Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China

---Manuscript Draft---

Abstract:
Background
Using social media surveillance data, this study aimed to assess public attention and awareness, risk perception, emotion, and behavioural response to the COVID-19 outbreak in real time.

Methods
We collected social media data from the three most popular platforms in China: Sina Weibo (microblog), Baidu search engine, and Ali e-commerce marketplace, from the beginning of the outbreak, 1 Dec 2019, to 15 Feb 2020. Quantitative behavioural data including Weibo post counts, Weibo Hot Search ranking, and Baidu searches were used to generate indices assessing public attention and awareness. Public intention and actual adoption of recommended personal protection measures (e.g. hand sanitisers) or panic buying triggered by rumours and misinformation (e.g. garlics) were measured by Baidu and Ali indices. Correlation analysis was performed to detect consistency among the three indices. Qualitative data from Weibo posts were collected and analysed by the Linguistic Inquiry and Word Count (LIWC) text analysis programme to assess public emotion responses to epidemiological events, governments’ announcements, and epidemic control measures.

Findings
We identified two missed windows of opportunity for early epidemic control during the early stages of the COVID-19 outbreak, one in Dec 2019 and the other between 31 Dec and 19 Jan, when public attention and awareness was very low despite the emerging outbreak. Delayed release of information ignited negative public emotions. The public responded quickly to government announcements and adopted recommended behaviours according to issued guidelines. We found rumours and misinformation regarding remedies and cures led to panic buying during the outbreak, and timely clarification of rumours effectively reduced irrational behaviour.

Interpretation
Social media surveillance can enable timely assessments of public reaction to risk communication and epidemic control measures, and the immediate clarification of rumours. This should be fully incorporated into epidemic preparedness and response systems.

Funding
National Natural Science Foundation of China.
Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China

Zhiyuan Hou, PhD,1 Fanxing Du, MSc,1 Hao Jiang, Undergraduate,1 Xinyu Zhou, Undergraduate,1 Leesa Lin, PhD2

Affiliations:
1. School of Public Health, Fudan University, Key Laboratory of Health Technology Assessment, National Health Commission, Shanghai, China
2. Department of Global Health and Development, Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, UK

Correspondence author: Zhiyuan Hou. School of Public Health, Fudan University, Key Laboratory of Health Technology Assessment, National Health Commission, Shanghai, China.
Address: 130 Dong’an Road, Shanghai 200032, China.
Email: zyhou@fudan.edu.cn

Running head: Social media surveillance on public responses to the COVID-19 outbreak in China

Abstract word limit: 299
Word limit: 3466
Summary

Background

Using social media surveillance data, this study aimed to assess public attention and awareness, risk perception, emotion, and behavioural response to the COVID-19 outbreak in real time.

Methods

We collected social media data from the three most popular platforms in China: Sina Weibo (microblog), Baidu search engine, and Ali e-commerce marketplace, from the beginning of the outbreak, 1 Dec 2019, to 15 Feb 2020. Quantitative behavioural data including Weibo post counts, Weibo Hot Search ranking, and Baidu searches were used to generate indices assessing public attention and awareness. Public intention and actual adoption of recommended personal protection measures (e.g. hand sanitisers) or panic buying triggered by rumours and misinformation (e.g. garlics) were measured by Baidu and Ali indices. Correlation analysis was performed to detect consistency among the three indices. Qualitative data from Weibo posts were collected and analysed by the Linguistic Inquiry and Word Count (LIWC) text analysis programme to assess public emotion responses to epidemiological events, governments’ announcements, and epidemic control measures.

Findings

We identified two missed windows of opportunity for early epidemic control during the early stages of the COVID-19 outbreak, one in Dec 2019 and the other between 31 Dec and 19 Jan, when public attention and awareness was very low despite the emerging outbreak. Delayed release of information ignited negative public emotions. The public responded quickly to government announcements and adopted recommended behaviours according to issued guidelines. We found rumours and misinformation regarding remedies and cures led to panic buying during the outbreak, and timely clarification of rumours effectively reduced irrational behaviour.

Interpretation

Social media surveillance can enable timely assessments of public reaction to risk communication and epidemic control measures, and the immediate clarification of rumours. This should be fully incorporated into epidemic preparedness and response systems.

Funding National Natural Science Foundation of China.

