Early characteristics of the COVID-19 outbreak predict the subsequent epidemic scope

Lei Zhang\textsuperscript{a,b,c,d,*}, Yusha Tao\textsuperscript{a,b}, Jing Wang\textsuperscript{a}, Jason J. Ong\textsuperscript{a,b,c}, Weiming Tang\textsuperscript{e}, Maosheng Zou\textsuperscript{a}, Lu Bai\textsuperscript{a}, Miao Ding\textsuperscript{a}, Mingwang Shen\textsuperscript{a}, Guihua Zhuang\textsuperscript{a,**}, Christopher K. Fairley\textsuperscript{a,b,c}

\textsuperscript{a} China-Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi’an Jiaotong University Health Science Center, Xi’an, Shaanxi, 710061, PR China
\textsuperscript{b} Melbourne Sexual Health Centre, Alfred Health, Melbourne, Australia
\textsuperscript{c} Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC, Australia
\textsuperscript{d} Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou 450001, Henan, China
\textsuperscript{e} University of North Carolina Project-China Dermatology Hospital, Southern Medical University, Guangzhou 510085, China

\textbf{Objective(s)}: The mostly-resolved first wave of the COVID-19 epidemic in China provided a unique opportunity to investigate how the initial characteristics of the COVID-19 outbreak predict its subsequent magnitude.

Methods: We collected publicly available COVID-19 epidemiological data from 436 Chinese cities from 16th January–15th March 2020. Based on 45 cities that reported >100 confirmed cases, we examined the correlation between early-stage epidemic characteristics and subsequent epidemic magnitude.

Results: We identified a transition point from a slow- to a fast-growing phase for COVID-19 at 5.5 (95\% CI, 4.6–6.4) days after the first report, and 30 confirmed cases marked a critical threshold for this transition. The average time for the number of confirmed cases to increase from 30 to 100 (time from 30–to–100) was 6.6 (5.3–7.9) days, and the average case-fatality rate in the first 100 confirmed cases (CFR-100) was 0.8\% (0.2–1.4\%). The subsequent epidemic size per million population was significantly associated with both of these indicators. We predicted a ranking of epidemic size in the cities based on these two indicators and found it highly correlated with the actual classification of epidemic size.

Conclusions: Early epidemic characteristics are important indicators for the size of the entire epidemic. © 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

The recent outbreak of the coronavirus SARS-COV-2 has led to a worldwide pandemic (Cucinotta and Vanelli, 2020) with substantial social, health, and economic costs (UN News, 2020). The epidemic spread quickly from Wuhan to neighboring cities and then to the rest of China (Wu et al., 2020), possibly exacerbated by the ’travel rush’ for the Lunar New Year. On 23rd January 2020, the Chinese government initiated an unprecedented move of introducing a ’metropolitan-wide quarantine’ of the city of Wuhan, by terminating all public transportation in the city and intercity links (Chen et al., 2020; Zhang et al., 2020b). Within days of implementing this quarantine, an additional 12 major cities in Hubei province were similarly quarantined (Sina news, 2020). Within a week of the quarantine, all 31 provinces of mainland China initiated the highest level of public health emergency response (Chinese Center for Disease Control and Prevention, 2020). The strict control nationwide has been highly effective (Chinazzi et al., 2020; Pan et al., 2020; Tian et al., 2020), with the daily reported confirmed cases significantly reduced from 3000 to 4000 at its peak to 10 cases or less a day in mid–March (World Health Organization, 2020a). The COVID-19 epidemic had evolved into a worldwide pandemic. As of 24th May, over 5,200,000 cases have been reported in over 200 countries. The epicenter has moved from China to Europe, to the United States, and now to South America (World Health Organization, 2020a).
The epidemic in China preceded epidemics in other countries and has now mostly resolved (Kupferschmidt and Cohen, 2020). The Chinese data provide a unique opportunity to understand how the early characteristics of an epidemic may predict its subsequent magnitude. This may provide a useful guide for understanding the development of the COVID-19 outbreaks in cities in other global settings.

