World Journal of Otorhinolaryngology-Head and Neck Surgery (2020) 6, S40—S48

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/WJOHNS; www.wjent.org

Research Paper

Using the internet search data to investigate symptom characteristics of COVID-19: A big data study

Hui-Jun Qiu [a,1], Lian-Xiong Yuan [b,1], Qing-Wu Wu [a,1], Yu-Qi Zhou [c,d], Rui Zheng [a], Xue-Kun Huang [a,d,**], Qin-Tai Yang [a,d,*]

[a] *Department of Otolaryngology-Head and Neck Surgery, The Third Affiliated Hospital of Sun Yat-senUniversity, Guangzhou, 510630, China*
[b] *Department of Science and Research, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, 510630, China*
[c] *Department of Pulmonary and Critical Care Medicine, The Third Affiliated Hospital of Sun Yat-senUniversity, Guangzhou, 510630, China*
[d] *Department of Allergy, The Third Affiliated Hospital of Sun Yat-senUniversity, Guangzhou, 510630, China*

Received 10 April 2020; received in revised form 30 April 2020; accepted 7 May 2020
Available online 19 May 2020

KEYWORDS
SARS-CoV-2;
COVID-19;
Baidu index;
Big data;
Internet

Abstract *Objective:* Analyzing the symptom characteristics of Coronavirus Disease 2019(COVID-19) to improve control and prevention.
Methods: Using the Baidu Index Platform (http://index.baidu.com) and the website of Chinese Center for Disease Control and Prevention as data resources to obtain the search volume (SV) of keywords for symptoms associated with COVID-19 from January 1 to February 20 in each year from 2017 to 2020 and the epidemic data in Hubei province and the other top 9 impacted provinces in China. Data of 2020 were compared with those of the previous three years. Data of Hubei province were compared with those of the other 9 provinces. The differences and

* Corresponding author. Department of Otolaryngology-Head and Neck Surgery, Department of Allergy, The Third Affiliated Hospital of Sun Yat-sen University, No. 600 Tianhe Road, Guangzhou, 510630, China.
** Corresponding authors. Department of Otolaryngology-Head and Neck Surgery, Department of Allergy, The Third Affiliated Hospital of Sun Yat-sen University, No. 600 Tianhe Road, Guangzhou, 510630, China.
E-mail addresses: xuekunhuang@163.com (X.-K. Huang), yang.qt@163.com (Q.-T. Yang).
Peer review under responsibility of Chinese Medical Association.

Production and Hosting by Elsevier on behalf of KeAi

[1] The authors contributed equally to this work

https://doi.org/10.1016/j.wjorl.2020.05.003

characteristics of the SV of COVID-19-related symptoms, and the correlations between the SV of COVID-19 and the number of newly confirmed/suspected cases were analyzed. The lag effects were discussed.

Results: Comparing the SV from January 1, 2020 to February 20, 2020 with those for the same period of the previous three years, Hubei's SV for cough, fever, diarrhea, chest tightness, dyspnea, and other symptoms were significantly increased. The total SV of lower respiratory symptoms was significantly higher than that of upper respiratory symptoms ($P < 0.001$). The SV of COVID-19 in Hubei province was significantly correlated with the number of newly confirmed/suspected cases ($r_{\text{confirmed}} = 0.723$, $r_{\text{suspected}} = 0.863$, both $p < 0.001$). The results of the distributed lag model suggested that the patients who searched relevant symptoms on the Internet may begin to see doctors in 2–3 days later and be confirmed in 3–4 days later.

Conclusion: The total SV of lower respiratory symptoms was higher than that of upper respiratory symptoms, and the SV of diarrhea also increased significantly. It warned us to pay attention to not only the symptoms of the lower respiratory tract but also the gastrointestinal symptoms, especially diarrhea in patients with COVID-19. Internet search behavior had a positive correlation with the number of newly confirmed/suspected cases, suggesting that big data has an important role in the early warning of infectious diseases.

Introduction

Since December 2019, a series of atypical pneumonia cases appeared, which were identified as Coronavirus Disease 2019 (COVID-19), an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2),[1] now became a global concern. As of March 5, more than 80 000 people had been infected in China, including more than 60 000 cases in Hubei province.[2] Besides, several foreign countries had confirmed cases.[2] On January 31, The World Health Organization (WHO) declared this epidemic to be a Public Health Emergency of International Concern(PHEIC).[3]

SARS-CoV-2 is a novel β-coronavirus that infects humans, which can bind to the human angiotensin-converting enzyme 2 (ACE2) to infect cells and replicate,[4] leading to a range of clinical symptoms. For the time being, this new disease is not well understood. The previous studies and the plan of diagnosis and treatment for COVID-19 (China trial version 7) indicated that its main symptoms included fever, cough, and fatigue, while the symptoms of the upper respiratory tract, such as nasal obstruction, runny nose, sore throat, and other nasal symptoms were rare.[5–8] The Internet big data is of great value to the monitoring of infectious diseases and a beneficial supplement to the traditional surveillance systems.[9] At the end of 2019,China had a total of 854 million web users,[10,11] most of whom (more than 90%) used the Baidu search engines,[12] which was the world's largest mandarin search engine and mostly used Chinese website to acquire medical information.[12] The previous studies indicated that the search volume (SV) of symptom keyword was highly correlated with the symptom of patients, which can reflect the real trend of public demand.[13–16] The Baidu Index Platform(http://index.baidu.com), a Google Trend equivalent to the most popular search engine (Baidu) in China, provides search data (SV) for many keywords in each city. To some extent, the SV of keywords in Baidu can reflect the real demands of citizens. Inorder to understand COVID-19's symptom characteristics, the Baidu search-engine data was conducted in our study.

