Using WeChat, a Chinese social media, to early detect the SARS-CoV-2 outbreak in 2019: a retrospective study

Wenjun Wang, Yikai Wang, Xin Zhang, Xiaoli Jia, Yaping Li, Shuangsuo Dang

Submitted to: JMIR mHealth and uHealth
on: April 25, 2020

Table of Contents

Using WeChat, a Chinese social media, to early detect the SARS-CoV-2 outbreak in 2019: a retrospective study

Wenjun Wang[1] MD; Yikai Wang[1] MD; Xin Zhang[1] MD; Xiaoli Jia[1] MD; Yaping Li[1] MD; Shuangsuo Dang[1] MD, PhD

[1]Second Affiliated Hospital of Xi'an Jiaotong University Xi'an CN

Corresponding Author:
Shuangsuo Dang MD, PhD
Second Affiliated Hospital of Xi'an Jiaotong University
157 Xiwu Road
Xi'an
CN

Abstract

Background: A novel virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has started a global pandemic of pneumonia since December 2019, when first cases were reported in Wuhan, China. It has caused 2.7 million confirmed cases and nearly 200,000 deaths as of April 24, 2020. Early warning systems with new technologies should be established to avoid such disasters.

Objective: This study aimed to early detect the SARS-CoV-2 outbreak in 2019 using social media.

Methods: WeChat Index is a data service that shows how frequent a specific keyword has appeared in posts, subscriptions, and search over a period of last 90 days on WeChat, the most popular Chinese social media. We plotted daily WeChat Index from Nov 17, 2019 to Feb 14, 2020 for keywords related to the SARS-CoV-2 disease.

Results: WeChat Index for “Feidian” that is SARS in Chinese language had stayed at low levels until 16 days ahead of the outbreak announcement by the local authority when the index increased significantly. Later, the index persisted at relative high levels and rose rapidly on the day before the announcement. WeChat Index also spiked or increased for keywords “SARS”, “coronavirus”, “novel coronavirus”, “shortness of breath”, “dyspnea”, and “diarrhea”, but not as meaningful as “Feidian” in early detection of the outbreak.

Conclusions: Using WeChat may detect the SARS-CoV-2 outbreak in 2019 about two weeks earlier than the traditional surveillance systems. WeChat offers a new approach to early detect disease outbreaks.

(JMIR Preprints 25/04/2020:19589)
DOI: https://doi.org/10.2196/preprints.19589

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Short Paper

Using WeChat, a Chinese social media, to early detect the SARS-CoV-2 outbreak in 2019: a retrospective study

Abstract

Background: A novel virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has started a global pandemic of pneumonia since December 2019, when first cases were reported in Wuhan, China. It has caused 2.7 million confirmed cases and nearly 200,000 deaths as of April 24, 2020. Early warning systems with new technologies should be established to avoid such disasters.
Objective: This study aimed to early detect the SARS-CoV-2 outbreak in 2019 using social media.
Methods: WeChat Index is a data service that shows how frequent a specific keyword has appeared in posts, subscriptions, and search over a period of last 90 days on WeChat, the most popular Chinese social media. We plotted daily WeChat Index from Nov 17, 2019 to Feb 14, 2020 for keywords related to the SARS-CoV-2 disease.
Results: WeChat Index for “Feidian” that is SARS in Chinese language had stayed at low levels until 16 days ahead of the outbreak announcement by the local authority when the index increased significantly. Later, the index persisted at relative high levels and rose rapidly on the day before the announcement. WeChat Index also spiked or increased for keywords “SARS”, “coronavirus”, “novel coronavirus”, “shortness of breath”, “dyspnea”, and “diarrhea”, but not as meaningful as “Feidian” in early detection of the outbreak.
Conclusions: Using WeChat may detect the SARS-CoV-2 outbreak in 2019 about two weeks earlier than the traditional surveillance systems. WeChat offers a new approach to early detect disease outbreaks.

Keywords: novel coronavirus; SARS; SARS-CoV-2; COVID-19; social media; WeChat; early detection; surveillance

Introduction

An outbreak of pneumonia of unknown cause in Wuhan, the capital of Hubei province of China, occurred in December 2019[1]. Shortly, the cause was identified as a novel coronavirus[1], which resembles severe acute respiratory syndrome coronavirus (SARS-CoV) and is named SARS-Cov-2[2, 3]. The outbreak has become a pandemic, with 2.7 million confirmed cases and nearly 200,000 deaths globally up to April 24, 2020[4]. Early warning systems should be established to avoid disasters like this.

