Spatial epidemic dynamics of the COVID-19 outbreak in China

Dayun Kang[a], Hyunho Choi[a], Jong-Hun Kim[b], Jungsoon Choi[c,d,*]

[a] *Department of Applied Statistics, Hanyang University, Seoul, Republic of Korea*
[b] *Department of Social and Preventive Medicine, Sungkyunkwan University School of Medicine, Suwon, Republic of Korea*
[c] *Department of Mathematics, Hanyang University, Republic of Korea*
[d] *Research Institute for Natural Sciences, Hanyang University, Republic of Korea*

A B S T R A C T

Background: On 31 December 2019 an outbreak of COVID-19 in Wuhan, China, was reported. The outbreak spread rapidly to other Chinese cities and multiple countries. This study described the spatio-temporal pattern and measured the spatial association of the early stages of the COVID-19 epidemic in mainland China from 16 January–06 February 2020.
Methods: This study explored the spatial epidemic dynamics of COVID-19 in mainland China. Moran's I spatial statistic with various definitions of neighbours was used to conduct a test to determine whether a spatial association of the COVID-19 infections existed.
Results: The spatial spread of the COVID-19 pandemic in China was observed. The results showed that most of the models, except medical-care-based connection models, indicated a significant spatial association of COVID-19 infections from around 22 January 2020.
Conclusions: Spatial analysis is of great help in understanding the spread of infectious diseases, and spatial association was the key to the spatial spread during the early stages of the COVID-19 pandemic in mainland China.

© 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

On 31 December 2019 the Chinese government first reported an outbreak of coronavirus disease (COVID-19) in Wuhan, the capital of Hubei Province in China. The outbreak rapidly spread from Wuhan into all provinces of China and at least 24 countries. As of 06 February 2020 31,481 cases of COVID-19 were officially confirmed in mainland China, including 639 deaths. A total of 22,112 cases were confirmed in Hubei Province, accounting for 70.89% of the total cases.

Until now, studies evaluating the spatial spread of the COVID-19 pandemic in China are limited. However, understanding the spatial spread of the COVID-19 outbreak is critical to predicting local outbreaks and developing public health policies during the early stages of COVID-19. Previous studies have described the spatial spread of severe acute respiratory syndrome (SARS) in Beijing and mainland China (Meng et al., 2005; Fang et al., 2009). One study also considered the different types of connections between cities to calculate the spatial association (Meng et al., 2005). Other studies have analyzed the epidemic data of the Middle East

respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia using various spatial approaches (Adegboye et al., 2017; Lin et al., 2018; Al-Ahmadi et al., 2019).

This study investigated the spatial epidemic dynamics of the COVID-19 outbreak in mainland China. It also measured and compared the spatial association of the daily epidemic data. Different spatial connection assumptions between the provinces regarding possible pathways for the spread of COVID-19 (Meng et al., 2005) were considered. The objective was to provide spatial dynamic information about the spread of COVID-19 for infection prevention and control.

Materials and methods

Data sources

The COVID-19 dataset was obtained from a Chinese website that provides real-time information on outbreaks of epidemic diseases (https://ncov.dxy.cn/ncovh5/view/pneumonia). The website updates data on newly confirmed cases in mainland China by province and date. There are 31 provinces in mainland China, and this study used 3 weeks' data from 16 January to 06 February 2020, which was during the early stages of COVID-19 in China. Data before 16 January 2020, the very early stage of COVID-19, was not examined because of data

---

* Corresponding author.
   E-mail address: jungsoonchoi@hanyang.ac.kr (J. Choi).

reliability concerns. Other datasets – such as population, population density, number of licensed doctors, and hospital and health centre beds per 1000 inhabitants by province – were acquired from a website (Statista, 2020). All population-related and medical resource datasets were collected in 2018; these were the most recent data that could be obtained.

