Temporal development of research publications on SARS-CoV-2 and COVID-19 [version 1; peer review: awaiting peer review]

Jonghoon Kang1, Erin Kang2, Matthew L. Cowan1, Manuel Orozco3

1Biology, Valdosta State University, Valdosta, Georgia, 31698, USA
2James S. Rickards High School, Tallahassee, Florida, 32301, USA
3Medicine, Philadelphia College of Osteopathic Medicine, Suwanee, Georgia, 30024, USA

Abstract
The coronavirus disease 2019 (COVID-19) pandemic has affected daily life throughout the world. The scientific community has globally responded to the pandemic with research on an unprecedented scale to help prevent disease spread and terminate the pandemic, resulting in a proliferation of scientific publications. In this article, the temporal trend of research on COVID-19 is analyzed to describe its development and inform a prediction of its future. Four other viruses are included in the analysis as negative or positive controls to illustrate that the concerns of the general public and/or the interest of the scientific community are major driving forces in the development of research. Our analysis predicts that COVID-19 and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will be major topics of research until at least 2025. We discuss the implications of our analysis for three sectors of community: researchers, epidemiologists, and young students.

Keywords
COVID-19, SARS-CoV-2, HCV, HIV, Ebola, Zika, PubMed, Scientometrics

This article is included in the Disease Outbreaks gateway.

This article is included in the Research on Research, Policy & Culture gateway.
This article is included in the Coronavirus collection.

Corresponding author: Jonghoon Kang (jkang@valdosta.edu)

Author roles: Kang J: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Kang E: Data Curation, Formal Analysis, Investigation, Validation, Writing – Original Draft Preparation; Cowan ML: Conceptualization, Data Curation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Orozco M: Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

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

Copyright: © 2021 Kang J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kang J, Kang E, Cowan ML and Orozco M. Temporal development of research publications on SARS-CoV-2 and COVID-19 [version 1; peer review: awaiting peer review] F1000Research 2021, 10:283 https://doi.org/10.12688/f1000research.42122.1

First published: 12 Apr 2021, 10:283 https://doi.org/10.12688/f1000research.42122.1
Introduction
The recent outbreak of coronavirus disease 2019 (COVID-19) has imposed an unprecedented and devastating burden on the world, including a serious encumbrance to health care systems. Collectively the scientific community has responded to the pandemic by researching the spread of the disease and its causative pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in order to understand and terminate the pandemic. These efforts have resulted in a vast amount of publications. We believe it would be worthwhile to analyze the trend of the publications in order to predict the future of research in this area.

We have previously demonstrated that the number of publications may be a reliable quantitative measure of the magnitude of research activity of a biological or biomedical science. In conjunction with regression analysis, the method of assessing research activity of a biological or biomedical discipline based on the number of publications in the field has been found to be effective in the prognostication of the future of biomedical fields by extrapolation of the best fit equation. The method has successfully been applied to various fields such as food sciences, epigenetics, metabolomics, and environmental sciences.

In this paper, we apply the method mentioned above to COVID-19 research to quantitatively describe the temporal development of the research and predict its future. We also include four other viruses in the study; hepatitis C virus (HCV) and HIV as negative controls without any apparent outbreaks in the period from January to November, 2020, and Ebola virus disease (EVD) and Zika virus (ZIKV) as positive controls of epidemiological outbreak during the period of examination from January 2014 to November 2020.

Methods
To quantitatively investigate the trend of research related to the five viruses (SARS-CoV-2, HCV, HIV, EVD, and ZIKV), we searched the PubMed database on December 23, 2020. Our search strategy was as follows for the different viruses: (The superscripts a and b in the search phrases represent month and year, respectively.)

SARS-CoV-2: (((COVID [Title/Abstract]) OR (COVID-19[Title/Abstract])) OR (SARS-CoV-2[Title/Abstract])) AND (("2020/MA"[Date - Publication]: "2020/MA"[Date - Publication]))

HCV: (((HCV [Title/Abstract]) OR (hepatitis C virus'[Title/Abstract])) AND (virus [Text Word])) AND (("2020/MA"[Date - Publication]: "2020/MA"[Date - Publication]))

HIV: (((HIV [Title/Abstract]) OR (human immunodeficiency virus'[Title/Abstract])) AND (virus [Text Word])) AND (("2020/MA"[Date - Publication]: "2020/MA"[Date - Publication]))

Ebola: ((Ebola [Title/Abstract]) AND (virus [Text Word])) AND (("Yb/MA"[Date - Publication]: "Yb/MA"[Date - Publication]))

Zika: ((ZIKA [Title/Abstract]) AND (virus [Text Word])) AND (("Yb/MA"[Date - Publication]: "Yb/MA"[Date - Publication]))

The number of publications on each virus was manually recorded on a monthly basis for eleven months for SARS-CoV-2, HCV, and HIV from January to November 2020, and for eighty three months for EVD and ZIKV from January 2014 to November 2020 for further investigation of data. Subsequent nonlinear regression analysis of the PubMed search results was conducted to obtain equation of best fit using SigmaPlot (version 11; Systat Software, Inc., San Jose, CA).

Results
We retrieved monthly publication numbers of the five viruses from the Pubmed database, and obtained the best fitting equation for each virus. Our results are summarized in Figure 1 and Table 1. We identified that temporal dynamics of publications related to the five viruses exhibit four characteristics.

