Early detection of COVID-19 pandemic: 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)
Anhui Medical University

Research article

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

Posted Date: July 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-44082/v1

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 the onset and progress of the pandemic with a novel perspective using Baidu Index.

Methods: We collected the confirmed data of COVID-19 infection between January 11, 2020, and April 22, 2020, from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Based on known literature, we obtained the search index values of the most common symptoms of COVID-19, including fever, cough, fatigue, sputum production, and shortness of breath. Spearman's correlation analysis was used to analyze the association between the Baidu index values for each COVID-19-related symptoms and the number of confirmed cases. Regional differences among 34 provinces/ regions were also analyzed.

Results: Daily growth of confirmed cases and Baidu index values for each 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 (fever: 22 days, cough: 19 days, fatigue: 20 days, sputum production: 19 days, and shortness of breath: 19 days). We discovered similar results in the top 10 provinces/regions, which had the highest cumulative cases.

Conclusion: Search terms of COVID-19- related symptoms on the Baidu search engine can be used to early warn the outbreak of the epidemic. Relevant departments need to pay more attention to areas with high search index and take precautionary measures to prevent these potentially infected persons from spreading further. Baidu search engine can reflect the public's attention to the pandemic and regional epidemics of viruses. Based on changes in the Baidu index value, we can predict the arrival of the peak confirmed cases. The clinical characteristics related to COVID-19- including fever, cough, fatigue, shortness of breath, deserve more attention during the pandemic.

1. Background

Since the outbreak of COVID-19 pandemic in late December 2019 [1], SARS-CoV-2 had attacked more than 188 countries and regions, resulted in over 9.6 million cumulative confirmed cases and 490 thousand deaths worldwide [2]. The astonishing spread speed of the epidemic, to some extent, is failing to monitor and manage potentially infected persons, which later confirmed, may pose a substantial infection control challenge [3]. 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 development of the epidemic disease is 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, Chinese Internet users reached 904 million, and the penetration rate of search engine use reached 83 percent [4]. Big data shows that, among all the internet users, 80% of them tend to use electronic devices to acquire the information they are interested in [5]. The high inclusion ratio makes network data representative. People can easily get 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 [6]. Search behavior has been used to predict some epidemic diseases, such as influenza [7], epidemic erythromelalgia [8], dengue [9], HIV/AIDS [10]. Many approaches have been applied to achieve near real-time surveillance of the emergence and spred of COVID-19, including official announcements, news reports, and mass media [11, 12]

The surveillance of network searches about clinical Characteristics of COVID-19 are more predictable and timely compared to previous detection methods [13]. Baidu serves as the largest search market in China, and more than 90% of Internet users search for the information interested using Baidu search engine [14]. In this study, we monitored the Baidu index of COVID-19-related symptoms and the confirmed cases of COVID-19 across China to explore the association between these variables. Moreover, we conducted a search-to-confirm study to explore whether the Baidu index can warn the peak of confirmed cases earlier. Therefore relevant departments have sufficient time to make preparations to control the spread of the epidemic.

2. Methods

2.1 Data from Badu Index

More than 90% of Chinese search engines user tend to use Baidu, which occupies a considerable market share in China to retrieve their interesting information [15, 16]. The weighted sum of the Baidu search values can describe the characteristics of people's search behaviors [17]. Baidu Index (BI) is defined by calculating the number of search terms of specific keywords input by the searchers [17]. Using the keywords analysis function, Baidu Index automatically matches its related words according to the keywords typed by users. According to previous researches, 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%) [18]. Thus, we also implemented those symptoms as the keywords in the current study. Based on the keyword analysis function, 26 search terms which represent the symptoms of COVID-19 were defined (Table S1). We add the search value of each keyword and its related words as the Baidu Index values of the keyword. Besides, we compared the search volume of keywords in other years vertically to investigate whether the change of Baidu Index was an accidental event during the outbreak (Figure S1).

2.2 Data of confirmed cases of COVID-19

We obtained information about confirmed cases of the 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]. We conducted a statistical analysis of confirmed cases in 34

provinces/regions in China. Since the epidemic situation in China tended to be stable in the later period, we divided the COVID-19 situation into a growth period (GP) and a decline period (DP). GP and DP are separated by the dates on February 10, 2020, when officials announce the road closures re-opened and resume production fully [21].

