Dynamic characteristics of the COVID-19 epidemic in China’s major cities

Ci Song, Tao Pei, Xi Wang, Yaxi Liu, Jia Ma, and Daojing Zhou

1. Introduction

Since December 2019, when the coronavirus disease of 2019 (COVID-19) first appeared in Wuhan, COVID-19 spread rapidly in cities and provinces across the country and invaded Asia, Europe, America, and the rest of the world, posing a serious threat to global public health (Desjardins, Hohl, and Delmelle 2020; X. Li et al. 2020). As of 14 July 2021, 187,717,255 people have been infected and 4,047,597 have died worldwide (Dong, Du, and Gardner 2020). Facing a pandemic of such a massive scale, governments around the world have implemented a substantial number and variety of policies (Koo et al. 2020; Kissler et al. 2020; Ding et al. 2020), including individual isolation, family quarantine (Tian et al. 2020; Chinaazzi et al. 2020), travel restrictions (Lai et al. 2020), contact reduction (Ferretti et al. 2020), etc. (Zhang, Litvinova, and Liang et al. 2020; Liu et al., 2020b). Although most policies slowed down the spread of COVID-19 to a certain extent, the effects are regionally different. Considering the endemic epidemic situations, zonal management and regional prevention and control measures are key to control the epidemic (Q. Li et al. 2020c).

In order to effectively achieve zonal management and endemic prevention and control policy implementation for mitigating the spread of COVID-19, it is crucial to understand the various patterns of the COVID-19 epidemic and the mechanisms influencing it in different places. Studies on the dynamic characteristics of the COVID-19 epidemic can be grouped into three categories: outbreak trend surveillance, spatial hotspot identification, and spread process analysis. In the early period of the epidemic, several outbreak indices, including basic reproduction number (R0), growth rate, and endemic case numbers, have been estimated based on several epidemic dynamic models for evaluating the epidemic severity in China (Wu, Leung, and Leung 2020; Li and Wang 2020; R. Li et al. 2020a), the United States (Yadav et al. 2020), Europe (Flaxman et al. 2020), and other countries (Adekunle et al. 2020). Other studies focused on the increasing (decreasing) trends in different epidemic stages and the identification of turning points through time-series analysis methods (Tian et al. 2020; Tang et al. 2020). In addition, epidemic risk analysis through prediction models has also been...
addressed (Chimmula and Zhang 2020). Studies on spatial hotspot identification of COVID-19 mainly focused on discovering severely hit areas at different scales. Common spatial analysis methods, including spatial scan statistics (Desjardins, Hohl, and Delmelle 2020), spatial clustering (Zhang et al. 2020) and local spatial autocorrelation (Xiong et al. 2020) were widely applied to different cities (Koo et al. 2020), provinces, and countries (Desjardins, Hohl, and Delmelle 2020). Studies on the spread process aim at discovering the source and sink areas and analysing potential transmission paths of COVID-19 through population flow between regions using human movement data (Hu et al. 2020). Although there are few studies on this topic, previous works have proposed many effective methods for spread process analysis of other infectious diseases, such as malaria (Ruktanonchai et al. 2016), dengue fever (Wesolowski et al. 2015), cholera (Finger et al. 2016), etc.

Studies on the mechanisms influencing regional COVID-19 outbreaks mainly focused on social factors, natural factors, and biological factors. For social activity factors, the population outflow from the source is believed to be the most important factor affecting the epidemic spread in the early period (Jia et al. 2020), and the local population density along with its activities has a large impact on endemic epidemic outbreaks (Kraemer et al. 2020). Natural factors of the COVID-19 epidemic mainly include temperature (Sajadi et al. 2020), humidity and sunshine (Paez et al. 2020). These factors may affect the transmission ability of COVID-19 by reducing the virus biological activity and survival time. The influence from biological factors is mainly due to the antiviral ability of individuals of different ages, gender, blood types and health levels (Dowd et al. 2020). For instance, many studies have found that older male patients suffer higher incidence and mortality of COVID-19 (Lighter et al. 2020).

