

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

3

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

6

7 **Affiliations:**

8 1. School of Public Health, NHC Key Laboratory of Health Technology Assessment, Fudan University,
9 Shanghai, China

10 2. Department of Global Health and Development, Faculty of Public Health and Policy, London School of
11 Hygiene & Tropical Medicine, UK

12

13 **Correspondence author:** Zhiyuan Hou. School of Public Health, Fudan University, Key Laboratory of Health
14 Technology Assessment, National Health Commission, Shanghai, China.

15 Address: 130 Dong'an Road, Shanghai 200032, China.

16 Email: zyhou@fudan.edu.cn

17

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

19

20 Abstract word limit: 247

21 Word limit: 3500

22

23 **Summary**

24 **Background**

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

27 **Methods**

28 We collected data from most popular social medias: Sina Weibo, Baidu search engine, and Ali e-commerce
29 marketplace, from 1 Dec 2019 to 15 Feb 2020. Weibo post counts and Baidu searches were used to generate
30 indices assessing public attention. Public intention and actual adoption of recommended protection measures or
31 panic buying triggered by rumours and misinformation were measured by Baidu and Ali indices. Qualitative
32 Weibo posts were analysed by the Linguistic Inquiry and Word Count text analysis programme to assess public
33 emotion responses to epidemiological events, governments' announcements, and control measures.

34 **Findings**

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

41 **Interpretation**

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

45 **Funding** National Natural Science Foundation of China.

46

47 **Key words:** COVID-19, social media, public response, emotion, behaviour, rumour.

48

49 **Research in context**

50 **Evidence before this study**

51 Preparedness experts and the World Health Organisation have urged the integration of social sciences into
52 outbreak response, identified as one of the key research priorities for the response to COVID-19. As a powerful
53 platform for public information sharing and emotional expression, social media's role in risk communication
54 during emerging infectious diseases outbreaks was first explored during the 2009/2010 A(H1N1) pandemic and
55 again in response to Ebola epidemics. We searched PubMed and EMBASE for all articles published until 9
56 March, 2020 with the keywords "novel coronavirus," "CoV," "COVID-19," and "SARS-CoV-2." Only two
57 commentaries called for social media surveillance and managing public panic and rumours with information. No
58 social media studies to date have conducted a comprehensive assessment of public attention and awareness, risk
59 perception, and emotion and behavioural responses as they emerge and evolve in real time during the
60 COVID-19 outbreak.

61 **Added value of this study**

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

72 **Implications of all the available evidence**

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

80

81 **Introduction**

82 In early Dec 2019, novel coronavirus disease 2019 (COVID-19) emerged in Wuhan city, and spread rapidly
83 across China.¹⁻³ Since 23 Jan 2020, the Chinese government has put Wuhan and several nearby cities under
84 quarantine and implemented containment measures to slow down community transmission. As of 25 Jan, 30
85 provinces (autonomous regions, municipalities), excluding Tibet, had launched level 1 major public health
86 emergency responses.⁴ On 31 Jan (Beijing Time), WHO declared the outbreak a public health emergency of
87 international concern (PHEIC). On 2 Mar, a total of 80,302 cases have been reported in China, and the outbreak
88 has spread to sixty-four other countries with 8,993 cases.^{4,5}

89

90 During an epidemic, understanding how critical information about the health threat is disseminated and how the
91 public accesses, processes and uses this information is crucial.^{6,7} The risks and uncertainties of emerging
92 infectious diseases may arouse public emotion and change behaviour in either constructive (e.g. taking up
93 personal hand hygiene and avoiding mass gatherings) or disruptive ways (e.g. inflated public fear, unnecessary
94 anxiety, and socio-economical unease).^{7,10} Governments should attempt to understand public emotions during a
95 public health emergency and relay accurate information to ease the public's fears.⁸ The power of social media is
96 increasingly recognized in public health as a means of disseminating information.^{11,12} Social media surveillance
97 can be used to assess the effectiveness of public risk communication, evaluate public reactions to existing
98 interventions, and alert decision-makers about emerging rumours and misinformation.¹³⁻¹⁶ Furthermore, in
99 contrast with traditional surveys, social media surveillance can systematically monitor public emotions and
100 responses to outbreaks in real-time and is less likely to be affected by recall bias or reporting bias.^{13,14} Despite
101 these benefits, social media surveillance during epidemic outbreaks has been under-explored, with only few
102 studies using Twitter and Facebook having been conducted on A(H1N1) and Ebola.^{14, 17-19}

103

104 With rapid advancement in mobile and search engine technology, e-commerce since 2010, China has its own
105 social media platforms, which cover 95% of its residents and are a major part of their daily lives;²⁰ however,
106 there has been very little evidence of these platforms use in large-scale public health emergencies. Using social
107 media surveillance data, this study aimed to assess real-time public attention, risk perception, emotion, and
108 behavioural responses to live events during the first 11 weeks of the COVID-19 outbreak.

109

110 **Methods**

111 **Data Collection**

112 *Time period: 1 Dec 2019-15 Feb 2020*

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

118

119 The National Health Commission issued a series of guidelines recommending residents adopt personal
120 protection measures consisting of four main aspects: respiratory protection, hand hygiene, home disinfection and
121 health monitoring. Accordingly, we identified "mask," "hand sanitizer," "disinfectant," and "thermometer" as
122 keywords to reflect the four above measures (Table 1). We also identified from the Weibo "Hot Search" ranking

123 (see below Sina Weibo) three keywords: “radix isatidis,” “Shuanghuanglian,” and “garlic,” which were widely
124 circulated rumours and misinformation regarding certain herbal medicines for personal health protection. For
125 comparison, we also gathered Baidu data from last year during the same time-period as our baseline data: 1 Dec
126 2018-15 Feb 2019.

127

128 **Sina Weibo** (aka “microblog”) is a popular social networking site in China with more than 500 million
129 registered users. Similar to Twitter, Sina Weibo is a free social media site that enables registered users to
130 communicate with others in real-time using posts. We programmed a crawler in Python to collect Weibo
131 statuses at the provincial level from the advanced search function of the Sina Weibo website from 1 Dec 2019 to
132 15 Feb 2020, collecting total of 6,216,572 Weibo posts (Appendix Figure 1). All posts gathered were sorted by
133 date and user location (i.e. being either in or out of Hubei province). A list of Weibo keywords (hashtags##)
134 about COVID-19 was manually scanned from the first five hundred identified posts and tested for sensitivity
135 and specificity before being used for data crawling. Crawled Weibo posts were used, individually or in
136 combination, to generate various Weibo indices of public attention and awareness on COVID-19, and rumours
137 and misinformation (Table 1, Figures 1 and 5). Furthermore, in order to pinpoint events that triggered public
138 reactions, we referred to Weibo’s “Hot Search” ranking to identify top Weibo topics that drew public attention
139 (Appendix Table 1).

140

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

147

148 **Data analysis**

149 **Public attention and awareness**, as well as **rumours and misinformation**, were assessed using both Weibo
150 and Baidu daily indices. To assess public attention and awareness, we constructed a Weibo index using the daily
151 number of “posts” with keywords related to COVID-19, where the proportions of Weibo posts that contained
152 these keywords were calculated and presented on a daily basis. **Public emotion** was assessed by importing and
153 processing all gathered Weibo posts in the Linguistic Inquiry and Word Count (LIWC) text analysis program,
154 ^{21,22} for assessing psychological constructs, particularly relating to emotion. We adopted the Chinese version
155 (SCLIWC), which has been previously adapted to the Chinese context and validated in Weibo posts.^{23,24} In this
156 study, we assessed the elected psycholinguistic features: negative emotion (i.e. anxiety, sad, anger), and risk
157 perception (i.e. drives). Examples of words included in these features are presented in Appendix Table 2. Using
158 the Baidu and Ali daily indices, we assessed the **intention and behaviours** of adopting recommended personal
159 protection measures and/or rumours about ineffective treatments during the COVID-19 outbreak.

160

161 Finally, we employed Spearman’s rank correlation analyses to detect the consistency of the Weibo, Baidu and
162 Ali indices and to assess the correlation between different indices in different aspects of public response: public
163 attention and awareness, risk perception and negative emotion, and intention and behaviours. We further

164 conducted regression models to estimate the provincial differences in risk perception and various negative
165 emotions across different periods.