Key words: COVID-19, social media, public response, emotion, behaviour, rumour.
Research in context

Evidence before this study

Preparedness experts and the World Health Organisation have urged the integration of social sciences into outbreak response, identified as one of the key research priorities for the response to COVID-19. As a powerful platform for public information sharing and emotional expression, social media’s role in risk communication during emerging infectious diseases outbreaks was first explored during the 2009/2010 A(H1N1) pandemic and again in response to Ebola epidemics. We searched PubMed and EMBASE for all articles published until March, 2020 with the keywords “novel coronavirus,” “CoV,” “COVID-19,” and “SARS-CoV-2.” Only two commentaries called for social media surveillance and managing public panic and rumours with information. No social media studies to date have conducted a comprehensive assessment of public attention and awareness, risk perception, and emotion and behavioural responses as they emerge and evolve in real time during the COVID-19 outbreak.

Added value of this study

Using social media surveillance data, we tracked and assessed real-time public emotional and behavioural responses to government risk communication messages, rumours and misinformation, and major epidemiological events during the first 11 weeks of the COVID-19 outbreak in China. We identified two missed windows of opportunity for early epidemic control in the early stages; we found the delayed release of information ignited negative public sentiment, and that early implementation of containment measures could address public anxiety. Official guidelines, epidemiological information, and rumours can all capture public attention and trigger emotional and behavioural reactions, and timely clarification of rumours can effectively reduce irrational behaviour. This study utilizes rigorous methods and procedures in conducting social media surveillance for COVID-19 with Chinese social media platforms, which can inform future research on social media surveillance for epidemic preparedness and response.

Implications of all the available evidence

During an outbreak, real-time social media surveillance can enable a timely assessment of public reactions to risk communication, epidemic control measures, and the timely clarification of rumours. Social media surveillance has the potential to track both disease activity before it reaches health care facilities and public concerns and risk perceptions as they emerge and evolve in real time. Future research should explore the epidemiological information shared on these platforms and examine the role of social media in early detection and early warning during epidemic preparedness and response. Findings from this study demonstrate that social media surveillance should be fully incorporated into epidemic preparedness and response system.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3551338
Introduction
In early Dec 2019, novel coronavirus disease 2019 (COVID-19) emerged in Wuhan city, and spread rapidly across China. Since 23 Jan 2020, the Chinese government has put Wuhan and several nearby cities under quarantine and implemented containment measures to slow down community transmission. As of 25 Jan, 30 provinces (autonomous regions, municipalities), excluding Tibet, had launched level 1 major public health emergency responses. On 31 Jan (Beijing Time), WHO declared the outbreak a public health emergency of international concern (PHEIC). On 2 Mar, a total of 80,302 cases have been reported in China, and the outbreak has spread to sixty-four other countries with 8,993 cases.

During an epidemic, understanding how critical information about the health threat is disseminated and how the public accesses, processes and uses this information is crucial. The risks and uncertainties of emerging infectious diseases may arouse public emotion and change behaviour in either constructive (e.g. taking up personal hand hygiene and avoiding mass gatherings) or disruptive ways (e.g. inflated public fear, unnecessary anxiety, and socio-economical unease). Governments should attempt to understand public emotions during a public health emergency and relay accurate information to ease the public’s fears. The power of social media is increasingly recognized in public health as a means of disseminating information. Social media surveillance can be used to assess the effectiveness of public risk communication, evaluate public reactions to existing interventions, and alert decision-makers about emerging rumours and misinformation. Furthermore, in contrast with traditional surveys, social media surveillance can systematically monitor public emotions and responses to outbreaks in real-time and is less likely to be affected by recall bias or reporting bias. Despite these benefits, social media surveillance is relatively under-explored, with only a few studies using Twitter and Facebook having been conducted on A(H1N1) and Ebola.

With rapid advancement in mobile and search engine technology, e-commerce since 2010, China has its own social media platforms, which cover 95% of its residents and are a major part of their daily lives; however, there has been very little evidence of these platforms use in large-scale public health emergencies. Using social media surveillance data, this study aimed to assess real-time public attention, risk perception, emotion, and behavioural responses to live events during the first 11 weeks of the COVID-19 outbreak.

