this is a weak proxy, because decades are sometimes used for long-term citations in other contexts, such as Stegehuis, Litvak, & Waltman, 2015).

On the day that a document is first findable in Dimensions, its tweeter count is the best indicator of likely long-term citation impact (Figure 11). Twitter users seem to be able to notice documents approximately on the date of first publication for their potential importance to COVID-19. After this date, the tweeter count does not increase much and its correlation with longer term Dimensions citations is stable. After about 3 weeks, Mendeley reader counts take over as a marginally better indicator of longer term citation impact. It is not clear whether the same would be true for more mature citation counts, however, such as after a year. It is possible that early Dimensions citations (and Mendeley readers) reflect more temporary interest and are themselves highly influenced by the news or social sharing on Twitter, for example. The most cited sets of five papers analyzed above suggest that highly recognized papers are particularly important for the disease, however. As above, this correlation ignores field differences and document type differences, although document differences seem to have little effect (Tables 8 and 9).

## 5. DISCUSSION

The results are limited by the range of factors mentioned in Section 3. In particular, the coverage figures for the sources are not directly comparable due to the different scopes of the queries. In addition, the count data has not been field-normalized, so the coverage comparisons do not reveal disciplinary differences. The correlations may also be exaggerated by not taking into account disciplinary differences. The results may show different patterns for earlier or later time periods. The properties of the scholarly databases and *Altmetric.com*'s strategies may evolve over time, rendering the results obsolete. They may also not be applicable for later stages of COVID-19 research or for future epidemics or pandemics.

The COVID-19 query results comparison confirms the previous finding that COVID-related academic publications are appearing rapidly (Torres-Salinas, 2020). In addition, it confirms that Dimensions finds many publications not in Scopus and WoS but that Scopus indexes nearly all relevant publications found in the WoS core collection with the Conference Proceedings Citation Index. Presumably the difference would be smaller if other parts of WoS were included, such as the Book Citation Index, although the core collection includes the Emerging Sources Citation Index (Clarivate, 2020).

The results are not directly comparable to studies from before COVID-19 due to the unprecedented speed and volume of publishing on the topic. For example, Dimensions citation counts accrue more rapidly than previously reported for any topic. For comparison, the Scopus citations of 12 subject categories (full journal articles only) were a maximum of 0.12 in the month of publication, whereas the COVID-19 mixed set averaged almost double this after 3 weeks. The results are also qualitatively different in some respects. Although correlation tests have previously found tweeter counts to have little value as a scholarly impact indicator due to very low correlation with citation counts, typically close to 0 or even negative (Costas et al., 2015; Haustein et al., 2014; Thelwall, Haustein, et al., 2013), the current study has found tweet counts to be reasonable academic impact indicators and the best early impact indicator for the first 3 weeks. This may be partly due to the set of articles here covering multiple disciplines, but the results for the top-cited documents suggest that altmetrics are effective at pointing to the documents that are most central to COVID-19 as a medical and public health issue. Thus, the unprecedented threat of COVID-19 seems to have led to an unprecedentedly high and focused level of

societal and academic attention being given to the most relevant research. There was some support for this from the correlation analysis for March 24, 2020 documents. This correlation analysis also suggested that the high correlation may also be at least partly due to an international issue: a relatively high amount of publishing from China not in prestigious journals coupled with greater interest in research concerning patients in Twitter-using countries (particularly the United States), and that research sometimes being published in more prestigious journals.

## 6. CONCLUSIONS

The confirmed rapid increase in COVID-19 academic publications is encouraging in terms of the academic community rapidly reacting to the need for relevant research and commentaries. The importance of short-form and quick contributions (viewpoints, correspondence, brief reports) is also evident in the highly cited papers, as is the importance of academic research for practical public health issues. Dimensions seems to be the most comprehensive database to find relevant literature, although Google Scholar might have wider coverage and be useful to those that do not mind its false matches.

Despite the apparent high medical and public health value of some academic papers, the huge number of publications returned by a relevant search will presumably make the most important publications more difficult to find. This should not be a problem for medical researchers trained to use MeSH queries effectively, but might be problematic for other researchers, end users, and the public, who may find bewilderingly many matches for their queries. The altmetric results suggest that altmetrics may be helpful for researchers needed to quickly identify the most useful new documents from the large number published daily. Altmetric counts may help to distinguish between core primary research and other contributions, such as editorial commentaries with narrower disciplinary or professional relevance (e.g., radiographers). Perhaps ironically, given that a core original goal for altmetrics was to develop indicators of societal impact that were different from scholarly impact indicators (Priem et al., 2010), their greatest value (as early impact indicators) seems to be occurring when the two concepts are most closely converging.