166

167 **Results**

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

169 The first COVID-19 case emerged on 8 Dec 2019, which was first announced by the Wuhan Municipal Health
170 Commission on 31 Dec. The Baidu search index for COVID-19 had no results throughout the entire month of
171 December, and jumped to 127,336 searches on 31 Dec 2019, when Wuhan Health Commission first reported a
172 total of 27 patients with pneumonia of unknown cause (Weibo Hot Search #4). However, public attention and
173 awareness immediately dropped to 30,504 searches the next day, and remained low between 1 and 19 Jan.
174 Starting on 20 Jan, when the National Health Commission confirmed human-to-human transmission (Weibo Hot
175 Search #1), it increased sharply and peaked at 3,039,324 on 25 Jan. It then steadily decreased with minor
176 fluctuations. Figure 1 highlighted the squandered windows of opportunity for early epidemic control.

177

178 The Weibo index presented a similar trend as the Baidu index, with an exceptional peak on 2 Feb, the day
179 before people were scheduled to return to work following the spring festival holiday (Weibo Hot Search #25).
180 Correlation analysis demonstrated a positive correlation of public attention and awareness between the Baidu
181 and Weibo indices (Spearman's rank correlation coefficient 0.897, $p < 0.001$).

182

183 ***Public risk perception and emotional response***

184 A total of 2,937,082 Weibo posts were collected, including those in Hubei Province (159,624, 5.4%) and outside
185 of Hubei (2,777,458, 94.6%). The daily proportions of Weibo posts expressing perceived risk and emotion (both
186 positive and negative) are presented in Figure 2 and because public awareness of COVID-19 was low before 20
187 Jan, we observed strong fluctuations in the three indicators during this period caused by a small data size (many
188 days Weibo posts < 1000).

189

190 On 31 Dec 2019, 32.0% of Weibo posts ($n=4,543$) in Hubei province, where Wuhan is located, started to
191 recognize the risk of this disease, and 29.1% expressed negative emotions. At the same time, 47% and 52% of
192 Weibo posts in Hubei and other provinces, respectively, expressed positive emotions in celebration of the New
193 Year 2020. Positive emotion has been consistently higher than negative emotion throughout the entire study
194 period in Hubei and across China. On 3 and 5 Jan, the Wuhan Health Commission repeatedly announced "there
195 was no obvious evidence of human-to-human transmission," and confirmed that "the virus was neither severe
196 acute respiratory syndrome coronavirus (SARS-CoV) nor Middle East respiratory syndrome coronavirus
197 (MERS-CoV)" (Weibo Hot Search #4). Along with these announcements, the proportions of those expressing
198 perceived risk and negative emotion showed steady reduction while positive emotion fluctuated across China for
199 several days, signalling a period of confusion. On 11-13 Jan when one death from COVID-19 and the first
200 confirmed case in Thailand were reported, respectively (Weibo Hot Search #29 and #7), the proportions of posts
201 that expressed perceived risk and negative emotions reached a peak of around 75% and 40%, respectively,
202 across all provinces in China. We observed a strong presence of public risk perception between the first death
203 reported on 11 Jan and the confirmation of human-to-human transmission on 20 Jan. Since then, positive
204 emotion took over and remained high in the 47-68% range in Hubei and 40-63% across China. One exception
205 was that all three indicators merged on the day after 31 Jan, when WHO declared the outbreak as a PHEIC

206 (Weibo Hot Search #7). On 2 Feb when public attention shifted to discussing whether work should resume the
207 next day as originally planned, risk perception and negative emotion plunged sharply to a low-mark and
208 remained low thereafter. Correlation analysis showed positive correlations between risk perception and negative
209 emotion (Spearman's rank correlation coefficient 0.500, $p < 0.001$ for Hubei; 0.351, $p = 0.016$ outside of Hubei).

210

211 Figure 3 presents a more detailed analysis of the three negative emotions expressed on Weibo: public anxiety,
212 sadness and anger; all reached their peak three times around 1 Jan, 13 Jan, and 6 Feb, and manifested within 24
213 hours after the triggering event took place. Reports of the first confirmed case abroad triggered more anxiety,
214 sadness and anger outside of Hubei province, and risk perception about imported cases from Hubei rose. On 13
215 Jan, 18.0% of Weibo posts outside of Hubei were angry with the outbreak, much higher than 8.3% in Hubei. On
216 6 Feb, Wenliang Li, the whistle-blowing doctor who first warned of the outbreak and who was accused by
217 police of spreading fake news, died from COVID-19.²⁵ This renewed public expression of negative emotions,
218 especially in Hubei. Around 23% of Weibo posts in Hubei expressed anxiety, sadness and anger in response to
219 Dr. Li's death, much higher than the 14% of users outside of Hubei.

220

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

233

234 ***Behavioural Responses***

235 National Health Commission first proposed respiratory protection and hand hygiene in the protection guidelines
236 issued on 21 Jan, and first proposed the use of home disinfection and health monitoring on 22 and 25 Jan,
237 respectively. The public responded quickly to the issuing of specific guidelines, and both intended and actual
238 purchasing behavior dramatically increased accordingly. The Baidu and Ali indices of the four recommended
239 personal protection measures, except for Ali index of thermometer, all started to increase sharply on 21 Jan, with
240 an exceptional drop on 24 Jan (Chinese New Year's Eve) (Figure 4). The Ali index of thermometer increased
241 rapidly from 25 Jan. The Baidu index indicating an intention to adopt all four measures increased earlier than
242 the actual behaviors presented in the Ali index. Both indices were steadily low during the same period last year
243 (baseline).

244

245 Figure 5 shows trends in the Baidu and Weibo indices on rumors, representing intended behavior and public
246 attention, respectively. Radix Isatidis is a traditional Chinese medicine for fever, and its effectiveness is still
247 widely believed in China. The Baidu index of "radix isatidis" started to increase when the outbreak was first

248 announced. It further increased sharply on 20 Jan, when human-to-human transmission was confirmed, and
249 reached a peak on 21 Jan. It started to decline from 21 to 24 Jan, during which the People's Daily issued three
250 reports in an attempt to refute rumors of its effectiveness against COVID-19. On 31 Jan, the People's Daily
251 reported that Shuanghuanglian, another Chinese medicine, could inhibit COVID-19, and both Baidu indices of
252 "Shuanghuanglian" and "radix isatidis" rapidly reached their peaks. Both indices decreased rapidly on 2 Feb,
253 when the People's Daily clarified this misinformation, stating Shuanghuanglian cannot prevent COVID-19. The
254 rumor that "garlic can prevent COVID-19" started to spread on Weibo on 21 Jan, and the Baidu index of "garlic"
255 increased accordingly. It reached a peak on 27 Jan, and declined after 28 Jan, when the People's Daily first
256 refuted rumors about the protective function of garlic. However, these three Baidu indices were still higher than
257 their previously observed levels, even during late Feb. Representing the baseline, these indices had been steadily
258 low during the same period last year. The Weibo index, representing public attention, demonstrated the same
259 pattern as the Baidu index, representing intended behavior, for rumors.

260

261 We further conducted correlation analysis (Appendix Table 3), which showed strong positive correlations
262 between the Baidu and Ali indices for all recommended measures (correlation coefficient range of 0.676-0.860,
263 $p < 0.001$) and positive correlations between the Baidu and Weibo indices for all rumors (correlation coefficient
264 range of 0.765-0.916, $p < 0.001$).

265

266 **Discussion**

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

276

277 Compared with traditional surveys, social media surveillance provides data that are real-time, longitudinal, and
278 dynamic in capturing public attention/awareness, risk perception, emotions, intentions, and behavioural
279 reactions, and can be an effective means to assess government's risk communication efforts and assist in
280 epidemic control and response. This study presented a rigorous methodological process for conducting social
281 media surveillance during epidemic outbreaks that can inform future studies, especially in the Chinese context.
282 There are several limitations. First, there is censorship on Weibo in China, and thusly there was a possibility that
283 we might not have captured all posts. To reduce this bias, we collected data (i.e. Weibo posts) daily, before
284 government intervention. Second, LIWC is a tool that detects emotions using the frequency of words instead of
285 the whole posts, which may lead to misclassification. We adopted its adapted Chinese version, which had been
286 validated locally and can better match with Weibo to reduce misclassification. Third, there is an inherent bias in
287 social media studies where data might misrepresent the real world because users might present themselves
288 differently online and/or represent a skewed population towards the young.