First, a sigmoidal equation (Equation 1) was found to be the best quantitative description of the publication trend of COVID-19 research:

\[ f_1(x) = \frac{a}{1 + \exp \left( \frac{x-x_0}{b} \right)} \]  

(Equation 1)
Figure 1. Number of research publications related to five viruses. The solid line in each graph represents the best fit. The corresponding year in the panel B is presented above the x-axis. SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.

Table 1. Fitting parameters ($a$, $b$, and $c$) and associated standard errors (SE) and the squared Pearson correlation coefficients ($R^2$).

<table>
<thead>
<tr>
<th>Equation</th>
<th>Virus</th>
<th>$a$ ± SE</th>
<th>$b$ ± SE</th>
<th>$c$ ± SE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eq. (1)</td>
<td>SARS-CoV-2</td>
<td>12900 ± 370</td>
<td>0.67 ± 0.12</td>
<td>4.1 ± 0.14</td>
<td>0.9803</td>
</tr>
<tr>
<td></td>
<td>Ebola</td>
<td>150 ± 65</td>
<td>1.8 ± 0.87</td>
<td>9.5 ± 1.9</td>
<td>0.8345</td>
</tr>
<tr>
<td></td>
<td>Zika</td>
<td>220 ± 3.9</td>
<td>0.96 ± 0.088</td>
<td>25.8 ± 0.1</td>
<td>0.9916</td>
</tr>
<tr>
<td>Eq. (2)</td>
<td>Ebola</td>
<td>150 ± 8.7</td>
<td>0.013 ± 0.001</td>
<td></td>
<td>0.5522</td>
</tr>
<tr>
<td></td>
<td>Zika</td>
<td>300 ± 22</td>
<td>0.01 ± 0.001</td>
<td></td>
<td>0.5466</td>
</tr>
</tbody>
</table>
The value of each parameter is listed in Table 1. The mathematical meaning of each parameter can be found in our previous publication. In brief, the parameter “a” represents an asymptotic maximum value of the function, “b” is related to the shape of the function, and “c” is the year when the value of the function is half of the asymptotic maximum value. The sigmoidal kinetics observed in the research trend of COVID-19 (Figure 1) is congruent with other areas of research such as bioinformatics, epigenetics, food sciences, and environmental sciences.

Second, there was no significant correlation between the temporal point and the number of research publications on HCV and HIV during the time period examined from January to November 2020 (p = 0.240 for HCV, and p = 0.367 for HIV) (Figure 1). This can be attributed to the absence of any significant outbreaks of HCV or HIV during the time period; while these viruses are important in a biomedical sense, those viruses have likely been endemic.

Third, two examples of outbreaks in the decade of 2010, EVD and ZIKV, exhibit biphasic kinetics in the publication trend (Figure 1). The phase of sharp increase in number of publications, which overlaps with the time of each outbreak, also follows sigmoidal kinetics (Equation 1 and Table 1) as does COVID-19. The second phase, a decreasing phase, shows a slow and gradual decline that can be described by an exponential decay function (Equation 2):

$$f_2(x) = \frac{a}{2} \exp(-bx)$$

(Equation 2)

Fourth, the exponential nature of the decay kinetics may be valuable for the prediction of the future of COVID-19 research. In the case of EVD, the publication number started to decrease, when x = 11 (Figure 1), where the publication number is 123 (see underlying data) corresponding to 82% of the asymptotic maximum value of 150 (Table 1). Zika research started to decrease, when x = 33 (Figure 1), where the publication number is 222 (see underlying data) corresponding to 101% of its asymptotic maximum value of 219 (Table 1). As of June, 2020, COVID-19 research reached 95% of its asymptotic maximum value of 12900 (Figure 1): 12288/12900 = 0.95 (underlying data9 and Table 1). The quantitative comparison between SARS-CoV-2 and the two viruses suggests that the case of ZIKV is a more appropriate model for the prediction of COVID-19 research. Despite the apparent similarity of the research trend between SARS-CoV-2 and ZIKV, one should note that there is a substantial difference in the asymptotic maximum value (a in Equation 1) between these two areas of research: SARS-CoV-2 has an almost 60 times larger value of a than ZIKV (Table 1).

**Discussion**

The results of our research have implications for three sectors of the global community. One is for the scientific community in that research on COVID-19 is predicted to be active for a long time, even after commencing a downward trend. According to our mathematical model of the research on ZIKV, it will take COVID-19 research approximately 5 years (65.8 months) to reach half of its maximum value: $f_2(98.8) = f_1(33)/2$ and 98.8 – 33 = 65.8. While it is not certain when the publications on COVID-19 will start to decline, we expect that it will remain a major topic of research until at least 2025. This prediction may serve as a guide in planning research on COVID-19. The second implication of our results is for researchers in epidemiology as the method introduced in this paper can be easily applied to other epidemics and pandemics. The third implication is for young students. Our analysis of the ongoing research on COVID-19 should show them that science is a valuable way of contributing to humanity by providing solutions for public concerns such as COVID-19.

**Data availability**

**Underlying data**

Figshare: Number of PubMed-indexed articles related to five viruses; SARS-CoV-2, HCV, HIV, Ebola, and Zika. https://doi.org/10.6084/m9.figshare.12958361.v3

This project contains the following underlying data:

- covid_figshare_kang.csv (spreadsheet of the number of research publications found relating to five viruses).

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

**Acknowledgments**

We would like to thank Kathi Canese, a Program Manager for the support department of PubMed, for the valuable advice in the search of PubMed.
References


The benefits of publishing with F1000Research:

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

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