Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Using Baidu search values to monitor and predict the confirmed cases of COVID-19 in China: – evidence from Baidu Index

Bizhi Tu
Anhui Medical University https://orcid.org/0000-0003-0665-9167
Laifu Wei
Anhui Medical University
Yaya Jia
Shanxi Medical University
Jun Qian (qjpaper@sina.cn)
The First Affiliated Hospital of Anhui Medical University

Research article

Keywords: COVID-19, web-based data, internet searching, Baidu Index

Posted Date: November 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-44082/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published on January 21st, 2021. See the published version at https://doi.org/10.1186/s12879-020-05740-x.

Abstract

Background: New coronavirus disease 2019 (COVID-19) poses a severe threat to human life and causes a global pandemic. The purpose of current research is to explore whether the search-engine query patterns could serve as a potential tool for monitoring the outbreak of COVID-19.

Methods: We collected the number of COVID-19 confirmed cases between January 11, 2020, and c, from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). The search index values of the most common symptoms of COVID-19 (e.g., fever, cough, fatigue) were retrieved from Baidu Index. Spearman's correlation analysis was used to analyze the association between the Baidu index values for each COVID-19-related symptom and the number of confirmed cases. Regional distributions among 34 provinces/ regions in China were also analyzed.

Results: Daily growth of confirmed cases and Baidu index values for each COVID-19 related symptoms presented a robust positive correlation during the outbreak (fever: r_s=0.705, p=9.623×10^{-6}; cough: r_s=0.592, p=4.485×10^{-4}; fatigue: r_s=0.629, p=1.494×10^{-4}; sputum production: r_s=0.648, p=8.206×10^{-5}; shortness of breath: r_s=0.656, p=6.182×10^{-5}). The average search-to-confirmed interval is 19.8 days in China. The daily Baidu Index value's optimal time lags were the fourth day for cough, third day for fatigue, firth day for sputum production, firth day for shortness of breath, and 0 days for fever.

Conclusion: Search terms of COVID-19-related symptoms on the Baidu search engine have significant correlations with confirmed cases. Since the Baidu search engine can reflect the Public's attention to the pandemic and regional epidemics of viruses, relevant departments need to pay more attention to areas with high searches of COVID-19-related symptoms and take precautionary measures to prevent these potentially infected persons from further spreading.

Background

The outbreak of new coronavirus disease 2019 (COVID-19) was characterized by fever, cough, fatigue, sputum production, and shortness of breath, receiving people's attention globally [1, 2]. Till April 22, 2020, COVID-19 had spread more than 188 countries and regions, resulted in over 9.6 million cumulative confirmed cases and 490 thousand deaths worldwide [3]. The astonishing spread speed of the epidemic, to some extent, is failing to monitor and manage potentially infected persons, which may pose a substantial infection control challenge [4]. Therefore, recognizing the potential quantity of infected persons timely and accurately for the control of COVID-19 is in urgent need.

Because of the unpredictability of international public health emergency, novel methods for monitoring the epidemic's development are substantial. Network real-time data can be easily obtained from the web due to the quick availability of the Internet. According to the 45th China statistical report on internet development, there were over 904 million Internet users in China, with the penetration rate of search

engine use reached 83 percent [5]. Almost 80% of Internet users tend to use electronic devices to acquire the information they are interested in [6].

Recently, people can easily get the health-related information via Internet search engines, which, to some extent, could greatly reflect the physical condition of the searchers or the relatives and friends the searches concerned [7]. Moreover, to interrupt the transmission of the epidemic, the Chinese government has put in place strong quarantine measures, which also influences the routinely outpatient service process. Public search behaviors have been used to predict some epidemic diseases, such as influenza [8], epidemic erythromelalgia [9], dengue [10], and HIV/AIDS [11].

The surveillance of network searches about clinical Characteristics of COVID-19 is more predictable and timely compared to previous detection surveillance (e.g., official announcements, news reports, and mass media) [12-14]. Baidu serves as the most popular search engine, occupies more than 90% of Internet users in China [15]. In this study, we obtained the Baidu index values of COVID-19-related symptoms and the data of confirmed cases of COVID-19 across China to analyze the association between these variables, and explore whether the Baidu index could act as a novel tool for monitoring and predicting the epidemic of COVID-19 in China.

Methods

Data from Baidu Index

More than 90% of Chinese search engine users tend to use Baidu to retrieve their interesting information [16, 17]. The weighted sum of the Baidu search values can describe the characteristics of people's search behaviors [18]. Baidu Index is obtained by calculating the number of search terms of specific keywords input by the searchers [18]. Using the keywords analysis function, Baidu Index automatically matches its related words according to the keywords typed by users. According to previous studies, the top five most common symptoms of the COVID-19 was fever (which accounted for 88.7% of the confirmed cases during hospitalization), cough (67.8%), fatigue (38.1%), sputum production (33.7%), and shortness of breath (18.7%) [1]. Thus, we also selected those symptoms as the keywords in the current study. Based on the keyword analysis function, 26 search terms representing the most common symptoms of COVID-19 were selected (Table S1). We added each symptoms' search values and its related keywords together to get the composite Baidu Index values to perform our research. Besides, we compared the search values of 5 keywords in other years vertically to investigate whether the change of Baidu Index was an accidental event during the outbreak (Figure S1). To explore whether people's search behaviors appear earlier than the epidemic of COVID-19, we defined a definition to examine our hypothesis: search-to-confirmed interval (STCI), which can be calculated by minus the peak growth rate (DBIV minus its previous day's values as the growth rate of the DBIV) of Baidu Index with the peak DGCC. The top ten provinces/regions ranked by the cumulative confirmed cases were selected for STCI analysis.

Confirmed cases of COVID-19

We obtained the data of confirmed cases of COVID-19 from accessible official channels, including the official website of Hopkins University [2], WHO [19], the National Health Commission of the People's Republic of China [20]. Since China's epidemic had been gradually controlled after April 22, 2020, we divided the COVID-19 pandemic (January 11, 2020, to April 22, 2020) into a growth period and a decline period. The cut-off date was set on February 10, 2020, when the government announced the road closures re-opened and fully resume production [21].