Traditional surveillance systems typically rely on clinical, virological, and microbiological data submitted by physicians and laboratories. Due to time and resource constraints, a lack of operational knowledge of reporting systems, and regulations in these systems, substantial lags between an outbreak event and its report are common[5].

With the popularization of Internet and smartphones, an increasing number of people use social media (e.g. Twitter and Facebook) to share information. An event may have been posted in social media for several days or even months before its report through health institutions and official reporting structures. Internet-based search engines are an important source for health information for people from all walks of life. Analyzing data on search behaviors provides a new approach for detection and monitoring of diseases and symptoms. Technologies using social media, search queries and other Internet resources offer novel and economic approaches for detecting and tracking

emerging diseases and have been successfully used in the cases of SARS[6], influenza[7], dengue[8], etc. Herein, we describe the use of such an approach with WeChat, a Chinese social media, to early detect the SARS-Cov-2 outbreak in China, 2019. Internet search queries from Hubei province were also explored.

Methods

WeChat (called Weixin in China), provided by Tencent Inc., is the largest social media in China with over 1 billion monthly active users. WeChat Index, accessed in WeChat app, is a publicly available data service that shows how frequent a specific keyword has appeared in posts, subscriptions, and search on WeChat over a period of last 90 days. Using WeChat Index, we obtained daily data from Nov 17, 2019 to Feb 14, 2020 for keywords related to the SARS-Cov-2 disease such as “SARS”, “Feidian (that is SARS in Chinese language)”, “pneumonia”, “fever”, “cough”, “shortness of breath”, “dyspnea”, “fatigue”, “stuffy nose”, “runny nose”, “diarrhea”, “coronavirus”, “novel coronavirus”, and “infection” (raw data in Multimedia Appendix 1). All the keywords except for “SARS” were used in Chinese.

Baidu is the dominant Chinese Internet search engine. Baidu Index (https://index.baidu.com), provided by Baidu Inc., can show how frequent a keyword has been queried over a time period from a region. The keywords as mentioned above were also explored through Baidu Index for Hubei province.

The daily data were plotted according to time for each of the keywords. As the outbreak is an isolated rather than recurrent event and the cut-off value to detect an outbreak based on social media and online search behavior is unknown, statistical analysis were not performed. The outbreak was announced by Wuhan Health Commission (WHC) on Dec 31, 2019 (D-day) and on this day Chinese Centers for Disease Control and Prevention (China CDC) involved in investigation and response[2]. If WeChat Index for a keyword spiked or increased before D-day, the index for the keyword was considered as a potential candidate for the outbreak sign[9].

Results

WeChat Index for “Feidian” stayed at low levels before Dec 15, 2019 when it increased significantly. Then, the index persisted at relative high levels till the day before D-day when it rose rapidly and reached a peak on D-day (Figure 1). The index for “SARS” was stable except in the first three days in December with a peak on Dec 1, 2019 (Figure 1). During the period from Nov 17, 2019 to Dec 30, 2019 (44 days), WeChat Index also spiked or increased for “coronavirus”, “novel coronavirus”, “shortness of breath”, “dyspnea”, and “diarrhea”, but not as meaningful as “Feidian” in early detection of the outbreak (Multimedia Appendix 2 and 3).