We hypothesized that the early characteristics of an outbreak would be good indicators to forecast the epidemic’s subsequent size. Previous studies demonstrated that early epidemiological indicators, such as the basic reproduction ratio, may determine the peak of an epidemic and the epidemic level in the long term (Holme and Masuda, 2015; Ridenhour et al., 2014). This study aims to investigate the association between the early characteristics of the COVID-19 epidemic and the magnitude of the epidemic assessed in its late stages. This knowledge could provide public health policymakers with an indication of the likely size of an epidemic and, therefore, the urgency of control measures that may be required.

Methods

Data source

We collected publicly available data from 436 Chinese cities, inclusive of prefectures and municipalities, that reported on cases of COVID-19 (number of confirmed cases, deaths, and recovered cases) from 16th January to 15th March 2020, at which time the epidemic in China had mostly resolved. Our primary data source was Dingxiangyuan (DXY) (https://ncov.dxy.cn/), an online platform built by Chinese medical community members, which integrated COVID-19 case information from both local media and government reports. We included a total of 45 cities with more than 100 confirmed cases in our analysis. In contrast, 391 cities reported less than 100 cases over the two months; the outbreaks in these cities were presumably controlled by the nationwide emergency public health response. These cities were not included in our analysis. The analysis was conducted in 31 provinces of mainland China, excluding the two Chinese special administrative regions (Macau and Hong Kong) and Taiwan. This is because these regions were under a different epidemic surveillance system, and the outbreaks had undergone a significant growth after 15th March. We also excluded the city of Jining, Shandong province, because most of its reported cases were due to an outbreak in a prison. For included cities, we obtained the population size of each of these cities from the statistical yearbooks of these cities or their respective provinces.

Selection of early and outcome indicators

For each of the 45 Chinese cities, we used Joinpoint software (https://surveillance.cancer.gov/joinpoint/) to identify the trend and transition point of the epidemic during the initial phase of the outbreak based on the first 100 confirmed cases. We used a maximum of one joinpoint (corresponding to two time intervals); a two-phase fit can be successfully determined through the Joinpoint software automatically (National Cancer Institute, 2020). We identified: (1) the time of the transition point between two phases; (2) the number of cases at the transition point; the growth rates of the (3) first (slow-growing) phase and (4) the second (fast-growing) phase. For each model’s calibration, we estimated the sum of squared errors (SSE) and mean squared error (MSE) for the fitness of Joinpoint models (Table S2). Datasets with less than six data points were automatically fitted with a single-phase model to avoid over-parameterization. We found that the majority of transition points occurred below 30 cases (Table 1, Table S1) and hence regarded 30 cases as a critical threshold for epidemic growth where the epidemic changed from a slow-growing to a fast-growing phase. We also estimated three additional predictors based on the first 100 confirmed cases, namely:

[1] The days required to increase from 30 to 100 cases (time from 30-to-100),
[2] The case-fatality rate among the first 100 confirmed cases (CFR-100), and
[3] The case recovery rate among the first 100 confirmed cases.

The ‘first 100 cases’ was taken as the number of confirmed cases on the day the 100th confirmed case was reported.

We defined the outcome indicator as the epidemic size per million population, which was the cumulative number of confirmed cases on 15th March divided by the population size then multiplied by one million in each of the corresponding Chinese cities.

Statistical analysis

We categorized the Chinese cities into three groups: (1) Wuhan city, (2) 15 neighboring cities of Wuhan in Hubei province, and (3) 29 cities in the rest of the country. We compared the indicators between two groups using the nonparametric Mann–Whitney tests and indicators across three groups using the Kruskal Wallis one-way analysis of variance (ANOVA).