Material and methods

Retrieval strategy

Step 1, determined the top 10 provincial-level regions (hereinafter referred to as provinces) of the total confirmed cases nationwide (as of February 20, 2020) and obtained the number of daily newly confirmed/suspected cases (as of March 5, 2020) through the data released by the Chinese Center for Disease Control and Prevention (http://www.chinacdc.cn/).

Step 2, searched keywords (Table 1.) in mandarin, which were identified based on symptoms recorded among COVID-19 patients and identified the Baidu Index Platform (http://index.baidu.com) as data resources to obtain the daily SV in the Baidu search-engine from January 1 to February 20 in each year from 2017 to 2020, geographically in the top 10 impacted provinces. Keywords for different expressions were combined (Table1.).

Step 3, data from 2020 were compared with those of the previous three years, and data from Hubei province were compared with those of the other 9 impacted provinces.

Statistical analysis

Compared the data from Hubei province and those from the other 9 impacted provinces and compared the SV of upper respiratory symptoms and those of lower respiratory symptoms. Because the Baidu index in 2020 and the Baidu index increment relative to 2017–2019 did not obey the normal distribution, the Wilcoxon Signed Rank Test was

Table 1 Keywords of retrieval strategy at the Baidu index platform.

Combined Keywords (English translation)	Combined Keywords (in Chinese)	Keywords (in Chinese)
COVID-19	新型冠状病毒肺炎	新型冠状病毒肺炎、新冠肺炎、新型肺炎、肺炎
Chronic obstructive pulmonary disease	慢性阻塞性肺疾病	慢性阻塞性肺疾病、慢阻肺
Rhinitis	鼻炎	鼻炎
Gastroenteritis	胃肠炎	胃肠炎、肠胃炎
Coronary heart disease	冠心病	冠心病
Cough	咳嗽	咳嗽、干咳
Sputum	咳痰	咳痰
Runny nose	流涕	流涕、流鼻涕
Nasal obstruction	鼻塞	鼻塞
Sneezing	喷嚏	喷嚏、打喷嚏
Sore throat	咽喉痛	咽喉痛、喉咙痛
Dyspnea	呼吸困难	呼吸不畅、呼吸困难
Shortness of breath	气短	气短
Chest tightness	胸闷	胸闷
Chest pain	胸痛	胸痛

Table 1 (*continued*)

Combined Keywords (English translation)	Combined Keywords (in Chinese)	Keywords (in Chinese)
Palpitations	心悸	心慌、心悸
Fever	发热	发热、发烧
Chills	寒颤	寒颤
Fatigue	乏力	乏力
Myalgia	肌肉酸痛	肌肉酸痛
Lumbago	腰痛	腰痛
Joint pain	关节痛	关节痛
Headache	头痛	头痛
Dizziness	头晕	头晕
Vertigo	眩晕	眩晕
Abdominal pain	腹痛	腹痛、肚子疼、肚子痛
Diarrhea	腹泻	腹泻、拉肚子
Vomiting	恶心呕吐	恶心、呕吐
Conjunctival congestion	结膜充血	结膜充血、眼球充血、眼睛红
Itchy eyes	眼睛痒	眼睛痒
Eyes pain	眼睛痛	眼睛痛、眼睛疼

conducted. Spearman Correlation Analysis was used to analyze the correlations between the Baidu index and the number of newly confirmed/suspected cases. Considered that there may be a lag effect between the newly confirmed/suspected and the network search behavior, and Spearman Correlation Analysis was conducted between the Baidu index on that day and the number of newly confirmed/suspected cases 7 days later. Additionally, a distributed lag model was used to analyze specific lag effects (the time that people who may be infected with SARS-CoV-2 will be suspected/confirmed after searching online). Before constructed a distributed lag model, took the natural logarithm of the number of newly confirmed/suspected cases and the Baidu index. Starting from the zero-order lag model (simple linear regression), it increased the lag effect step by step. For example, the third-order lag model simultaneously considered the regression effect of the Baidu index on the day and the first 1–3 days on newly confirmed cases. All data analyses were performed using R 3.6.2 software (R Foundation for Statistical Analysis), in which the distributed lag model used the DLNM function package. Two-tailed p value < 0.05 indicated as statistical significance. Figures were designed using GraphPad Prism 6 (GraphPad Software, San Diego, CA, USA).