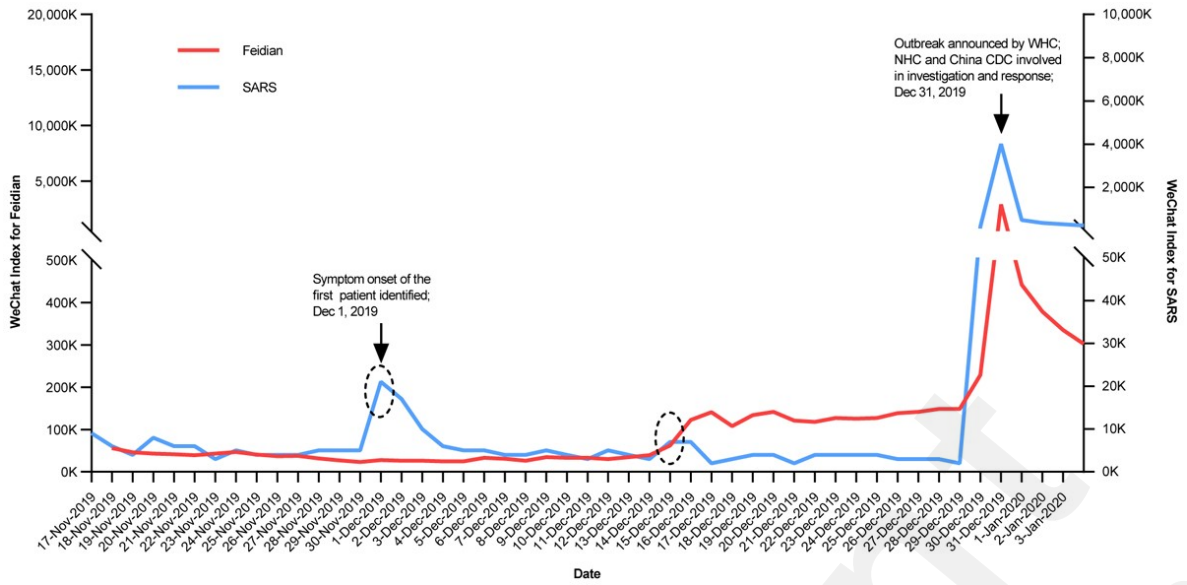


Figure 1. WeChat Index for "Feidian" and "SARS". The index for "Feidian" began to rise on Dec 15, 2019 (dashed circle), persisted at relative high levels till Dec 29, 2019 and rose rapidly on Dec 30, 2019 with a peak on Dec 31, 2019. The index for "SARS" behaved abnormally in the first three days in December with a peak on Dec 1, 2019 (dashed circle). China CDC: Chinese Centers for Disease Control and Prevention; Feidian: Chinese abbreviation for severe acute respiratory syndrome; NCIP: novel coronavirus-infected pneumonia; NHC: National Health Commission of the People's Republic of China; SARS: severe acute respiratory syndrome.

Baidu Index for "Feidian", "SARS", "pneumonia", and "coronavirus" rose rapidly on Dec 30, 2019, the day just before D-day. For other keywords Baidu Index showed no obvious increase during the period from Nov 17, 2019 to Dec 30, 2019 (Multimedia Appendix 4).

Discussion

Principal Results

By exploring daily data from WeChat, a Chinese social media, we found that the frequencies of several keywords related to the SARS-CoV-2 disease behaved abnormally during a period ahead of the outbreak in China, 2019. Of these keywords, "Feidian" is especially worthy of attention. In 2003, the SARS outbreak caused mass panic among people in China and approximately half of the victims were health care workers[10]. Since then, Chinese physicians are alert to SARS as well as similar diseases[11]. If the manifestations and chest images indicate viral pneumonia and similar cases occur in a region during a short time, they naturally think of SARS that is called "Feidian" in Chinese language. When suspected cases admitted to hospitals, involved physicians may mention "Feidian" and communicate in WeChat using this word. As showed in the current study, the frequency of "Feidian" in WeChat began to rise on Dec 15, 2019. According to publications of early cases with laboratory-confirmed SARS-Cov-2 infection, five to eleven patients had symptom onset by this day, with the first onset on Dec 1, 2019[1, 2]. Furthermore, the index for "Feidian" persisted at levels higher than those ahead of Dec 15, 2019 and it dramatically rose to a peak on the day of the outbreak announcement. Altogether, WeChat Index for "Feidian" is a strong warning sign for the SARS-Cov-2 outbreak. Using it may early detect the outbreak ahead of two weeks compared with the traditional surveillance systems in China.

The frequency of "SARS" in WeChat was abnormally high in Dec 1 to 3, 2019 compared to the days ahead and after. According to Huang et al, the symptom onset date of the first patient identified was Dec 1, 2019. It is not clear whether this frequency abnormality has something to do with early cases. If it has, it indicates the existence of cases ahead of the first reported one. The frequency of "novel coronavirus" in WeChat was abnormally high on Dec 11, 2019 with an index value of 400. However,

its baseline value (0 or 50) was very low so that the index was sensitive to noise (Multimedia Appendix 3). As for keywords related to symptoms, these symptoms are not specific to SARS-Cov-2 infection. Their increases in frequency might have some association with the emergence of the disease, or just a coincidence. Though lack inner link or not performed that well compared with “Feidian”, other keywords explored in the study as well as keywords not explored in the study (e.g. the names of drugs treating SARS-CoV-2 infection), may still have value in future outbreak detection and monitoring. Experience from Google Flu Trends showed that a combination of several keywords other than a single keyword predicted better[7].

