

# Evaluation of the Dynamics of Large Scale COVID-19 Related Literature through Bibliometric Analysis from a Mathematical Standpoint

Biplab Sarkar<sup>1,2</sup>, Anusheel Munshi<sup>3</sup>, Bhaswar Ghosh<sup>4</sup>, Tharmarnadar Ganesh<sup>3</sup>, Arjunan Manikandan<sup>5</sup>, Subhra Snigdha Biswal<sup>1</sup>, Tanweer Shahid<sup>1</sup>, Rajagopalan Bhaskar<sup>1</sup>, Sinjini Sengupta<sup>1</sup>, Sandipan Roy Chowdhury<sup>1</sup>, Jibak Bhattacharya<sup>1</sup>, Mahasin Gazi<sup>1</sup>, Arundhati De<sup>1</sup>, Kirubha George<sup>1</sup>, Tanmoy Ghosh<sup>1</sup>, Jawed Akhtar<sup>1</sup>, Sourav Mandal<sup>1</sup>, Mukti Mukherjee<sup>1</sup>, Rosalima Gupta<sup>6</sup>, Soumya Roy<sup>2,3</sup>, Kanan Jassal<sup>7</sup>, Suchanda Goswami<sup>1</sup>, Suparna Ghosh<sup>1</sup>, Prasenjit Chatterjee<sup>1</sup>, Litan Naha Biswas<sup>1</sup>, Syamasis Bandyopadhyay<sup>1</sup>, Anirudh Pradhan<sup>2</sup>, Bidhu Kalyan Mohanti<sup>8</sup>

<sup>1</sup>Apollo Multispeciality Hospitals, Kolkata, West Bengal, INDIA.

<sup>2</sup>GLA University, Mathura, Uttar Pradesh, INDIA.

<sup>3</sup>Manipal Hospitals, Dwarka, New Delhi, INDIA.

<sup>4</sup>International Institute of Information Technology, Hyderabad, Telangana, INDIA.

<sup>5</sup>Apollo Proton Cancer Center, Chennai, Tamilnadu, INDIA.

<sup>6</sup>Patient Information Centre, Kolkata, West Bengal, INDIA.

<sup>7</sup>Fortis Memorial Research Institute, Gurgaon Haryana, INDIA.

<sup>8</sup>Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, INDIA.

## ABSTRACT

This study aims to analyze the dynamics of the published articles and preprints of Covid-19 related literature from different scientific databases and sharing platforms. The PubMed, ScienceDirect, and ResearchGate (RG) databases were under consideration in this study over a specific time. Analyses were carried out on the number of publications as (a) function of time (day), (b) journals and (c) authors. Doubling time of the number of publications was analyzed for PubMed "all articles" and ScienceDirect published articles. Analyzed databases were (1A) PubMed (01/12/2019-12/06/2020) "all\_articles" (1B) PubMed Review articles) and (1C) PubMed Clinical Trials (2) ScienceDirect all publications (01/12/2019- 25/05/2020) (3) RG (Article, Pre Print, Technical Report) (15/04/2020 – 30/4/2020). Total publications in the observation period for PubMed, ScienceDirect, and RG were 23000, 5898 and 5393 respectively. The average number of publications/day for PubMed, ScienceDirect and RG were  $70.0 \pm 128.6$ ,  $77.6 \pm 125.3$  and  $255.6 \pm 205.8$  respectively. PubMed shows an avalanche in the number of publications around May 10, the number of publications jumped from  $6.0 \pm 8.4$ /day to  $282.5 \pm 110.3$ /day. The average doubling time for PubMed, ScienceDirect, and RG was  $10.3 \pm 4$  days, 20.6 days, and  $2.3 \pm 2.0$  days respectively. The average number of publications per author for PubMed, ScienceDirect, and RG was  $1.2 \pm 1.4$ ,  $1.3 \pm 0.9$ , and  $1.1 \pm 0.4$  respectively. Subgroup analysis, PubMed review articles mean review  $<0|17 \pm 17|77>$  days; and reducing at a rate of  $-0.21$  days (count)/day. The number of publications related to the COVID-19 until now is huge and growing very fast with time. It is essential to rationalize and limit the publications.

**Keywords:** COVID-19, Corona virus, Pandemic, Publications, Doubling time.

## INTRODUCTION

A cluster of viral pneumonia cases of unknown cause, subsequently identified as a novel coronavirus, named as COVID-19 or 2019-nCoV, was detected on December 31, 2019 in Wuhan, China.<sup>[1]</sup> Subsequently, WHO declared it as a global pandemic on March 11, 2020.<sup>[2]</sup> This virus has rapidly crossed borders and led to a major healthcare crisis and

economic slowdown around the globe. Most international health organizations have stated an urgent need to stop, control and reduce the impact of the virus at every opportunity.<sup>[3]</sup> Healthcare systems, various sectors of industry, and overall economy of the globe have been hit severely by this pandemic. As of July 23, 2020, there were 15 million confirmed cases worldwide with the number of fatalities in excess of 600, 000 so far.<sup>[3]</sup>

How has the scientific world reacted to this pandemic? This pandemic has brought to fore the gaping incoherence in opinions expressed by agencies the world over. Government bodies, non-government institutions, pharmaceutical industry, researchers have all made their sounds, but without much unison.