We used the Spearman correlation test to examine the correlation between the epidemic size per million population of COVID-19 and each of the seven predictors as defined previously. We found that the epidemic size was associated with most of the proposed predictors (Table S3). However, most of these predictors,

---

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Days required to increase from 30 to 100 cases</th>
<th>CFR in the first 100 confirmed cases</th>
<th>Slow growing phase (cases/day)</th>
<th>Fast-growing phase (cases/day)</th>
<th>Day of the phase transition point</th>
<th>Number of cases at the transition point</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>China (45 cities)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wuhan city</td>
<td>6.6 (5.3–7.9)</td>
<td>0.8% (0.2–1.4%)</td>
<td>3.3 (2.6–4.1)</td>
<td>16.1 (12.3–19.8)</td>
<td>5.5 (4.6–6.4)</td>
<td>17.7 (11.9–23.6)</td>
</tr>
<tr>
<td>15 cities in Hubei province (excluding Wuhan)</td>
<td>4.4 (3.1–5.7)</td>
<td>1.8% (0.6–2.9%)</td>
<td>4.8 (3–6.6)</td>
<td>24 (18.4–29.6)</td>
<td>3.8 (2.9–4.8)</td>
<td>16.9 (9.9–23.8)</td>
</tr>
<tr>
<td>29 cities in China (excluding Hubei) (Kruskal–Wallis One-way ANOVA Mann–Whitney test (p-value))</td>
<td>7.8 (6.7–9)</td>
<td>0.2% (0–0.4%)</td>
<td>3 (2.3–3.7)</td>
<td>11.3 (9.4–13.3)</td>
<td>5.9 (5–6.8)</td>
<td>17.9 (11.7–24.1)</td>
</tr>
<tr>
<td></td>
<td>0.0005*</td>
<td>0.0006*</td>
<td>0.1076</td>
<td>0.0001*</td>
<td>0.0275*</td>
<td>0.7664</td>
</tr>
</tbody>
</table>
except the case-fatality rate within the first 100 confirmed cases, were collinear with the time from 30-to-100. Some predictors were not representative of all 45 cities. We, therefore, chose ‘time from 30-to-100’ and CFR-100 for the subsequent prediction of the ranking of the risk of a major epidemic (details in Supplementary materials).

We compared the predicted ranking of the epidemic size per million population based on the ‘time from 30-to-100’ and CFR-100 with the actual ranking of the epidemic size. We first ranked the two indicators ‘time from 30-to-100’ and CFR-100 independently, then summed the corresponding indexes of the indicators. We then ranked again the sum of the indexes to produce a predicted ranking of the epidemic size. This index was then compared with the actual classification of the epidemic size in each city using Wilcoxon matched-pairs signed-rank tests.

Results

We identified a total of 45 Chinese mainland cities that reported more than 100 cases of COVID-19. Apart from the epicenter Wuhan, there were 15 neighboring cities in Hubei and 29 cities in other Chinese provinces. Most (n = 34) cities demonstrated a successful two-phase fit, whereas eleven cities had only one-phase identified (Figure 1, Figure S2). Notably, the number of cases at the phase transition point in these Chinese cities was (17.7 [95% confidence intervals, 11.9–23.6]), Table 1), and 88.2% (30/34) of Chinese cities had their phase transition points below 30 cases (Figure 1, Figure S2). We regarded the 30 confirmed cases as a critical threshold after which the COVID-19 epidemic started to increase rapidly.

Table 1 reports the epidemiological characteristics for the first 100 confirmed cases. In the 45 Chinese cities, the days required for the number of confirmed cases to increase from 30 to 100 was 6.6 (5.3–7.9) days. In Wuhan, the number of days to rise from 30 to 100 cases was two days, compared to 4.4 (3.1–5.7) days for the 15 other cities in Hubei province and 7.8 (6.7–9.0) days for the 29 Chinese cities outside Hubei province. The difference was significant (Kruskal–Wallis one-way ANOVA, p = 0.0005). The average case-fatality rate in the first 100 confirmed cases across the 45 Chinese cities was 0.8% (0.2–1.4%). In particular, the CFR-100 in Wuhan was 2.5%, followed by 1.8% (0.6–2.9%) in 15 other cities in Hubei province and 0.2% (0.0–0.4%) for the 29 Chinese cities outside Hubei province. The slow-growing phase was relatively short (5.5 [4.6–6.4] days) with a growth rate of 3.3 (2.6–4.1) cases/day, whereas the growth rate in the fast-growing phase was about five times higher (16.1 [12.3–19.8] cases/day).