Figure 1 shows a map of cumulative cases by province. The number of cumulative cases is the sum of the newly confirmed cases from 16 January–06 February 2020. The largest number of cases was in Hubei Province, of which Wuhan is the capital city. Figure 2 presents the population and population density (populaiton/km^2) for each province in 2018. Guangdong and Shanghai had the highest population and population density, respectively. Hubei ranked ninth in population and thirteenth in population density. As shown in Figure 3, Shandong has the highest number of doctors and hospital beds (per 1000 inhabitants), whereas Hubei ranks ninth and seventh for the number of doctors and hospital beds, respectively. Table 1 shows detailed information for each province.

Spatial analysis

To show the spatial association of COVID-19, Moran's I statistic was used for each day with various types of neighbourhoods (Li and Calder, 2007). Moran's I statistic measures the spatial autocorrelation and is calculated as follows:

$$I = \frac{n \sum_{i,j} W_{ij}(Y_i - \overline{Y})(Y_j - \overline{Y})}{\sum_{i \neq j} W_{ij} \sum_i (Y_i - \overline{Y})^2},$$

where i and j were the region indexes and W_{ij} indicated the adjacency between area i and area j. This study considered different types of adjacency. Y_i and Y_j denoted the number of

newly confirmed cases in areas i and j, respectively, and $\overline{Y}$ was the average of the number of newly confirmed cases in the entire region. A value of 0 indicated that there was no spatial autocorrelation in the data. A positive Moran's I value indicated the clustering of similar values, whereas a negative Moran's I value indicated the clustering of dissimilar values. The larger the absolute Moran's I value, the stronger the spatial autocorrelation.

The number of cases in this study was skewed, so the spatial dependency may not have been properly captured. Therefore, to adjust for the skewness, logarithmic transformation of the newly confirmed cases was used instead of the number of cases itself. Because there were many 0 in the dataset, 0.01 was added to the data for log transformation.

Similar to that in a previous study (Meng et al., 2005), six different types of neighbourhoods were used. In Model 1, two provinces were considered adjacent if they shared a border. In Model 2, the distance between two provinces was used. In this case, the centroid for each province was determined using the gCentroid function in the rgeos package of statistical software R (Bivand et al., 2019). Thined as the Euclidean distance between the centroids of these provinces. The extent to which the two provinces were adjacent was defined as the inverse of the distance. In Models 1 and 2, spatial adjacency was defined by geographical information, which is the usual method for examining spatial relationships. As COVID-19 is spread from person to person, population and population density were the key foci. Thus, Models 3 and 4 considered population and population density. The population (population density) for each province was ranked. A province was defined as adjacent to both the previous and following ranked provinces. Thus, the first-ranked and last-ranked provinces only had one adjacent neighbour. In terms of medical care resources, Models 5 and 6 considered the number of doctors and hospitals or medical centre beds. The definition of an adjacent neighbour was the same as that in Models 3 and 4. Moran's I
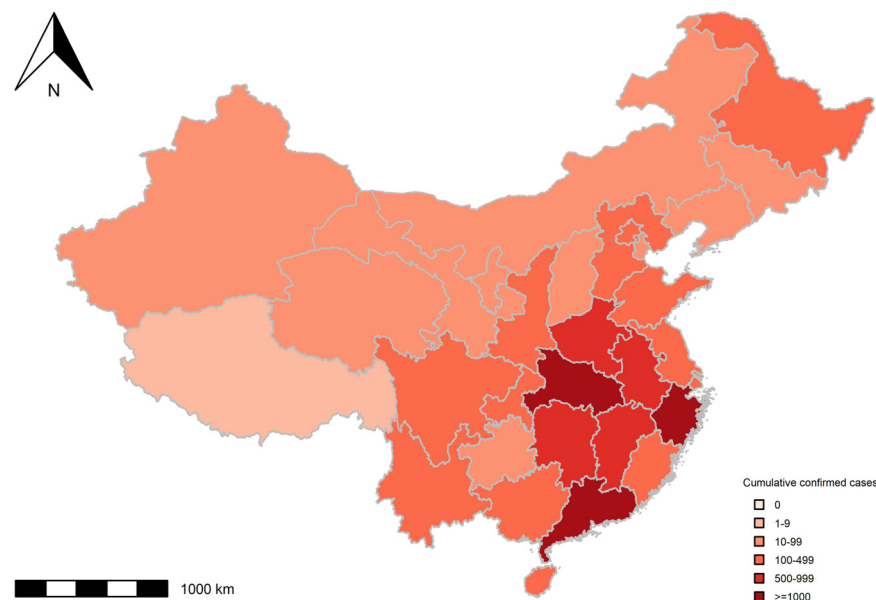

Figure 1. Map of the cumulative cases of COVID-19 in mainland China.