Statistical analysis

Using SPSS (version 23.0), we applied a Spearman correlation analysis to explore the relationships between daily growth of confirmed cases (DGCC) and daily Baidu index values (DBIV) of COVID-19-related symptoms from January 11, 2020, to April 22, 2020. Using the same statistical methods, we also explored the time lag pattern between DGCC and DBIV of the symptoms-related to COVID-19. $P < 0.05$ was set as the significant statistical difference between variables (two-sided test). Besides, GraphPad Prism 8.2 was used to draw figures.

Results

Correlation analysis among search values of Baidu Index, cumulative confirmed cases and DGCC in China

Nationwide cumulative confirmed cases have a strong negative correlation with DBIV (fever: r_s=-0.455, p=1.206×10^{-6}; cough: r_s=-0.923, p=4.985×10^{-44}; fatigue: r_s=-0.425 p=7.041×10^{-6}; sputum production: r_s=-0.794, p=8.585×10^{-24}; shortness of breath: $r_{s=}$-0.428, p=5.786×10^{-6}) (Figure 1). Taking the cut-off date (February 10, 2020) as the demarcation point, the cumulative confirmed cases and DBIV of fever (r_s=0.705, p=9.623×10^{-6}), cough (r_s=0.592, p=4.485×10^{-4}), fatigue ($r_{s=}$0.629 p=1.494×10^{-4}), sputum production ($r_{s=}$0.648, p=8.206×10^{-5}), shortness of breath ($r_{s=}$0.656, p=6.182×10^{-5}) had a strong positive correlation during the grow period and a significantly negative correlation during the decline period (fever: r_s=-0.971, p =5.850×10^{-46}; cough: r_s=-0.967, p=8.601×10^{-44}; fatigue: r_s=-0.937 p =3.948 ×10^{-34}; sputum production: r_s=-0.770, p=1.604 x10^{-15}; shortness of breath: $r_{s=}$-0.930, p=5.786×10^{-32}) (Figure S2, S3).

Table 1 and Figure 2 shows that there was strong statistically positive correlations among the DGCC and Baidu search values of fever ($r_{s=}$0.786, p=8.013×10^{-23}), cough (r_s=0.556, p=1.087×10^{-9}), fatigue (r_s=0.763, p=7.930×10^{-21}), sputum production (r_s=0.665, p=1.793×10^{-14}), and shortness of breath (r_s=0.780, p=2.673 x10^{-22}), nationwide. Among the 34 provinces/regions in China, we found significant correlations between DGCC and DBIV, and observed that the number of daily confirmed cases tend to increase when the Baidu searches for terms related to fever, cough, fatigue, and shortness of breath increasing (Table 1). For Hong Kong, Macao, Taiwan, and Tibet, there no correlation was detected between DGCC and Baidu search

values of COVID-19-related symptoms (p>0.05 for all). However, DBIV of cough in Shanghai did not show correlations with DGCC (r_s=0.133, p=0.184). Besides, the correlation between sputum production and DGCC in several provinces/regions is inconspicuous (e.g., Beijing: r_s=0.249, p=0.012; Guangdong: r_s=0.262, p=0.008, Hunan: r_s=-0.244, p=0.014) (Table 1).

STCI analysis for people's search behaviors of COVID-19-related symptoms and the epidemic of COVID-19

Figure 3 shows that the peak of the growth rate of the Baidu Index occurred 19-22 days earlier than the peak of DGCC across china (STCI for fever: 22 days; cough: 19 days; fatigue: 20 days; sputum production: 19 days; shortness of breath: 19 days). And the top 10 provinces/regions ranked by confirmed cases presented similar results except for sputum production (Figure 3). However, the peak of the growth rate of the Baidu Index occurred 17 days lag compare with the peak of DGCC in Heilongjiang.

Lag correlation between the DGCC and search index values of COVID-19 related symptoms

Figure 4 and table S2 manifest the lag correlation between DBIV of the different keywords and DGCC, We found the highest lag correlation with DBIV for cough is 4 days earlier (r_s =0.574, p=1.826x10^{-10}), fatigue is 2 days earlier *(*r_s =0.778, p=2.434x10^{-22}), sputum production is 3 days earlier (r_s =0.664, p=1.630x10^{-14}), shortness of breath is 1 day earlier (r_s =0.804), p=9.707x10^{-25}), and fever is 0 days earlier compared with the number of DGCC (r_s =0.791, p=1.623x10^{-23}).

Discussion

People with the travel and exposure history of high-risk areas with COVID-19 patients will be required quarantined. Without a clear understanding of the new coronavirus's characteristics and effective treatments, people usually compare COVID-19 with the SARS, which outbroke in 2003 in China with a mortality rate of 11% [19, 22]. Due to the separate isolation precautions policy and the fear of an unknown virus, people with exposure history are likely to conceal their own and their family's high-risk behaviors, which undermines the government's early attempts to control the suspected cases of COVID-19 [23]. Using Internet search engines, we could predict the potential quantity of affected persons, and the real-time data of the Baidu Index helps monitor the epidemic development and formulates the corresponding government policies.

The control of the COVID-19 pandemic in China had achieved preliminary success by April 22, 2020. The correlation analysis between Chinese public searches of COVID-19-related symptoms and actual number of confirmed cases will benefits us to explore the relationships between Internet search values and COVID-19 pandemic, and further to provide new ideas for control the epidemic of COVID-19.

The current research showed that the related DBIV reached a peak earlier than the DGCC, and the dynamic changes of DBIV were also earlier than DGCC. We also noticed the higher the search values, the higher the cumulative confirmed cases will be during the growth period, which indicates that the

searchers could be the potential infectors of the virus. Besides, DGCC and DBIV presented with a positive correlation during the whole observation period, even in the decline period, which implies the DBIV declines with a decreased number of DCGG. However, the number of cumulative cases continues to increase (when DGCC is declining), which could be an explanation for the negative correlation between cumulative cases and DBIV during the decline period. The Public's search behaviors about health-related symptoms can reflect their potential physical and psychological problem [7, 24]. The decline of search values of COVID-19-related symptoms indicates that the Public's mentality may tend to be more relaxing in the decline period compared with the growth period.