### Copyright

© The Author(s). 2022 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Correspondence

**Biplab Sarkar**

Department of Radiation Oncology,  
Apollo Multispeciality Hospitals,  
West Bengal-700054 Kolkata, INDIA.  
Email id: biplabphy@gmail.com

Received: 30-05-2021;

Revised: 19-12-2021;

Accepted: 20-06-2022;

**DOI:** 10.5530/jscires.11.1.5

COVID-19 has attracted tremendous interest from researchers and clinicians worldwide, resulting in an sudden outbreak in the number of publications observed during the early March-April 2020.<sup>[4-7]</sup>

Given the urgent need for evidence to support clinical and public health decisions, investigators have begun summarising and analysing the published literature to collective existing evidence in the form of systematic reviews. Different bibliometric analyses of the COVID-19 literature are available for limited time period for different data bases like PubMed, Scopus, Web of Science and Google Scholar.<sup>[6,8-12]</sup>

These bibliometric analyses have provided overviews of the COVID-19 research landscape. They have primarily focused on authorship, keywords, and collaboration patterns, country of origin, determined the top-cited publications, etc.,<sup>[6,9]</sup>

By one estimate, the COVID-19 literature published since January has reached more than 23,000 papers and is doubling every 20 days - among the biggest explosions of scientific literature ever.<sup>[4,5,13]</sup> Another study says the COVID-19 literature acceptance time and doubling time is 3 and 14 days respectively.<sup>[14]</sup>

A quick run of the total count of publications with the keyword “COVID-19” on the seventh known strain of the coronavirus that can infect humans on Google Scholar showed about 20,500 results for December 2019-May 2020. For the same period, PubMed yielded 23,480 results for the keyword “COVID-19”[All Fields] [as on 12-06-2020]. The other 6 strains of the same virus yielded around 10 thousand publications between 2000-2020.<sup>[10]</sup> In addition to these staggering numbers, the pandemic has also affected the quality and content of publications. All editorial platforms and the publishing houses have created structures to promote and expedite the publication of COVID-19 related articles.<sup>[7]</sup>

Professional bodies, both national and international, of every specialty have attempted to bring out their own “guidelines” and “recommendations” to deal with patients of their specialty in the “COVID situation”. Researchers seemed to write articles endlessly, often repeating and restating what was already written.<sup>[15]</sup> This also raises questions on the quality of data and the authenticity of the results and typically low peer review time.<sup>[4,5,14,15]</sup>

On April 29, 2020, after a hurriedly conducted trial, the National Institute of Allergy and Infectious Diseases proposed Remdesivir, as an effective drug for this virus.<sup>[16,17]</sup> A contrary opinion as well as serious criticism was given to these studies and some trials showed no benefit of using Remdesivir.<sup>[18,19]</sup> The use of chloroquine or hydroxychloroquine is another example invited widespread criticism.<sup>[21-25]</sup>

Bibliometric analysis of COVID-19 scientific literature showed that one-third of the published papers were on clinical management with poor adherence to research priorities, and over sixty percent were opinion pieces not reporting original data.<sup>[26,27]</sup> The booming number of publications in short time requires a systemic review for its scientific content as well as bibliometric analysis in time.<sup>[4-5,15]</sup>

Present study is limited to bibliometric analysis of temporal distribution (publication vs. days), publication doubling time, spatial distribution (publication vs. journal), and distribution of authorship (publication vs. author) in three different large medical databases. Further, we evaluated the dynamics of the growth characteristics of different parameters through the relevant mathematical formulation. The databases considered were PubMed, ScienceDirect and Research Gate.

## REVIEW OF LITERATURE

Coronavirus disease 2019 (COVID-19) has impelled the global scientific community to take unprecedented interest on a single subject aimed at learning more about the disease, sharing knowledge immediately, and undertaking concerted, evidence-based efforts to mitigate the disease condition. Calls for ensuring that all research findings relevant to COVID-19 be made available openly and promptly had begun to appear as early as in January.<sup>[28]</sup> In March, UNESCO mobilized countries to promote open science and data sharing to manage this crisis. Simultaneously, a global research roadmap for COVID-19 was issued by the World Health Organization (WHO) which pointed to the existing knowledge gaps. It has also drawn up timelines for the implementation of specific research actions.

Based on the declaration from welcome trust, all journals have made the COVID-19 literature free to access.<sup>[28]</sup> With the rapid increase in the number of publications, some of the journals have already crossed their December 2022 issue, and most of the journals have filled until the December 2021 issue (as on 12-06-2020).

Different authors have done bibliometric analyses of the COVID-19 literature from different databases, however, to the best of our knowledge, these studies are for only limited time or analyzed only limited number of articles.<sup>[6,8-12,29]</sup>

Chahrour *et al.* analyzed the publications between 16/12/2019-16/03/2020 in PubMed for the country wise research output. With the specific inclusion criteria, they could identify 564 articles from 39 different countries, with China producing highest number of articles (67%) followed by United States (7%); total 43% of all articles were “case report”.<sup>[8]</sup>

Aristovnik *et al.* and Zyoud *et al.* analyzed the bibliometric results of the research output in the early stage of outbreak of the pandemic for the Scopus database.<sup>[9,29]</sup> They have

identified different characteristics of COVID-19 related articles like top 10 countries, journals, institution, research topic cluster and 20 highest cited articles. They identified 19,044 publications which were distributed as: Articles – 9,140 (48.0%); correspondence – 4,192 (22.0%); reviews – 1,797 (9.4%), editorials – 1,754 (9.2%); notes – 1,728 (9.1%) and 433 (2.3%) were miscellaneous.