(a)

Population
- ☐ 3440000-25105000
- ☐ 25105000-38640000
- ☐ 38640000-61205000
- ■ 61205000-113460000

1000 km

(b)

Population density
(population per km²)
- ☐ 2.80-133.43
- ☐ 133.43-278.49
- ☐ 278.49-508.61
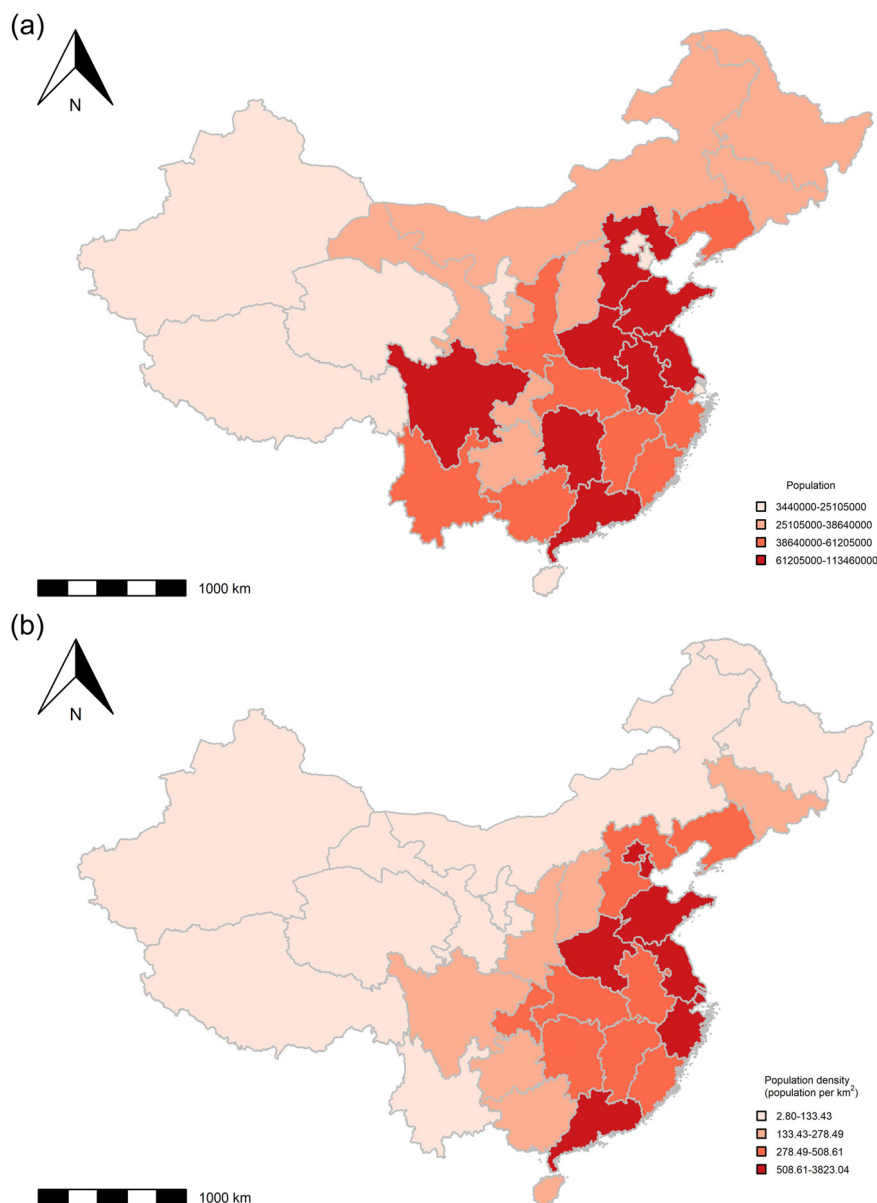- ■ 508.61-3823.04

1000 km

Figure 2. Choropleth map of population information. (a) population; (b) population density.

function was used in the ape package of the statistical software R (Paradis et al., 2019). The significance level of Moran's I test was 0.05.