We can tell from Baidu's time plots for COVID-19-related symptoms and the number of confirmed cases that the former dynamic changes appeared earlier than the later. Among 34 provinces/regions in China, although most areas in this research showed statistically correlations of the DBIV and DGCC (except sputum production), Hong Kong, Macao, Taiwan, and Tibet did not present with such correlations. This is probably owing to the Baidu search engine is not the primary search tool in these places [4]. Additionally, there are few cumulative confirmed cases in Tibet (only one cumulative case), therefore, there are insufficient cases to perform the statistical analysis. However, there was no correlation between DGCC and DBIV for cough in Shanghai, probably due to the incompleteness of search words related to keywords. Besides, the relationship between DBIV for sputum production and DGCC was not observed. The reasonable explanation could be that sputum production is more common in the elderly with chronic respiratory diseases and tends to possess a strong connection with seasonal influenza that occurs every year in the late autumn to early spring [25]. Based on our research, the increase in the DBIV of COVID-19-related symptoms could be treated as an abnormal signal, which is worthy of the corresponding action by government departments in advance.

The increased number of relevant searches indicates there are more potentially infected candidates. Around 97.5% of people with identifiable exposure history will develop symptoms within 11.5 days, and 1 % of them have a longer incubation period with more than 14 days [26]. We found that the average maximum of DBIV's growth rate was 20 days earlier than DGCC in most areas except Heilongjiang. On May 10, 2020, the Heilongjiang government reported that the pandemic had relapsed, so the apex of DBIV appeared later compared with other provinces [27]. Compared with the traditional diagnosis and treatment process, most potential patients are inclined to search the Internet for help, indicating the difference to publicly reported overrepresent severe cases of COVID-19 [7, 28, 29]. Those potential infectors are likely to use search engines (usually Baidu) to search for the related information, so the Baidu index could serve as an original way to reflect the approximate number of these potential infectors. Since the mild potential infectors may possess a more extended incubation period theoretically on account of several days lags before confirmed [30], the soaring DBIV of COVID-19-related symptoms in a certain area might be an indicator for the forthcoming outbreak of the epidemic. The STCI analysis showed that the peak DBIV of COVID-19-related symptoms appears 19-22 days earlier than the peak DGCC. However, the results of the time-lag correlation analysis shown a shorter lag than STCI. The STCI study only compares the interval between the peak DBIV of COVID-19-related symptoms and DGCC, did not take other data into account. Therefore, time-lag correlation analysis could be better to explore the lag

patterns of DBIV and DGCC. We found that the optimal time lag of DBIV for fever, cough, fatigue, sputum production, and shortness of breath was 0, 4, 2, 3, 1 day/days, respectively. According to Cuilian et al, the peak of Internet searches about COVID-19 appeared 10-14 days earlier than the peak of reported daily growth cases in China [31], and 10 days earlier in America [32]. People who search the terms of "□□" or "□□□□" (keywords in Cuilian's study) were more likely to experience the incubation period, while the searchers querying the COVID-19-related symptoms are likely those who were infected, and had already experienced the incubation period. As there is no time-lag for "fever", this may attribute to the body temperature reporting mechanism adopted by both the Chinese government and local institutions. This reporting system required that people with fever should be actively isolated and quarantined immediately to obstruct the source of infection [33, 34]. Therefore, people shown the symptom of fever would be isolated and confirmed subsequently. As a result, no time lag was observed in the DBIV for fever.

Limitations

There are some limitations needed to be recognized. Firstly, we only utilized the data form Baidu to perform our research, other search engines, such as Weibo, Twitter were not included. Secondly, some keywords related to the symptoms of COVID-19 were not included in the current study, and the keywords utilized in the current work could not guanine the consistency and efficiency of the long-term prediction in the future. Therefore, future studies are suggested to add or delete the corresponding keywords of COVID-19-related symptoms to confirm that the time lag patterns exist between DBIV and DGCC. Thirdly, the detailed information about the individual searchers remains unclear, so it is impossible to targeted monitoring the potential infectors. Besides, there were several documented issues with predictability of disease incidence trends using search engines. To avoid the failure of predicting an epidemic with the utilization of Internet search engine, a random forest regression model is suggested in the future study to facilitate our observing results [35].

Conclusion

Our research suggested that there is a significant correlation between DBIV of COVID-19-related symptoms and DGCC. DGCC dynamic change showed several days lags compare with the DBIV. Besides, DBIV for COVID-19-related symptom could serve as a potential indicator for predicting the epidemic of emerging infectious diseases, and could guide targetable intervention and prevention of COVID-19 to further assist in the overall control of the pandemic.

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent to publish:

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

Author(s) declare(s) that there is no conflict of interest

Funding

This work was supported by the grant from the National Natural Science Foundation of China (grant numbers 9101054002), the Foundation of Supporting Program for the Excellent Young Faculties in the University of Anhui Province in China. Grants for Scientific Research of BSKY from the First Affiliated Hospital of Anhui Medical University; and Grants for Outstanding Youth from the First Affiliated Hospital of Anhui Medical University.

Author contribution

JQ conceived the study idea. BZ collected the data. BZ, YY, and LF contributed to the analysis of the data as well as wrote the initial draft with all authors providing critical feedback and edits to subsequent revisions. All authors approved the final draft of the manuscript. All authors are accountable for all aspects of the work in ensuring related questions accuracy or integrity. Any parts of the work are appropriately investigated and resolved. JQ is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Acknowledgments

We thank all the people who offer help for this study.