Although the above studies have analysed dynamic characteristics of the COVID-19 epidemic, along with factors influencing it, in many regions, as yet no research has been carried out to categorize regional epidemic dynamic situations and summarize the evolution patterns. Therefore, this study aims to propose a time-series clustering method to discover dynamic patterns of the COVID-19 epidemic by categorizing regional epidemic situations into groups and analysing the driving factors among the city background conditions for each category. Through this work, endemic prevention and control measures for containing COVID-19 can be effectively recommended in areas with ongoing (or upcoming) outbreaks according to the city background conditions. The epidemic dynamic processes in Chinese cities function as an example because the outbreaks in those cities have been under control since March, with a total of over 80,000 confirmed cases (84,047 cases, including 4,638 deaths, reported on 17 May 2020).

2. Data and methods

2.1. Data

COVID-19 case data of 362 cities (or autonomous prefectures or special administrative regions) in China were collected from the DXY.cn website. This data set includes daily reported numbers of confirmed cases at the prefecture-level. In order to analyse the dynamic process of the epidemic, we chose the data from 11 January to 27 March 2020, when the epidemic had been brought under control for the most part in most cities, and adjusted the records in some cities by a simple reassignment process. Based on this adjustment, sudden increases in the number of reported cases due to a change in testing are proportionally reassigned to the days of the previous week (Figure 1); the data could thus be better used for further time-series analysis. We also extracted human mobility data from the Baidu Smart Eye Map (Baidu, Inc., Beijing, China). This data contains two indices: the Baidu migration index (BMI) and the intra-city travel intensity (ITI). The BMI represents the relative population flow from one city to another in per cent, whereas the ITI is the ratio of travellers to the local resident population in a city. The Baidu migration data have been widely used in many studies on the COVID-19 epidemic (Wu, Leung, and Leung 2020; Xiong et al. 2020; Kraemer, Yang, and Gutierrez et al. 2020). We collected these data for 83 major cities in China from 17 January to 30 January 2020, 1 week before and after the quarantine of Wuhan, and calculated several indices representing

Figure 1. Data of daily confirmed cases of COVID-19 in China. Both raw and adjusted data are shown.
city background conditions for further analysis. In addition, the statistical data for the residential population and number of hospitals were obtained from the China Statistical Yearbook 2018. Table 1 shows some descriptive statistics for the yearbook data.

| Table 1. Descriptive statistics table for the yearbook data. |
|-------------------------------------------------------------|
|                | Urban population (million) | Number of hospitals | GDP (billion yuan) |
| Descriptive     |                            |                      |                   |
| statistics      |                            |                      |                   |
| Number of cities| 297                        | 4.4                  | 107.6             |
| (0.19–33.9)     |                            | (47–127.3)           | (34.3–134.3)      |

Values in the table are mean values (First quartile–Third quartile).

2.2. Time-series analysis of epidemic characteristics

In order to analyse the epidemic characteristics, we counted the cumulative confirmed cases and analysed the date of the first case occurrence and outbreak duration in each city. The date of the first case occurrence is the date when the reported confirmed cases number was greater than zero. The outbreak duration is the number of days between the date of the first case occurrence and the time when the outbreak was brought under control, which is defined as the first date that the new confirmed cases number was zero for 7 consecutive days.

2.3. Time-series clustering for categorizing cities of different patterns

We also implemented a time-series clustering method to categorize several cities with larger numbers of confirmed cases (greater than 50) into different types according to the local epidemic dynamic in each city. We first construct a synthesized time series, which is the linear combination of the normalized pattern and the cumulative number of confirmed cases. Then, to determine different types of patterns, we use a semi-supervised K-means clustering method. Thus, the effect of clustering result is determined by a silhouette coefficient (Rousseeuw 1987).

The process of clustering is divided into the following three steps: (1) Construct the synthesized time series that combines the normalized pattern and the cumulative number of confirmed cases. (2) Cluster the time series of confirmed cases using K-means and determine the optimal clustering number based on silhouette coefficient. (3) Post-process the clustering result by assigning clusters to cities of different patterns. Each of these steps is now described in detail.

2.3.1. Construct the synthesized time series

The synthesized time-series data of each city used for clustering combines three parts. The first is the total cumulative number of confirmed cases from 11 January 2020 to 27 March 2020, which reflects the outbreak size of the local epidemic. The second is the normalized time series of daily new confirmed cases, which indicates the dynamic variation of the local epidemic. The third is the cumulative number of confirmed cases when the epidemic was under control in most cities (from 5 March 2020 to 27 March 2020), which reflects the resurgence situation of local epidemics. This synthesized time series, given by

\[ Y_i = \langle \ln(N_i + 1), Z_i, \ln(N_i^f + 1) \rangle, \]

where \( Y_i \) is the synthesized time series for clustering in city \( i \), \( N_i \) is the cumulative number of confirmed cases in city \( i \), and \( Z_i \langle Z_i \rangle \) is the standardized time series of daily new confirmed cases in city \( i \) defined as

\[ z_{ij} = n_{ij}/N_i, \]

where \( n_{ij} \) is the number of daily new confirmed cases and \( N_i \) is the cumulative number of confirmed cases in city \( i \) from 5 March 2020 to 27 March 2020.