Results

Figure 4 shows the time series plot of the newly confirmed cases for each day. The number of cases for each day was the sum of all cases in mainland China. As shown in Figure 4, the number of cases increased almost exponentially.

To prevent an exponential spread over mainland China, it was important to detect the spatial spread in the early stages. Because COVID-19 spread from Hubei Province, the epicentre of the outbreak, the number of newly confirmed cases in the provinces neighbouring Hubei was investigated. The provinces of Hunan, Sichuan and Tianjin were selected as representative areas of first-order, second-order and third-order neighbouring provinces, respectively. The daily number of confirmed cases in Hubei is shown in the upper panel of Figure 5. The lower panel of Figure 5 shows the daily number of confirmed cases in Hunan, Sichuan and
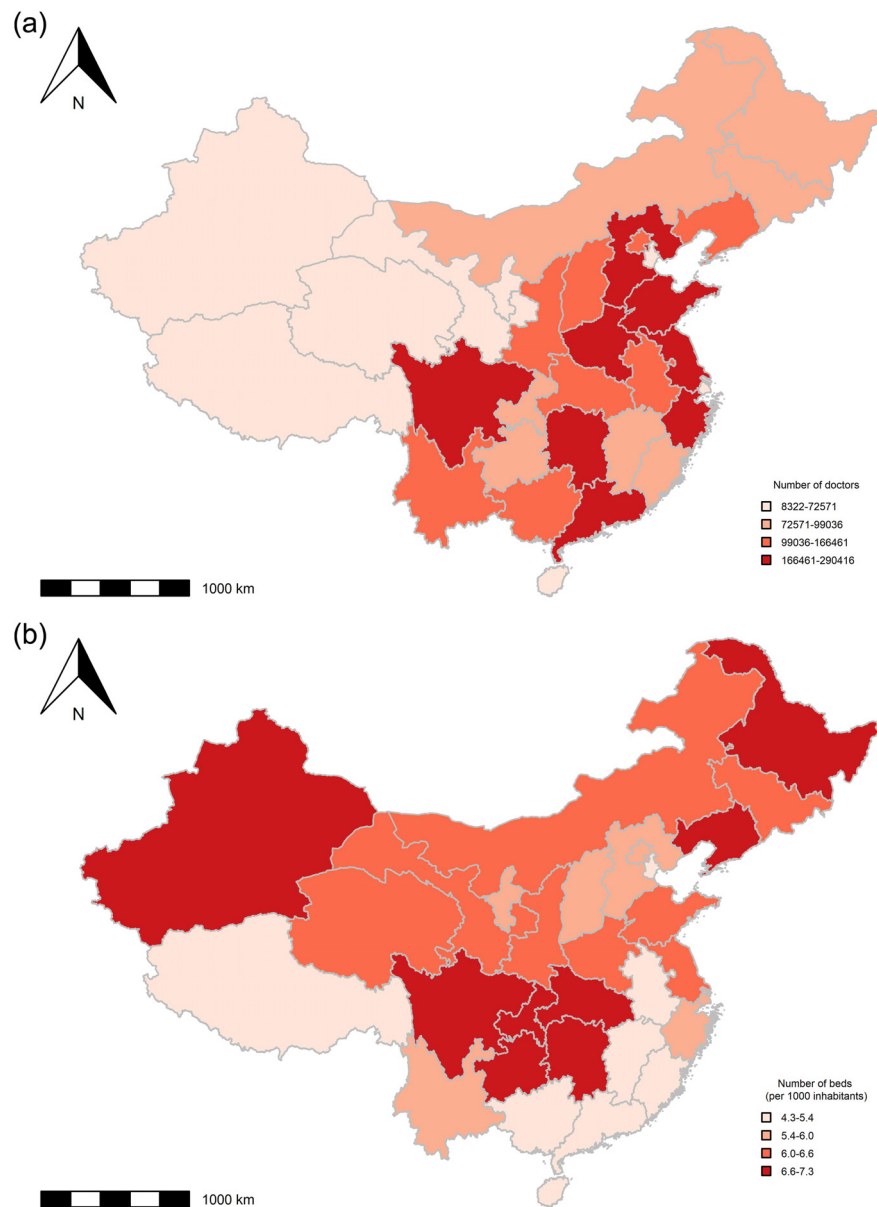
Figure 3. Choropleth map of medical care information. (a) Number of doctors; (b) Number of hospital beds.