Results

Comparison of the Baidu Index of diseases between Hubei province and the other 9 provinces

As of February 20, 2020, the top 10 provinces in China with the number of total confirmed cases were Hubei, Guangdong, Henan, Zhejiang, Anhui, Jiangxi, Hunan, Jiangsu, Chongqing, and Shandong. During the period from January 1 to February 20, 2020, compared with the average of the other 9 impacted provinces, the people in Hubei were more concerned about COVID-19($p = 0.002$), which began to increase sharply on January 20, reached its peak on January 23, and gradually fell back (Fig. 1). On the contrary, during this period, compared with the average of the other 9 impacted provinces, Hubei province did not

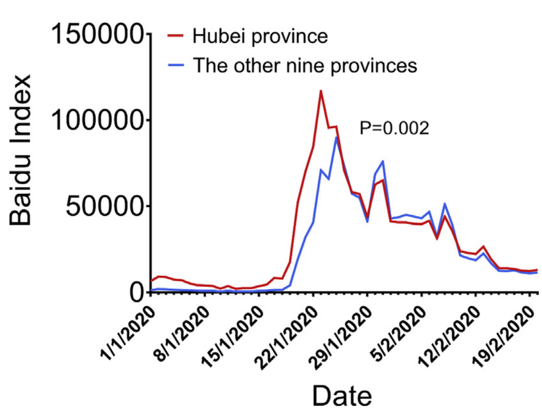

Fig. 1 The Baidu index trend of COVID-19 in Hubei province and the other 9 impacted provinces (average) from January 1, 2020 to February 20, 2020 ($p = 0.002$).

have a significant SV for diseases such as chronic obstructive pulmonary disease (COPD), rhinitis, gastroenteritis, and coronary heart disease (CHD), whose symptoms were like COVID-19. Moreover, the SV of these diseases in Hubei province were even lower than those in the other 9 impacted provinces ($p = 0.318$, <0.001, <0.001, <0.001, respectively, Fig. 2).

Comparison of the Baidu Index of symptoms between Hubei province and the other 9 provinces

From January 20 to February 20, 2020, compared with the average daily SV of the same period in previous years (2017–2019), the average daily SV of the people in Hubei province for the following symptoms increased significantly (Fig. 3A, negative value not shown): cough, fever, diarrhea, chest tightness, dyspnea, palpitations, sneezing, fatigue, runny nose, shortness of breath, chest pain, dizziness, nasal obstruction, headache, conjunctival congestion, sputum, myalgia, chills, itchy eyes, dizziness, and joint pain.

According to the Baidu Index increment, the systems (organs) infected with the SARS-CoV-2included respiratory system, digestive system, circulatory system, locomotion system, nervous system, and eye (Fig. 3B). During this period, the increment in the total SV of lower respiratory tract symptoms (chest tightness, dyspnea, shortness of breath, and chest pain) among people in Hubei province was significantly higher than that of upper respiratory symptoms (sneezing, runny nose, nasal obstruction, and sore throat) ($p < 0.001$, Fig. 3C).

Comparison of the Baidu Index of the top 5 symptoms over the years

Compared with the same period of 2017–2019, the SV of the top 5 symptoms (cough, fever, diarrhea, chest tightness, and dyspnea) began to increase sharply from January 20, 2020, and gradually fell after reaching the peak on January 23 (Fig. 4A), which were consistent with a sea of local cases. The overall trends were consistent with the SV of COVID-19 (Fig. 2A) but were inconsistent with the SV of diseases with similar symptoms of COVID-19, such as COPD, rhinitis, gastroenteritis, and CHD (Fig. 2B). Compared with the average of the other 9 impacted provinces, it could be seen that the SV of the top 5 symptoms in Hubei province had also increased significantly (all $p < 0.001$, Fig. 4B), which had a positive relationship with the actual number of patients.

Correlation analysis between newly confirmed/suspected cases and the Baidu index

The curves of newly confirmed/suspected cases were closely correlated to the Baidu index curve in Hubei province (Fig. 5A, $r_{confirmed} = 0.723$, $r_{suspected} = 0.863$, all $p < 0.001$). The results of the distributed lag models showed that for the newly confirmed, the coefficients of determination(R^2) of the 0 to 4 order lag models were 0.3012, 0.4307, 0.5359, 0.6495, 0.7668, respectively. The R^2 were all increased by 0.1 or more, and the corresponding 0, 1, 2, 3, and 4-order lag effect items in each model were statistically significant. Starting from the 5-order lag