Figure 2 demonstrates a significant negative correlation between the epidemic size per million population and time to rise from 30-to-100 cases in 45 Chinese cities (Spearman correlation, p < 0.0001, r = −0.6653). The case-fatality rate in the first 100 cases was positively correlated with the epidemic size (Spearman correlation, p = 0.0168, r = 0.3472). Other predictors were also significantly associated with the epidemic size, but they were found to be significantly collinear were from 30-to-100 (Table S3). When stratified by the ‘time from 30-to-100’, the epidemic size per million population was the lowest among cities with a ‘time from 30-to-100’ of nine days or greater (46 [25–88] cases/million), followed by those that required 5–8 days (70 [46–107] cases/million), 3–4 days (329 [154–703] cases/million) and those of two days or less (1895 [481–7469] cases/million, Figure 3). When stratified by CFR-100, the epidemic size was the

---

**Figure 1.** Joinpoint two-phase fitting for 45 Chinese cities, showing the transition point below a threshold of 30 cases. The abbreviations next to the city names represent the names of the corresponding Chinese provinces (AH: Anhui; BJ: Beijing; CQ: Chongqing; GD: Guangdong; HJ: Heilongjiang; HN: Henan; HB: Hubei; Hunan; JX: Jiangxi; SX: Shaanxi; SH: Shanghai; SC: Sichuan; TJ: Tianjin; ZJ: Zhejiang). Fittings for individual cities are shown in Figure S2.

**Figure 2.** A significant correlation between the outbreak morbidity rate and (a) the time from 30-to-100 (r = −0.6653, p < 0.0001); (b) CFR-100 (r = 0.3472, p = 0.0168).
least among cities with zero fatalities among the first 100 confirmed cases (71 [49–103] cases/million), followed by cities with CFR-100 between 0.01–1.00% (207 [61–704] cases/million), 1.01–2.00% (511 [76–3419] cases/million) and ≥2.00% (1186 [412–3414] cases/million).

Table 2 demonstrates a highly significant correlation between the predicted ranking of epidemic size per million population based on the 'time from 30-to-100' and CFR-100 and the actual ranking of epidemic size (Wilcoxon matched-pairs signed-rank test, p < 0.0001, r = 0.7627).

Discussion

Our study identified that among 45 cities in China, 30 cases could be a critical threshold for switching from a relatively slow-growing phase to a fast-growing phase, which grows five times faster. Of the seven early-stage epidemic characteristics we assessed, we found that the time from the 30th to 100th case and the case-fatality rate in the first 100 cases were strong indicators of the size of the future epidemic. These early-stage 'indicators' may be useful to public health officials in other settings to identify appropriate control measures.
Our study of Chinese cities provides a unique opportunity to understand the COVID-19 epidemic in cities with quite different reproductive rates when the virus first spread to the cities. This may provide guidance to other cities with different reproductive rates worldwide. We argue that before the outbreak was detected, it is likely that most Chinese cities had similar reproductive rates, but the nationwide response dramatically lowered the reproductive rates at a time when many Chinese cities were in different stages of the epidemic. This allowed an observational study of many similar cities with different reproductive rates to determine what factors predicted large outbreaks and therefore allowed the identification of cities that were likely to have large epidemics.

The first 30 cases appear to be an essential indicator for the initiation of a fast-growing phase of COVID-19. It is possible that the detection of 30 cases may represent a time when the epidemic shifts from one associated primarily with imported cases to one primarily driven by local transmission. Once it reaches this critical mass, local transmissions start to dominate, and a large number of domestic transmissions begin to surface. This pattern may not be evident if a city is very closely associated with another major outbreak, and this may be the reason that most of Wuhan’s neighboring cities do not have a slow-growing phase. Besides, the duration of the slow-growing phase is only about six days, which is consistent with the incubation period of SARS-CoV-2 infection (5–6 days [Lauer et al., 2020; Li et al., 2020; Zhang et al., 2020a]). This suggests a large number of pre-symptomatic cases in the incubation period have now become symptomatic and detectable, which marks the beginning of a rapid increase in the number of cases.