Tianjin. From 22 January the number of newly confirmed cases in Hunan clearly increased. The infection first increased in Hubei and then in the first-order neighbouring provinces such as Hunan, and the second-order neighbouring provinces such as Sichuan. The infection finally spread to the third-order neighbouring provinces, including Tianjin. This supports the fact that COVID-19 spread spatially and that investigation of spatial dependency is very essential.

This study examined whether a spatial association existed in the cases of COVID-19 in China. It used Moran's I statistic, a measure of spatial association, for the number of confirmed cases with different types of neighbourhoods. Figure 6 shows Moran's I statistic and its p-value for each day in Models 1–6. Overall, the p-values in Figure 6 are very close to a value of 0 in Models 1–5, except for the first few days. On 22 January Models 1-4 first detected a significant spatial dependency on the number of newly confirmed cases. Since approximately 24 January, the number of newly confirmed cases show significant spatial dependency in Models 1 and 2. The maximum value of Moran's I statistic in Models 1 and 2 is 0.4598 and 0.0841, respectively. The further the statistic is from 0, the stronger the spatial dependency. Therefore, the numbers 0.4598 and 0.0841 are significant, with p-values < 0.05. For population-related

Table 1
Data from the provinces in mainland China.

Province name	Number of cumulative cases	Population (×10000)	Population density (population/km²)	Number of doctors	Number of hospital beds (per 1000 inhabitants)
Anhui	665	6324	453.66	126,824	5.19
Beijing	297	2154	1,312.53	99,807	5.74
Chongqing	411	3102	376.46	76,379	7.1
Fujian	224	3941	318.08	91,110	4.88
Gansu	67	2637	61.93	59,560	6.17
Guangdong	1018	11,346	631.39	276,361	4.56
Guangxi	172	4926	207.32	105,979	5.2
Guizhou	75	3600	204.31	81,475	6.82
Hainan	111	934	264.19	22,289	4.8
Hebei	171	7556	400.21	211,387	5.58
Heilongjiang	263	3773	82.96	89,489	6.63
Henan	914	9605	575.15	235,649	6.34
Hubei	22,112	5917	318.29	152,040	6.65
Hunan	772	6899	325.73	180,882	6.99
Jiangsu	408	8051	784.7	233,263	6.11
Jiangxi	661	4648	278.49	87,304	5.37
Jilin	65	2704	144.29	77,108	6.18
Liaoning	94	4359	293.73	120,431	7.21
Inner Mongolia	50	2534	21.42	73,563	6.27
Ningxia	43	688	103.61	19,435	5.96
Qinghai	18	603	8.35	16,153	6.49
Shaanxi	184	3864	187.76	99,036	6.57
Shandong	379	10,047	639.53	290,416	6.06
Shanghai	269	2424	3,823.04	71,580	5.74
Shanxi	96	3718	237.27	99,490	5.6
Sichuan	344	8,341	171.59	204,956	7.18
Tianjin	81	1,560	1,309.05	43,105	4.37
Xinjiang	39	2,487	14.94	63,312	7.19
Tibet	1	344	2.8	8322	4.88
Yunnan	135	4830	122.56	99,669	6.03
Zhejiang	1006	5737	563.56	190,782	5.79

neighbourhoods, both Models 3 and 4 show a spatial clustering tendency since 22 January, except for 2 and 4 days, respectively. Among the days with significant spatial dependency, the maximum value is 0.6991 and 0.7336 in Models 3 and 4, respectively. Models 3 and 4 also have significant p-values < 0.05. For medical-care-based neighbourhoods, Model 5 shows a spatial association since 23 January except 2 days. However, no spatial association exists in Model 6. Since 23 January, the average p-value is 0.0129 and 0.6638 in Models 5 and 6, respectively, which show a significant difference.