2.3.2. Cluster the time series for categorizing

After generating the synthesized epidemic time series, we applied a K-means clustering algorithm (Jain 2010) for categorizing cities with different epidemic characteristics. The number of clusters is determined by the silhouette coefficient (Rousseeuw 1987) as a quantitative index, which is defined as

\[ S = \sum_{i} S_i = \sum_{i} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \]

Here, \( a(i) \) stands for average distance between cluster centre \( i \) and other samples in the same cluster. \( b(i) \) stands for minimum value of average distances between cluster centre \( i \) and samples in other clusters. Large silhouette coefficient means a good clustering result. Thus in this process, we try different number of clusters and determine the clustering result with the optimal silhouette coefficient.

2.3.3. Post-processing to assign clusters to cities of different patterns

Once the clusters have been determined, we perform post-processing to assign each cluster to cities with specific patterns. Each pattern is determined by a decision tree based on three quantification indices including cumulative cases number, duration and turning point (Figure 2a,b). A cluster is assigned to a specific
pattern type (quantified by these three indices) whose centre is close to the cluster centre $i$ in step 2. Considering the spatial effect, the final clustering result also need to be adjusted to make the neighbouring cities into the same category as possible. In this case, we select the cities surrounded by cities of other types, and compared their pattern quantification indices. If each index of the sample city is within one standard deviation of mean value in the cluster, we can adjust the type of the city to the same type of the surrounding cities.

The complexity of this entire process is mainly determined by the clustering step, which is at most $\sum n_{\text{Cluster}} \times O(n \times n_{\text{Iteration}})$. Here, $n$ is the number of cities. $m$ is the length of the synthesized epidemic time series. $n_{\text{Iteration}}$ is the number of iterations in the K-means algorithm. $n_{\text{Cluster}}$ is the number of clusters to be determined. The overall complexity of the algorithm can be simplified as $O(n)$, namely, linear complexity. This algorithm is coded and performed in a Matlab software, whereas the result is presented in ArcMap.

3. Results

3.1. Dynamic characteristics of COVID-19 in China’s major cities

Figure 3 provides the daily confirmed cases in China. As can be seen, Wuhan was the most severely hit area during the COVID-19 epidemic in China. With a sharply increasing trend in January, daily confirmed cases there reached the peak value in mid-February and then gradually decreased until the pandemic was under control.
in March. In contrast, the outbreak trends in Hubei Province (excluding Wuhan) and other places are relatively flat, and the peaks of daily confirmed cases appeared earlier. Figure 4a shows the date of the first confirmed case occurrence for each city analysed in this study. In addition to Wuhan and Huanggang, the earliest occurrences of confirmed cases are mainly distributed in megacities (or major cities) in China, including Beijing, Shanghai, Tianjin, Chongqing, Guangzhou, etc. These cities are densely populated and have large population flows from Wuhan. Therefore, while the epidemic spread in Hubei Province, infected individuals also travelled to other major cities across the country through the inter-regional transportation network and the disease spread to the surrounding urban agglomerations, which led to outbreaks of the epidemic all over China.

Figure 4b shows the cumulative number of confirmed cases for each city analysed in this study. It can be seen that confirmed cases of COVID-19 are mainly distributed throughout the central and eastern regions in China. The areas with severe outbreaks are centred on Wuhan and are mainly distributed in Hubei Province and its neighbouring major cities, including Chongqing, Nanchang, Changsha and Xinyang. The number of cumulative confirmed cases in these cities exceeded about 1,000 and the number in Wuhan reached more than 50,000. Other endemic areas away from Wuhan include Beijing, Shanghai, Wenzhou, Guangzhou and Shenzhen. More than 400 cases were confirmed in each of these cities.
The durations of the respective outbreaks are shown in Figure 4c. It can be seen that Hubei Province and its neighbouring cities (Chongqing, Chengdu, Nanyang, etc.) are the areas with the longest outbreak durations. Durations in provinces in central and eastern China (Hunan, Jiangxi, Anhui, Henan, etc.) were mainly up to 1 month only. In other provinces, the outbreaks lasted no more than 3 weeks. Although the epidemic was under control in most cities by the end of March, there are still a number of imported infection cases in Beijing, Shanghai, and Guangzhou. In these cities, along with Wuhan, Hong Kong and Taiwan, the COVID-19 epidemic is still ongoing.