2.3 Statistical analysis

First, we compared the correlation between cumulative confirmed cases in China and the daily search volume during the pandemic, GP, DP, respectively. Spearman correlation analysis of SPSS (version 23.0) was applied to explore the relationships between daily growth of confirmed cases (DGCC) and daily Baidu index values (DBIV). DBIV for keywords was treated as an independent variable, while the cumulative confirmed cases and DGCC as the dependent variable, respectively. We analyzed the association between DBIV and DGCC nationwide and 34 provinces/regions, respectively. DBIV minus its previous day's value as the growth of the Baidu Index, which indicates that compared to the previous day, the more growth of the Baidu Index, the more searchers and potential affecters will be. In order to explore whether people's search behaviors are ahead of the outbreak, we compared the day when the growth of Baidu Index reached the apex, and the maximum of DGCC to explore whether the former is ahead of the latter. And we defined this as a search-to-confirmed interval (STCI). We only selected the top ten provinces/regions ranked by the cumulative confirmed cases for STCI study due to the lack of accumulated confirmed cases in other provinces/regions. $P < 0.05$ was set as the significant statistical difference between variables (two-sided test). Besides, GraphPad Prism 8.2 was used to draw figures.

3. Results

3.1 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 high-speed unblocking 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}) has a strong positive correlation during GP and a significantly negative correlation during DP (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 2 and figure 2 shows that there is a strong statistically positive correlation between the DGCC and search values of Baidu-Index-related to fever ($r_{s=}$0.786, p=8.013 x10^{-23}), cough (r_s=0.556, p=1.087 x10^{-9}), fatigue (r_s=0.763, p=7.930 x10^{-21}), sputum production (r_s=0.665, p=1.793 x10^{-14}), shortness of breath

($r_{s=}$0.780, p=2.673 x10^{-22}), nationwide. For 34 provinces/regions in China, we observed that the number of daily confirmed cases increased when the Baidu searches for terms related to fever, cough, fatigue, and shortness of breath increasing. Except for Hong Kong, Macao, Taiwan, and Tibet. However, DBIV of cough in Shanghai does 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.

Table 2. Correlation between daily growth of confirmed cases (DGCC) across China and Values of Baidu index (BI)

Daily growth of CC across China		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

3.2 Baidu Index maximum growth rate earlier than DGCC

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: 9 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 related to fatigue in Heilongjiang lags behind the peak of DGCC by 17 days.

4. Discussion

The current study observed that Big data of the Internet could be used to warn the outbreak of epidemic diseases. In this research, analytical research into the correlation between search behavior of COVID-19-related keywords and the number of confirmed cases is conducted according to the Internet's big data. We discovered that the search volume of several COVID-19-related keywords has a strong correlation to the number of confirmed cases. And STCI research predicts the onset of epidemic peaks earlier than previous big data monitoring (usually a week in advance), longer than the incubation period for epidemic diseases.

China, which had reached prideful successes for the control of the COVID-19 pandemic, as one of the few countries, had resumed production in the whole society. The search behavior of Chinese citizens during the epidemic help analyze the correlation between clinical symptoms of affected people and retrieved values. People with the travel history of highly regulated areas and exposure history with the confirmed patients will be required quarantined. Without a clear understanding of the characteristics and effective treatments of the new coronavirus, 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, people tend to conceal their own and their family's high-risk behaviors, such as a case of Zhengzhou poison King [23], who intentionally conceal his exposure history, resulted in hundreds of people's isolation. A deep fear of an unknown virus undermines the government's early attempts to control COVID-19. Using Baidu Index, we can figure out the potential quantity of affected people. Moreover, real-time data of the Baidu Index is particularly crucial for the monitor of epidemic development and the formulation of corresponding government policies.

Our research suggests that DGCC dynamic change lags behind the public Baidu index values of COVID-19-related symptoms. When the search volumes increased, the cumulative confirmed cases increased as well, which indicates that the searchers could be the potential infector of the virus. Although the number of confirmed cases is increasing, the public's attention has dropped significantly manifested by the declined Baidu Index value during DP. Those presents the related daily Baidu index values reached a peak earlier than the DGCC, and has a priority to decline in the later period. Based on this result, we can hypothesize related DBIV can be used as an indicator of epidemic development. Public search behavior can reflect potential physical and psychological problems [6, 24]. The decline of search values also indicates that the public's attention to COVID-19 is lighter in the later stage of the pandemic compared with the former stage. We can use the Baidu index to supervise the epidemic situation as well as the public attention to COVID-19.