The characteristics of the epidemic in the early stage may reflect the healthcare system’s potential epidemic development and response. The number of days required to increase from 30 to 100 cases represents a rough measure of a localized epidemic’s initial growth rate. A short duration implies a fast and probably uncontrolled expansion of the epidemic, which is likely associated with either high transmissibility of the virus or a delayed diagnosis. The high transmissibility is likely related to the absence of effective prevention or intervention strategies at this stage. Besides, when symptomatic patients present themselves to the hospitals in hundreds, this likely means a lot more pre-asymptomatic cases are yet to be diagnosed, and the epidemic is far more severe than what has been observed. Consistently, a high case-fatality rate in the first 100 diagnoses also implies a missed opportunity for earlier diagnosis. Considering the incubation period (5–6 days) and the time from symptom onset to death (2–8 weeks [World Health Organization, 2020b]), a high case-fatality rate in the early stage suggests that the surveillance system was too slow in responding to the epidemic to prevent infected individuals from progressing to severe disease.

Our analysis showed that based on these two simple early-stage indicators, we can rank the predicted epidemic size per million
population, and this ranking is highly consistent with the ranking of the actual epidemic size at a later stage of the epidemic. This may provide useful insights into the potential severity of the COVID-19 epidemic in its later development. Therefore, a fast-growing phase and a high case-fatality rate are early warning signs for the healthcare system to appropriately react to the epidemic. This is comparable to previous studies where a novel framework has been developed to assess the epidemic severity for influenza based on its transmissibility and clinical severity (e.g., case-fatality rate) (Carrie et al., 2013; Shrestha et al., 2011). However, these studies did not investigate characteristics of the early epidemic and their implications.

Our findings need to be interpreted with caution. First, our results are not a quantification of the actual size of the epidemic, but rather a comparison between the predicted and actual rank of epidemic size. Second, since China had implemented rigorous control strategies to curb the epidemic in most parts of the country, the epidemic size may be under-represented in comparison with settings in other countries. Third, we regarded Wuhan’s epidemic as a major outbreak, yet, the outbreaks in many cities worldwide are already comparable to or exceed Wuhan’s level. The implications of a Chinese city ranking may be interpreted differently in these settings. Therefore, whether the prediction of ranking may be applied to other countries is uncertain and warrants further investigations.

Conclusions

The first 30 cases may mark a critical threshold for the transition from a slow to a fast-growing phase of the COVID-19 epidemic. Early epidemic characteristics may be regarded as important indicators for later epidemic development and size.

Conflict of interest

None declared.

Ethical approval

Ethical approval was not required.

Acknowledgment

The work was supported by the Bill & Melinda Gates Foundation. Lei Zhang is supported by the National Natural Science Foundation of China (Grant number: 819560410639); Outstanding Young Scholars Funding (Grant number: 31115000001); Xi’an Jiaotong University Basic Research and Profession Grant (Grant number: xtr022019003 and xzy032020032) and Xi’an Jiaotong University Young Talent Support Grant (Grant number: YXJJ04). Mingwang Shen was supported by the National Natural Science Foundation of China [grant number: 11801435 (MS)]. China Postdoctoral Science Foundation [grant number: 2018M631134]; the Fundamental Research Funds for the Central Universities [grant number: jxj012019055, xzy032020026]; Natural Science Basic Research Program of Shaanxi Province [Grant number: 2019JQ-187]; and Xi’an Special Science and Technology Projects on Prevention and Treatment of Novel Coronavirus Pneumonia Emergency [Grant number: 20200005XY005].

We are thankful for the input and discussion from Guoqiang Li and Xinghui Li.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: https://doi.org/10.1016/j.ijid.2020.05.122.

References