Discussion

This study is the first to provide information on the spatial and temporal patterns of the COVID-19 pandemic in mainland China. In the early stage of the COVID-19 outbreak, new cases occurred intensively in the Hubei province. Over time, the cases spread to provinces neighbouring Hubei; the first-order neighbouring provinces showed a particularly increased number of confirmed cases after 22 January. Then, the second-order and third-order provinces showed a steeply increasing number of cases from 23 January and 24 January, respectively. This shows the spread of COVID-19. Eventually, the impact spread to all provinces in mainland China.

This study investigated the spatial dependency through Moran's I with different types of spatial connections. Except for the medical centre bed-based neighbourhood, a spatial clustering tendency was observed in every neighbourhood type from approximately 22 January. The regions connected by express trains to Wuhan – such as Shenzhen, Shanghai and Beijing – had five, two and two confirmed cases, respectively on 21 January, the early stage of COVID-19 in China. This implies the possibility that
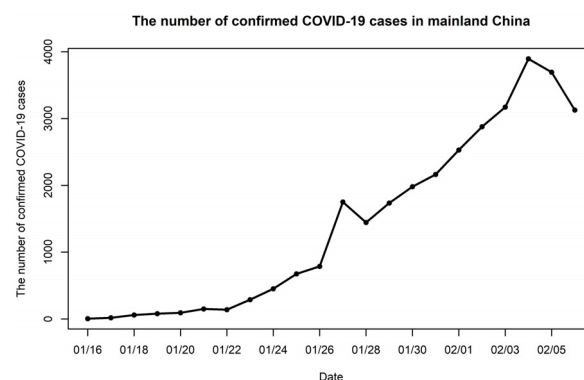

Figure 4. Time series plot of the number of newly confirmed COVID-19 cases in mainland China.