289

290 The lack of transparent, timely, and effective risk communication from health authorities on an emerging
291 infectious disease in its early stages failed to bring about the appropriate level of public awareness and
292 behavioral responses, such as avoidance of mass gatherings. The government did not provide any actionable
293 advice for personal protection until 21 Jan. There were two missed windows of opportunity for early epidemic
294 control of COVID-19 that we identified during the first 11 weeks of the outbreak: 1) the first COVID-19 case
295 emerged on 8 Dec 2019, more than three weeks before 31 Dec, when the Wuhan Health Commission finally
296 announced the outbreak;¹⁻³ 2) between 31 Dec and 19 Jan, the Wuhan Health Commission made four public
297 announcements emphasizing that “there was no obvious evidence for human-to-human transmission” without
298 providing any actionable advice on personal health protection, and the eight physicians who attempted to raise
299 awareness of the outbreak had been warned by the police to keep silent.²⁵ This series of government
300 announcements kept public attention and awareness low, preventing the public from realizing the risk of the
301 disease and from taking personal protection measures at an earlier time. This period overlapped with the
302 celebration of the arrival of 2020 and preparation for the Chinese New Year (25 Jan), the most important period
303 of the year for family reunion and social gathering in China. Instead of raising public awareness about the
304 possible outbreak, on 18 Jan local Wuhan officials and residents held a traditional social gathering, where tens
305 of thousands of residents attended a joint celebration banquet that was promoted on the front-page of local
306 newspapers.²⁶

307
308 Despite uncertainty and incomplete information about the risk, the early release of information on emerging
309 infectious diseases and early implementation of containment measures might effectively address public
310 sentiment and control the COVID-19 epidemic and future outbreaks. Our data showed public risk perception
311 and negative emotions were highly correlated to outbreak news events, government announcements, and the
312 implementation of containment measures. Acknowledging the knowns and unknowns about an emerging health
313 threat and taking proactive mitigation measures help ease public fear and build public trust. Our previous studies
314 showed that, although the prevalence of anxiety was high at 33% among Wuhan residents, 92% supported
315 quarantining Wuhan, and 75% were confident that governmental containment measures would quell the
316 outbreak.²⁸⁻²⁹ Interestingly, the proportion of Weibo posts that expressed positive emotion were significantly
317 higher than negative emotion throughout the entire study period, especially after the implementation of
318 restrictive public health measures. Looking closer at the data showed the positive emotion was boosted by a
319 strong sense of solidarity for Wuhan and for health care workers who worked at the frontline. The timing of the
320 lockdown coincided with the Chinese New Year with unexpected extended family time also lifting public spirits.

321
322 By incorporating social media in the emergency response system, health authorities can detect misinformation,
323 confront inaccurate messages in real-time and assess the effectiveness of current epidemic control measures for
324 immediate improvement.³⁰ As an example, during this outbreak, several incidents of panic buying were
325 triggered by both official guidelines and rumours (e.g. masks, Shuanghuanglian, or garlic), and these behaviours
326 were dramatically reduced after clarification. The most popular social media platforms in China, such as
327 WeChat and Weibo, and several health professionals’ online communication platforms (e.g. Dingxiangyuan)
328 established their own platforms to clarify rumours. Working with the private sector to ensure sufficient stock
329 and reasonable pricing for recommended personal protection products, such as hand sanitizer and masks, is
330 critical for strengthening epidemic prevention and control and avoiding public panic. As we collected social
331 media data, we found that the public shared information about their disease symptoms, their health care seeking

332 experiences, and the social/cultural aspects of risks and containment strategies. Social media surveillance has
333 the potential to track both disease activity before it reaches health care facilities and public concerns and risk
334 perceptions as they emerge and evolve in real time.¹³⁻¹⁵ Future research should explore the epidemiological
335 information shared on these platforms and examine the role of social media in early detection and early warning
336 during epidemic preparedness and response.

337

338 **Conclusion**

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

342

343 **Contributors**

344 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.
345 interpreted results and wrote the manuscript. All authors approved the final draft of the manuscript.

346

347 **Ethics**

348 All data are publicly available.

349

350 **Declaration of interests**

351 We declare no competing interests.

352

353 **Acknowledgements**

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

355

356 **Role of the funding source**

357 The funders played no part in any process of the study.

358 The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the
359 report. The corresponding author had full access to all the data in the study and had final responsibility for the
360 decision to submit for publication.

361

362 **Table 1: Keywords to measure public attention, risk perception, emotion, and behaviour**

Data sources	Public attention and awareness of COVID-19	Risk and Emotion	Behaviour	
			Recommended protection measures	Rumours
Weibo	“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).	NA	NA	“radix isatidis”, “Shuanghuanglian”, and “garlic”
Baidu	“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).	NA	"mask", "hand sanitizer", "disinfectant", and "thermometer"	“radix isatidis”, “Shuanghuanglian”, and “garlic”
Ali	NA	NA	"mask", "hand sanitizer", "disinfectant", and "thermometer"	NA

Dates: 1 December, 2019 to 15 February, 2020.

363
364

365 **Table 2: Association between location and emotional responses separately before and after 20 Jan**

	Risk perception	Positive emotion	Negative emotion	Anxiety	Sadness	Anger
Subsample before 20 Jan						
Location						
Non-Hubei provinces	0(ref)	0(ref)	0(ref)	0(ref)	0(ref)	0(ref)
Hubei province	0.0074 (0.0409)	0.0040 (0.0273)	0.0176 (0.0275)	0.0271 (0.0244)	-0.0009 (0.0007)	-0.0150 (0.0150)
Subsample for 20 Jan and after						
Location						
Non-Hubei provinces	0(ref)	0(ref)	0(ref)	0(ref)	0(ref)	0(ref)
Hubei province	0.0484** (0.0194)	0.0431** (0.0163)	0.0466*** (0.0163)	0.0261** (0.0103)	0.0003* (0.0002)	0.0166** (0.0081)

366 Coefficient and standard errors (in parentheses) were obtained from linear regressions. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

367

368 **Figure legend**

369

370 **Figure 1: Public attention and awareness of COVID-19 by Baidu and Weibo index**

371 Notes: Key events related to COVID-19 outbreak were identified on Weibo “Hot Search” ranking. PHER:
372 public health emergency responses; PHEIC: public health emergency of international concern.

373

374 **Figure 2: Daily proportions of Weibo posts expressing risk perception and emotions**

375 Notes: PHER: public health emergency responses; PHEIC: public health emergency of international concern.

376

377 **Figure 3: Daily proportions of Weibo posts expressing anxiety, sadness and anger**

378 Notes: PHER: public health emergency responses; PHEIC: public health emergency of international concern.

379

380 **Figure 4 Trend of Baidu and Ali index on recommended personal protection measures**

381 Notes: Ali index can be only available within one year, with no data on Dec 2018 and Jan 2019.

382

383 **Figure 5 Trend of Baidu and Weibo index on rumors**

384

385 **Appendix Figure 1. Flowchart for Weibo posts collecting process**

386 Notes: Ineligible posts: A Python program is designed to exclude status messages that were not considered to be
387 public opinion expression such as news, advertisements, and government announcements, etc.