Methods

Data Collection

Time period: 1 Dec 2019-15 Feb 2020

Baidu index: Baidu search engine, Google equivalent’s in China, has more than 1 billion Chinese users. The Baidu index, stemming from search frequency on the Baidu search engine, is powered by Baidu statistics and exhibited as pre-set keywords. The Baidu index reflects users’ attention on specific keywords over days, thereby assessing public awareness. For this study, we manually scanned and identified all Baidu index keywords that include COVID-19, recommended personal protection measures, and rumours and misinformation.

The China Centre for Disease Control and Prevention (China CDC) issued a series of guidelines recommending residents adopt individual protection measures consisting of four main aspects: respiratory protection, hand hygiene, home disinfection and health monitoring. Accordingly, we identified "mask," "hand sanitizer," "disinfectant," and "thermometer" as keywords to reflect the four above measures (Table 1). We also identified...
Sina Weibo (aka “microblog”) is a popular social networking site in China with more than 500 million registered users. Similar to Twitter, Sina Weibo is a free social media site that enables registered users to communicate with others in real-time using posts. We programmed a crawler in Python to collect Weibo statuses at the provincial level from the advanced search function of the Sina Weibo website from 1 Dec 2018 to 15 Feb 2020. All Weibo posts gathered were sorted by date and user location (i.e. being either in or out of Hubei province, Appendix Figure 1). A list of Weibo keywords (hashtags ##) about COVID-19 was manually scanned from the first five hundred identified posts and tested for sensitivity and specificity before being used for data crawling. Crawled Weibo posts were used, individually or in combination, to generate various Weibo indices of public attention and awareness on COVID-19, and rumours and misinformation (Table 1, Figures 1 and 5). Furthermore, in order to pinpoint events that triggered public reactions, we referred to Weibo’s “Hot Search” ranking to identify top Weibo topics that drew public attention (Appendix Table 1).

Ali index: Ali platform, Amazon’s equivalent in China, is a powerful and popular internet-based e-commerce marketplace with more than 1 billion users, and its purchasing index reflects the number of specific products purchased online. We employed the same keywords relevant to recommended individual protection measures: "mask," "hand sanitizer," "disinfectant," and "thermometer," to assess public behavioural responses to the COVID-19 outbreak (Understandably, Ali platform did not generate an index for rumour-related items). We similarly collected behavioural data during the current outbreak and in the year prior.

Data analysis

Public attention and awareness, as well as rumours and misinformation, were assessed using both Weibo and Baidu daily indices. To assess public attention and awareness, we constructed a Weibo index using the daily number of “posts” with keywords related to COVID-19, where the proportions of Weibo posts that contained these keywords were calculated and presented on a daily basis. Public emotion was assessed by importing and processing all gathered Weibo posts in the Linguistic Inquiry and Word Count (LIWC) text analysis program.21 LIWC is an internationally-recognised software licensed to analyse texts and output scores for various psychological and other types of features based on narrative content, and has proven reliability and validity as a method for assessing psychological constructs, particularly relating to emotion.21,22 Each feature in LIWC is scored based on specific dictionaries, with words aggregated by linguistic and psychological experts. LIWC computes numerical scores, which count and calculate the percentage of words for categories relating to specific features. We adopted the Chinese version (SCLIWC), which has been previously adapted to the Chinese context and validated in Weibo posts.23,24 In this study, we assessed the elected psycholinguistic features: negative emotion (i.e. anxiety, sad, anger), and risk perception (i.e. drives). Examples of words included in these features are presented in Appendix Table 2. Using the Baidu and Ali daily indices, we assessed the intention and behaviours of adopting recommended personal protection measures and/or to rumours about ineffective treatments during the COVID-19 outbreak.

From the Weibo “Hot Search” ranking (see below Sina Weibo) three keywords: “radix isatidis,” “Shuanghuanglian,” and “garlic,” which were widely circulated rumours and misinformation regarding certain herbal medicines for individual health protection. For comparison, we also gathered Baidu data from last year during the same time-period as our baseline data: 1 Dec 2018-15 Feb 2019.
Finally, we employed Spearman’s rank correlation analyses to detect the consistency of the Weibo, Baidu and Ali indices and to assess the correlation between different indices in different aspects of public response: public attention and awareness, risk perception and negative emotion, and intention and behaviours. We further conducted regression models to estimate the provincial differences in risk perception and various negative emotions across different periods.