References

1. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC *et al*: **Clinical Characteristics of Coronavirus Disease 2019 in China**. *N Engl J Med* 2020, **382**(18):1708-1720.
2. **Wuhan Municipal Health Commission** [http://wjw.wuhan.gov.cn/] (accessed June 26 2020)
3. **COVID-19 Dashboard by Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)** [https://www.arcgis.com] (accessed June 26 2020)
4. Al-Tawfiq JA: **Asymptomatic coronavirus infection: MERS-CoV and SARS-CoV-2 (COVID-19)**. *Travel Med Infect Dis* 2020, **35**:101608.
5. **Search engines in China - statistics & facts.** Available at: https://www.statista.com/topics/1337/search-engines-in-china/ (accessed June 26 2020)

6. **Fox S (2005) Health information online.** Pew Internet & American Life Project, Washington, DC
7. Cervellin G, Comelli I, Lippi G: **Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings**. *J Epidemiol Glob Health* 2017, **7**(3):185-189.
8. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS: **Monitoring influenza epidemics in china with search query from baidu**. *PLoS One* 2013, **8**(5):e64323.
9. Gu YZ, Chen FL, Liu T, Lv XJ, Shao ZM, Lin HL, Liang CB, Zeng WL, Xiao JP, Zhang YH *et al*: **Early detection of an epidemic erythromelalgia outbreak using Baidu search data**. *Sci Rep-Uk* 2015, **5**.
10. Guo P, Liu T, Zhang Q, Wang L, Xiao JP, Zhang QY, Luo GF, Li ZH, He JF, Zhang YH *et al*: **Developing a dengue forecast model using machine learning: A case study in China**. *Plos Neglect Trop D* 2017, **11**(10).
11. He GY, Chen YS, Chen BW, Wang H, Shen L, Liu L, Suolang DJ, Zhang BY, Ju GD, Zhang LL *et al*: **Using the Baidu Search Index to Predict the Incidence of HIV/AIDS in China**. *Sci Rep-Uk* 2018, **8**.
12. Freifeld CC, Mandl KD, Ras BY, Bronwnstein JS: **HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports**. *J Am Med Inform Assn* 2008, **15**(2):150-157.
13. van de Belt TH, van Stockum PT, Engelen LJLPG, Lancee J, Schrijver R, Rodriguez-Bano J, Tacconelli E, Saris K, van Gelder MMHJ, Voss A: **Social media posts and online search behaviour as early-warning system for MRSA outbreaks**. *Antimicrob Resist In* 2018, **7**.
14. Li CL, Chen LJ, Chen XY, Zhang MZ, Pang CP, Chen HY: **Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020**. *Eurosurveillance* 2020, **25**(10):7-11.
15. **China Internet Network Information Center.** Available at: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/. (accessed June 26 2020)
16. **China Search Engine Market Overview (2015).** Available at: https://www.chinaInternetwatch.com/17415/search-engine-2012-2018e/ (accessed June 26 2020).
17. **China Internet Network Information Center**. Chinese Internet users search behavior study. Beijing, China; 2014. URL: http://www.cnnic.cn/hlwfzyj/hlwmrtj/201410/t20141017_49359. Htm (accessed 26 Jun 2020) [WebCite Cache ID 75IWlqvRn]
18. **Baidu. Baidu Index** URL: https://index.baidu.com/ (accessed 26 Jun 2020) [WebCite Cache ID 6yOtOa7p9]
19. **World Health Organization**. Available at: https://www.who.int/ (accessed June 26 2020)
20. National Health Commission of the People's Republic of China Available at: http://www.nhc.gov.cn/ (accessed June 26 2020)
21. **State Council of the PRC**. Available at: http://www.gov.cn/guowuyuan/ (accessed June 26 2020)
22. Ashraf H: **Investigations continue as SARS claims more lives**. *The Lancet* 2003, **361**(9365).
23. **China Central Television**. Available at: https://www.cctv.com/ (accessed June 26 2020)

24. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, Zou Y, Liu J, Sun G, Wang F: **More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends**. *J Glob Health* 2020, **10**(1):011003.
25. Uyeki TM, Bernstein HH, Bradley JS, Englund JA, File TM, Fry AM, Gravenstein S, Hayden FG, Harper SA, Hirshon JM *et al*: **Clinical Practice Guidelines by the Infectious Diseases Society of America: 2018 Update on Diagnosis, Treatment, Chemoprophylaxis, and Institutional Outbreak Management of Seasonal Influenzaa**. *Clin Infect Dis* 2019, **68**(6):895-902.
26. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J: **The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application**. *Ann Intern Med* 2020, **172**(9):577-582.
27. **Health Commission of Heilongjiang Province**. Available at: http://wsjkw.hlj.gov.cn/ (accessed October 3 2020)
28. Carneiro HA, Mylonakis E: **Google trends: a web-based tool for real-time surveillance of disease outbreaks**. *Clin Infect Dis* 2009, **49**(10):1557-1564.
29. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron AJ: **More Diseases Tracked by Using Google Trends**. *Emerg Infect Dis* 2009, **15**(8):1327-1328.
30. **Press Conference of the Joint Prevention and Control Mechanism of the State Council.** Available at: http://www.gov.cn/xinwen/gwylflkjz18/index.htm (accessed June 26 2020)
31. Cuilian Li,Li Jia Chen,Xueyu Chen,Mingzhi Zhang,Chi Pui Pang,Haoyu Chen. **Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020.** Euro Surveill,2020 Mar;25(10):2000199. doi: 10.2807/1560-7917.ES.2020.25.10.2000199.
32. Cousins HC, Cousins CC, Harris A, Pasquale LR: **Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns**. *J Med Internet Res* 2020, **22**(7):e19483.
33. **National Health Commission of the People's Republic of China** [http://www.nhc.gov.cn/xcs/zcwj2/new_zcwj.shtml] (accessed October 3 2020)
34. Lin RT, Cheng Y, Jiang YC: **Exploring Public Awareness of Overwork Prevention With Big Data From Google Trends: Retrospective Analysis**. *J Med Internet Res* 2020, **22**(6):e18181.
35. Sasikiran Kandula,Jeffrey Shaman: **Reappraising the utility of Google Flu Trends**. PLoS Comput Biol,2019 Aug 2;15(8):e1007258. doi: 10.1371/journal.pcbi.1007258. eCollection 2019 Aug.