Gong *et al.* calculated the evidence map of COVID-19 related medical articles from PubMed and China National Knowledge Infrastructure databases from January 1, 2020 to March 8, 2020 and Liu *et al.* on PubMed and Embase database for the period 1 January to 24 March 2020.<sup>[6,11]</sup> Gong *et al.* found that by mid-February 2020, the number of articles in Chinese were 2.5 times than the English articles with highest contribution from Peoples Republic of China (PRC) comprising mainland China, Hong Kong, Macao, and Taiwan, followed by United States (97 articles), and other European countries, such as the UK (27 papers), Germany (15 papers) and Italy (15 papers).

In all these bibliometric analyses, almost all research groups have accepted the unprecedented growth in the article numbers even as early as February 2020, independently and as well as from the other strain of SARS-CoV and MERS-CoV.<sup>[6,10-11]</sup>

Excessive number of literature on a short span of time calls for addressing some pertinent questions: (1) how is it possible to bring out such a large volume of literature in such a short time.<sup>[27]</sup> (2) what is the reliability of the data collection, analysis, and literature review.<sup>[7]</sup> (3) when articles, including randomized trials, are being accepted in a very short time, some showing same day acceptance, what is the quality and reliability of the peer review process.<sup>[14,15]</sup> (4) how will it be possible to read and comprehend this large volume of literature to extract useful and actionable information.<sup>[30]</sup> (5) whether this plethora of literature is actually translating or will it ever translate into any medical, social, or economic benefit.<sup>[4,5]</sup> (6) how much of the information is real progress and how much is mere repetition? (7) how have the biomedical journals maintained due diligence for such large volumes of submissions and acceptances.<sup>[4,5,13,15,31-33]</sup>

Probably the answer for most of the above questions shall be in negative. To further analyses the growth pattern of the Covid-19 related literature following study has been designed.

## METHODOLOGY

Different medical databases, publication houses and sharing platforms were analyzed, each over a certain period of time to obtain the different characteristics and dynamics of the published/uploaded articles. All the databases created a separate section to tackle the COVID-19 related literature. The keywords in PubMed were “COVID-19” or “Coronavirus” or “Corona virus” or “Coronaviruses”.<sup>[34]</sup> ScienceDirect and

Research Gate (RG) have a separate database for COVID-19 specific research.<sup>[35,36]</sup> The data for different databases were extracted as a spreadsheet (PubMed) or saved first as a html (ScienceDirect and RG) file, which was then converted into a spreadsheet. Several articles and write-ups with only meager relationship to the disease were not further counted in the analysis.

Analyzed databases were: (1a) PubMed “all articles” (13/01/2020–12/06/2020), (1b) PubMed review articles (13/01/2020–12/06/2020), with a subgroup analysis of total review time, (1c) PubMed clinical trials (January–June 2020); (2) ScienceDirect all publications except Erratum (December 2019– 25-05-2020); (3) Research Gate (Published article, Pre-print, Technical Report) (15/04/2020 – 30/04/2020).

The dynamics of COVID-19 related documents in PubMed, ScienceDirect, Research Gate databases were analyzed for number of publications as (i) temporal distribution (publication vs. days), (ii) spatial distribution (publication vs. journal), and (iii) authorship (publication vs. author). Number of publications as a function of doubling time was analyzed for PubMed “all articles” and ScienceDirect published articles.

Analysis of total review time (submission date to acceptance date) was done for 150 review articles published in ScienceDirect and COVID-19 related control trials presented in PubMed.

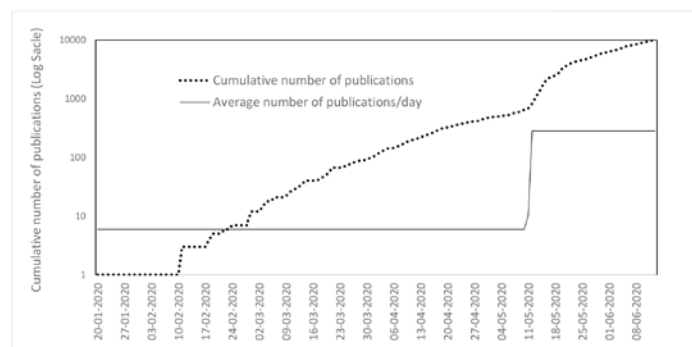
Average values quoted in this article as <minimum leverage ± standard deviation/ maximum>. All time (t) reported in days.

## FINDINGS

**PubMed: “all articles”:**

**Temporal distribution: Number of Publications (NOP) Vs Days**

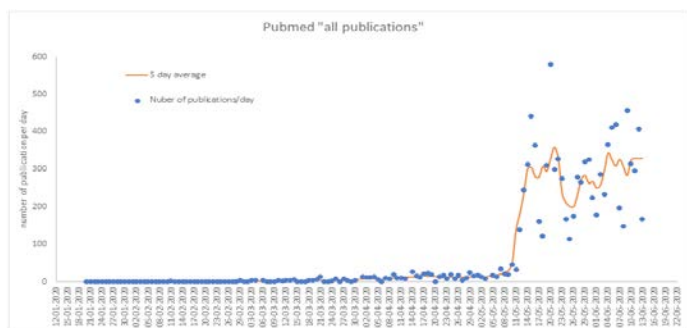
Total number of publications in PubMed: “all articles” was 23,480. Nonetheless, PubMed display is limited to 10,000 articles only therefore bibliometric analysis was limited to the displayed number of articles.<sup>[34]</sup> Figure 1a shows the number of publications until 13/06/2020 as a function of date and a



**Figure 1a:** Cumulative and average number of publications 20/01/2020–13/06/2020 as a function of time (date).

5-days moving average. The overall average ( $\pm$  standard deviation) number of publications is  $70.0 \pm 128.6$  /day.