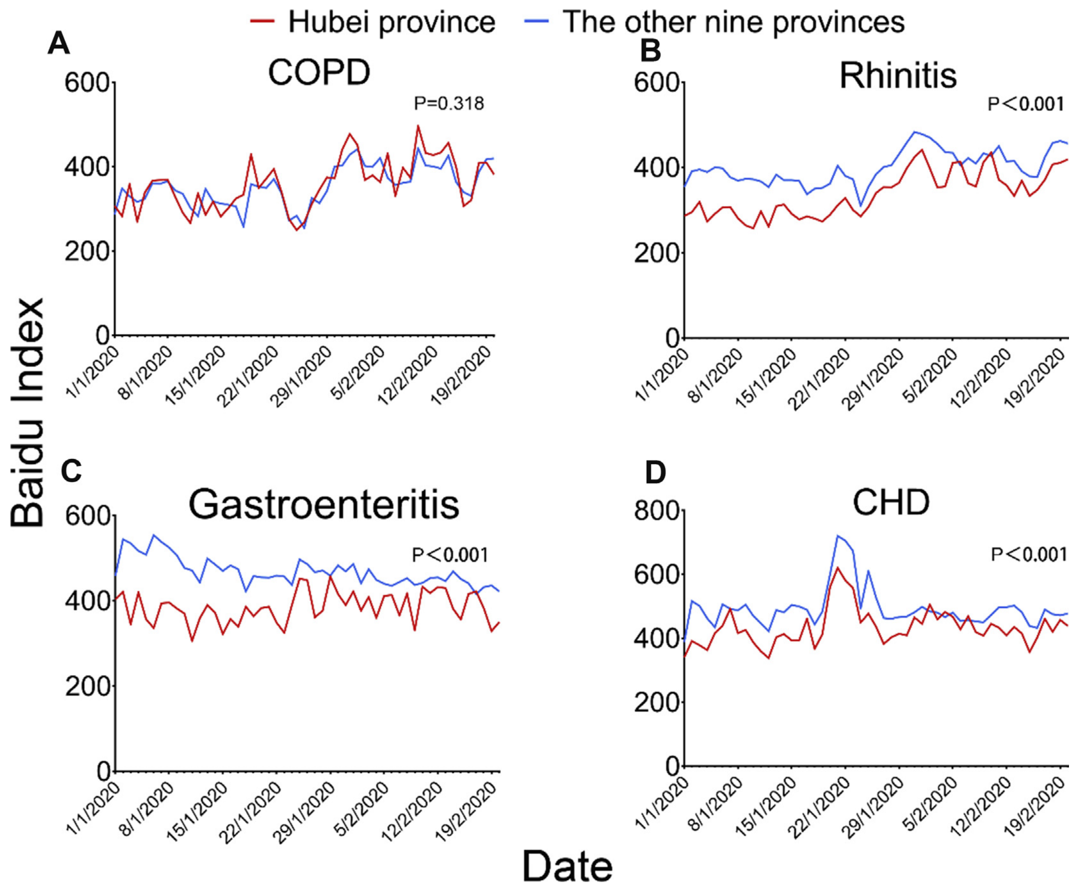

Fig. 2 Comparison of the Baidu index trends of "chronic obstructive pulmonary disease", "rhinitis", "gastroenteritis", and "coronary heart disease" in Hubei province and the other 9 impacted provinces (average) from January 1, 2020 to February 20, 2020 A: Chronic obstructive pulmonary disease ($p = 0.318$); B: Rhinitis ($p < 0.001$); C: Gastroenteritis ($p < 0.001$); D: Coronary heart disease ($p < 0.001$).

model, the R^2 only increased by 0.02—0.04, and the effects were not obvious. For the newly suspected, the R^2 of the 0 to 3-order lag models were 0.0327, 0.4162, 0.7252, and 0.8041, respectively. The R^2 increased significantly, and the corresponding 0, 1, 2, and 3-order lag effect items in each model had statistical significance. Starting from the 4-order lag model, the increases in R^2 were less than 0.03, and the effects were not obvious.

The results of the other 9 impacted provinces were similar. The Baidu index of the day was analyzed with the number of newly confirmed/suspected cases after 7 days by Spearman Correlation Analysis (Fig. 5B). The coefficients of correlation were 0.727 and 0.828, respectively, and the p values were all <0.001. After taking the natural logarithm, distributed lag models were established. For the newly confirmed, the R^2 of the 0 to 3-order lag models were 0.278, 0.6643, 0.7786, and 0.8239, respectively, and the lag effect items were statistically significant. Starting from the 4-order lag model, R^2 only increased by about 0.01, and the effects were not obvious. For the newly suspected, the R^2 of the 0 to 3-order lag models were 0.1524, 0.5343, 0.7758, and 0.8816, respectively. The R^2 increased significantly, and the corresponding lag effect items in each model had statistical significance. Starting from the 4-order lag model, the increases in R^2 were less than 0.03, and the effects were not obvious.

Discussion

The internet big data helps recognize and monitor new diseases

With the continuous development of the Internet and the substantial increase of network coverage in China, search engines had increasingly become the most important channel for people to query medical information. "Searching online before visiting doctors" had become a habit of patients and their families in China. The search keywords directly reflect the real intentions and needs of the patients,[13—16] and the big data of search could be real-time statistics and easy to obtain. Therefore, the internet big data, which are large enough, could become an ideal data source for us to recognize the characteristics of new infectious diseases and to monitor them. During this epidemic, a series of "big data report on the search for COVID-19" were jointly released by People's Daily Online and Baidu, which indicated the average daily number of users searching for information related to COVID-19 exceeded 1 billion on Baidu. The focuses of these reports were epidemic progress and daily preventative methods. Unfortunately, the reports did not cover the search trend related to symptoms of COVID-19. Therefore, we