Table

Table 1. Correlation between daily growth of confirmed cases (DGCC) across China and Values of Baidu index (BI)

Region		Values of BI				
		Fever	Cough	Fatigue	Sputum production	Shortness of breath
China	r_s	0.768	0.556	0.763	0.665	0.780
	p	8.013×10^{-23}	1.087×10^{-9}	7.930×10^{-21}	1.793×10^{-14}	2.673×10^{-22}
Anhui	r_s	0.801	0.770	0.760	-0.028	0.775
	p	5.39×10^{24}	3.131×10^{-21}	2.172×10^{-20}	0.782	1.205×10^{-21}
Beijing	r_s	0.657	0.431	0.582	0.249	0.610
	p	6.336×10^{-14}	6.502×10^{-6}	1.358×10^{-10}	0.012	1.040×10^{-11}
Chongqing	r_s	0.796	0.769	0.740	0.572	0.738
	p	1.542×10^{-23}	1.647×10^{-23}	6.057×10^{-19}	3.389×10^{-10}	8.809×10^{-19}
Fujian	r_s	0.588	0.471	0.705	0.367	0.537
	p	8.473×10^{-11}	5.677×10^{-7}	1.388×10^{-16}	1.485×10^{-4}	5.809×10^{-9}
Gansu	r_s	0.527	0.444	0.373	-0.150	0.484
	p	1.277×10^{-8}	3.008×10^{-6}	1.112×10^{-4}	0.133	2.586×10^{-7}
Guangdong	r_s	0.535	0.336	0.527	0.262	0.506
	p	7.113×10^{-9}	1.564×10^{-4}	1.287×10^{-8}	0.008	5.598×10^{-8}
Guangxi	r_s	0.766	0.754	0.760	0.287	0.731
	p	7.075×10^{-21}	5.780×10^{-20}	1.904×10^{-20}	0.004	6.872×10^{-8}
Guizhou	r_s	0.673	0.657	0.622	0.355	0.629
	p	9.182×10^{-15}	6.433×10^{-14}	2.921×10^{-12}	2.555×10^{-4}	1.388×10^{-12}
Hainan	r_s	0.717	0.735	0.694	-0.354	0.693
	p	2.474×10^{-17}	1.468×10^{-18}	6.080×10^{-16}	2.673×10^{-4}	6.597×10^{-16}
Hebei	r_s	0.731	0.635	0.662	0.040	0.705
	p	2.622×10^{-18}	7.392×10^{-13}	3.396×10^{-14}	0.691	1.297×10^{-16}
Heilongjiang	r_s	0.413	0.201	0.453	0.089	0.345
	p	1.590×10^{-5}	0.042	1.710×10^{-6}	0.375	2.669×10^{-4}
Henan	r_s	0.771	0.766	0.728	0.655	0.759
	p	2.652×10^{-21}	6.291×10^{-21}	4.647×10^{-18}	7.887×10^{-14}	2.288×10^{-20}
Hong Kong	r_s	-0.094	-0.514	-0.282	0.517	-0.085
	p	0.349	3.394×10^{-8}	0.004	2.676×10^{-8}	0.398
Hubei	r_s	0.709	0.745	0.631	0.614	0.704
	p	7.410×10^{-17}	2693×10^{-19}	1.131×10^{-12}	6.640×10^{-12}	1.640×10^{-16}
Hunan	r_s	0.813	0.797	0.738	-0.244	0.759
	p	2.942×10^{-25}	1.256×10^{-23}	9.111×10^{-19}	0.014	2.300×10^{-20}
Inner Mongolia	r_s	0.322	0.129	0.369	0.385	0.316
	p	0.001	0.197	1.384×10^{-4}	6.326×10^{-5}	0.001
Jiangsu	r_s	0.695	0.565	0.629	0.502	0.630
	p	5.378×10^{-16}	5.918×10^{-10}	1.441×10^{-12}	7.609×10^{-8}	1.306×10^{-12}
Jiangxi	r_s	0.692	0.672	0.686	-0.317	0.640
	p	7.678×10^{-16}	1.052×10^{-14}	1.861×10^{-15}	0.001	4.433×10^{-13}
Jilin	r_s	0.538	0.446	0.626	0.323	0.355
	p	5.415×10^{-9}	2.646×10^{-6}	1.925×10^{-12}	0.001	2.472×10^{-4}
Liaoning	r_s	0.575	0.425	0.486	-0.221	0.513
	p	2.685×10^{-10}	8.698×10^{-6}	2.179×10^{-7}	0.026	3.436×10^{-8}
Macau	r_s	0.105	0.016	0.093	0.204	0.015
	p	0.293	0.872	0.354	0.040	0.882
Ningxia	r_s	0.696	0.649	0.541	-0.389	0.503
	p	4.495×10^{-16}	1.656×10^{-13}	4.279×10^{-9}	5.317×10^{-5}	7.051×10^{-8}
Qinghai	r_s	0.461	0.465	0.428	0.297	0.396
	p	1.115×10^{-6}	8.234×10^{-7}	7.029×10^{-6}	0.002	3.833×10^{-5}