**Results**

**Public attention and awareness of COVID-19**

The first COVID-19 case emerged on 8 Dec 2019, which was first announced by the Wuhan Municipal Health Commission on 31 Dec. The Baidu search index for the keywords relating to COVID-19 had been 0 throughout the entire month of December, and jumped to 127,336 searches on 31 Dec 2019, when Wuhan Health Commission first reported a total of 27 patients with pneumonia of unknown cause (Weibo Hot Search #4) (Figure 1). However, public attention and awareness immediately dropped to 30,504 searches the next day, and remained at a low level between 1 and 19 Jan. Starting on 20 Jan, when the National Health Commission confirmed human-to-human transmission (Weibo Hot Search #1), it increased sharply and peaked at 3,039,324 on 25 Jan. It then steadily decreased with minor fluctuations.

The Weibo index presented a similar trend as the Baidu index, with an exceptional peak on 2 Feb, the day before people were scheduled to return to work following the spring festival holiday (Weibo Hot Search #25).

Correlation analysis demonstrated a positive correlation of public attention and awareness between the Baidu and Weibo indices (Spearman’s rank correlation coefficient 0.897, p < 0.001). This figure indicated two missed windows of opportunity for early epidemic control of COVID-19: 1) between 8 Dec 2019 with the first case’s onset, and 31 Dec, when the outbreak was first announced; 2) between 31 Dec and 19 Jan, the date before the confirmation of human-to-human transmission.

**Public risk perception and emotional response**

A total of 2,937,082 Weibo posts were collected, including those in Hubei Province (159,624, 5.4%) and outside of Hubei (2,777,458, 94.6%). The daily proportions of Weibo posts expressing perceived risk and negative emotion are presented in Figure 2. On 31 Dec 2019, 32.0% of Weibo posts started to recognize the risk of this disease, and 29.1% in Hubei province, where Wuhan is located, expressed negative emotions. The proportions of those expressing perceived risk and negative emotion in Hubei increased to 37.0% and 42.4% on 1 Jan 2020, respectively. However, the two proportions held steady around 33% across other provinces in China. On 3 and 5 Jan, the Wuhan Health Commission repeatedly announced “there was no obvious evidence of human-to-human transmission,” and confirmed that “the virus was neither severe acute respiratory syndrome coronavirus (SARS-CoV) nor Middle East respiratory syndrome coronavirus (MERS-CoV)” (Weibo Hot Search #4). Along with these announcements, the proportions of those both expressing perceived risk and negative emotion kept falling across China until 11-13 Jan, when one death from COVID-19 and the first confirmed case in Thailand were reported, respectively (Weibo Hot Search #29 and #7). The death and confirmed case abroad led to high public risk perception about COVID-19, and proportions of perceived risk and negative emotions reached a peak of around 75% and 40%, respectively, across all provinces in China, including Hubei. On 14 Jan, the Wuhan Health Commission reported the first household cluster case, implying the possibility of limited person-to-person transmission, but again announced “no obvious evidence of human-to-human transmission”
(Weibo Hot Search #32). Both proportions declined again, and then slightly rose as the case count increased.

After the confirmation of human-to-human transmission, both proportions kept steady with minor fluctuations. One exception was that both proportions reached another peak on the day after 31 Jan, when WHO declared the outbreak as a PHEIC (Weibo Hot Search #7), and sharply fell to a low-mark on 2 Feb when public attention shifted to discussing whether work should resume the next day as originally planned. Additionally, correlation analysis showed positive correlations between risk perception and negative emotion (Spearman’s rank correlation coefficient 0.500, p <0.001 for Hubei; 0.351, p=0.016 outside of Hubei).

Figure 3 presents a more detailed analysis of the three negative emotions expressed on Weibo: public anxiety, sadness and anger; all reached their peak three times around 1 Jan, 13 Jan, and 6 Feb, and manifested within 24 hours after the triggering event took place. Reports of the first confirmed case abroad triggered more anxiety, sadness and anger outside of Hubei province, and risk perception about imported cases from Hubei rose. On 13 Jan, 18.0% of Weibo posts outside of Hubei were angry with the outbreak, much higher than 8.3% in Hubei. On 6 Feb, Wenliang Li, the whistle-blowing doctor who first warned of the outbreak and who was accused by police of spreading fake news, died from COVID-19. 25 This renewed public expression of negative emotions, especially in Hubei. Around 23% of Weibo posts in Hubei expressed anxiety, sadness and anger in response to Dr. Li’s death, much higher than the 14% of users outside of Hubei.