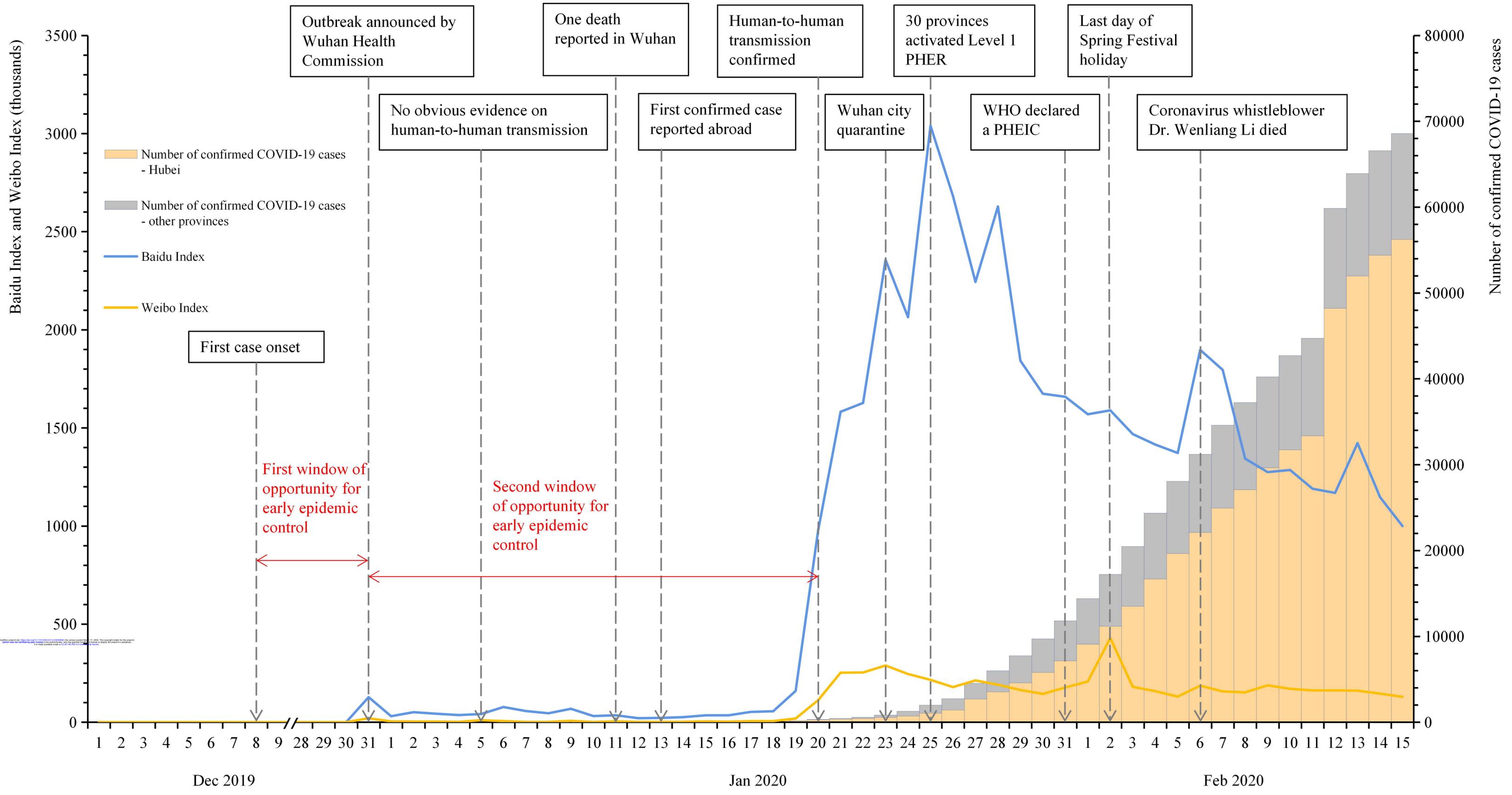
388

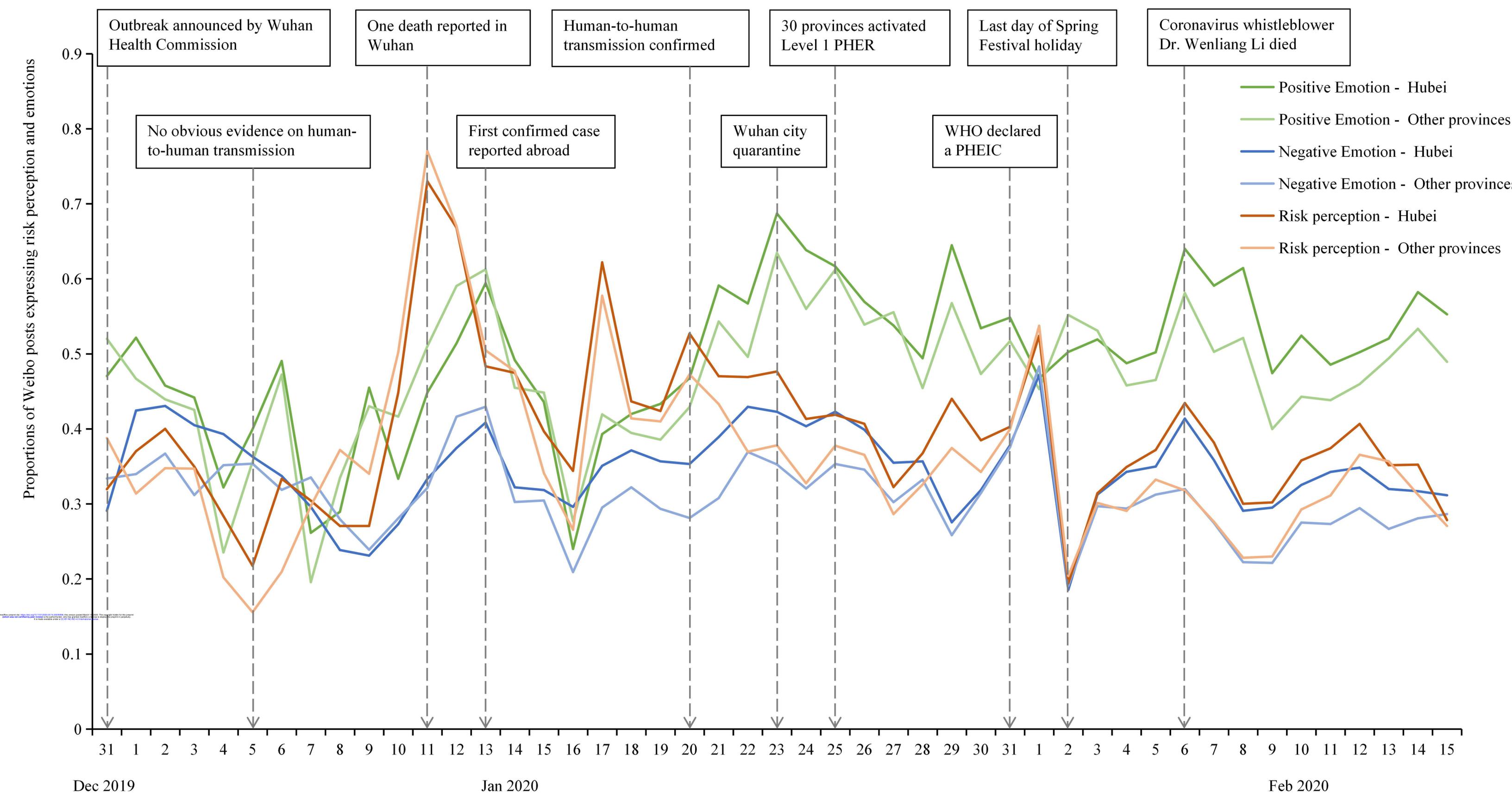
389

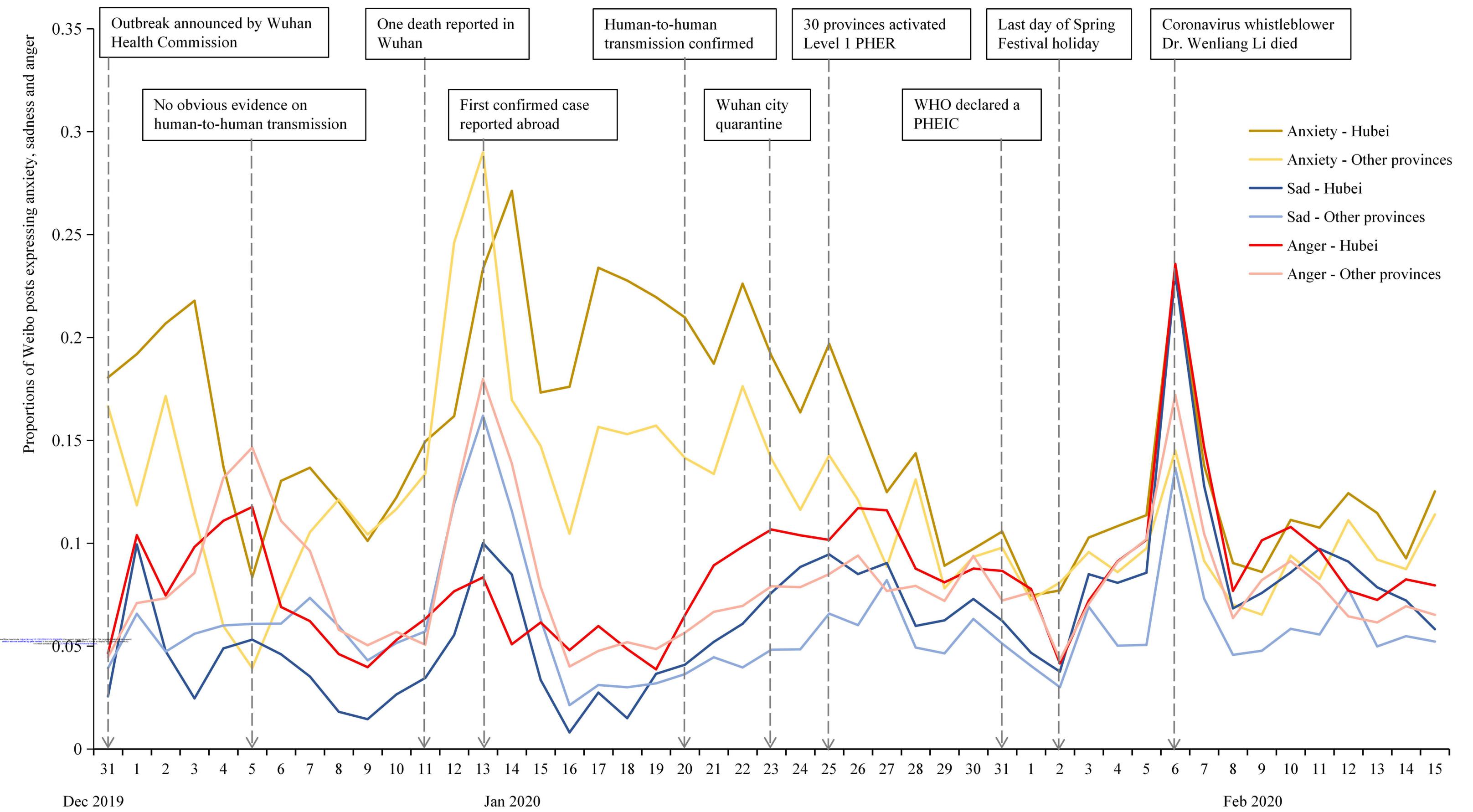
390 **References**

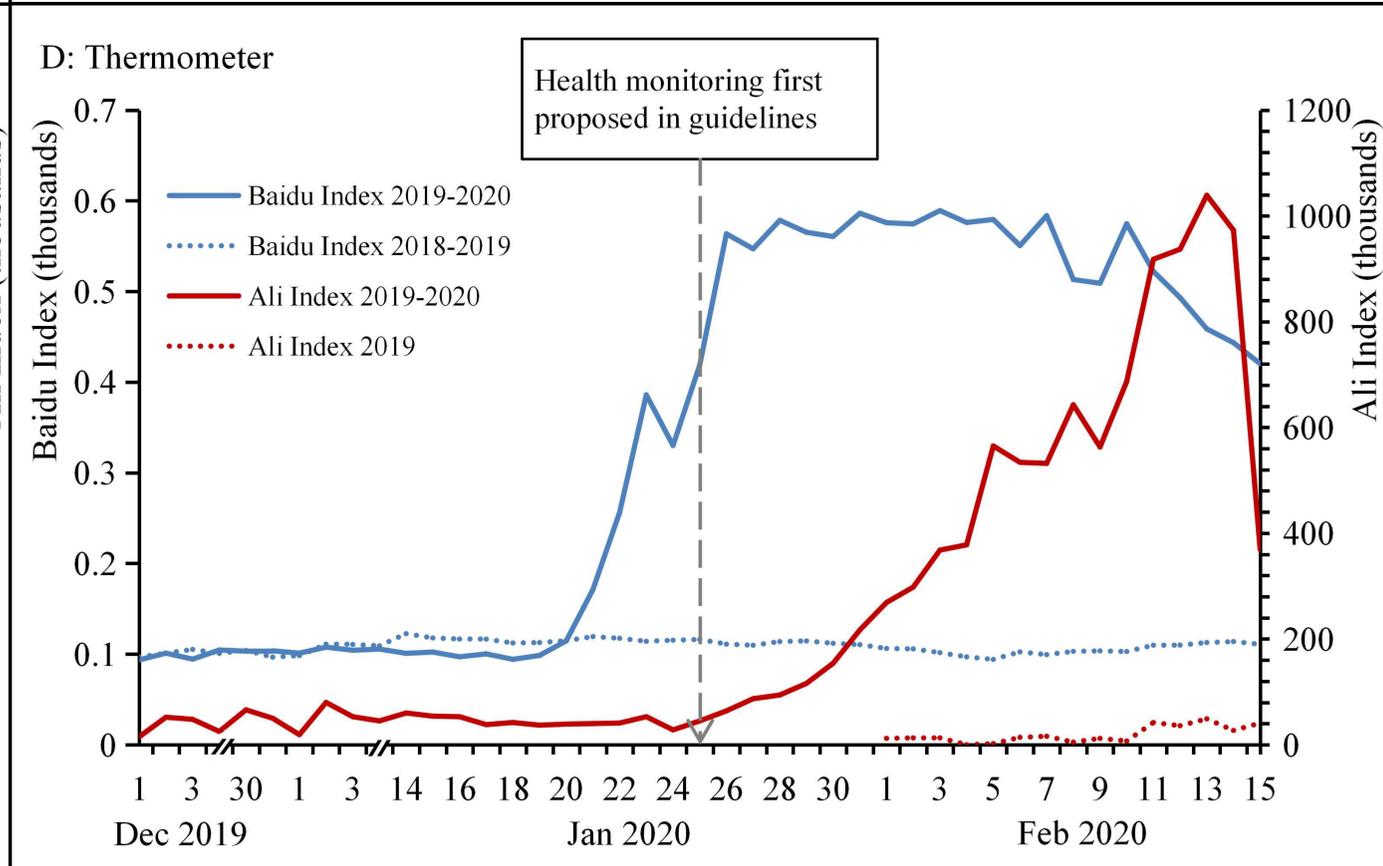
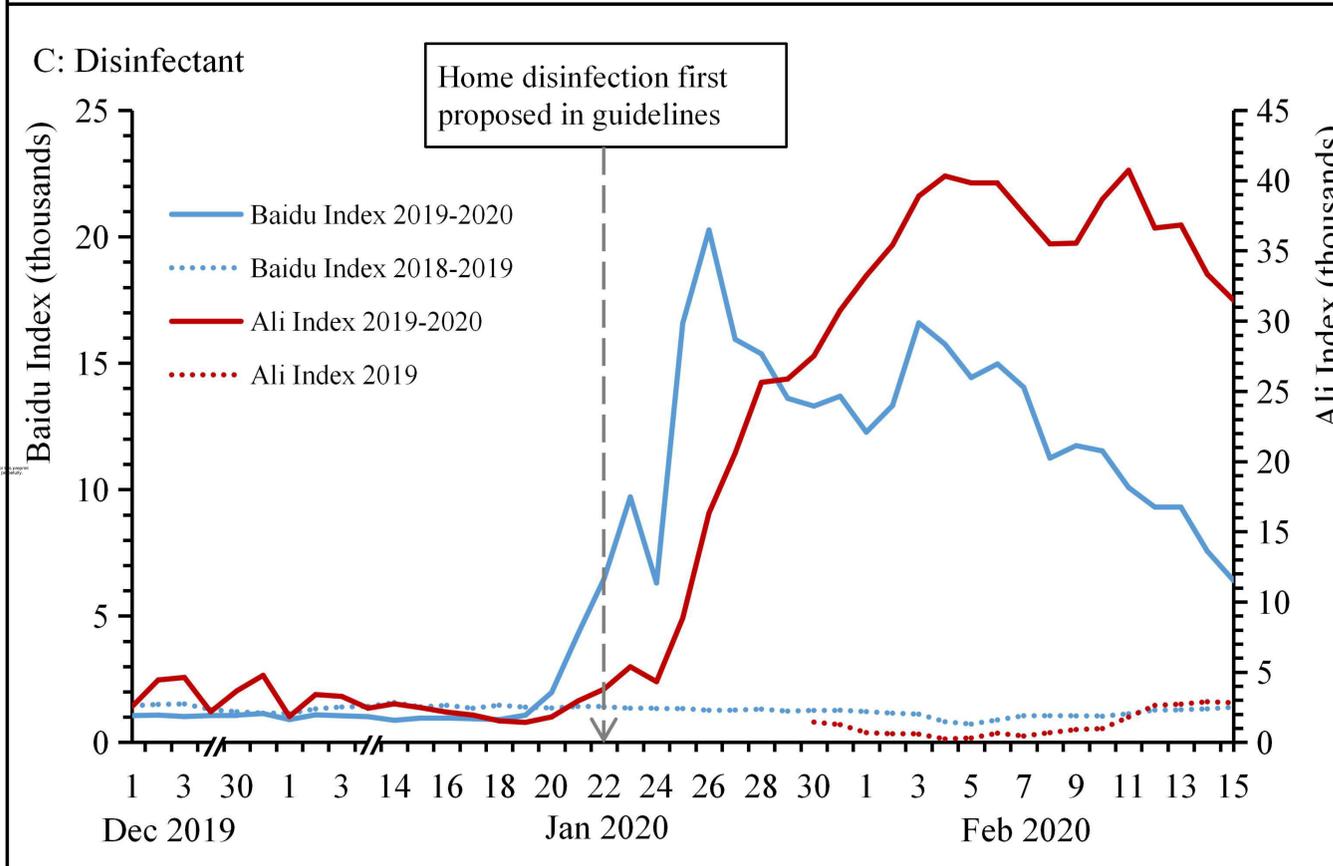
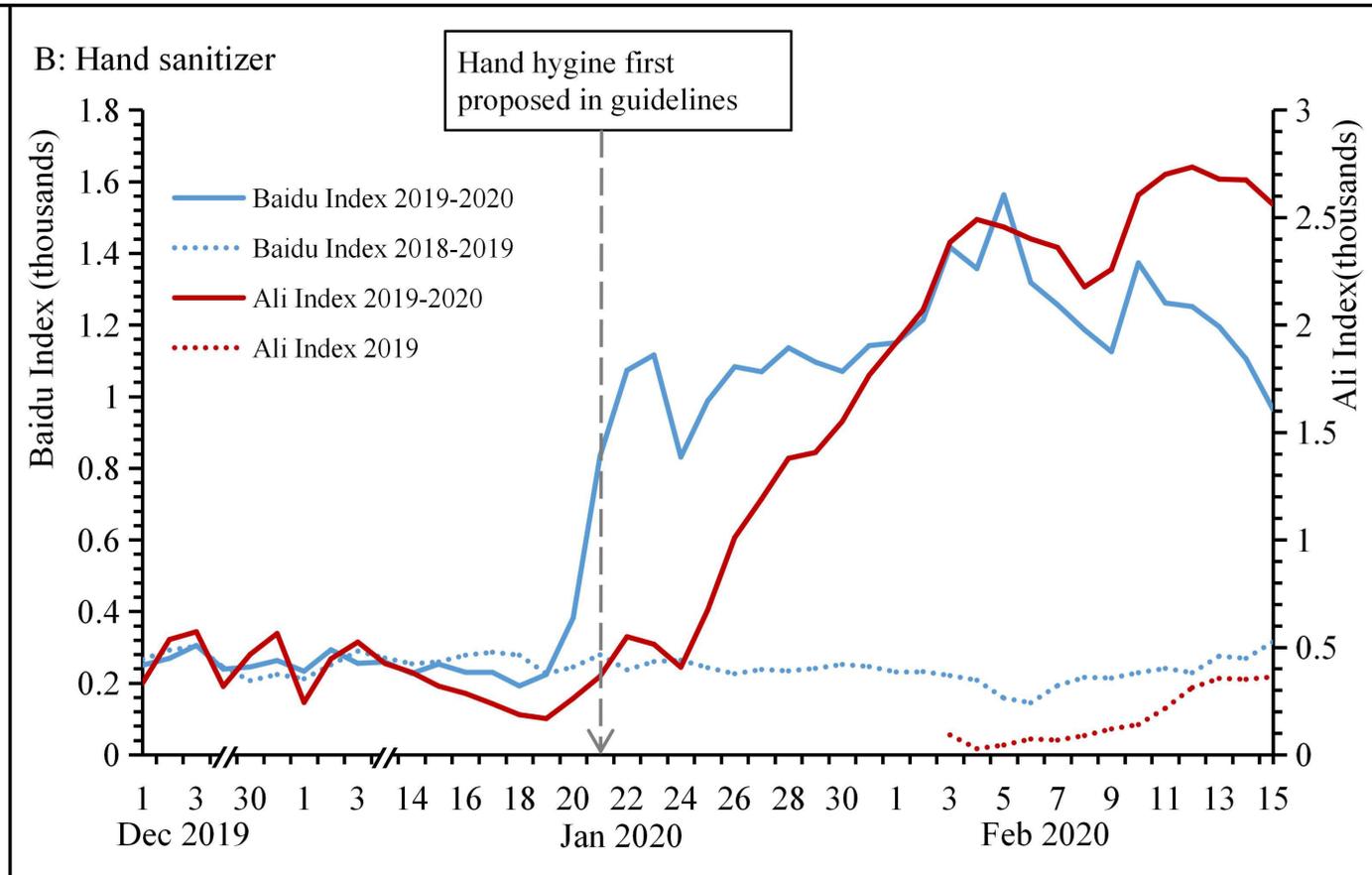
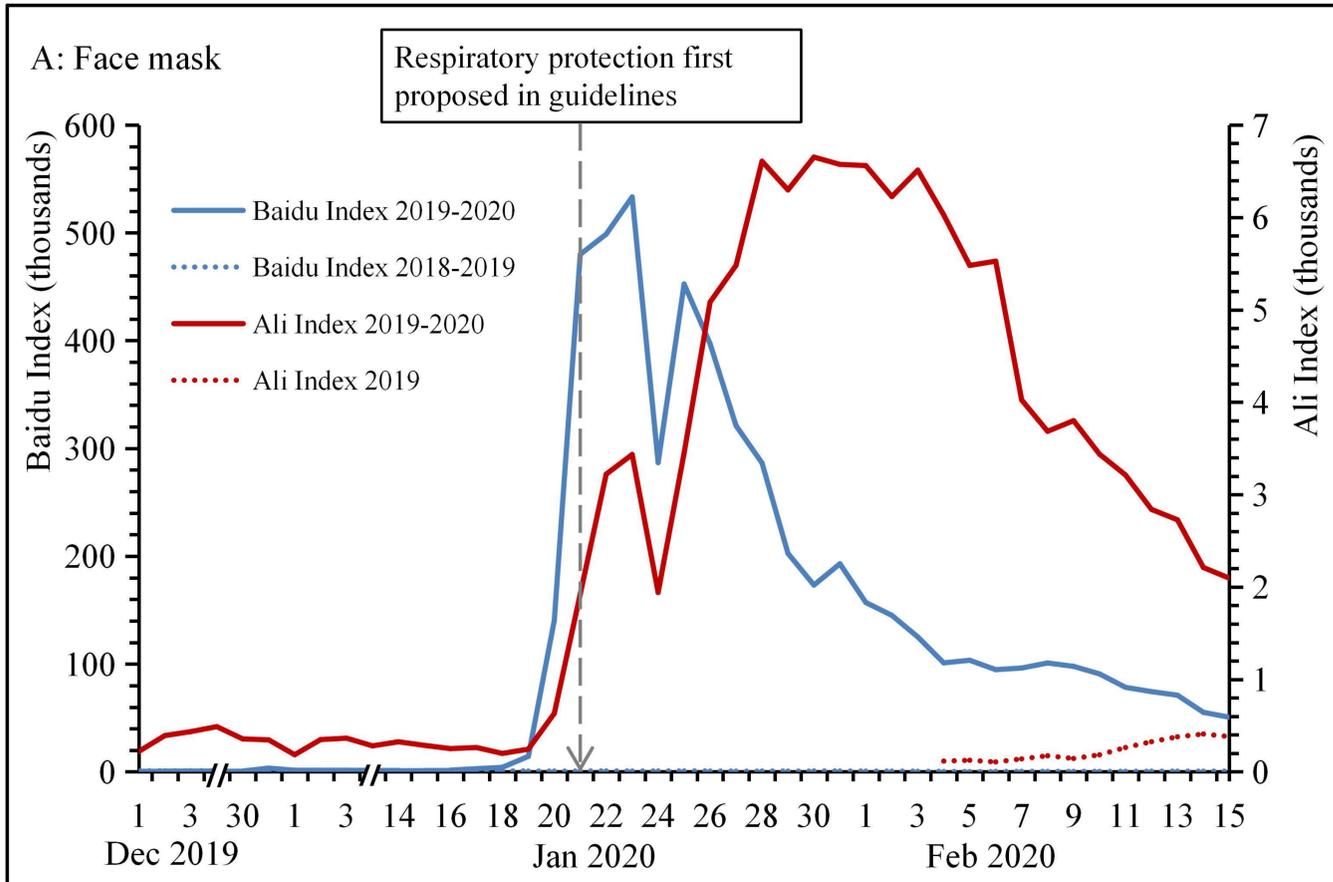
- 391 1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.
392 *Lancet* 2020; **395**(10223): 497-506.
- 393 2. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel
394 coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395**(10223): 507-513.
- 395 3. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected
396 Pneumonia. *N Engl J Med* 2020.
- 397 4. National Health Commission of the People's Republic of China. <http://www.nhc.gov.cn>. (accessed Mar 2, 2020).
- 398 5. Wu Z, McGoogan JM. Characteristics of and important lessons from the Coronavirus Disease 2019 (COVID-19)
399 outbreak in China: summary of a report of 72□314 cases from the Chinese Center for Disease Control and
400 Prevention. *JAMA* 2020.
- 401 6. Lin L, Savoia E, Agboola F, Viswanath K. What have we learned about communication inequalities during the
402 H1N1 pandemic: a systematic review of the literature. *BMC public health* 2014; **14**: 484-8.