3.2. City clusters of different epidemic dynamic patterns

The spatial distribution of each type of epidemic situation is presented in Figure 5, and six epidemic indices in different types of epidemic situation are presented in Figure 6. On the basis of these, Table 2 summarizes the different types of epidemic situations with different COVID-19 dynamic patterns in 83 major cities. Type 1, represented by Wuhan, is categorized as ‘extreme outbreak area’ for the earliest appearance of cases, largest number of cumulative cases, and longest duration. The type 2 cities around Wuhan in Hubei Province, such as Xiaogan, Huanggang, Ezhou, etc. are categorized as ‘large spread areas’ with thousands of confirmed cases and outbreak durations of over 1 month. Type 3 are mainly megacities with large population and traffic flows, including Beijing, Tianjin, Shanghai and Shenzhen. These cities are categorized as ‘potential resurgence areas’ with fewer confirmed cases compared to type 2 cities. However, due to large population flows, a few imported infection cases persist despite the fact that the epidemic is under control in most cities. Type 4 includes two small cities in Hubei Province and some major cities away from Hubei, including Harbin, Dongguan and Tangshan. These cities are categorized as ‘middle spread areas’ because the cumulative confirmed cases and outbreak durations of these cities are all at mid-levels. Cities of type 5 include several major cities in south Henan, west Anhui, and Chongqing. These cities are categorized as ‘controlled outbreak areas’, because the outbreak durations were relatively short and the turning points were earlier due to effective prevention and control measures. Cities of type 6 are mainly distributed in north Hunan, east Jiangxi, west Jiangsu, etc. These cities are categorized as ‘limited growth areas’ based on small numbers of cumulative confirmed cases and no turning points. Cities of type 7 include Hong Kong, Macau and Taiwan, where daily new confirmed cases began to rise in the late period of the epidemic. These cities are categorized as ‘delayed outbreak areas’ Type 8, Jining, is a special case.
where 200 confirmed cases suddenly emerged in a prison on 20 February 2020. This city is categorized as ‘Lag report areas’.

4. Discussion

4.1. City background conditions of different types of epidemic dynamic patterns

In order to understand the driving factors among the city background conditions for each category of epidemic dynamic pattern, we calculated five indices using the Baidu migration and China Statistical yearbook data, including the population flows from Wuhan, distance from Wuhan, local residential populations, activity ratio (ratio between 1-week cumulative ITI after and before 23 January, which equals to cumulative ITI from 17 to 23 January divided by cumulative ITI from 24 to 30 January), and number of local hospitals. The first two indices represent interactions with the extreme outbreak area. The third index suggests potentially susceptible population sizes. The fourth index reflects the government emergency response capability, and the last index the medical resource conditions. Table 3 shows the statistics of the city background condition indices for each type of epidemic situation. It is clear that in most cases extreme outbreak areas are likely to appear in metropolises with large populations, such as Wuhan (Hu et al. 2020), Milan (La Maestra, Abbondandolo, and De Flora 2020) and New York City (Desjardins, Hohl, and Delmelle 2020). These cities generally hold the most or earliest confirmed cases in the early period of the epidemic. The large spread areas are generally the cities around extreme outbreak areas, because these cities have large population flows and thus import cases from there. In addition, limited hospital resources are also one of the reasons for epidemic outbreaks in such cities. The potential resurgence areas tend to be international cities or port cities (including Beijing, Shanghai and Shenzhen) with large populations and flows of domestic or foreign people; thus, population flows are also believed to be a potential driver of the COVID-19 epidemic spread (Jia et al. 2020; Ren et al. 2020). The middle spread areas are also cities with relatively high population flows from extreme outbreak areas (not as high as cities of large spread areas), but have a lower number of

Figure 6. Dynamic characteristics of the COVID-19 epidemic in different types of epidemic situations.
local residents. Controlled outbreak areas are mainly cities with strong government emergency response capabilities, such as Chongqing (Heet al. 2020) and Wenzhou (Han et al. 2020). In these cities, once the government realized the severity of the situation, strict prevention and control policies were implemented immediately to contain the spread of the epidemic. The limited areas are generally cities far away from extreme outbreak areas. These cities have a limited population flow from areas severely hit by the epidemic and thus only a few occasional cases are imported. The delayed outbreak areas are special administrative regions where new confirmed cases are more likely to be imported from overseas. The lag report areas are very likely to be cities with poor medical treatment capabilities, where new confirmed cases are reported late due to limited testing capacities. However, this case was a speciality case in China because most of the confirmed cases occurred in a prison.