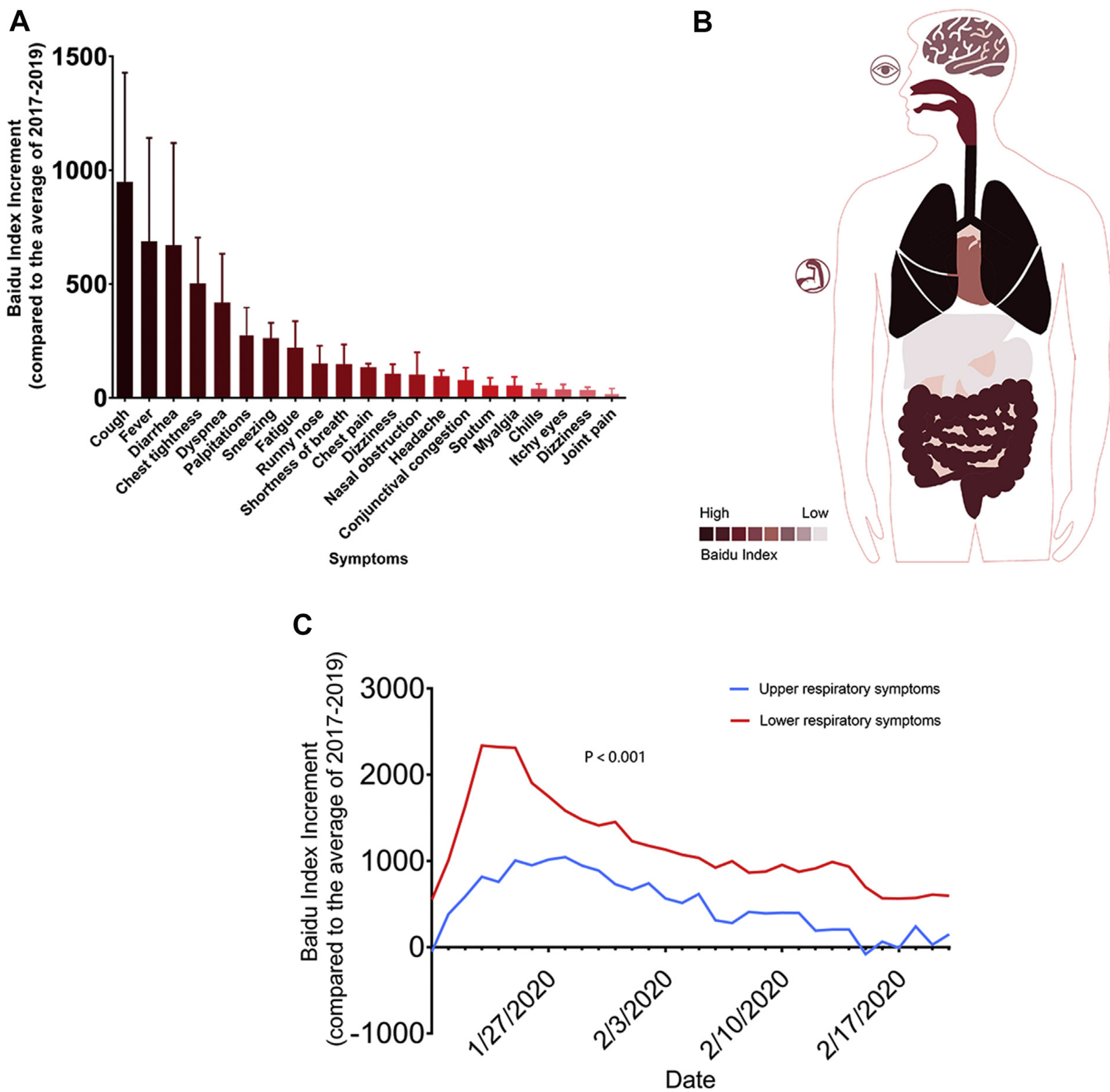

Fig. 3 Comparison of the Baidu Index of COVID-19 Related Symptoms in Hubei Province from January 20, 2020 to February 20, 2020 and the same period in 2017—2019 A: Comparison of daily average increment of the Baidu index related symptoms of COVID-19 in Hubei province; B: The higher the increment in the Baidu index of COVID-19 relative symptoms, the darker the area of the corresponding organ; C: The Baidu index trend of upper and lower respiratory symptoms related to COVID-19.

investigated the symptom characteristics of COVID-19 by analyzing the Baidu search-engine data and identified possible awareness gaps if possible. The results of this study showed that the Baidu search-engine truly reflected the user 's search behavior and the prevalence and symptom characteristics of COVID-19, which could help us to understand this new disease.

Analysis of symptom characteristics of COVID-19 based on the internet big data

COVID-19 was thought to be mainly a disease transmitted through the respiratory tract. However, it was found that unlike the common cold or flu, and the upper respiratory tract symptoms were few. In this study, we found that cough and fever were the most searched symptoms for people in Hubei province based on the big data from the Baidu Index (Fig. 3A). The SV of upper respiratory tract symptoms such as nasal obstruction, runny nose, sneezing, and sore throat were lower, while the SV of lower respiratory tract symptoms such as chest tightness, dyspnea, shortness of breath and chest pain were higher (Fig. 3C), which was consistent with the symptoms of COVID-19 reported in the previous studies and the plan of diagnosis and treatment for COVID-19 (China trial version 7).[5—8] The reason why the upper respiratory tract symptoms were less than the lower respiratory tract may be related to the fact that ACE2 receptors were widely distributed in lung tissues and less distributed in the