The avalanche on number of publications that occurred after May 10, 2020 is obtained from the number of average daily publications that jumped from  $6.0 \pm 8.4$  to  $282.5 \pm 110.3$  (Figure 1b) yielding a growth of 47 times. As per the characteristics of the cumulative number, the publications were divided into two groups: from 13/01/2020 to 10/05/2020 (Group-1) and from 11/05/2020 to 13/06/2020 (Group-2). The mathematical relationship between cumulative number of publication and time (days) for group-1 shows a parabolic relationship (supplementary Figure 1) with a major coefficient

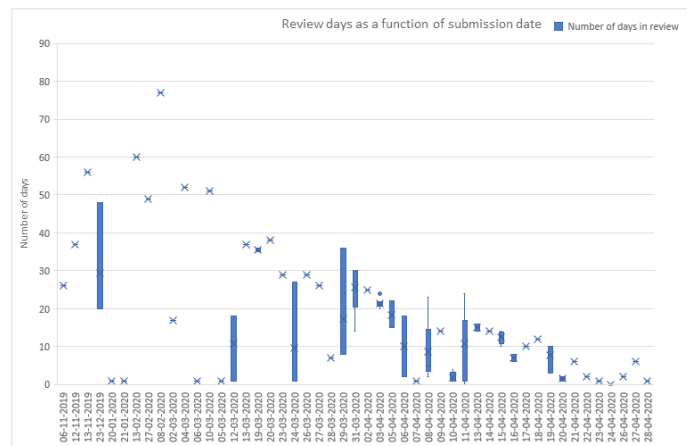


**Figure 1b:** Publication as a function of date for all articles presented in PubMed.

$a=0.09$ . In the post-avalanche Group-2, (supplementary Figure 2) it changed to a linear relationship with a slope of  $m=280.2$  pointing to the average number of publications per day (282.5) during that period.

**PubMed: "all articles": spatial distribution (NOP vs. journals)**

The analysis was limited to the last 10,000 articles published in 1868 different journals, as per the availability from the PubMed.

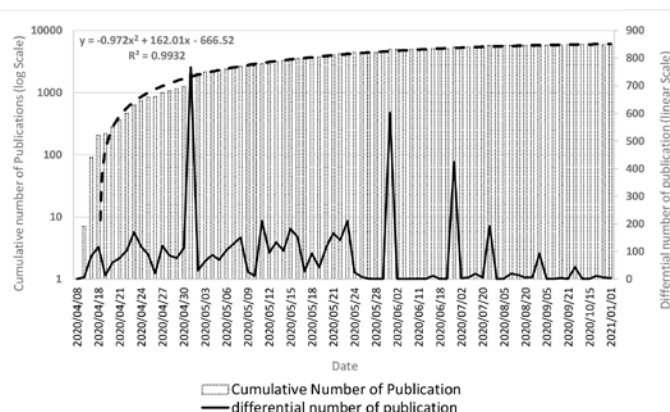


**Figure 2:** PubMed Review Articles: Analysis of days in review (acceptance-received date) days as a function of the submission date (PubMed review) - (The bigging date is 06-11-2019 earlier to the acceptance).

Number of articles per journal was  $\langle 1|15.2 \pm 10.3|191 \rangle$ , median=2. While 750 journals published a single article, 695 journals published 2 to 5 articles (2336 articles in total) and the rest 423 journals published 6914 articles. The highest number of articles (191) was published by British Medical Journal (BMJ), followed by Journal of Medical Virology (138) and Dermatologic Therapy (113). Supplementary Table 1 lists the number of publications as a function of the journals.

**PubMed: Doubling time (d)**

The doubling time for the NOP was calculated considering 3 publications on February 11, 2020 as the base. Between 11/02/2020- 13/06/2020 (123 days) total of 12 doublings were observed ( $2^{12}$ ) and presented in Supplementary Figure 3. The mathematical relationship between number of DT vs. time shows a linear relationship ( $=10.3t-0.7$ ; t in days, fitting



**Figure 3:** ScienceDirect: Cumulative and differential number of publications as a function of date.

accuracy of 98.9%) with average doubling time of 10.3 ( $\pm 4$ ) days; ranging between 2-15 days.

**PubMed: Review articles: Analysis of review days**

In all 150 review articles appeared between 06/11/2019 to 28/04/2020 were analyzed for the number of days in review (DIR) as a dependent function of submission date and presented in Figure 2. Three articles were accepted on the same date of submission, 18 in 1 day, 9 in 2 days, 5 in 3-5 days, 22 in 6-10 days, 41 in 11-20 days, 33 in 21-30 days. Only 19 articles were reviewed in excess of a month. The average and median number of days in review were  $\langle 1|17 \pm 17|77 \rangle$  days and 13.9 days respectively. Mathematical relation between number of DIR and days is a straight line ( $DIR = -0.2144t + 9433$ , t in days) with the days in review reducing at a rate of -0.21 days (count)/day.