Overall, five keywords of DBIV were positively correlated with DGCC during the outbreak. From dynamic fluctuations, we can identify the coordinated changes of DBIV and DGCC, with the former keep ahead of the later. For 34 provinces/regions, although most areas in this research showed statistically essential correlations of the DBIV with DGCC (except sputum production), Hong Kong, Macao, Taiwan, and Tibet did not show that correlation. This is probably owing to the Baidu search engine is not the primary search tools in non-mainland areas, such as Hong Kong, Macao, Taiwan [4]. There are few cumulative confirmed cases in Tibet (only one cumulative case), which leads to insufficient cases to calculate the correlation

using SPSS 23.0. However, there is no correlation between DGCC and DBIV for cough in Shanghai. This is probably owing to the incompleteness of search words related to keywords. Based on our research, the increase in the related DBIV value can be treated as an abnormal signal, compared with a period of past time, which is worthy of the corresponding action by government departments in advance. 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].

The growth rate of the Baidu Index represents the newly increased searchers compared to the previous day. The increased number of relevant searchers indicates more potentially exposed persons. Around 97.5% of people with identifiable exposure history will develop symptoms within 11.5 days; more than 14 days occupy 2% (99th percentile) [26]. We found that the maximum of DBIV's growth rate was 20 days earlier than DGCC on average in most areas except Heilongjiang. The abnormality in Heilongjiang may suggest the possibility of insufficient preparation for the pandemic. People used the Internet to search for symptoms rather than going to the hospital, indicating the difference to publicly reported overrepresent severe cases [6, 27, 28]. Since the government implemented the isolation measures during the epidemic, the standard medical treatment process is slower and more complicated [29]. Moreover, many community hospitals cannot prescribe medicine for fever patients result in omission to potential patients with minor symptoms. These people are likely to use search engines (usually is Baidu) for related information, so the Baidu index provided an original way to reflect the number of these potential infecters. Those mild potential infectors may possess a more extended incubation period theoretically on account of a lag of several days in confirmed cases [30]. The longer search-to-confirmed interval, the more time for relevant departments to make adequately prepare. The results mean that the big data of public search behavior can detect the COVID-19 pandemic situation in advance, to some extent, highlighting the importance of including search engine data for follow-up prevention and control. We can derive a vital message that the network search value about Clinical Characteristics of COVID-19 using the Baidu Index can monitor the development of the epidemic. The results will be more convincing if all mainstream search engine data is included

5. Strengths And Limitations

Previous researches based on search engine data are focused on the assessment of epidemic burden [31] and progression [7-10, 27, 28]. This is the first research to explore the public search behavior of search terms for COVID-19-related symptoms to warning the outbreak and popularity of COVID-19.

However, there remains some limitations need to be recognized. Other search engines also occupy a prominent market share, such as Weibo, Twitter. We did not combine the search values of all search engines to get a more representative database. People who never searched on the Internet are also worthy of incorporation. Search engine values are unavoidably disturbed by some mainstream media owing to public search behavior is primarily guided by the media promotion [32]. The Baidu Index does not provide specific search information, such as gender, age, and position, so it is difficult to targeted monitor

potential infectors. It is a substantial urgent need to finding a useful model to overcome the shortcomings of a single mainstream search engine and the unavailability of obtaining retrieval information.

6. Conclusion

The public search behavior shows that Baidu Index can provide practical real-time information to predict the development of the epidemic during the outbreak of COVID-19. Baidu Index could guide more effective and targetable intervention and prevention of COVID-19, assist in the overall control of this pandemic.

7. Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Data availability statement