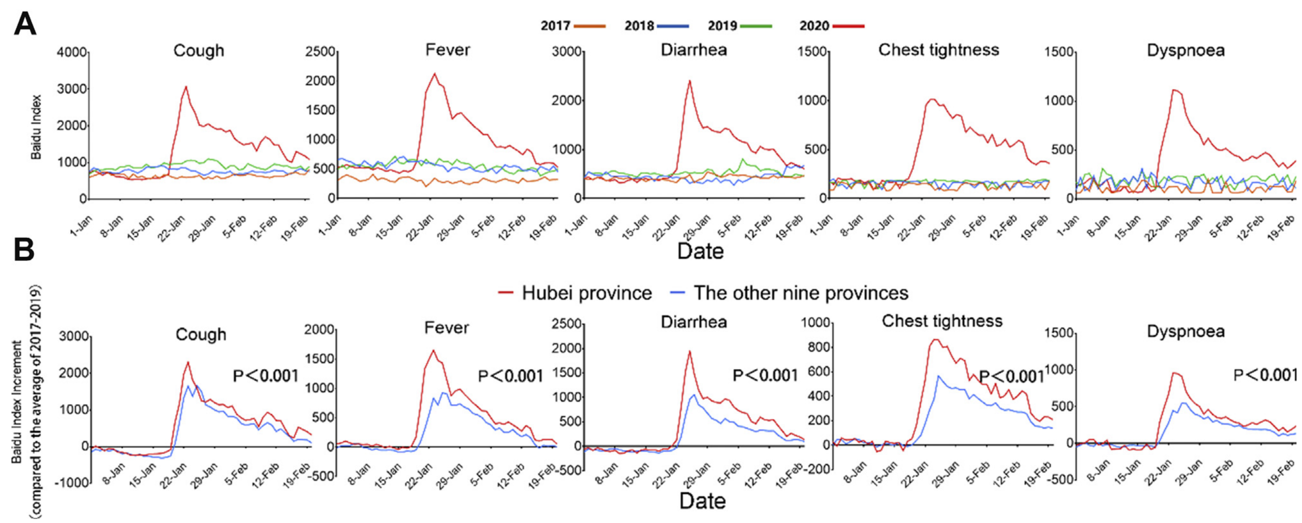

Fig. 4 The trend of the top 5 symptoms of the Baidu index increment from 2017 to 2020 A: The trend of the top 5 symptoms in Hubei province; B: Comparison of the trend of the top 5 symptoms in Hubei province and the other 9 impacted provinces (average) from January 1, 2020 to February 20, 2020 (all $p < 0.001$).

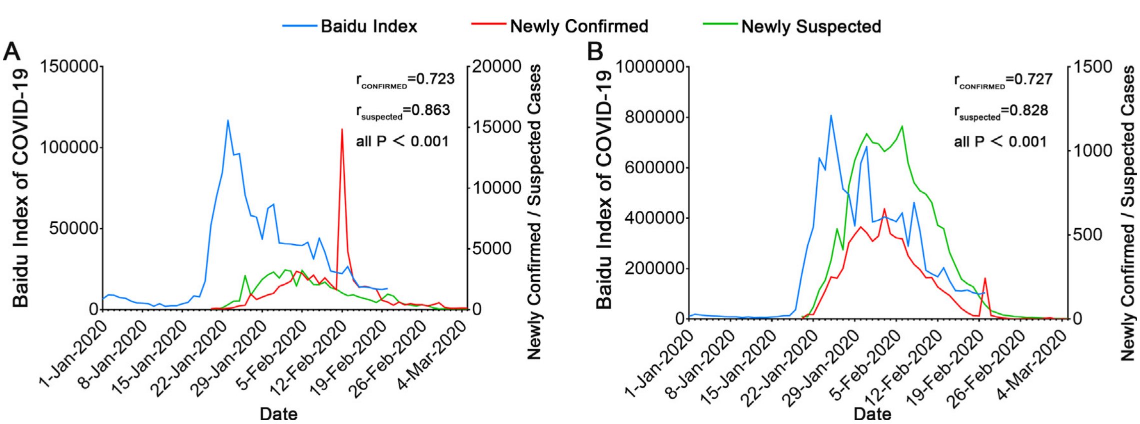

Fig. 5 Correlation between the Baidu Index of COVID-19 and the number of newly confirmed/suspected cases from January 1 to March 5, 2020 A: Comparison of the Baidu Index of COVID-19 and the number of newly confirmed/suspected cases in Hubei province ($r_{confirmed} = 0.723$, $r_{suspected} = 0.863$, both $p < 0.001$); B: Comparison of the sum of the Baidu index of COVID-19 and the sum of the number of newly confirmed/suspected cases in the other 9 impacted provinces ($r_{confirmed} = 0.727$, $r_{suspected} = 0.828$, both $p < 0.001$).

nasal cavity and nasopharyngeal mucosa. Single-cell sequencing data analysis showed that the average proportion of ACE2-positive type II alveolar cells was only about 1%. The respiratory epithelial cells from the lower respiratory tract sample contained about 2% ACE2 positive cells while almost no cells obtained from nasal and bronchial samples showed high ACE2 expression, which indicated that the lower respiratory tract was considered as high risk, while the nasal cavity was in low risk.[17] Therefore, we speculate that after the SARS-CoV-2 reaches the respiratory tract and lungs, it binds to the ACE2 receptor and replicates, damaging the respiratory tract and lungs and causing viremia, which manifests as cough and fever. The virus may then reach and infect other organs with ACE2. In severe cases, it can trigger an "immune inflammation storm", which can lead to multiple organ failure and even death.