- 403 7. Lin L, Jung M, McCloud RF, Viswanath K. Media use and communication inequalities in a public health
404 emergency: a case study of 2009-2010 pandemic influenza A virus subtype H1N1. *Public health reports*
405 (*Washington, DC : 1974*) 2014; **129** Suppl 4(Suppl 4): 49-60.
- 406 8. The Lancet. COVID-19: fighting panic with information. *Lancet* 2020; **395**(10224): 537.
- 407 9. Brooks SK, Webster RK, Smith LE, et al. The psychological impact of quarantine and how to reduce it: rapid
408 review of the evidence. *Lancet* 2020. pii: S0140-6736(20)30460-8.
- 409 10. John Sumo et al. Risk and Outbreak Communication: Lessons from Taiwan's Experiences in the Post-SARS Era.
410 *Health Secur* 2017; **15**(2): 165-169.
- 411 11. St Louis C, Zorlu G. Can Twitter predict disease outbreaks? *BMJ* 2012; **344**: e2353.
- 412 12. The Lancet. The medium and the message of Ebola. *Lancet* 2014; **384**: 1641.
- 413 13. Gabriel J Milinovich, Gail M Williams, Archie C A Clements, Wenbiao Hu. Internet-based surveillance systems
414 for monitoring emerging infectious diseases. *Lancet Infect Dis* 2014; **14**(2): 160-8.
- 415 14. Tang L, Bie B, Park SE, Zhi D. Social media and outbreaks of emerging infectious diseases: A systematic
416 review of literature. *Am J Infect Control* 2018; **46**(9): 962-972.