### 4.2. Endemic prevention and control measure recommendations

Based on the above analysis, we recommend different policy guidelines according to the empirically predicted epidemic dynamic patterns and their city background conditions along with the early epidemic situation in Table 4. Then, we also list some specific endemic prevention and control measures for each type of cities. For the primary source city and its surrounding cities, strict control policies are indispensable for containing the outbreak situation, especially for metropolitan areas with large populations (type 1 cities) and cities with large population flows from those regions (type 2, 3 and 4 cities). Because preventing the spread of infected cases and containing the epidemic is the primary task in these cities. In this respect, quarantine of the outbreak area and travel restrictions from neighbouring cities with large population flows along with family isolation are recommended as the most effective measures (Liu et al., 2020; Flaxman et al. 2020; Tian et al. 2020; Jia et al. 2020). Moreover, rapid surveillance based on efficient virus testing capacities and adequacy of medical resources are also considered as key measures for these cities once outbreaks have started (R. Li et al. 2020; Liu et al., 2020; Zhang, Litvinova, and Liang et al. 2020). As we knew, the medical staff and resources assistance from other cities to Hubei Province strongly support COVID-19 containment later (Chinazzi et al. 2020; R. Li et al. 2020). At the same time, the establishment of new hospitals, including Leishenshan Hospital and Huoshenshan Hospital and Fangcang shelter hospitals also play an important role in the COVID-19 pandemic (Chen et al., 2020; Z. Li et al. 2020e). These measures have

### Table 2. COVID-19 dynamic patterns in different types of epidemic situations.

| Type | Category characteristics | Number | Cumulative Cases | Duration | Turning point | Cities |
|------|--------------------------|--------|------------------|----------|---------------|--------|
| 1    | Extreme outbreak area    | 1(1.2%)| 50006            | 76 (Still ongoing) | Feb 7 | Wuhan |
| 2    | Large spread area        | 12 (14.5%) | 1403 | 37 | Feb 4 | Cities around Wuhan with a large number of cases |
| 3    | Potential resurgence area| 8(9.7%) | 313 | 49 (Still ongoing) | Jan 31 | Beijing, Shanghai, Tianjin, Guangzhou, Shenzhen, Chengdu |
| 4    | Middle spread area       | 6(7.2%) | 188 | 32 | Feb 9 | Tianmen, Qianjiang, Harbin, Dongguan |
| 5    | Controlled outbreak area | 23(27.7%) | 181 | 28 | Feb 1 | Chongqing, Wenzhou, Xinyang, Changsha, Zhengzhou, Nanchang |
| 6    | Limited growth area      | 29(34.9%) | 75 | 26 | No sign. | Fuzhou, Qingdao, Bangbu, Shangrao, Xian, Nanjing |
| 7    | Delayed outbreak area    | 3(3.6%) | 273 | 65 (Still ongoing) | Feb 3 | No sign. | Taiwan, Hong Kong, Macau |
| 8    | Lag report area          | 1(1.2%) | 260 | 38 | Feb 1 | No sign. | Jining |