**Analysis of PubMed Clinical Trials**

Between January-June 2020, PubMed showed a total of 17 Randomized clinical trials (RCT) in (16 in English and 1



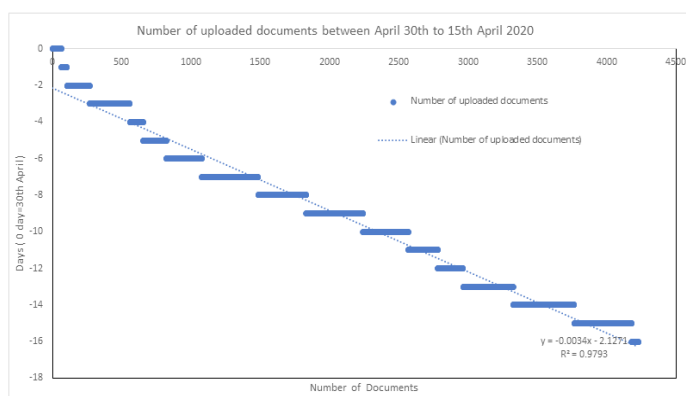
in Chinese language) published in 16 journals, with Lancet alone publishing 2 trials. However, after carefully reading each article, it was found that the publications from France and Italy were not RCTs and the total was reduced to 15; out of which 13 articles from China and 2 from Brazil. A total of 300 independent authors were involved in these articles, the median number of authors per article is 13 and the range is from 6 to 65. Of these, only 6 articles have information on submission and accepted dates. The average and median review time was  $\langle 10.7 \pm 15.3 | 4.1 \rangle$  days, 5.5 days respectively. Two review articles were from the same research group, Hainan General Hospital, Haikou, China which were accepted in 0 and 2 days respectively in a single journal (Complementary therapies in clinical practice).<sup>[37,38]</sup>

## ScienceDirect

### Temporal distribution: NOP Vs Days

ScienceDirect had published 5898 articles (excluding Erratum) for the observation period between 30/12/2019 to 25/05/2020.<sup>[35]</sup> The average NOP/day and the range were  $\langle 177.6 \pm 125.3 | 767 \rangle$ , median number of publications/day=27. Figure 3 shows the cumulative NOP (in log scale) and the differential NOP (in linear scale) as a function of time. Cumulative NOP varies in quadrature  $= -0.972t^2 + 162.01t - 666.52$ ;  $t$  in days with a fitting accuracy of 99.3%.

**Doubling time:** The cumulative number of publications in ScienceDirect encountered 13 doublings in number, with an average doubling time of 20.6 days. Mathematical relationship between cumulative number of articles as a function of the number of DT instances, fits with exponential growth  $0.7198e^{0.6931n}$  ( $n$  is the number of DTs) presented in supplementary Figure 4.



**Figure 4:** Shows number of publication vs days in last two weeks of the April 2020 (Research Gate). NOP increases at a rate of  $294.1/\text{day} \left( = \frac{1}{0.0034} \right)$ . Supplementary Material (all) Public document available from GitHub Link: [https://github.com/biplabphy/COVID-19\\_Literature.git](https://github.com/biplabphy/COVID-19_Literature.git)

**NOP vs authors:** ScienceDirect identifies a total of 27,845 authors, with 22,675 authors having lone articles, 3,849 authors having 2 articles, 1297 authors having 3-10 articles and 23 authors having 10+ articles. The maximum number of articles for single authors was 34. Mean and range of the article is  $\langle 112.3 \pm 0.9 | 34 \rangle$ , median =1 article. Supplementary Table 2 shows the number of publications as a function of the author (ScienceDirect).

## ResearchGate (RG)

Like all other medical database RG also created a special section for all COVID-19 related literatures.<sup>[36]</sup> This string direct the researchers and readers to all the available COVID-19 literature without any key word search between 15/04/2020 and 30/04/2020, in all 5,395 COVID-19 related documents were uploaded to RG. After carefully scrutinizing all entries, eliminating repetitions, erratum, presentations and comments, the number of useful documents was reduced to 4,180. A total of 1,986 documents had full text. With different key word (multiple option available) 635, 1623, 744, 928 2180 documents were tagged as basic science, diagnosis, drug and vaccination, social and economic impact, and public health respectively.

### Temporal distribution: NOP Vs Days

Temporal distribution of the uploaded documents, shown in Figure 4, yielded a linear relationship when plotted against time. The mean and median number publications/day were  $\langle 40 | 258.4 \pm 130.8 | 441 \rangle$  and 273/day respectively. The mathematical relationship between NOP as a function of days shows a straight line with coefficient of increase for NOP/day  $= 294.1 \left( = \frac{1}{0.0034} \right)$

The documents related to basic science, diagnosis, drug and vaccine development, social and economic impact, public health, and treatment (with multiple option) through key word search were 635, 1622, 743, 927, 2179, and 1318 items respectively.

**Doubling time:** During the last half of April 2020, RG encountered 6 doublings in the number of uploaded documents with an average DT was  $\langle 112.3 \pm 2.0 | 6 \rangle$  days.

**NOP vs. authors:** Supplementary Table 3 shows the number of documents against the authors. A total of 3,537 authors in RG have single article, where as 257 people have 2, 18 people have 3, and another 18 people have 4 to 9 documents.

## DISCUSSION

The growth pattern presented in this article along with all previous articles indicate the number of literature is a singular and perhaps an unprecedented event, at least at this scale:

Doubling every 10.3 days in PubMed with a daily increase of  $282.5 \pm 110.3$  articles.