It is worth noting that the confirmation of human-to-human transmission on 20 Jan marked the beginning of a steady decline in public anxiety; during the same period, public sadness and anger increased, but started to decline after the implementation of containment measures on 23-25 Jan. Wuhan city implemented a series of epidemic control and quarantine measures starting 23 Jan: public transportation in the city was shuttered, and all flights and trains leaving from Wuhan were cancelled (Weibo Hot Search #13). On 25 Jan, 30 provinces (autonomous regions, municipalities) in mainland China, excluding Tibet, activated major public health emergency responses (Weibo Hot Search #11). To assess the impact of the 20 Jan confirmation of human-to-human transmission on public perceived risk and negative emotions, we conducted regressions for the subsamples before and after the communication (Table 2), and found that there was no evidence suggesting differences in public perceived risk and all three negative emotions between residents in Hubei and outside of Hubei before 20 Jan; however, evidence showed Hubei residents were significantly more likely to have higher perceived risk and negative emotions than those outside of Hubei after 20 Jan.

**Behavioural Responses**

China CDC first proposed respiratory protection and hand hygiene in the protection guidelines issued on 21 Jan, and first proposed the use of home disinfection and health monitoring on 22 and 25 Jan, respectively. The public responded quickly to the issuing of specific guidelines, and both intended and actual purchasing behavior dramatically increased accordingly. The Baidu and Ali indices of the four recommended individual protection measures, except for Ali index of thermometer, all started to increase sharply on 21 Jan, with an exceptional drop on 24 Jan (Chinese New Year’s Eve) (Figure 4). The Ali index of thermometer increased rapidly from 25 Jan. The Baidu index indicating an intention to adopt all four measures increased earlier than the actual behaviors presented in the Ali index. Both indices were steadily low during the same period last year (baseline).
Figure 5 shows trends in the Baidu and Weibo indices on rumors, representing intended behavior and public attention, respectively. Radix Isatidis is a traditional Chinese medicine for fever, and its effectiveness is still widely believed in China. The Baidu index of “radix isatidis” started to increase when the outbreak was first announced. It further increased sharply on 20 Jan, when human-to-human transmission was confirmed, and reached a peak on 21 Jan. It started to decline from 21 to 24 Jan, during which the People's Daily issued three reports in an attempt to refute rumors of its effectiveness against COVID-19. On 31 Jan, the People's Daily reported that Shuanghuanglian, another Chinese medicine, could inhibit COVID-19, and both Baidu indices of “Shuanghuanglian” and “radix isatidis” rapidly reached their peaks. Both indices decreased rapidly on 2 Feb, when the People's Daily clarified this misinformation, stating Shuanghuanglian cannot prevent COVID-19. The rumor that “garlic can prevent COVID-19” started to spread on Weibo on 21 Jan, and the Baidu index of “garlic” increased accordingly. It reached a peak on 27 Jan, and declined after 28 Jan, when the People's Daily first refuted rumors about the protective function of garlic. However, these three Baidu indices were still higher than their previously observed levels, even during late Feb. Representing the baseline, these indices had been steadily low during the same period last year. The Weibo index, representing public attention, demonstrated the same pattern as the Baidu index, representing intended behavior, for rumors.

We further conducted correlation analysis (Appendix Table 3), which showed strong positive correlations between the Baidu and Ali indices for all recommended measures (correlation coefficient range of 0.676-0.860, p <0.001) and positive correlations between the Baidu and Weibo indices for all rumors (correlation coefficient range of 0.765-0.916, p <0.001).

**Discussion**

This study is the first social media surveillance analysis to track and assess real-time public emotional and behavioural responses to government announcements, rumours and misinformation, and major epidemiological events during the first 11 weeks of the COVID-19 outbreak in China. Delayed release of information ignited negative public sentiment. Public risk perception and anxiety started to decline after the confirmation of human-to-human transmission, whereas sadness and anger increased until the implementation of containment measures. We found the public responded quickly to government announcements and followed official guidelines accordingly, but rumours and misinformation regarding herbal remedies were found to have led to irrational panic buying during the outbreak. The timely clarification of rumours effectively reduced irrational behaviour.