Gathering and analyzing data from social media, Internet search queries, and web sites represents a novel approach to early warning and detection of disease outbreaks and is a supplementary to traditional surveillance systems[5]. Such a tool, the Global Public Health Intelligence Network (GPHIN), identified the early SARS outbreak in China in 2003 more than two months earlier and first alerted the outbreak of Middle East respiratory syndrome coronavirus (known as MERS-CoV) in 2012[6]. As far as we know, GPHIN and other established tools do not gather data from WeChat, the dominant Chinese social media. The current study shows that gathering and analyzing data from WeChat may be amazing to early detect disease outbreaks. Considering over 1 billion monthly active users in China, WeChat has advantage in detecting outbreaks especially in China. In addition, using WeChat data, as the current study indicated, may perform better than using Baidu search query data because people may communicate first in WeChat in the times of WeChat as a lifestyle of Chinese people[12].

Limitations

The main limitation of the study lies on its retrospective nature. The outbreak is a solitary case. Using WeChat data to early detect outbreaks like this should be valued in the future. In addition, WeChat Index data 90 days ago is unavailable and how the index is calculated is not open.

Conclusions

In summary, using data of WeChat, a Chinese social media, may early detect the SARS-CoV-2 outbreak in China in 2019. If WeChat Index for the keyword “Feidian” were monitored, the outbreak would be detected about two weeks earlier. Future studies can prospectively gather and analyze data from WeChat to early detect disease outbreaks in China. Tracing the source of keywords that behave abnormally in frequencies in WeChat, following rapid response, may become a promising approach to control a disease outbreak at its very early stage.

Acknowledgements

This work was supported by Science and Technology Research and Development Program of Shaanxi Province (2020ZDXM-SF-005).

Conflicts of Interest

None declared.

Abbreviations

China CDC: Chinese Centers for Disease Control and Prevention
GPHIN: the Global Public Health Intelligence Network
MERS-CoV: Middle East respiratory syndrome coronavirus
NCIP: novel coronavirus-infected pneumonia
NHC: National Health Commission of the People's Republic of China
SARS: severe acute respiratory syndrome

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHC: Wuhan Health Commission

Multimedia Appendix 1

Raw data of WeChat Index for keywords related to SARS-CoV-2.

Multimedia Appendix 2

Keywords for which WeChat Index spiked or increased during the period from Nov 17, 2019 to Dec 30, 2019.

Multimedia Appendix 3

WeChat Index curves for keywords related to SARS-CoV-2 except for “Feidian” and “SARS”.

Multimedia Appendix 4

Baidu Index curves for keywords related to SARS-CoV-2.