Shaanxi	r_s	0.637	0.607	0.606	-0.157	0.670
	p	5.969×10^{-13}	1.319×10^{-11}	1.494×10^{-11}	0.115	1.406×10^{-14}
Shandong	r_s	0.706	0.584	0.702	0.528	0.708
	p	1.230×10^{-16}	1.217×10^{-10}	2.135×10^{-16}	5.238×10^{-7}	9.317×10^{-17}
Shanghai	r_s	0.331	0.133	0.379	-0.020	0.391
	p	0.001	0.184	8.633×10^{-5}	0.841	4.810×10^{-5}
Shanxi	r_s	0.380	0.275	0.313	0.001	0.365
	p	8.102×10^{-5}	0.005	0.001	0.991	2.382×10^{-4}
Sichuan	r_s	0.775	0.687	0.720	0.681	0.771
	p	1.247×10^{-21}	1.565×10^{-15}	1.588×10^{-17}	3.530×10^{-15}	2.517×10^{-21}
Tianjin	r_s	0.483	0.424	0.517	0.295	0.453
	p	2.675×10^{-7}	9.050×10^{-6}	2.624×10^{-8}	0.003	1.755×10^{-6}
Tibet	r_s	0.167	0.139	0.173	-0.003	0.043
	p	0.093	0.165	0.082	0.973	0.670
Xinjiang	r_s	0.737	0.704	0.593	-0.284	0.504
	p	9.948×10^{-19}	1.642×10^{-16}	4.944×10^{-11}	0.004	6.872×10^{-8}
Yunnan	r_s	0.689	0.616	0.635	-0.340	0.638
	p	1.274×10^{-15}	5.308×10^{-12}	7.636×10^{-13}	4.776×10^{-14}	5.252×10^{-13}
Zhejiang	r_s	0.592	0.530	0.628	0.349	0.618
	p	5.553×10^{-11}	1.026×10^{-8}	1.569×10^{-12}	3.250×10^{-4}	4.461×10^{-12}
Taiwan	r_s	-0.111	-0.428	-0.242	0.523	-0.019
	p	0.269	7.105×10^{-6}	0.014	1.699×10^{-8}	0.854

Figures

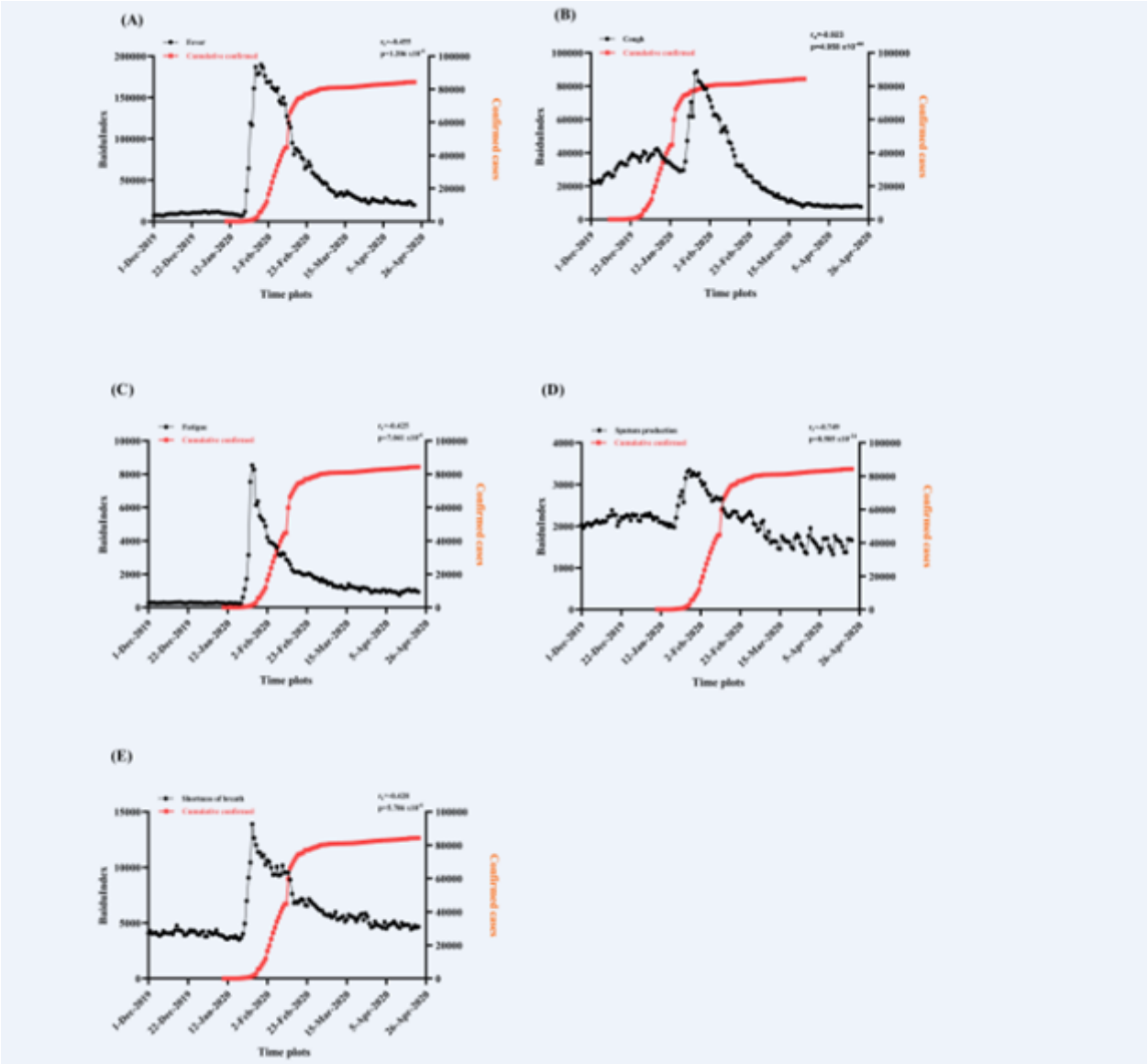

Figure 1. Correlation and time plots among cumulative confirmed cases and each keyword of COVID-19-related symptoms. (A), (B), (C), (D), and (E) represent Baidu searches for "fever," "cough," "fatigue," "sputum production," and "shortness of breath," respectively."

Figure 1

Legend in Figure.