This tendency of somehow to get an article published, is seen across the board, from established researchers of the most reputed medical schools to undergraduate scholars from little known universities. It is therefore not surprising that reputed journals had to do one of the biggest ever retraction of scientific papers in the modern history, marking a grim testimony to the chinks in the peer review process<sup>[23-25]</sup> and robustness of quality checks and audits of the submissions. Even more concerning is that the authors of these articles are well-known researchers.<sup>[23-25]</sup>

These retracted papers are most often fabricated and serve as examples of how the COVID-19 has overwhelmed the peer review process with a huge rush for publications.<sup>[21,23-25]</sup> Making use of the prevailing situation, some new predator journals have also popped up, usually paid journals, to present these pseudoscientific facts.<sup>[39,40]</sup>

Other groups also have investigated the sudden outbreak in the COVID-19 publications, authorship, and clinical trial registrations.<sup>[26,27,32,40-43]</sup>

Bonini *et al.* analyzed 11,000 publications until 12<sup>th</sup> May 2020 in PubMed and found a COVID-19 related publication every 6 min. They analyzed the clinical trials associated with different monoclonal antibodies and found 147 trials for Chloroquine/Hydroxychloroquine, 39 for the efficacy of tocilizumab, 23 for the use of systemic Corticosteroids, and multiple other studies analysing the same aspect of different drugs. They concluded such an overwhelming number of similar trials are originated due to the absence of guidelines for more harmonized clinical research.<sup>[32]</sup> Similar work on the analysis of the clinical trials was also done by Hillel *et al.*<sup>[42,43]</sup> They analysed 285 registered interventional clinical trials in 18 large trial registry databases associated the treatment and prevention of COVID-19. They concluded in the first 6 months of the pandemic many trials were registered and apparently completed. However, failed to yield rapid results in the literature or on clinical trial registries.

Similar to our study Ioannidis *et al.* analysed the rapid growth of COVID-19 publications in terms of engagement of scientific workforce on Scopus database for 2 months 1<sup>st</sup> January to 1<sup>st</sup> March 2020.<sup>[41]</sup> They found a total of 129,570 COVID-19 papers which is 3.3% of the total of all scientific papers (3,963,55) across all scientific disciplines published and indexed in Scopus until 1<sup>st</sup> January 2020.<sup>[41]</sup>

Odone *et al.* analysed published articles until May 2020. It was found as high as 60% of the published literature “opinion articles” without proper scientific evidence. Only 150 journals published half of the 10,000 published papers which substantiate our findings.<sup>[26]</sup> Another study by slim *et al.* analyzed 15,909

COVID-19 articles in the top 100 surgery journals between March-June 2020 and found 83.4% of articles were “opinion articles” and 40% of COVID-19 related articles published in the top 10 Journals.<sup>[27]</sup>

All these investigating groups found this as an unprecedented editorial situation caused by a pandemic. 60–80% of articles are opinion pieces like editorials or viewpoints, narrative reviews, surveys, letters, guidelines without proper scientific analysis or evidence. Only around 8% of articles were proper scientific articles, i.e., randomized trials, original articles (including meta-analysis) with structured methods and results.

As established by several authors in their early bibliometric analysis of COVID-19 literature, and in our study as well, PRC is the leading contributor of the scientific articles.<sup>[6,8,9,11,12,41]</sup> A total of 87% of the controlled trials in PubMed has originated from PRC; 90% investigators in controlled trials are of Chinese origin. This kind of runaway publication raises concerns regarding nexus and pressure, for which the biomedical community will need to reorient the approaches to peer reviewed publication.<sup>[5,13,26,32,33]</sup> The other alarming tendency among the journals and reviewers are very short review time, which was observed as decreasing at a rate of -0.21 days with every passing day. Articles were often accepted in the same day or within a week.<sup>[5,14]</sup>

A shorter review period is not acceptable as it is rather impossible to judge the any article is such short time. In PubMed RCT section, we found two randomized controlled trials which were accepted within an average of 1 day.<sup>[37,38]</sup> Average review date for a review article in ScienceDirect journals reduced to 2 weeks instead of usual 5–6 weeks time. With a very short review time quality of studies will obviously be misjudge or highly compromised. Such under reviewed publication will produce more clinical concern in future.

We have presented the data from a closed data sharing platform PubMed operated by National Institute of Health with no role of end users, data from one publishing house (ScienceDirect) and an open data sharing platform Research Gate operated by the end users. This mixed approach of bibliometric analysis, presented in this study, helps in truly reflecting the overall scenario of all kinds of closed and open scientific data sharing platforms on the characteristics dynamics of COVID-19 data.

### Mathematical Analysis

Other bibliometric analyses limit their reporting only to the mean value or graphically representing the temporal or spatial distribution of different parameters.<sup>[6,8,9,11,12,30]</sup> Unlike other bibliometric analysis as and when possible, we have introduced a mathematical function to identify the functional form of COVID-19 publication pattern. Typical forms are function of time or as a function of instances with time as

implicit. The mathematical forms to characterize the growth pattern presented in this article either reflected by a conic section (straight line and parabola) or by a known series like exponential; fitting accuracy was quite good, varying between 97.9%–100%. This helps in two ways: first it mathematically categorizes the distribution and second it shows the tendency of the variation and allows calculating other parameters like slope. For example, cumulative number of publications as a function of doubling time for ScienceDirect (Supplementary Figure 4) shows an exponential increase; without fitting with a mathematical function, it is not possible to categorize the distribution. Represented by a known mathematical function (conic section or series) it provides an analytical form of the distribution which helps to comprehend the distribution. Nonetheless, it may not be always possible to present a suitable analytical form of the distribution. For example, Figure 1a, distribution is random, it may be possible to fit a complex form like B Spline curve with a low accuracy; however, under current conditions it is difficult to predict the tendency or comprehend the distribution. In such scenario the distribution is grossly represented by a mean change as a function of time.

Has the interest in publishing more on COVID-19 taken undue precedence over other subjects? The answer to this pertinent question can be judged from the fact that the number of publications on COVID-19 stands at 15,354 that far exceeds the number of publications on cancer (in its all derivatives) which is a paltry 4,718 (as on 14/07/2020).