## AUTHOR CONTRIBUTIONS

Kayvan Kousha: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing. Mike Thelwall: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing.

## COMPETING INTERESTS

The authors have no competing interests.

## FUNDING INFORMATION

This research was not funded.

## DATA AVAILABILITY

The processed data used to produce the graphs are available in the supplementary material (<https://doi.org/10.6084/m9.figshare.12301475>).

## REFERENCES

- Adie, E., & Roe, W. (2013). Altmetric: Enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11–17.
- Allen Institute. (2020). COVID-19 Open Research Dataset (CORD-19). <https://pages.semanticscholar.org/coronavirus-research>
- Chen, Q., Allot, A., & Lu, Z. (2020). Keep up with the latest coronavirus research. *Nature*, 579(7798), 193–193.
- Clarivate. (2020). Web of Science: Emerging Sources Citation Index. <https://clarivate.com/webofsciencegroup/solutions/webofscience-esci/>
- Clark, F., & Illman, D. L. (2006). A longitudinal study of the *New York Times* Science Times section. *Science Communication*, 27(4), 496–513.
- Colavizza, G., Costas, R., Traag, V. A., Van Eck, N. J., Van Leeuwen, T., & Waltman, L. (2020). A scientometric overview of CORD-19. *BioRxiv*. <https://doi.org/10.1101/2020.04.20.046144>
- Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019.
- de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292–306.
- Delwiche, F. A. (2018). Bibliometric analysis of scholarly publications on the Zika virus, 1952–2016. *Science & Technology Libraries*, 37(2), 113–129.
- Fairclough, R., & Thelwall, M. (2015). More precise methods for national research citation impact comparisons. *Journal of Informetrics*, 9(4), 895–906.
- Gunn, W. (2014). Mendeley: Enabling and understanding scientific collaboration. *Information Services & Use*, 34(1–2), 99–102.
- Harzing, A. W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, 120(1), 341–349.
- Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter. *Journal of the Association for Information Science and Technology*, 67(1), 232–238.
- Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLOS ONE*, 10(3), e0120495.
- Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? *Information Technology*, 56(5), 207–215.
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), 387–395.
- Holmberg, K., & Vainio, J. (2018). Why do some research articles receive more online attention and higher altmetrics? Reasons for online success according to the authors. *Scientometrics*, 116(1), 435–447.
- Ioannidis, J. P. (2020). Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. *European Journal of Clinical Investigation*. <https://doi.org/10.1111/eci.13222>
- Kostoff, R. N., & Morse, S. A. (2011). Structure and infrastructure of infectious agent research literature: SARS. *Scientometrics*, 86(1), 195–209.
- Kousha, K., & Thelwall, M. (2019a). Can Google Scholar and Mendeley help to assess the scholarly impacts of dissertations? *Journal of Informetrics*, 13(2), 467–484.
- Kousha, K., & Thelwall, M. (2019b). An automatic method to identify citations to journals in news stories: A case study of the UK newspapers citing Web of Science journals. *Journal of Data and Information Science*, 4(3), 73–95.
- Lewis, G. (2008). The reporting of the risks from severe acute respiratory syndrome (SARS) in the news media, 2003–2004. *Health, Risk and Society*, 10(3), 241–262.
- medRxiv. (2020). COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. <https://connect.medrxiv.org/relate/content/181>
- Merchant, R. M., & Lurie, N. (2020). Social media and emergency preparedness in response to novel coronavirus. *Journal of the American Medical Association*, 323(20), 2011–2012. <https://doi.org/10.1001/jama.2020.4469>
- Mohammadi, E., Barahmand, N., & Thelwall, M. (2019). Who shares health and medical scholarly articles on Facebook? *Learned Publishing*, 33(1), 111–118.
- Mohammadi, E., Gregory, K., Thelwall, M., Barahmand, N. (2020). Which health and biomedical topics generate the most Facebook interest and the strongest citation relationships? *Information Processing and Management*, 57(3). <https://doi.org/10.1016/j.ipm.2020.102230>
- Mohammadi, E., Thelwall, M., & Kousha, K. (2016). Can Mendeley bookmarks reflect readership? A survey of user motivations. *Journal of the Association for Information Science and Technology*, 67(5), 1198–1209.
- Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. (2018). Academic information on Twitter: A user survey. *PLOS ONE*, 13(5), e0197265.
- Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*, 66(9), 1832–1846.
- Orduña-Malea, E., & Delgado-López-Cózar, E. (2018). Dimensions: re-discovering the ecosystem of scientific information. *El Profesional de la Información*, 27(2), 420–431.
- Ortega, J. L. (2018a). Reliability and accuracy of altmetric providers: A comparison among Altmetric.com, PlumX and Crossref Event Data. *Scientometrics*, 116(3), 2123–2138.
- Ortega, J. L. (2018b). The life cycle of altmetric impact: A longitudinal study of six metrics from PlumX. *Journal of Informetrics*, 12(3), 579–589.
- Ovadia, S. (2015). More than just cat pictures: Reddit as a curated news source. *Behavioral & Social Sciences Librarian*, 34(1), 37–40.
- PMC. (2020). Public Health Emergency COVID-19 Initiative. <https://www.ncbi.nlm.nih.gov/pmc/about/covid-19/>
- Pouris, A., & Ho, Y. (2016). A bibliometric analysis of research on Ebola in Science Citation Index expanded. *South African Journal of Science*, 112(3–4). <https://doi.org/10.17159/sajs.2016/20150326>
- Pouris, A., & Pouris, A. (2011). Scientometrics of a pandemic: HIV/AIDS research in South Africa and the world. *Scientometrics*, 86(2), 541–552.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. <http://altmetrics.org/manifesto/>
- Rethlefsen, M. L., & Livinski, A. A. (2013). Infectious diseases citation patterns: Mapping the literature 2008–2010. *Journal of the Medical Library Association*, 101(1), 55.
- Robinson-García, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLOS ONE*, 12(8), e0183551.