References

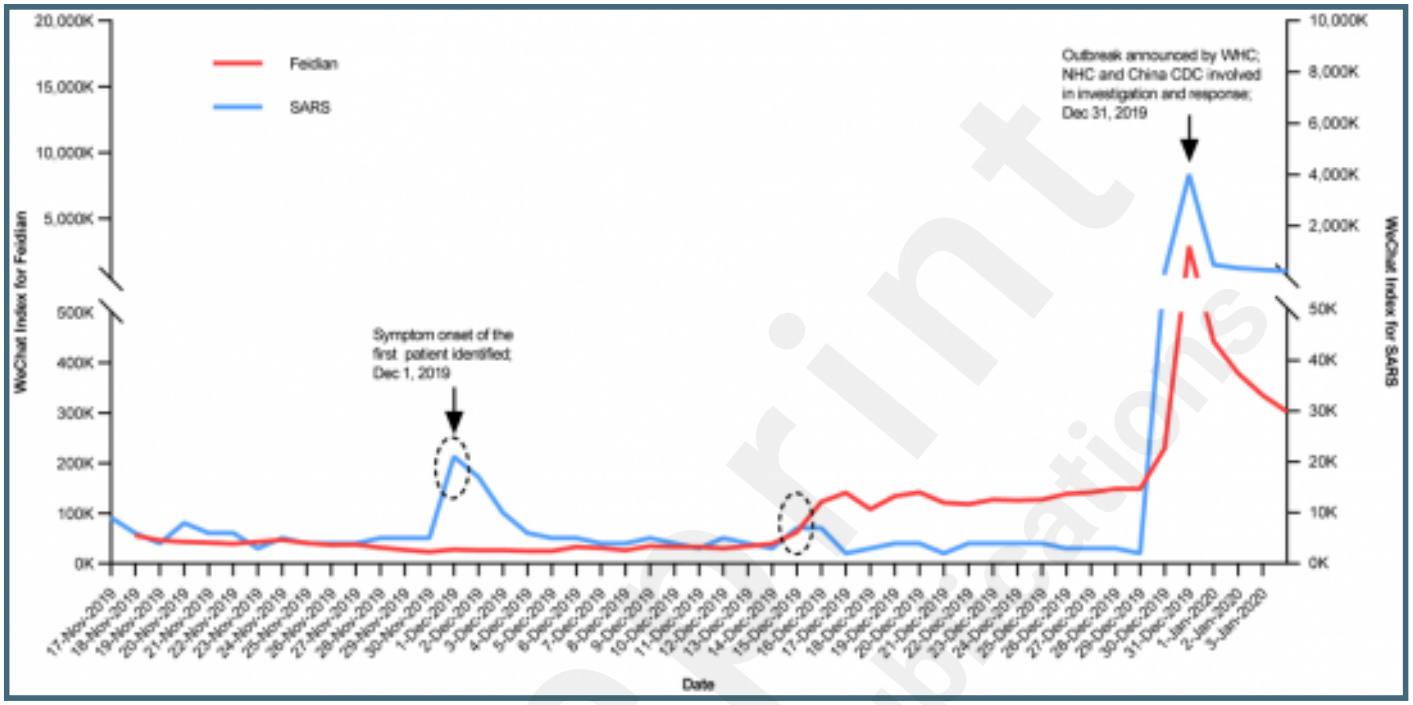
1. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020:S0140-6736(20)30251-8. http://dx.doi.org/10.1016/S0140-6736(20)30251-8
2. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020; 382:727-33. http://dx.doi.org/10.1056/NEJMoa2001017
3. Gorbalenya AE, Baker SC, Baric RS, et al. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. BioRxiv 937862 [Preprint]. February 11, 2020. URL: https://doi.org/10.1101/2020.02.07.937862. [Accessed April 14, 2020]
4. World Health Orgnization. Coronavirus disease (COVID-2019) situation reports. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [Accessed April 26, 2020]
5. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014; 14:160-8. http://dx.doi.org/10.1016/S1473-3099(13)70244-5
6. Dion M, AbdelMalik P, Mawudeku A. Big Data and the Global Public Health Intelligence Network (GPHIN). Can Commun Dis Rep 2015; 41:209-14. http://dx.doi.org/10.14745/ccdr.v41i09a02
7. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009; 457:1012-4. http://dx.doi.org/10.1038/nature07634
8. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 2011; 5:e1206. http://dx.doi.org/10.1371/journal.pntd.0001206
9. Mohsin MFM, Hamdan AR, Bakar AA. Review on anomaly detection for outbreak detection. In: International Conference on Information Science and Managemen (ICoCSIM). North Sumatra, Indonesia: 2012; 22-8.
10. Wenzel RP, Bearman G, Edmond MB. Lessons from severe acute respiratory syndrome (SARS): implications for infection control. Arch Med Res 2005; 36:610-6. http://dx.doi.org/10.1016/j.arcmed.2005.03.040
11. Zhong NS, Zeng GQ. Pandemic planning in China: applying lessons from severe acute

respiratory syndrome. Respirology 2008; 13 Suppl 1:S33-S5. http://dx.doi.org/10.1111/j.1440-1843.2008.01255.x
12. Tu FJ. WeChat and civil society in China. Communication and the Public 2016; 1:343-50. http://dx.doi.org/10.1177/2057047316667518

Supplementary Files

Figures

WeChat Index for "Feidian" and "SARS". The index for “Feidian” began to rise on Dec 15, 2019 (dashed circle), persisted at relative high levels till Dec 29, 2019 and rose rapidly on Dec 30, 2019 with a peak on Dec 31, 2019. The index for “SARS” behaved abnormally in the first three days in December with a peak on Dec 1, 2019 (dashed circle). China CDC: Chinese Centers for Disease Control and Prevention; Feidian: Chinese abbreviation for severe acute respiratory syndrome; NCIP: novel coronavirus-infected pneumonia; NHC: National Health Commission of the People's Republic of China; SARS: severe acute respiratory syndrome.



Multimedia Appendixes

Raw data of WeChat Index for keywords related to SARS-CoV-2.
URL: https://asset.jmir.pub/assets/47f4b5869d623c72af603859ece4ee18.xlsx

Keywords for which WeChat Index spiked or increased during the period from Nov 17, 2019 to Dec 30, 2019.
URL: https://asset.jmir.pub/assets/8fac7c309f2d003730a4dc77f57cb7e3.docx

WeChat Index curves for keywords related to SARS-CoV-2 except for “Feidian” and “SARS”.
URL: https://asset.jmir.pub/assets/c5086ae242e0590999bb5ccf86b1780a.pdf

Baidu Index curves for keywords related to SARS-CoV-2.
URL: https://asset.jmir.pub/assets/b7631e452142472b8351047bad30e4cc.pdf