- 417 15. Mollema L, Harmsen IA, Broekhuizen E, et al. Disease Detection or Public Opinion Reflection? Content
418 Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the
419 Netherlands in 2013. *J MED INTERNET RES* 2015; **17**(5): e128.
- 420 16. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ* 2014;
421 **349**: g6178.
- 422 17. Fung IC, Duke CH, Finch KC, Snook KR, Tseng PL, Hernandez AC, Gambhir M, Fu KW, Tse ZTH. Ebola
423 virus disease and social media: A systematic review. *Am J Infect Control* 2016; **44**(12): 1660-1671.
- 424 18. Fung IC, Tse ZT, Cheung CN, Miu AS, Fu KW. Ebola and the social media. *Lancet* 2014; **384**(9961): 2207.
- 425 19. Signorini A, Segre AM, Polgreen PM (2011). The Use of Twitter to Track Levels of Disease Activity and Public
426 Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLOS ONE* 2011; **6**(5): e19467.
- 427 20. Isaac Chun-Hai Fung et al. Chinese Social Media Reaction to Information about 42 Notifiable Infectious
428 Diseases. *PLOS ONE* 2015; **10**(5): e126092.
- 429 21. Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015,
430 Austin, TX: University of Texas at Austin;2015.
- 431 22. Tausczik Y, Pennebaker J. The psychological meaning of words: LIWC and computerized text analysis methods.