- Robinson-García, N., Torres-Salinas, D., Zahedi, Z., & Costas, R. (2014). New data, new possibilities: Exploring the insides of Altmetric.com. *El Profesional de la Información*, 23(4), 359–366.
- Song, P., & Karako, T. (2020). COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. *BioScience Trends*, 14(1). <https://doi.org/10.5582/BST.2020.01056>
- Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of Informetrics*, 9(3), 642–657.
- Stoddard, G. (2015). Popularity and quality in social news aggregators: A study of Reddit and Hacker News. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 815–818). New York: Association for Computing Machinery.
- Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131–1143.
- Thelwall, M. (2017a). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, 113(3), 1721–1731.
- Thelwall, M. (2017b). Are Mendeley reader counts high enough for research evaluations when articles are published? *Aslib Journal of Information Management*, 69(2), 174–183. <https://doi.org/10.1108/AJIM-01-2017-0028>
- Thelwall, M. (2017c). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa? *Journal of Librarianship & Information Science*, 49(2), 144–151.
- Thelwall, M. (2018a). Early Mendeley readers correlate with later citation counts. *Scientometrics*, 115(3), 1231–1240.
- Thelwall, M. (2018b). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2), 430–435.
- Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics*, 17(1). <https://digital.csic.es/handle/10261/174572>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLOS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>
- Thelwall, M., & Wilson, P. (2016). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. *Journal of the Association for Information Science and Technology*, 67(8), 1962–1972.
- Tian, D., & Zheng, T. (2015). Emerging infectious disease: Trends in the literature on SARS and H7N9 influenza. *Scientometrics*, 105(1), 485–495.
- Torres-Salinas, D. (2020). Ritmo de crecimiento diario de la producción científica sobre Covid-19. Análisis en bases de datos y repositorios en acceso abierto. *El Profesional de la Información*, 29(2), e290215. <https://doi.org/10.3145/epi.2020.mar.15>
- Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature News*, 512(7513), 126–130.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2009). Modeling a century of citation distributions. *Journal of Informetrics*, 3(4), 296–303.
- WHO (2020). *Rolling updates on coronavirus 2019*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
- Yu, H., Xu, S., Xiao, T., Hemminger, B. M., & Yang, S. (2017). Global science discussed in local altmetrics: Weibo and its comparison with Twitter. *Journal of Informetrics*, 11(2), 466–482.
- Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, 68(10), 2511–2521.
- Zahedi, Z., & Haustein, S. (2018). On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications. *Journal of Informetrics*, 12(1), 191–202. <https://doi.org/10.1016/j.joi.2017.12.005>
- Zahedi, Z., Haustein, S., & Bowman, T. (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. In *SIG/MET Workshop, ASIS&T 2014 Annual Meeting, Seattle*. <http://www.asis.org/SIG/SIGMET/data/uploads/sigmet2014/zahedi.pdf>