### Table 3. Indices of city background conditions in each type of epidemic situation.

| Type | Population flows from Wuhan (%) | Residential Population (million) | Distance from Wuhan (km) | Activity ratio after and before Jan. 23 | Number of hospitals |
|------|--------------------------------|----------------------------------|--------------------------|------------------------------------------|---------------------|
| 1    | 5.96(3.57–6.39)                | 8.54                             | 199.5(69.2–278)          | 0.36(0.34–0.37)                          | 56.2(38.5–67.8)     |
| 2    | 0.32(0.1–0.45)                 | 9.76(5.95–13.97)                 | 581.2(369.4–677.9)       | 0.35(0.33–0.38)                          | 430(257.8–598.5)    |
| 3    | 0.64(0.02–1.26)                | 4.37(1.48–8.05)                  | 770.7(178.2–1593.1)      | 0.50(0.37–0.59)                          | 193(116–275.8)      |
| 4    | 0.55(0.12–0.79)                | 8.07(5.27–9.49)                  | 359.4(159.1–604.7)       | 0.45(0.4–0.51)                           | 166.5(79–199.5)     |
| 5    | 0.12(0.05–0.17)                | 5.56(3.98–6.84)                  | 498.5(249.5–601.4)       | 0.49(0.4–0.58)                           | 117(459–138)        |
| 6    | 0.05                            | 883                              | 528.1(380.4–710)         | 0.47                                      | 164                 |

*aValues in the table are mean values (First quartile–Third quartile). Data in type 7 cities are vacant. Bold values indicate the characteristics of the city background conditions for each type.*
Table 4. Recommendations of prevention and control measures for different types of epidemic situations.

| Type | City background conditions | Categories | Epidemic characteristics | General guideline | Key measures |
|------|-----------------------------|------------|--------------------------|-------------------|-------------|
| 1    | Primary source cities of epidemic with large populations | Extreme outbreak area | Earliest outbreak, tens of thousands of cases, long duration | Extremely strict control | Quarantine, rapid surveillance, family isolation, supplementation of medical resources |
| 2    | Cities close to extreme outbreak areas and holding large population flows from there | Large spread area | Thousands of cases, long duration, large uptrend of new cases | Strict control | Travel restrictions, rapid surveillance, family isolation, supplementation of medical resources |
| 3    | International (and port) cities with large populations and flows of domestic and foreign people | Potential resurgent area | Large number of population flows along with infected cases, potential resurgence | Strict control | Travel restrictions, improvement of testing, family isolation, restriction lifting delay |
| 4    | Cities with high population flows from extreme outbreak areas, but having fewer local residents | Middle spread area | Hundreds of cases, one month duration, late turning point | Strict control | Travel restrictions, family isolation, social distancing |
| 5    | Cities with strong government emergency response capabilities | Controlled outbreak area | Hundreds of cases, short duration, early turning point | Semi-control and gradually resumption | Zonal management, family isolation, social distancing, personal contract-tracing |
| 6    | Cities far away from extreme outbreak areas with less population flows | Limited growth area | Less than one hundred cases, short duration, no significant uptrend of cases | Semi-control and gradually resumption | Rapid surveillance, social distancing, lifting of restrictions by the end of epidemic |
| 7    | Special administrative cities with cases imported from overseas | Delayed outbreak area | Few cases in the early period, outbreak around March | Surveillance and control | Travel restrictions, maintained surveillance restriction lifting delay |
| 8    | Cities with poor medical treatment or socioeconomics | Lag reported area | Sudden emergence of a high number of cases | Surveillance and control | Travel restrictions, improvement of testing |
effectively contained the COVID-19 epidemic in the outbreak areas. For cities with strong government emergency response capabilities and cities far away from extreme outbreak areas where the epidemic is not serious, semi-control and gradually resumption policies are recommended. Because, during the middle period of the epidemic, when quarantine and travel restrictions are well implemented and most cities hit by the epidemic have been isolated, gradually resumption of work under the epidemic control is also an important task. In this respect, mild control policies including zonal management, personal contact-tracing, family isolation and other social distancing measures (such as workplace distancing, bans of public gatherings) may be more efficient for controlled outbreak areas (type 5 cities) and the limited growth areas (type 6 cities) (Ferretti et al. 2020). For delayed outbreak areas (type 7 cities) and cities with poor medical treatment or socioeconomics (type 8 cities), however, the government should maintain surveillance to ensure that local infected cases have disappeared and to prevent a resurgence of the epidemic. Therefore, surveillance and control policies are considered as appropriate measures. It is also recommended to delay loosening restrictions and improve testing capacities for these areas even if the outbreaks are under control.