COVID-19 had spread from Wuhan to other areas via the transportation (Zhao et al., 2020). This possibility also supports the spatial dependency we detected in this study. On 23 January, the Chinese government closed off Wuhan City to prevent the spread of COVID-19. These findings could link with such a government policy. The results of the evaluation using geographical and distance-based neighbourhoods showed that COVID-19 is highly likely to spread between geographically adjacent regions. This may be because people in adjacent regions tend to interact with each other. In addition, Moran's I using population-based neighbourhoods also showed a strong spatial association. More
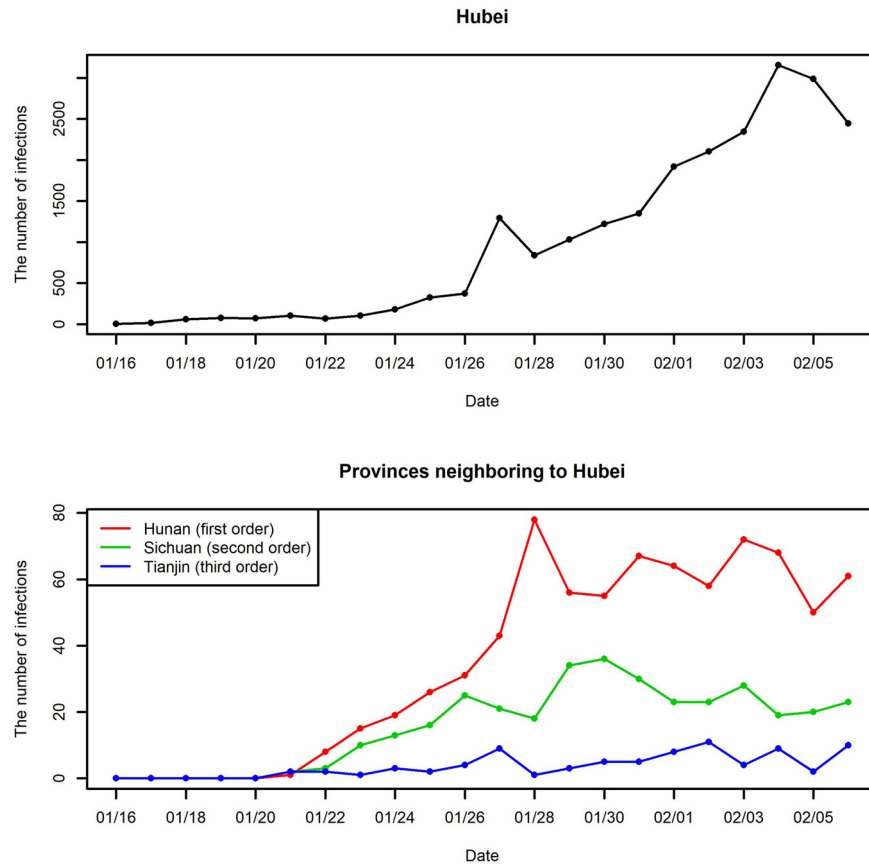
Figure 5. Plots of the incidence in Hubei (upper panel) and in provinces neighbouring Hubei (lower panel).
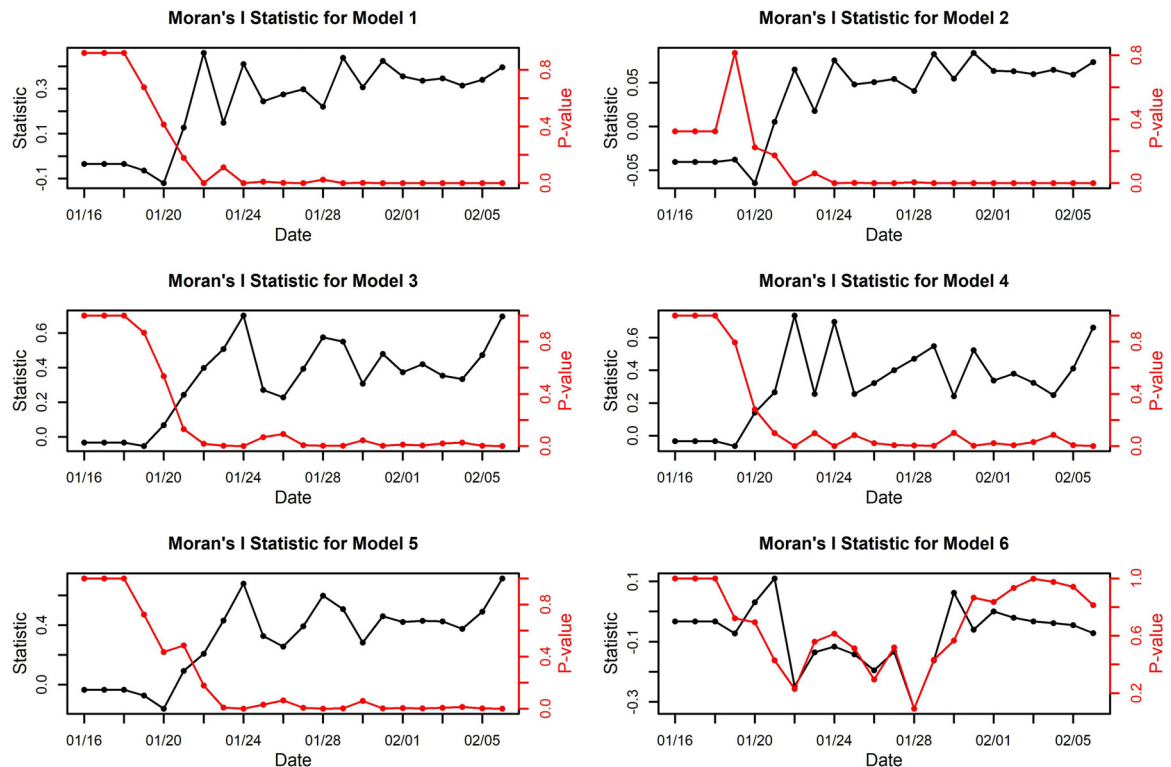

Figure 6. Plots of Moran's I statistic and *p*-values.

people are likely to be infected with the virus in densely populated regions, which leads to the active spread of COVID-19 to other areas. Finally, having many doctors in a region indicates that the region can accommodate many severely ill patients, which can lead to the spread of the virus. This result is consistent with that of a previous study (Meng et al., 2005).