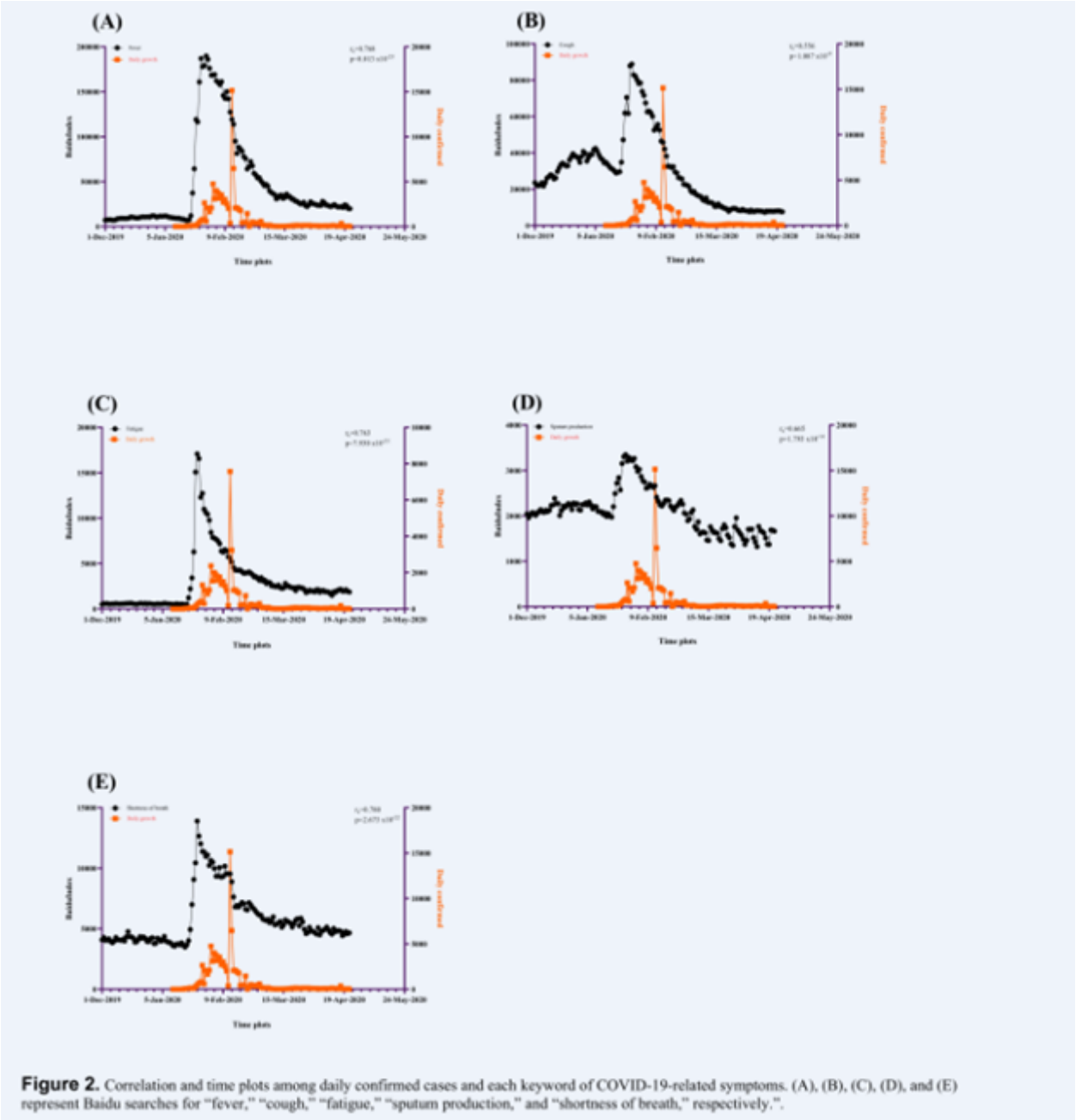

Figure 2. Correlation and time plots among daily confirmed cases and each keyword of COVID-19-related symptoms. (A), (B), (C), (D), and (E) represent Baidu searches for "fever," "cough," "fatigue," "sputum production," and "shortness of breath," respectively.".

Figure 2

Legend in Figure.