- 432 *J Lang Soc Psychol* 2010; **29**:24–54.
- 433 23. 黃金蘭, Chung C K, Hui N, et al. 中文版「語文探索與字詞計算」詞典之建立. *中華心理學刊* 2012,
- 434 **54**(2):185-201.
- 435 24. Zhao N, Jiao D, Bai S, et al. Evaluating the Validity of Simplified Chinese Version of LIWC in Detecting
- 436 Psychological Expressions in Short Texts on Social Network Services. *PLOS ONE* 2016, **11**(6):1-15.
- 437 25. West China Metropolis Daily. From "rumor maker" to "respectable person" Doctor Li Wenliang's 2020 New
- 438 Year. <http://news.jxntv.cn/2020/0203/9329454.shtml/> (accessed Feb 3, 2020).
- 439 26. Chutian Metropolis Daily. More than 40,000 families gathered for a group dinner to celebrate Chinese New
- 440 Year at Baibuting community in Wuhan.
- 441 https://ctdsbepaper.hubeidaily.net/pc/column/202001/19/node_A03.html/ (accessed Jan 19, 2020).
- 442 27. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, Jiang Z, Zheng D, Jiang J. Importance of Internet surveillance
- 443 in public health emergency control and prevention: evidence from a digital epidemiologic study during avian
- 444 influenza A H7N9 outbreaks. *J Med Internet Res* 2014; **16**(1):e20.
- 445 28. Hou Z, Lin L, Lu L, et al. Public Exposure to Live Animals, Behavioural Change, and Support in Containment
- 446 Measures in response to COVID-19 Outbreak: a population-based survey in China. *medRxiv*
- 447 2020.02.21.20026146.
- 448 29. The Lancet Infectious Diseases. Challenges of coronavirus disease 2019. *Lancet Infect Dis* 2020;20(3):261.
- 449 30. Depoux A, Martin S, Karafillakis E, Bsd RP, Wilder-Smith A, Larson H. The pandemic of social media panic
- 450 travels faster than the COVID-19 outbreak. *J Travel Med.* 2020. pii: taaa031. doi: 10.1093/jtm/taaa031.