4.3. Guidance for the global COVID-19 epidemic from China’s situation

Although the number of infected cases is relatively smaller and the duration is relatively short compared to global situation, China’s experience on COVID-19 provides the earliest and complete case-study (Yang et al. 2020; Huang et al. 2021). This dynamic process of COVID-19 epidemic in China, including a whole ‘emergence – outbreak–under control–gradually vanish’ epidemic period, is believed to provide important guidance for global epidemic prevention and control (Liu et al. 2021). First, the epidemic indices proposed according to China situation at different stages can be useful to describe dynamic process of COVID-19 epidemic in foreign cities, and some indices such as first occurrence, turning point, trend of infected cases are of important guiding for the implementation of epidemic prevention and control policies (Li, Feng, and Quan 2020; Tang et al. 2020). Second, through classifying the different dynamic patterns in Chinese cities, COVID-19 epidemic situation in some foreign cities with similar characteristics can be categorized, and then some epidemic indices may be predicted, such as duration, infected cases number, etc. More importantly, different endemic prevention and control policies can be recommended according to the categorized dynamic patterns for foreign cities (Lei et al. 2020; Fricke et al. 2021) so that we could effectively contain the epidemic spread and achieve resumption of work as soon as possible. In general, China’s epidemic dynamic process has many similarities with global epidemic, and during this period, the epidemic dynamic patterns, medical response, and endemic prevention and control policies have provided valuable experience and reference for the global COVID-19 epidemic.

5. Conclusions

In this study, we categorized eight types of epidemic situations with different epidemic dynamic characteristics in Chinese cities, based on a time-series clustering method, and analysed the driving factors among the city background conditions, including population flows, local resident numbers, government emergency response capabilities, and medical resource conditions. We suggest that different endemic prevention and control measures for containing COVID-19 epidemic should be implemented depending on the epidemic situation type. Our work may provide effective recommendations for cities with ongoing (or upcoming) COVID-19 outbreaks. We emphasize the importance of focusing on the primary outbreak areas (source cities) of the epidemic and their surrounding areas as well as international (and port) cities with large populations and flows of domestic and foreign people. Moreover, we believe that contact-tracing and testing capacity will play an important role in some cities during future outbreaks of COVID-19.

This study is subject to some limitations. First, the data used for analysis are daily reported numbers of confirmed cases, which differ from the actual numbers of infected people because of delay time from disease onset to diagnosis. Therefore, some dynamic characteristics in cities, such as rise duration, turning point, may shift a few days. Second, up trends of cases may be over-estimated in cities of Hubei Province, because testing capacities of cities in Hubei Province have been rapidly improved in the early period of epidemic due to the assistance from other provinces and the establishment of ‘Fangcang’ hospital. Nevertheless, we believe that the final clustering results of major cities will be little affected and the endemic prevention and control measures are still instructive. Our study is solely based on the characteristics of the epidemic in China’s major cities. Considering that the situation in foreign cities may be different, future work should focus on worldwide epidemic dynamics with a more appropriate data set.
Notes

1. Data source: https://ncov.dxy.cn/ncovh5/view/pneumonia.
2. Data source: http://qianxi.baidu.com/.
3. Data source: https://www.chinayearbooks.com/china-statistical-yearbook-2018.html.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was funded through support from the National Natural Science Foundation of China [Grant Nos. 42041001, 42071435, 41525004, and 41421001].

ORCID

Tao Pei http://orcid.org/0000-0002-5311-8761
Yaxi Liu http://orcid.org/0000-0002-3261-829X

References

Adekunle, I. A., A. T. Onanuga, O. O. Akinola, O. W. Ogbonjanjo. 2020. "Modelling Spatial Variations of Coronavirus Disease (COVID-19) in Africa." Science of the Total Environment 729: 138998. doi:10.1016/j.scitotenv.2020.138998.

Chen, S., Z. Zhang, J. Yang, J. Wang, X. Zhai, T. Bärnighausen, C. Wang, et al. 2020. "Fangcang Shelter Hospitals: A Novel Concept for Responding to Public Health Emergencies." The Lancet 395 (10232): 1305–1314. doi:10.1016/S0140-6736(20)30744-3.

Chimmula, V. K. R., and L. Zhang. 2020. "Time Series Forecasting of Transmission in Canada Using LSTM Networks." Chaos, Solitons & Fractals 135: 109864. doi:10.1016/j.chaos.2020.109864.

Chinazzi, M., J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore Y Piontti, et al. 2020. "The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak." Science 368 (6489): 395–400. doi:10.1126/science.aba9757.

Desjardins, M. R., A. Hohl, and E. M. Delmelle. 2020. "Rapid Surveillance of COVID-19 in the United States Using a Prospective Space-time Scan Statistic: Detecting and Evaluating Emerging Clusters." Applied Geography 118: 102202. doi:10.1016/j.apgeog.2020.102202.