Compared with traditional surveys, social media surveillance provides data that are real-time, longitudinal, and dynamic in capturing public attention/awareness, risk perception, emotions, intentions, and behavioural reactions, and can be an effective means to assess government’s risk communication efforts and assist in epidemic control and response. This study presented a rigorous methodological process for conducting social media surveillance during epidemic outbreaks that can inform future studies, especially in the Chinese context. There are several limitations. First, there is censorship on Weibo in China, and thusly there was a possibility that we might not have captured all posts. To reduce this bias, we collected data (i.e. Weibo posts) daily, before government intervention. Second, LIWC is a tool that detects emotions using the frequency of words instead of the whole posts, which may lead to misclassification. We adopted its adapted Chinese version, which had been validated locally and can better match with Weibo to reduce misclassification.
The lack of transparent, timely, and effective risk communication from health authorities on an emerging infectious disease in its early stages failed to bring about the appropriate level of public awareness and behavioral responses, such as avoidance of mass gatherings. The government did not provide any actionable advice for personal protection until 21 Jan. There were two missed windows of opportunity for early epidemic control of COVID-19 that we identified during the first 11 weeks of the outbreak: 1) the first COVID-19 case emerged on 8 Dec 2019, more than three weeks before 31 Dec, when the Wuhan Health Commission finally announced the outbreak;2,3 2) between 31 Dec and 19 Jan, the Wuhan Health Commission made four public announcements emphasizing that “there was no obvious evidence for human-to-human transmission” without providing any actionable advice on personal health protection, and the eight physicians who attempted to raise awareness of the outbreak had been warned by the police to keep silent.25 This series of government announcements kept public attention and awareness low, preventing the public from realizing the risk of the disease and from taking personal protection measures at an earlier time. This period overlapped with the celebration of the arrival of 2020 and preparation for the Chinese New Year (21 Jan), the most important period of the year for family reunion and social gathering in China. Instead of raising public awareness about the possible outbreak, on 18 Jan local Wuhan officials and residents held a traditional social gathering, where tens of thousands of residents attended a joint celebration banquet that was promoted on the front-page of local newspapers.26

Despite uncertainty and incomplete information about the risk, the early release of information on emerging infectious diseases and early implementation of containment measures might effectively address public sentiment and control the COVID-19 epidemic and future outbreaks. Our data showed public risk perception and negative emotions were highly correlated to outbreak news events, government announcements, and the implementation of containment measures. Acknowledging the knowns and unknowns about an emerging health threat and taking proactive mitigation measures help ease public fear and build public trust. Our previous studies showed that, although the prevalence of anxiety was high at 33% among Wuhan residents, 92% supported quarantining Wuhan, and 75% were confident that governmental containment measures would quell the outbreak.28-29

By incorporating social media in the emergency response system, health authorities can detect misinformation, confront inaccurate messages in real-time and assess the effectiveness of current epidemic control measures for immediate improvement.30 As an example, during this outbreak, several incidents of panic buying were triggered by both official guidelines and rumours (e.g. masks, Shuanghuanglian, or garlic), and these behaviours were dramatically reduced after clarification. The most popular social media platforms in China, such as WeChat and Weibo, and several health professionals’ online communication platforms (e.g. Dingxiangyuan) established their own platforms to clarify rumours. Working with the private sector to ensure sufficient stock and reasonable pricing for recommended personal protection products, such as hand sanitizer and masks, is critical for strengthening epidemic prevention and control and avoiding public panic. As we collected social media data, we found that the public shared information about their disease symptoms, their health care seeking experiences, and the social/cultural aspects of risks and containment strategies. Social media surveillance has the potential to track both disease activity before it reaches health care facilities and public concerns and risk perceptions as they emerge and evolve in real time.13-15 Future research should explore the epidemiological
information shared on these platforms and examine the role of social media in early detection and early warning
during epidemic preparedness and response.

Conclusion
During an outbreak, real-time social media surveillance can enable a timely assessment of public reactions to
risk communication and epidemic control measures and the timely clarification of rumors; as such, it should be
fully incorporated into epidemic preparedness and response systems.

Contributors
Z.H. and L.L. conceptualized the study design. X.Z., H.J., F.D. collected and analysed data. Z.H., L.L., F.D, H.J.
interpreted results and wrote the manuscript.

Ethics
All data are publicly available.

Declaration of interests
We have no conflict of interest.