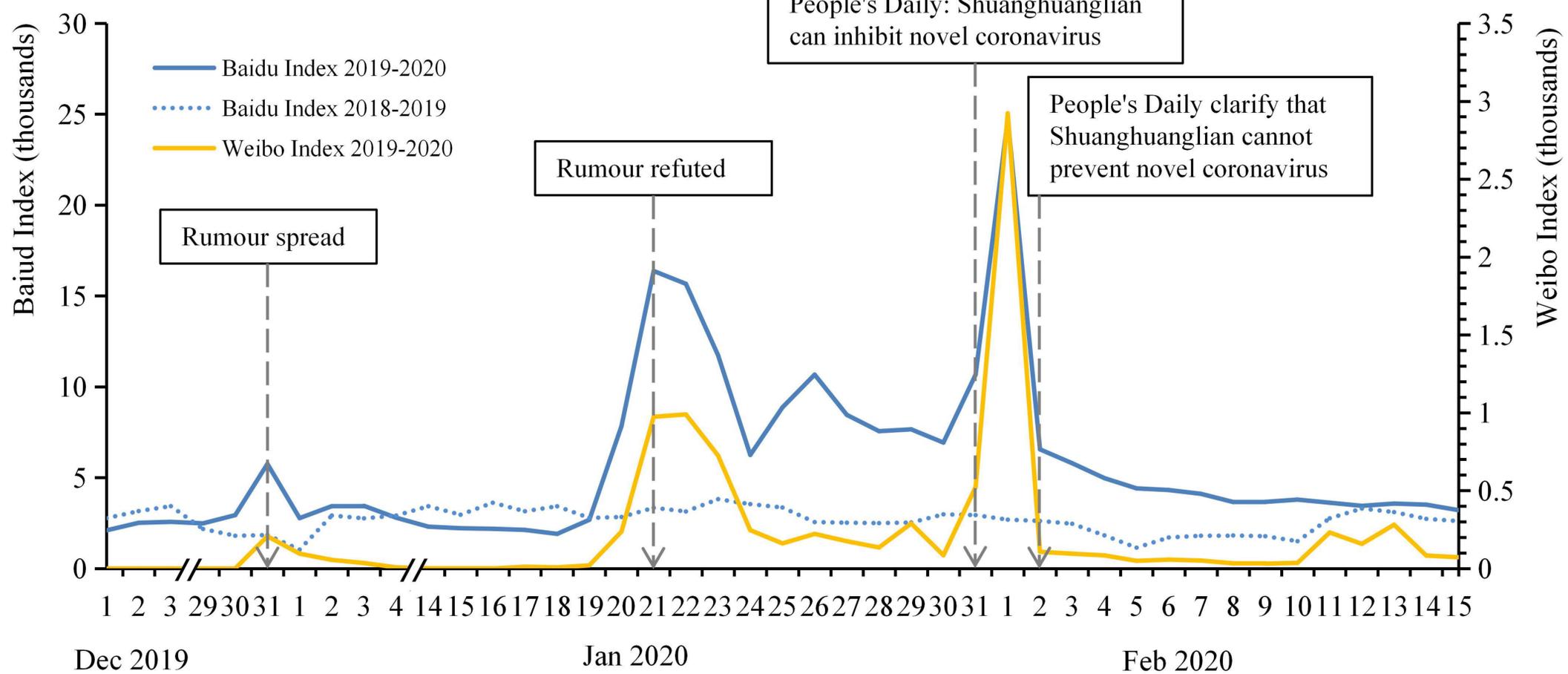




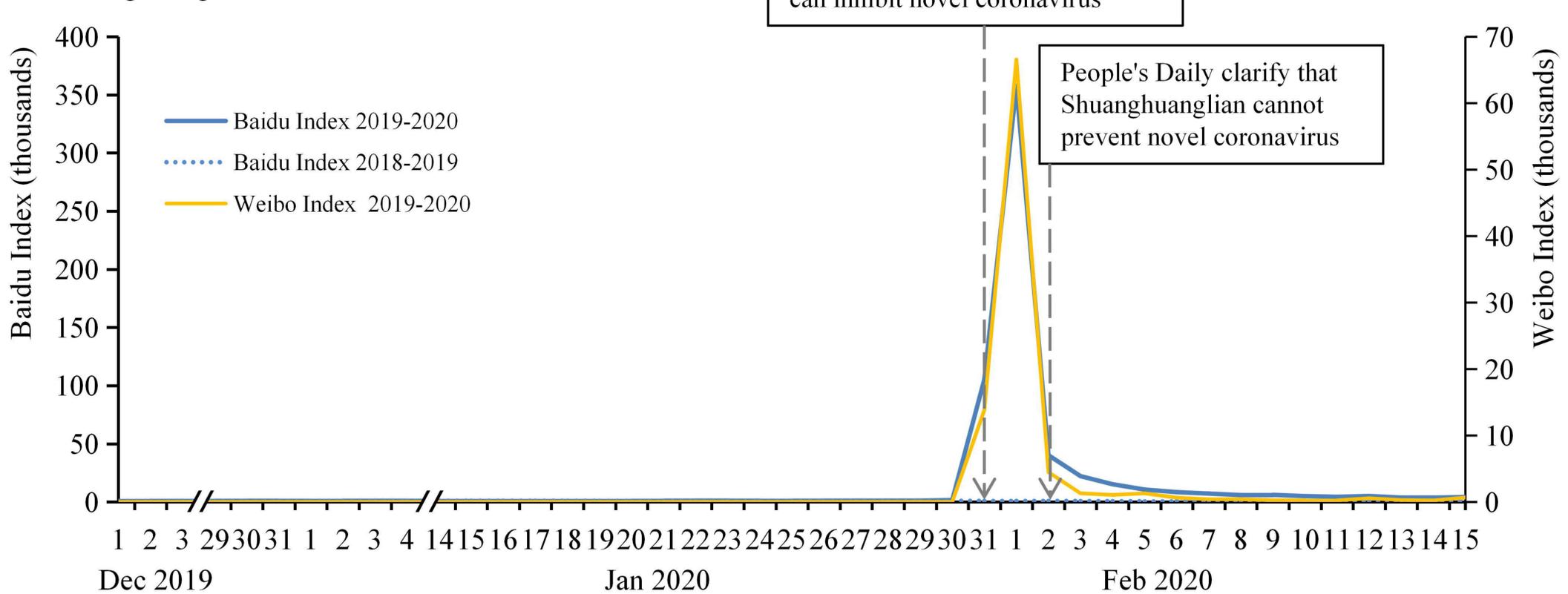




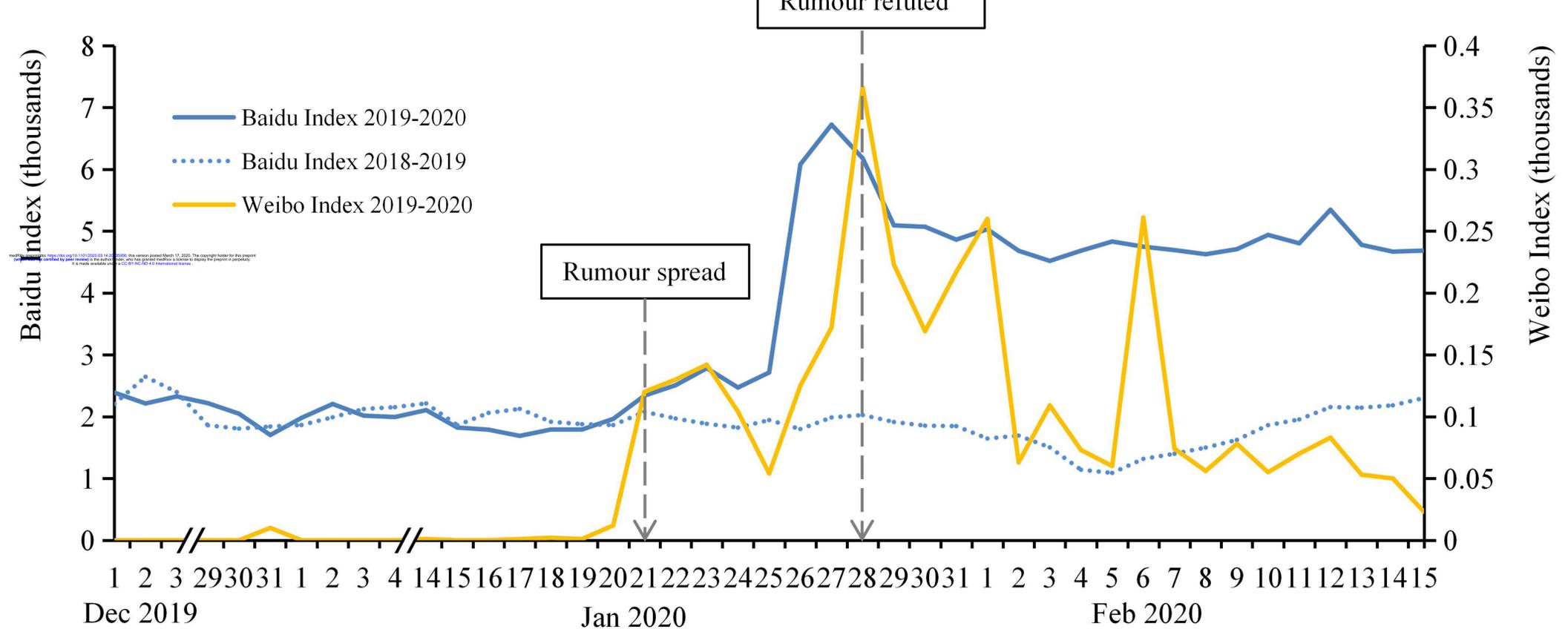
A: Radix isatidis - Chinese medicine



B: Shuanghuanglian - Chinese medicine



C: Garlic



Weibo posts relating
to COVID-19
from 1 Dec 2019 to
15 Feb 2020
n=6,216,572

Duplicates
n=962,333
(19.6%)

Unique posts
n=5,254,239 (80.4%)

Ineligible posts
n=2,317,157
(44.1%)

Eligible posts
n=2,937,082 (55.9%)

Posts from
other provinces
n=2,777,458
(94.6%)

Posts from
Hubei Province
n=159,624
(5.4%)