In addition, this study conducted the same spatial analysis using the ranks of the newly confirmed cases in a nonparametric approach because the data are quite skewed. The results were almost the same, except that there was a spatial association for a few more days.

Conclusions

COVID-19 has been affecting countries worldwide, and the World Health Organization has declared the COVID-19 outbreak a public health emergency of international concern. This study demonstrated that in the early stages of the COVID-19 pandemic, the disease dramatically spread from region to region in mainland China. Examining the spatial spread in the early stages is very important to prevent further transmission. It is believed that this study is the first to investigate the virus's spatial spread to various types of neighbourhoods in mainland China.

Although this study was conducted in the early stages of the COVID-19 outbreak to determine whether there was a spatial association, it did have a few limitations. First, it used the reported dataset for the daily number of newly confirmed cases in the 31 provinces of China. This did not include the number of suspected cases, so it was difficult to understand the spatio-temporal transmission of COVID-19. However, it was important to investigate the spatial and temporal characteristics of the COVID-19 outbreak at an early stage. Second, it considered six types of neighbourhoods; other types of neighbourhoods were not covered in this study, such as the urban-rural relationship, which might have also been significant (Meng et al., 2005). Third, it only investigated spatial spread in mainland China. As infections have also occurred in other countries, investigating the global spatial spread of COVID-19 might be important to manage COVID-19.

Future research, such as a study examining the spatial tendencies of the deaths and recoveries from COVID-19, will contribute to the control and prevention of this disease. Through such work, it will be able to be determined which factors affect death and recovery.

Ethics approval and consent to participate: No human or animal samples were included in the research presented in this article; therefore, ethical approval was not necessary.

Availability of data and materials: The datasets used and analysed during the current study are available from the websites https://ncov.dxy.cn/ncovh5/view/pneumonia and http://statista. com.

Acknowledgements

References

Adegboye OA, Gayawan E, Hanna F. Spatial modelling of contribution of individual level risk factors for mortality from Middle East respiratory syndrome coronavirus in the Arabian Peninsula. PLoS One 2020;12(7):e0181215, doi: http://dx.doi.org/10.1371/journal.pone.0181215.

Al-Ahmadi K, Alhmadi S, Al-Zahrani A. Spatiotemporal clustering of Middle East respiratory syndrome coronavirus (MERS-CoV) incidence in Saudi Arabia, 2012–2019. Int J Environ Res Public Health 2019;16(14):2520.

Roger Bivand, Colin Rundel, Edzer Pebesma, Rainer Stuetz, Karl Ove Hufthammer, Patrick Giraudoux, et al. rgeos: Interface to Geometry Engine. 2019 Available from: https://CRAN.R-project.org/package=rgeos. [Accessed 13 April 2020].

Fang LQ, Vlas SJ, Feng D, Liang S, Xu YF, Zhou JP, et al. Geographical spread of SARS in mainland China. Trop Med Int Health 2009;14:14–20.

Li H, Calder CA, Cressie N. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. Geogr Anal 2007;39(4):357–75.

Lin Q, Chiu AP, Zhao S, He D. Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. Stat Methods Med Res 2018;27(7): 1968–78.

Meng B, Wang J, Liu J, Wu J, Zhong E. Understanding the spatial diffusion process of severe acute respiratory syndrome in Beijing. Public Health 2005;119 (12):1080–7.

Emmanuel Paradis, Simon Blomberg, Ben Bolker, Joseph Brown, Julien Claude, Hoa Sien Cuong, et al. ape: Analyses of Phylogenetics and Evolution. 2019. Available from: https://CRAN.R-project.org/package=ape. [Accessed 13 April 2020].

Zhao S, Zhuang Z, Ran J, Lin J, Yang G, Yang L, et al. The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: a data-driven correlational report. Travel Med Infect Dis 2020;101568, doi:http://dx.doi.org/10.1016/j.tmaid.2020.101568.

http://statista.comStatista. 2020 (accessed 21 February 2020).