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

Conflict of interest statement

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. Wuhan Municipal Health Commission (2020). Available at: http://wjw.wuhan.gov.cn/ (accessed June 26 2020).
2. COVID-19 Dashboard by Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available at: https://www.arcgis.com (accessed June 26 2020)
3. Al-Tawfiq JA. Travel Med Infect Dis Asymptomatic coronavirus infection: MERS-CoV and SARS-CoV-2 (COVID-19). 2020 Feb 27:101608. doi: 10.1016/j.tmaid.2020.101608. [Epub ahead of print]
4. Search engines in China - statistics & facts. Available at: https://www.statista.com/topics/1337/search-engines-in-china/ (accessed June 26 2020)
5. Fox S (2005) Health information online. Pew Internet & American Life Project, Washington, DC
6. Cervellin G , Comelli I , Lippi G. J Epidemiol Glob Health, Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. 2017 Sep;7(3):185-189. doi: 10.1016/j.jegh.2017.06.001. Epub 2017 June 9.
7. Yuan Q , Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. PLoS One, Monitoring influenza epidemics in china with search query from baidu. 2013 May 30;8(5):e64323. doi: 10.1371/journal.pone.0064323. Print 2013.
8. Gu Y, Chen F, Liu T, Lv X, Shao Z, Lin H et al. Sci Rep, Early detection of an epidemic erythromelalgia outbreak using Baidu search data. 2015 July 28;5:12649. doi: 10.1038/srep12649.
9. Guo P, Liu T , Zhang Q , Wang L , Xiao J , Zhang Q , et al. PLoS Negl Trop Dis, Developing a dengue forecast model using machine learning: A case study in China. 2017 Oct 16;11(10):e0005973. doi: 10.1371/journal.pntd.0005973. eCollection 2017 Oct.
10. He G , Chen Y , Chen B , Wang H , Shen L , Liu L , et al. Sci Rep, Using the Baidu Search Index to Predict the Incidence of HIV/AIDS in China. 2018 Jun 13;8(1):9038. doi: 10.1038/s41598-018-27413-1.
11. Freifeld CC , Mandl KD, Reis BY, Brownstein JS. J Am Med Inform Assoc, HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. 2008 Mar-Apr;15(2):150-7. Epub 2007 December 20.
12. Tom H van de Belt, Pieter T van Stockum, Lucien J L P G Engelen, Jules Lancee, Remco Schrijver, Jesús Rodríguez-Baño, et al. Antimicrob Resist Infect Control, Social Media Posts and Online Search Behaviour as Early-Warning System for MRSA Outbreaks. 2018 May 30;7:69. doi: 10.1186/s13756-018-0359-4. eCollection 201

13. Li C, Chen LJ , Chen X , Zhang M , Pang CP , Chen H . Euro Surveill, Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. 2020 Mar;25(10). doi: 10.2807/1560-7917.ES.2020.25.10.2000199.
14. China Internet Network Information Center. Available at: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/. (accessed June 26 2020)
15. China Search Engine Market Overview (2015). Available at: https://www.chinaInternetwatch.com/17415/search-engine-2012-2018e/ (accessed June 26 2020).
16. 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]
17. Baidu. Baidu Index URL: https://index.baidu.com/ (accessed 26 Jun 2020) [WebCite Cache ID 6yOtOa7p9]
18. Guan WJ , Ni ZY , Hu Y , Liang WH , Ou CQ , He JX , et al. N Engl J Med, Clinical Characteristics of Coronavirus Disease 2019 in China. 2020 Apr 30;382(18):1708-1720. doi: 10.1056/NEJMoa2002032. Epub 2020 February 28.
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. Haroon Ashraf. Lancet, Investigations Continue as SARS Claims More Lives. 2003 April 12;361(9365):1276. doi: 10.1016/S0140-6736(03)13036-X.
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, et al. J Glob Health, More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. 2020 Jun;10(1):011003. doi: 10.7189/jogh.10.011003.
25. Timothy M Uyeki, Henry H Bernstein, John S Bradley, Janet A Englund, Thomas M File, Alicia M Fry, et al. Clin Infect Dis, Clinical Practice Guidelines by the Infectious Diseases Society of America: 2018 Update on Diagnosis, Treatment, Chemoprophylaxis, and Institutional Outbreak Management of Seasonal Influenzaa. 2019 March 5;68(6):895-902. doi: 10.1093/cid/ciy874.
26. Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, et al. Ann Intern Med, The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. 2020 May 5;172(9):577-582. doi: 10.7326/M20-0504. Epub 2020 March 10.
27. Herman Anthony Carneiro, Eleftherios Mylonakis. Clin Infect Dis, Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks. 2009 Nov 15;49(10):1557-64. doi: 10.1086/630200.