Ding, J., H. Fu, Y. Liu, J. Gao, Z. Li, X. Zhao, J. Zheng, et al. 2020. "Prevention and Control Measures in Radiology Department for COVID-19." European Radiology 30 (7): 3603–3608. doi:10.1007/s00330-020-06850-5.

Dong, E., H. Du, and L. Gardner. 2020. "An Interactive Web-based Dashboard to Track COVID-19 in Real Time." The Lancet Infectious Diseases 20 (5): 533–534. doi:10.1016/S1473-3099(20)30120-1.

Dowd, J. B., L. Andriano, D. M. Brazel, V. Rotondi, P. Block, X. Ding, Y. Liu, et al. 2020. "Demographic Science Aids in Understanding the Spread and Fatality Rates of COVID-19." Proceedings of the National Academy of Sciences 117 (18): 9696–9698. doi:10.1073/pnas.2004911117.

Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, et al. 2020. "Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing." Science 368 (6491): eabb6936. doi:10.1126/science.abb6936.

Finger, F., T. Genolet, L. Mari, G. C. de Magny, N. M. Manga, A. Rinaldo, E. Bertuzzo, et al. 2016. "Mobile Phone Data Highlights the Role of Mass Gatherings in the Spreading of Cholera Outbreaks." Proceedings of the National Academy of Sciences 113 (23): 6421–6426. doi:10.1073/pnas.1522305113.

Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, et al. 2020. "Estimating the Effects of Non-pharmaceutical Interventions on COVID-19 in Europe." Nature 584 (7820): 257–261. doi:10.1038/s41586-020-2284-y.

Fricke, L. M., S. Glöckner, M. Dreier, B. Lange. 2021. "Impact of Non-pharmaceutical Interventions Targeted at COVID-19 Pandemic on Influenza Burden - a Systematic Review." Journal of Infection 82 (1): 1–35. doi:10.1016/j.jinf.2020.11.039.

Han, Y., Y. Liu, L. Zhou, E. Chen, P. Liu, X. Pan, Y. Lu, et al. 2020. "Epidemiological Assessment of Imported Coronavirus Disease 2019 (COVID-19) Cases in the Most Affected City outside of Hubei Province, Wenzhou, China." JAMA Network Open 3 (4): e206785. doi:10.1001/jamanetworkopen.2020.6785.

He, S., S. Chen, L. Kong, et al. 2020. "Analysis of Risk Perceptions and Related Factors Concerning COVID-19 Epidemic in Chongqing, China." Journal of Community Health: 1–8. doi:10.1007/s10900-019-00710-0.

Hu, B., J. Qiu, H. Chen, V. Tao, J. Wang, H. Lin. 2020. "First, Second and Potential Third Generation Spreads of the COVID-19 Epidemic in Mainland China: An Early Exploratory Study Incorporating Location-based Service Data of Mobile Devices." International Journal of Infectious Diseases 96: 489–495. doi:10.1016/j.ijid.2020.05.048.

Huang, B., J. H. Wang, J. X. Cai, S. Yao, P. K. S. Chan, T. H.-W. Tam, -Y.-Y. Hong, et al. 2021. "Integrated Vaccination and Physical Distancing Interventions to Prevent Future COVID-19 Waves in Chinese Cities." Nature Human Behaviour 5 (6): 695–705. doi:10.1038/s41562-021-01063-2.

Jain, A. K. 2010. "Data Clustering: 50 Years beyond K-means." Pattern Recognition Letters 31 (8): 651–666. doi:10.1016/j.patrec.2009.09.011.

Jia, J. S., X. Lu, Y. Yuan, G. Xu, J. Jia, N. A. Christakis. 2020. "Population Flow Drives Spatio-temporal Distribution of COVID-19 in China." Nature 582 (7812): 389–394. doi:10.1038/s41586-020-2284-y.

Kissler, S. M., C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch. 2020. "Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period." Science 368 (6493): 860–868. doi:10.1126/science.abb5793.

Koo, J. R., A. R. Cook, M. Park, Y. Sun, H. Sun, J. T. Lim, C. Tam, et al. 2020. "Interventions to Mitigate Early Spread of SARS-CoV-2 in Singapore: A Modelling Study." The Lancet Infectious Diseases 20 (6): 678–688. doi:10.1016/S1473-3099(20)30162-6.