## CONCLUSION

We have presented the bibliometric analyses of the dynamics of COVID-19 related publications using different parameters through relevant mathematical formulation of three large databases. Our analysis, and few previous articles, undoubtedly proved the skewed bias toward excessive, unwarranted publishing on COVID-19 pandemic.<sup>[4,6,8,9,13,30,40]</sup>

Though the disease has been declared as a global pandemic, it does not warrant unwanted literature piling up continuously. Research groups, publishing houses, journals, reviewers and all associates need to realize that the race for publications related to COVID-19 needs to be pragmatic, and not a blind race for one-upmanship. Publishing articles based on scientifically unimportant, pseudoscientific, fabricated, unreliable and harmful facts just to increase the number of publications and citations is actually harmful to society and disservice to the scientific community.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## Supplementary Material (all)

Public document available from GitHub Link: [https://github.com/biplabphy/COVID-19\\_Literature.git](https://github.com/biplabphy/COVID-19_Literature.git)

## REFERENCES

- Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, et al. Singapore 2019 Novel Coronavirus Outbreak Research Team. Investigation of three clusters of COVID-19 in Singapore: Implications for surveillance and response measures. *Lancet*. 2020;395(10229):1039-46. doi: 10.1016/S0140-6736(20)30528-6, PMID 32192580.
- WHO. Timeline - COVID-19 [cited May 6, 2020]. Available from: <https://www.who.int/news-room/detail/27-04-2020-whatimeline-covid-19>.
- COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [cited May 31, 2020]. Available from: <https://coronavirus.jhu.edu/map.html>.
- Nowakowska J, Sobocinska J, Lewicki M, Lemanska Z, Rzymiski P. When science goes viral: The Research response during three months of the COVID-19 outbreak. *Biomedicine and Pharmacotherapy*. 2020;129:110451. doi: 10.1016/j.biopha.2020.110451.
- Van Norman GA. "Warp Speed" Operations in the COVID-19 Pandemic: Moving Too Quickly?? *JACC Basic to Translational Science*. 2020;5(7):730-4. doi: 10.1016/j.jacbts.2020.06.001, PMID 32760859.
- Gong Y, Ma TC, Xu YY, Yang R, Gao LJ, Wu SH, et al. Early research on COVID-19: A bibliometric analysis. *The Innovation*. 2020;1(2):100027. doi: 10.1016/j.xinn.2020.100027.
- Matias-Guiu J. The role of scientific journal editors during the COVID-19 pandemic. *Neurologia (Engl Ed)*. 2020;35(4):223-5. doi: 10.1016/j.nrl.2020.05.005, PMID 32507233.
- Chahrour M, Assi S, Bejjani M, Nasrallah AA, Salhab H, Fares MY, et al. A bibliometric analysis of Covid-19 research activity: A call for increased output. *Cureus*. 2020;12(3). doi: 10.7759/cureus.7357.
- Sa'ed HZ, Al-Jabi SW. Mapping the situation of research on coronavirus disease-19 (COVID-19): A preliminary bibliometric analysis during the early stage of the outbreak. *BMC Infectious Diseases*. 2020;20(1):1-8.
- Tao Z, Zhou S, Yao R, Wen K, Da W, Meng Y, et al. COVID-19 will stimulate a new coronavirus research breakthrough: A 20-year bibliometric analysis. *Annals of Translational Medicine*. 2020;8(8):528. doi: 10.21037/atm.2020.04.26, PMID 32411751.
- Liu N, Chee ML, Niu C, Pek PP, Siddiqui FJ, Ansah JP, et al. Coronavirus Disease 2019 (COVID-19): An evidence map of medical literature. *BMC Medical Research Methodology*. 2020;20(1):177. doi: 10.1186/s12874-020-01059-y, PMID 32615936.
- Zocchi J, Pietrobon G, Moretto S, Bonsembiante A, Mazzola F, Petrucci G, et al. Literature in the time of COVID-19: The "phase two". *Oral Oncology*. 2020;109:104837. doi: 10.1016/j.oraloncology.2020.104837.
- Teixeira da Silva JA. An alert to COVID-19 literature in predatory publishing venues. *Journal of Academic Librarianship*. 2020;46(5):102187. doi: 10.1016/j.acalib.2020.102187.
- Kun Á. Time to acceptance of 3 days for papers about COVID-19. *Publications*. 2020;8(2):30. doi: 10.3390/publications8020030.
- Pericàs JM, Torrallardona-Murphy O, Arenas A, Valero H, Nicolás D. Profile and quality of published reviews on COVID-19. *European Journal of Clinical Investigation*. 2020 30:e13293. doi: 10.1111/eci.13293, PMID 32473068.
- Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kaili AC, et al. Remdesivir for the Treatment of Covid-19-Final Report. *The New England Journal of Medicine*. 2020;383(19):1813-26. doi: 10.1056/NEJMoa2007764.
- Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate use of remdesivir for patients with severe Covid-19. *New England Journal of Medicine*. 2020;382(24):2327-36.
- Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: A randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet*. 2020;395(10236):1569-78. doi: 10.1016/S0140-6736(20)31022-9, PMID 32423584.
- Bonovas S, Piovani D. Compassionate use of Remdesivir in Covid-19. *The New England Journal of Medicine*. 2020;382(25):e101. doi: 10.1056/NEJMc2015312, PMID 32412705.
- WHO recommends against the use of Remdesivir in COVID-19 patients [cited Jan 09, 2021]. Available from: <https://www.who.int/news-room/feature-stories/detail/who-recommends-against-the-use-of-remdesivir-in-covid-19-patients>.
- Mehra MR, Desai SS, Ruschitzka F, Patel AN. Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis. *The Lancet*. 2020 May 22.
- Tselios K, Skendros P. Hydroxychloroquine against COVID-19: A critical appraisal of the existing evidence. *European Journal of Rheumatology*. 2020;7(Suppl 2):S110-6. doi: 10.5152/eurjrheum.2020.2064. PMID 32432527.
- Funck-Brentano C, Nguyen LS, Salem JE. Retraction and republication: Cardiac