28. Camille Pelat, Clément Turbelin, Avner Bar-Hen, Antoine Flahault, Alain- Jacques Valleron. Emerg Infect Dis,More Diseases Tracked by Using Google Trends. 2009 Aug;15(8):1327-8. doi: 10.3201/eid1508.090299.
29. Dingtao Hu, Xiaoqi Lou, Zhiwei Xu, Nana Meng, Qiaomei Xie, Man Zhang, et al. J Glob Health,More Effective Strategies Are Required to Strengthen Public Awareness of COVID-19: Evidence From Google Trends. 2020 Jun;10(1):011003. doi: 10.7189/jogh.10.011003.
30. 中国政府网. 国务院联防联控机制新闻发布会:介绍医疗救治工作进展情况[EB/OL]. (2020-02-17)[2020-3-5]. Available at: http://www.gov.cn/xinwen/gwylflkjz18/index.htm (accessed 26 Jun 2020)
31. Google Trends. URL: https://trends.google.com/trends/?geo=US [accessed 2019-01-23] [WebCite Cache ID 75e1rdRBY]
32. Ro-Ting Lin, Yawen Cheng, Yan-Cheng Jiang. J Med Internet Res, Exploring Public Awareness of Overwork Prevention With Big Data From Google Trends: Retrospective Analysis. 2020 June 26;22(6):e18181. doi: 10.2196/18181.

Figures

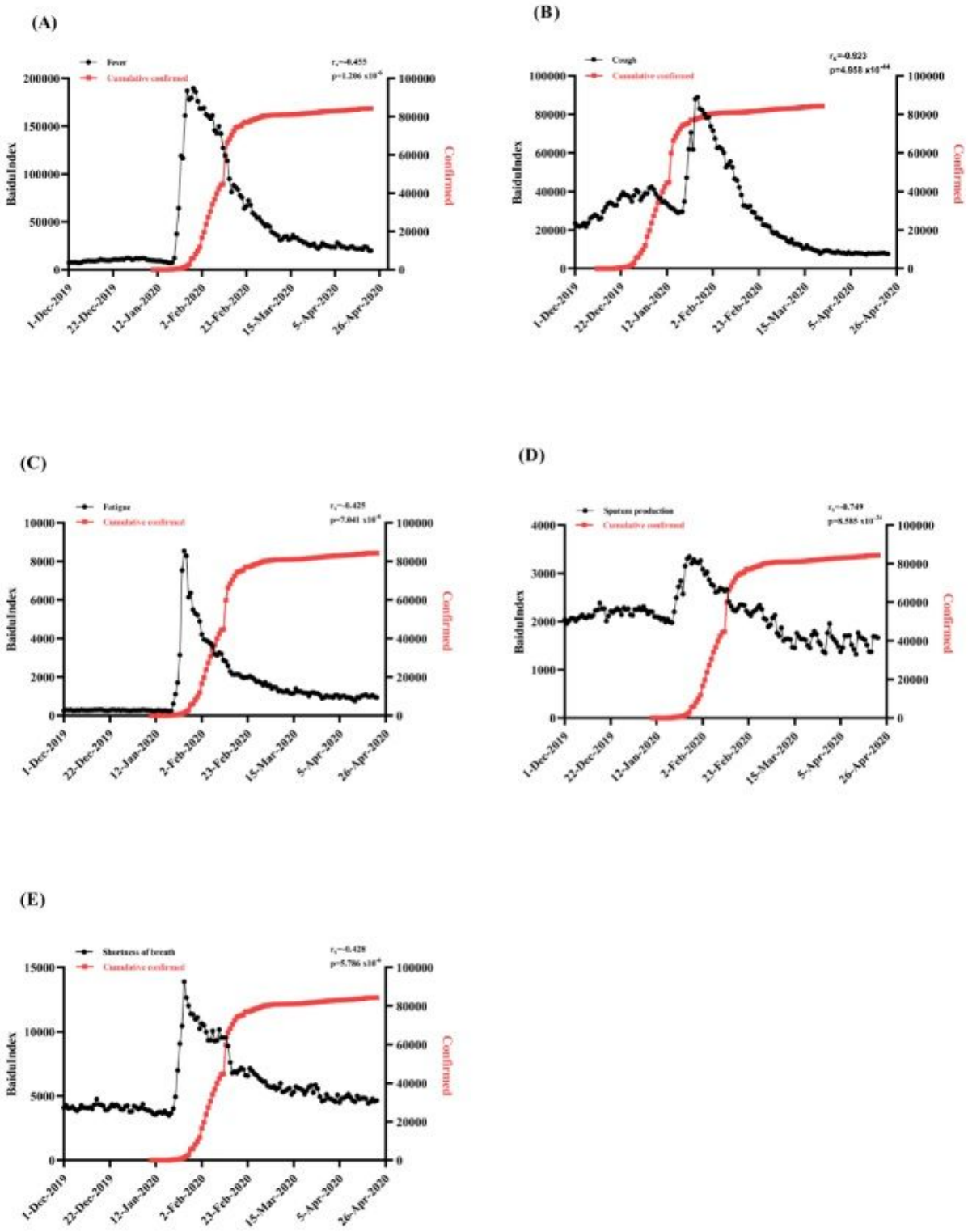

Figure 1

Correlation and time plots among cumulative confirmed cases and Baidu Index of Covid-19-related symptoms.