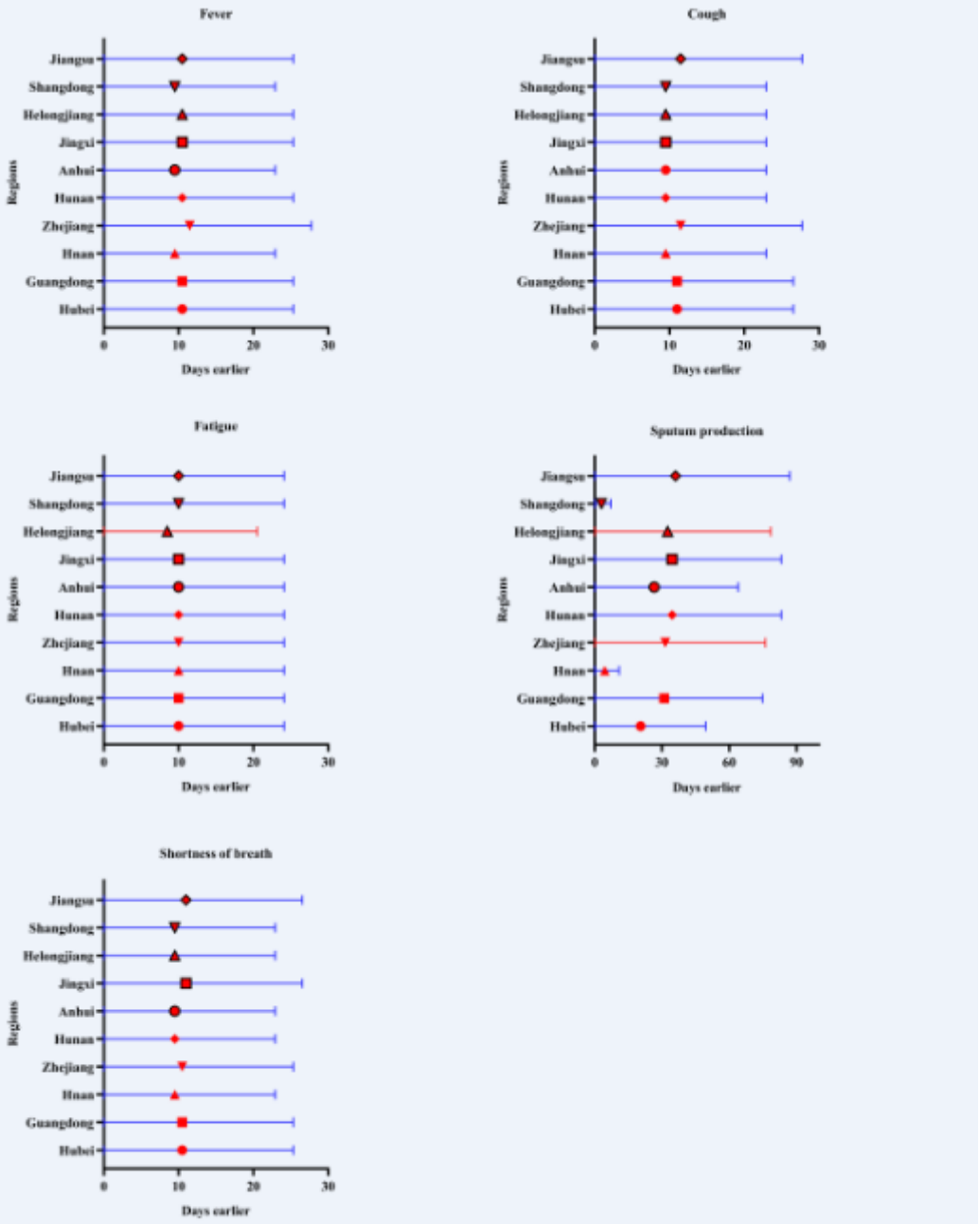

Figure 3. Days earlier of serarch-to-confirmed interval (STCI) among the top ten provinces in the cumulative number of confirmed cases. e.g., The red line represents the absolute value of the negative value

Figure 3

Legend in Figure.

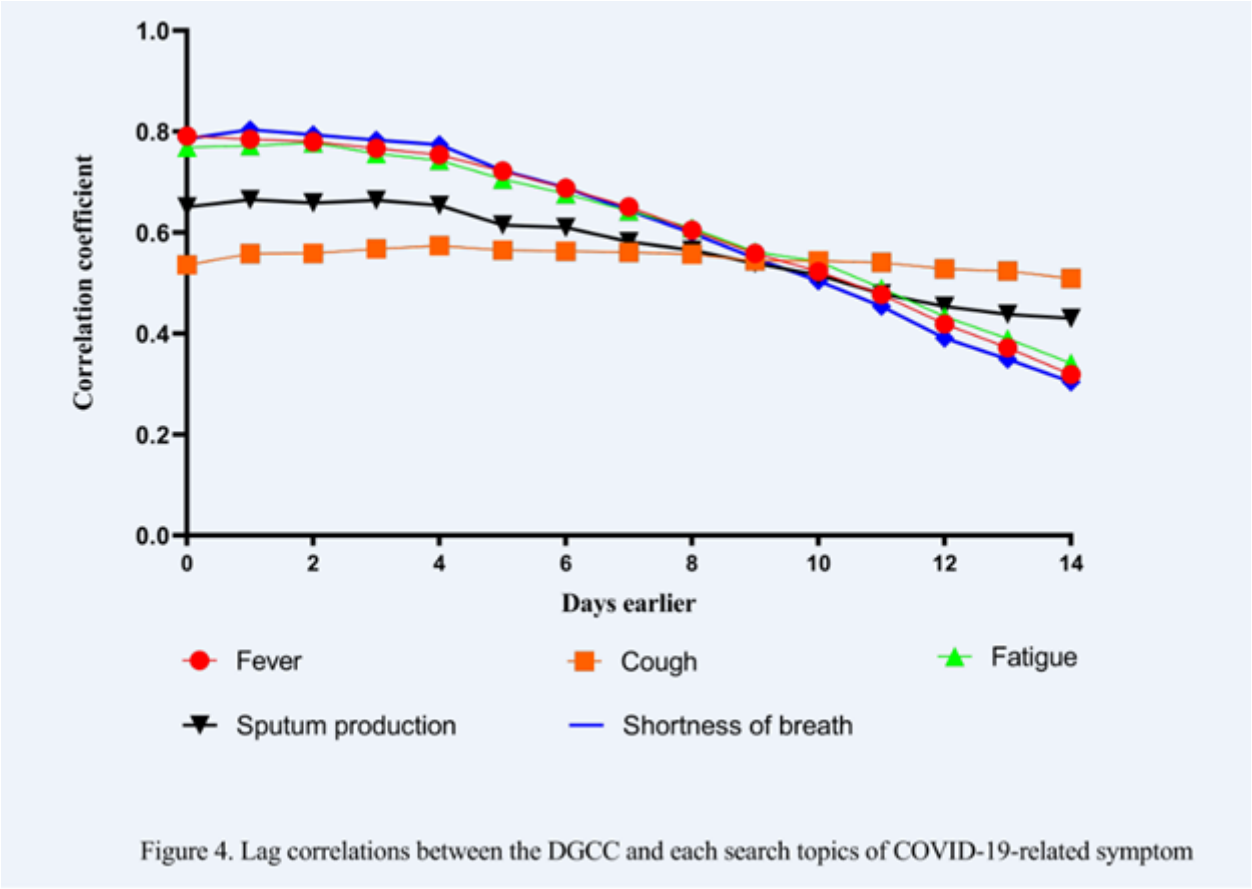

Figure 4. Lag correlations between the DGCC and each search topics of COVID-19-related symptom

Figure 4

Legend in Figure.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.PNG
- FigureS2.tif
- FigureS3.tif
- TableS1.docx
- TableS2.docx