- toxicity of hydroxychloroquine in COVID-19. *The Lancet*. 2020;396(10245):e2-3. doi: 10.1016/S0140-6736(20)31528-2, PMID 32653079.
24. Watson J, Adler A, Agweyu A. Open letter to MR Mehra. *Lancet*, and AN Patel, authors of "Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis." *The Lancet*. 2020:31180-6.
  25. Rubin EJ. Expression of concern: Mehra MR *et al.* *Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19*. *New England Journal of Medicine*. 2020;382(25):2464. doi: 10.1056/NEJMe2020822, PMID 32484612.
  26. Odone A, Salvati S, Bellini L, Buccini D, Capraro M, Gaetti G, *et al.* The runaway science: A bibliometric analysis of the COVID-19 scientific literature. *Acta Bio Medica: Atenei Parmensis*. 2020;91(9-S):34-9. doi: 10.23750/abm.v91i9-S.10121, PMID 32701915.
  27. Slim K, Mattevi C, Badon F, Lecomte C, Selvy M. The wave of "opinion articles" in the coverage of COVID-19 in surgical literature. *Langenbeck's Archives of Surgery*. 2020;405(6):877-8. doi: 10.1007/s00423-020-01932-w, PMID 32676739.
  28. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak Available from "https://wellcome.org/coronavirus-covid-19/open-data -Accessed on 07th January, 2021.
  29. Aristovnik A, Ravšelj D, Umek L. A bibliometric analysis of COVID-19 across science and social science research landscape. *Sustainability*. 2020 Jan;12(21):9132.
  30. Treanor L, Radonjic A. Assessing the reading level of online resources on COVID-19. *Canadian Journal of Public Health*. 2020;111(4):484-5. doi: 10.17269/s41997-020-00363-w, PMID 32588336.
  31. De Felice FR, Polimeni A. Coronavirus disease (COVID-19): A machine learning bibliometric analysis. *In vivo*. 2020;34(3);Suppl:1613-7. doi: 10.21873/invivo.11951, PMID 32503819.
  32. Bonini S, Maltese G. COVID-19 Clinical trials: Quality matters more than quantity. *Allergy*. 2020;75(10):2542-7. doi: 10.1111/all.14409, PMID 32434272.
  33. Cevik M, Bamford CGG, Ho A. COVID-19 pandemic—A focused review for clinicians. *Clin Microbiol Infect*. 2020;26(7):842-7. doi: 10.1016/j.cmi.2020.04.023, PMID 32344166.
  34. PubMed.gov. [cited Jun 12, 2020] Available from: <https://pubmed.ncbi.nlm.nih.gov/?term=COVID19+OR+Coronavirus+OR+%22Corona+virus%22+OR+Coronaviruses>
  35. Elsevier.com. Novel Coronavirus Information Center [cited May 25, 2020] Available from: <https://www.elsevier.com/connect/coronavirus-information-center>.
  36. ResearchGate. [cited Apr 30, 2020] Available from: <https://www.researchgate.net/community/COVID-19>.
  37. Liu K, Zhang W, Yang Y, Zhang J, Li Y, Chen Y. Respiratory rehabilitation in elderly patients with COVID-19: A randomized controlled study. *Complementary Therapies in Clinical Practice*. 2020;39:101166. doi: 10.1016/j.ctcp.2020.101166.
  38. Liu K, Chen Y, Wu D, Lin R, Wang Z, Pan L. Effects of progressive muscle relaxation on anxiety and sleep quality in patients with COVID-19. *Complementary Therapies in Clinical Practice*. 2020;39:101132. doi: 10.1016/j.ctcp.2020.101132.
  39. Caulfield T. Pseudoscience and COVID-19 - we've had enough already. *Nature*. 2020. doi: 10.1038/d41586-020-01266-z, PMID 32341556.
  40. Sarkar B, Munshi A, Ghosh B, Ganesh T, Manikandan A, Biswal SS, *et al.* Dynamics of the COVID-19 related publications. *Biorxiv*. 2020 Jan 1.
  41. Ioannidis JP, Salholz-Hillel M, Boyack KW, Baas J. The rapid, massive growth of COVID-19 authors in the scientific literature. *Biorxiv*. 2021;8(9):210389. doi: 10.1098/rsos.210389. PMID 34527271.
  42. Salholz-Hillel M, Grabitz P, Pugh-Jones M, Strech D, DeVito NJ. Results availability and timeliness of registered COVID-19 clinical trials: A cross-sectional study. *medrxiv*. 2021 Jan 1.
  43. Salholz-Hillel M, Grabitz P, Pugh-Jones M, Strech D, DeVito NJ. Results availability and timeliness of registered COVID-19 clinical trials: Interim cross-sectional results from the Direct study. *BMJ Open*. 2021 Nov 1;11(11):e053096. doi: 10.1136/bmjopen-2021-053096, PMID 34810189.