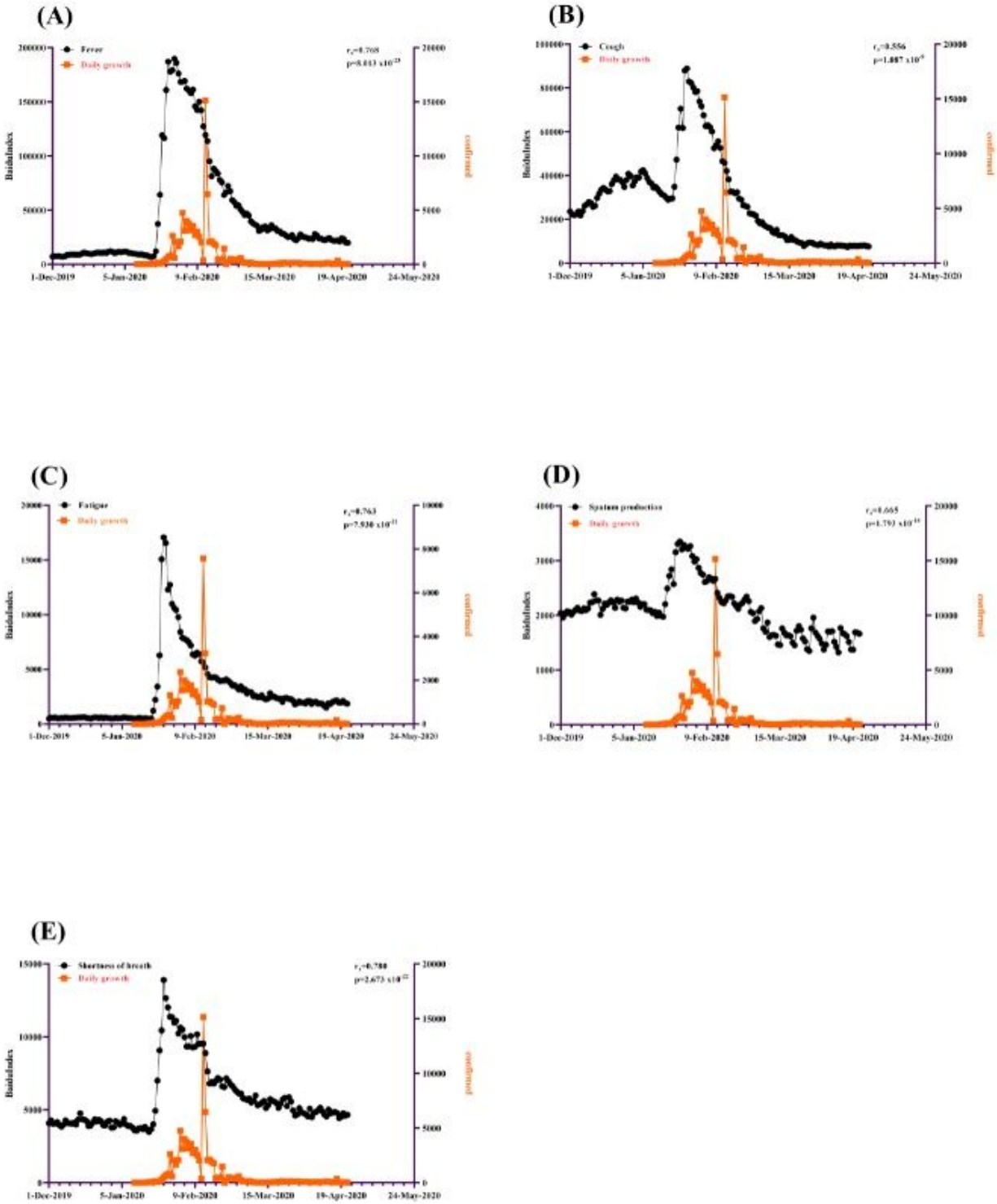

Figure 2

Correlation and time plots among daily confirmed cases and Baidu Index of Covid-19-related symptoms