In the early stage of the SARS epidemic in 2003, diarrhea was present in only 18.2% of the patients with SARS,[18] but later retrospective studies reported that the diarrhea symptom was as high as 45.6%,[19] which indicated that our understanding of new diseases has limitations. Our study found that the SV of gastrointestinal symptoms such as diarrhea was significantly higher than the same period in 2017–2019. What's more, the proportion of patients with gastrointestinal symptoms especially diarrhea and anorexia accounted for 1/3, based on the actual statistics of our team in Wuhan, and the late patients with gastrointestinal symptoms were increasing, which was consistent with the internet big data but higher than the reports of the previous studies (2%–3.7%).[5–7] Recently, it reported that SARS-CoV-2 was detected in stool,[20] and surface of the epithelial cells of the digestive tract expressed ACE2 in large amounts.[4,17,21] This strongly reminds us to pay attention to the gastrointestinal symptoms, especially diarrhea in patients with COVID-19 and alert the fecal-oral transmission. We should include the symptom of diarrhea in the screening and diagnosis of patients with COVID-19, which would help to improve early diagnosis, early isolation, and early treatment.

The role of the internet big data in predicting potential patients infected SARS-CoV-2

The newly confirmed/suspected curves were closely correlated to the Baidu index curve. It was suggested that there is a correlation between the big data of search and the newly confirmed/suspected cases, which has a certain early warning effect for infectious diseases. A distributed lag model is a model for time series data in which a regression equation is used to predict current values of a dependent variable based on both the current values of an explanatory variable and the lagged (past period) values of this explanatory variable.[22,23] The distributed lag model has been used in economics after Almon popularized them in 1965 and has been widely used in environmental health studies.[24–26] We believe that some people who search for symptoms related to COVID-19 on the internet may be potential patients. Therefore, there may be a close correlation between the popularity of online search and the confirmed/suspected cases. However, this correlation is not immediate. The potential patients who search for symptoms related to COVID-19 on the internet will gradually become suspected/confirmed cases in a subsequent period, which means there is a lag effect. In the current study, the results of the distribution lag models indicated that the infected people who search for related symptoms on the internet might start to see doctors and become suspected cases in about 2–3 days later, and be confirmed in about 3–4 days, which was similar to the National Health Commission's announcement that the national average time from onset to confirmation was 4.95 days.[27] Besides, we know that the number of the newly confirmed is affected by the diagnostic criteria, while the SV is not affected by this. However, especially, due to the changes of diagnostic criteria, the number of newly confirmed cases in Hubei province suddenly increased on February 12, rapidly decreased on February 13, and then gradually decreased in the following days, suggesting that the overall trend of newly confirmed was stable. Therefore, we believe that these changes will not affect the correlation between the overall trend of the Internet big data and the number of newly confirmed cases.

Limitations and prospects

The Baidu Index and other Chinese internet big data had played an important role in the prevention and control of this epidemic. The changing trend of SV for symptoms not only reflected people's real need but also had significant implications for our health departments to recognize the characteristics of new diseases, predict the epidemic trend and formulate strict guidelines for diagnosis and treatment as soon as possible. However, we must be clear that this correlation between the data of search and those of real world is not equal to causation, and it may be influenced by many factors, such as the differences in education and economic conditions of netizens. Besides, there are some other limitations to our study. Firstly, the SV of keywords could be influenced by the continuous changes of individual search behavior, and the Baidu Index Platform still has some keywords not included, which may lead to underestimation of the correlation. Secondly, the keywords of symptoms in this study were identified based on the symptoms recorded among COVID-19 patients. As the public knowledge of the pandemic and the awareness of COVID-19 gradually increases, the included search keywords should be increased accordingly, and even the search keywords for reasonably suspected symptoms should be monitored to fully utilize big data in preventing and controlling new infectious diseases. Finally, using internet search data to assist epidemic surveillance depends on the number of internet visits. The primary users of the internet search engines in China are young and middle-aged, and different regions have different population sizes so that the internet training for all people, especially the aged, should be strengthened in the future to provide more accurate and valuable data.

In the future, we should use advanced methods to achieve at least the following two ideal models for the prevention and control of infectious diseases: firstly, using big data and artificial intelligence to build an early warning system for infectious disease; secondly, using smartphones to lock down the activity trajectories of infectious sources, and realize the intelligence of cognition, early warning, prevention, and control. Compared with the data of traditional surveillance systems, the internet big data has the characteristics of large sample size, fast response, easy access, and low cost. However, digital disease surveillances of the internet still face many challenges, such as the uncertainty of network user behavior, the inaccuracy of search keyword acquisition, and the incompleteness of network coverage. Therefore, in the future, more researches should focus on how to improve the accuracy of the internet search data and explore how to combine search engine data with actual medical information systems. In short, big data will help medical development.