Acknowledgements
Z.H. acknowledges financial support from the National Natural Science Foundation of China (No. 71874034).

Role of the funding source
Z.H. acknowledges financial support from the National Natural Science Foundation of China (No. 71874034).
The funders played no part in the study design; in the collection, analysis, and interpretation of data; in the
writing of the report; and in the decision to submit the article for publication. The views expressed in this
publication are those of the authors and not necessarily those of their funders or employers.

Data sharing statement
All data are publicly available.
<table>
<thead>
<tr>
<th>Data sources</th>
<th>Public attention and awareness of COVID-19</th>
<th>Risk and Emotion</th>
<th>Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibo</td>
<td>“武汉肺炎” (Wuhan pneumonia), “武汉不明原因肺炎” (Wuhan unexplained pneumonia), “华南海鲜市场” (Huanan Seafood Market), “华南海鲜城” (Huanan Seafood City), “新型冠状病毒” (Novel Coronavirus), “新冠肺炎” (Novel Coronavirus Pneumonia), “疑似SARS” (Suspicious SARS), and “疑似非典” (Suspicious SARS pneumonia).</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Baidu</td>
<td>“武汉肺炎” (Wuhan pneumonia), “武汉不明原因肺炎” (Wuhan unexplained pneumonia), “华南海鲜市场” (Huanan Seafood Market), “华南海鲜城” (Huanan Seafood City), “新型冠状病毒” (Novel Coronavirus), “新冠肺炎” (Novel Coronavirus Pneumonia), “疑似SARS” (Suspicious SARS), and “疑似非典” (Suspicious SARS pneumonia).</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Ali</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>


This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3551338
Table 2: Association between location and emotional responses separately before and after 20 Jan

<table>
<thead>
<tr>
<th>Location</th>
<th>Risk perception</th>
<th>Negative emotion</th>
<th>Anxiety</th>
<th>Sadness</th>
<th>Anger</th>
<th>Coefficient and standard errors (in parentheses) were obtained from linear regressions. Significance level: *** p&lt;0.01, ** p&lt;0.05, * p&lt;0.1.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subsample before 20 Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hubei provinces</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td></td>
</tr>
<tr>
<td>Hubei province</td>
<td>0.0074</td>
<td>0.0176</td>
<td>0.0271</td>
<td>-0.0009</td>
<td>-0.0150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0409)</td>
<td>(0.0275)</td>
<td>(0.0244)</td>
<td>(0.0007)</td>
<td>(0.0150)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subsample for 20 Jan and after</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hubei provinces</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td>0(ref)</td>
<td></td>
</tr>
<tr>
<td>Hubei province</td>
<td>0.0484**</td>
<td>0.0466***</td>
<td>0.0261**</td>
<td>0.0003*</td>
<td>0.0166**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0194)</td>
<td>(0.0163)</td>
<td>(0.0103)</td>
<td>(0.0002)</td>
<td>(0.0081)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Public attention and awareness of COVID-19 by Baidu and Weibo index
Key events related to COVID-19 outbreak were identified on Weibo “Hot Search” ranking.

Figure 2: The daily proportions of Weibo posts expressing risk and negative emotion

Figure 3: The daily proportions of Weibo posts expressing anxiety, sadness and anger

Figure 4 Trend of Baidu and Ali index on recommended individual protection measures
Ali index can be only available within one year, with no data on Dec 2018 and Jan 2019.

Figure 5 Trend of Baidu and Weibo index on rumors

Appendix Figure 1. Flowchart for Weibo posts collecting process
Ineligible posts: A Python program is designed to exclude status messages that were not considered to be public opinion expression such as news, advertisements, and government announcements, etc.
References


   outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and 

6. Lin L, Savoia E, Agboola F, Viswanath K. What have we learned about communication inequalities during the 

7. Lin L, Jung M, McCloud RF, Viswanath K. Media use and communication inequalities in a public health 
   emergency: a case study of 2009-2010 pandemic influenza A virus subtype H1N1. Public health reports 


   Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the 

   349: g6178.


19. Signorini A, Segre AM, Polgreen PM (2011). The Use of Twitter to Track Levels of Disease Activity and Public 

20. Isaac Chun-Hai Fung et al. Chinese Social Media Reaction to Information about 42 Notifiable Infectious 

   Austin, TX: University of Texas at Austin;2015.


This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3551338