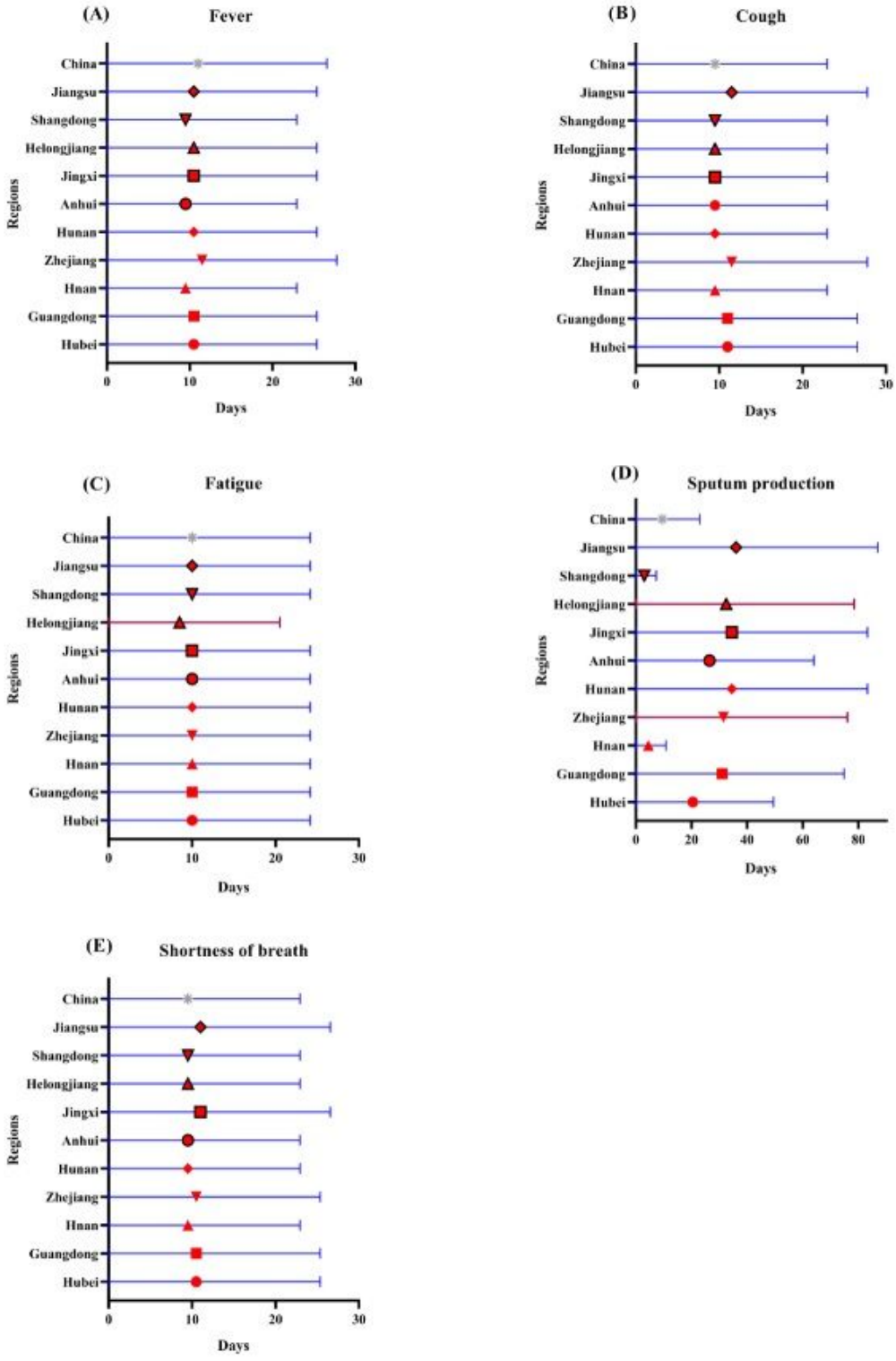

Figure 3

Search-to-confirmed interval of Baidu Index for Covid-19-related symptoms in the top den provinces/regions with the most confirmed cases. Eg. The purple line represents the absolute value of a negative value

Supplementary Files

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

- TableS1.docx
- FigureS1.PNG
- FigureS2.jpg
- FigureS3.jpg