Funding information

This work was supported by the National Natural Science Foundation of China (Grant No. 81670912 and 81870704), the Industry-Academic Cooperation Foundation of Guangzhou (No. 201704030046), and Sun Yat-sen University Clinical Research 5010 Program (NO. 2019006).

Conflict of interest

All authors declare that they have read and understood the policy on declaration of interests of this magazine and have no competing interests. This study was presented as an article at Chinese Journal of Otorhinolaryngology Head and Neck Surgery in Chinese.

Acknowledgment

We wish to thank Dili Daer, MD, PhD of Department of Medical Affairs, Chiesi Pharmaceutical (Shanghai) Co., Ltd, for the support and assistance in this research.

References

1. Wuhan Municipal Health Commission. *Situation Report on Unexplained Viral Pneumonia*. EB/OL; 2020-01-11 [2020-2-24] http://wjw.wuhan.gov.cn/front/web/showDetail/2020011109035.
2. Chinese Center for Disease Control and Prevention. *Pandemic Mapping of Coronavirus Disease 2019*. EB/OL; 2020-03-5 [2020-3-5] http://2019ncov.chinacdc.cn/2019-nCoV/.
3. World Health Organization. *Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV)*. EB/OL; 2020-1-30 [2020-2-24] https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov.
4. Zhang H, Kang Z, Gong H, et al. *The Digestive System Is a Potential Route of 2019-nCov Infection: A Bioinformatics Analysis Based on Single-Cell Transcriptomes*. EB/OL; 2020-1-31. https://doi.org/10.1101/2020.01.30.927806 [2020-2-25].
5. Chen NS, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395:507—513.
6. Guan WJ, Ni ZY, Hu Y, et al. *Clinical Characteristics of 2019 Novel Coronavirus Infection in China*. EB/OL; 2020-02-09. https://doi.org/10.1101/2020.02.06.20020974 [2020-2-24].
7. Huang CL, Wang YM, Li XW, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497—506.
8. National Health Commission of the People's Republic of China. *Diagnosis and Treatment of Pneumonia with New Coronavirus Infection (Trial Version 7)*. EB/OL; 2020-03-03 [2020-04-05] http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964.pdf.
9. Milinovich GJ, Williams GM, Clements AC, et al. Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis*. 2014;14:160—168.
10. Milinovich GJ, Magalhães RJ, Hu W. Role of big data in the early detection of Ebola and other emerging infectious diseases. *Lancet Glob Health*. 2015;3:e20—e21.
11. China Internet Network Information Center. *The 44th Statistical Report on Internet Development in China. 2020*; 2020-2-24 (2019-08-30)[2020-2-24] http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/.
12. China Internet Network Information Center. Research report on search engine usage of Chinese netizens in 2019. 2020 (2019-10-25)[2020-2-24] http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/; 2020-2-24.
13. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis*. 2009;49:1557—1564.
14. Li K, Liu M, Feng Y, et al. Using Baidu search engine to monitor AIDS epidemics inform for targeted intervention of HIV/AIDS in China. *Sci Rep*. 2019;9:320.
15. Dong X, Li L, Xu W, et al. Correlation between Baidu index of specific keywords and influenaz epidemics. *Chin J Public Health*. 2016;32:1543—1546.
16. Wang J, Zou Y, Peng Y, Li K, Jiang T. On prediction of dengue epidemics based on Baidu index. *Comp Appl Softw*. 2016;33:42—46.
17. Zou X, Chen K, Zou JW, Han PY, Hao J, Han ZG. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med*. 2020;12:1—8.
18. Li X, Jiang R, Guo J, Wang Q. Clinical analysis of SARS: 27 cases report. *Chin Med J*. 2003;83:910—912.
19. Chen SY, Chiang WC, Ma MH, et al. Sequential symptomatic analysis in probable severe acute respiratory syndrome cases. *Ann Emerg Med*. 2004;43:27—33.
20. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382:929—936.
21. Liang WC, Feng ZJ, Rao ST, et al. Diarrhea may be underestimated: a missing link in 2019 novel coronavirus. *Gut*. 2020;69:1141—1143.
22. Cromwell JB, Hannan MJ, Labys WC, Terraza M. *Multivariate Tests for Time Series Models*. London: Sage Publications; 1994:104.
23. Dielman TE. The theory and practice of econometrics. *Int J Forecast*. 1986;2:245—246.
24. Almon S. The distributed lag between capital appropriations and expenditures. *Econometrica*. 1965;33:178.
25. Xu QQ, Li RZ, Rutherford S, et al. Using a distributed lag non-linear model to identify impact of temperature variables on haemorrhagic fever with renal syndrome in Shandong Province. *Epidemiol Infect*. 2018;146:1671—1679.
26. Zhang YQ, Li CL, Feng RJ, et al. The Short-term effect of ambient temperature on mortality in Wuhan, China: a time-series study using a distributed lag non-linear model. *Int J Environ Res Public Health*. 2016;13:722.
27. National Health Commission of the People's Republic of China. *Press Conference of the Joint Prevention and Control Mechanism of the State Council: Introducing the Progress of Medical Treatment*; 2020-02-17 [2020-3-5] http://www.gov.cn/xinwen/gwylflkjz18/index.htm.

Edited by Yu-Xin Fang