Spatial epidemic dynamics of the COVID-19 outbreak in China

Dayun Kang¹, Hyunho Choi¹, Jong-Hun Kim², Jungsoon Choi³,⁴*

¹ Department of Applied Statistics, Hanyang University, Seoul, Republic of Korea; dayun4927@hanyang.ac.kr (D.K.); gusgh4950@hanyang.ac.kr (H.C.)
² Department of Social and Preventive Medicine, Sungkyunkwan University School of Medicine, Suwon, Republic of Korea; kimjh32@skku.edu (J.H.K.)
³ Department of Mathematics, Hanyang University, Republic of Korea; jungsoonchoi@hanyang.ac.kr (J.C.)
⁴ Research Institute for Natural Sciences, Hanyang University, Republic of Korea

* Corresponding author’s e-mail address: jungsoonchoi@hanyang.ac.kr; Tel.: +82-2-2220-2621

Highlights

- Spatial dynamics of the COVID-2019 in mainland China
- Moran’s I spatial statistic with different types of neighborhoods
- Consideration of population and medical-care related information
- Need for spatial analysis to prevent the spread of the infectious diseases

Abstract

Background: On December 31, 2019, an outbreak of COVID-19 in Wuhan, China, was reported. The outbreak spread rapidly to other Chinese cities and to multiple countries. We describe the spatio-temporal pattern and measure the spatial association of the early stages of the COVID-19 epidemic in mainland China from January
16 to February 6, 2020.

**Methods:** We explored the spatial epidemic dynamics of COVID-19 in mainland China. Moran’s I spatial statistic with various definitions of neighbors was used to conduct a test to determine whether a spatial association of the COVID-19 infections existed.

**Results:** We observed the spatial spread of the COVID-19 pandemic in China. 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 January 22, 2020.

**Conclusions:** Spatial analysis is of great help in understanding the spread of infectious diseases, and spatial association is the key to the spatial spread during the early stages of the COVID-19 pandemic in mainland China.

**Keywords:** COVID-19, Spatial autocorrelation, Spatial analysis, China

**Background**

On December 31, 2019, the Chinese government first reported an outbreak of coronavirus disease (COVID-19) in Wuhan, the capital of Hubei Province in China. The outbreak spread rapidly from Wuhan into all provinces of China and at least 24 countries. As of February 6, 2020, 31481 cases of COVID-19 were officially confirmed in mainland China, including 639 deaths. Particularly, 22112 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).
We investigated the spatial epidemic dynamics of the COVID-19 outbreak in mainland China. We also measured and compared the spatial association of the daily epidemic data. We considered different spatial connection assumptions between the provinces regarding possible pathways for the spread of COVID-19 (Meng et al., 2005). Our objective was to provide spatial dynamic information about the spread of COVID-19 for infection prevention and control.

**Materials and methods**

*Data sources*

We obtained the COVID-19 dataset 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 the newly confirmed cases in mainland China by province and date. There are 31 provinces in mainland China, and we used 3 weeks’ data from January 16 to February 6, 2020, the early stage of COVID-19 in China. In this study, we did not consider the data before January 16, the very early stage of COVID-19, because of data reliability concerns. Other datasets, such as population, population density, number of licensed doctors, and hospital and health center 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 we could obtain.

Figure 1 shows a map of cumulative cases by province. The number of cumulative cases in Figure 1 is the sum of the newly confirmed cases from January 16 to February 6, 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 for each province in 2018. Guangdong and Shanghai have the highest population and population density, respectively. Hubei ranks 9th in population and 13th in population density. As shown in Figure 3, Shandong has the highest number of doctors and hospital beds, whereas Hubei ranks 9th and 7th 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, we used Moran’s I statistic for each day with various types of neighborhoods (Li et al., 2007). Moran’s I statistic measures the spatial autocorrelation and is calculated as follows:

\[
I = \frac{n \sum_{i,j} W_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i,j} W_{ij} \sum_i (Y_i - \bar{Y})^2},
\]

where \(i\) and \(j\) are the region indexes and \(W_{ij}\) indicates the adjacency between area \(i\) and area \(j\). In this study, we considered different types of adjacency. \(Y_i\) and \(Y_j\) denote the number of newly confirmed cases in areas \(i\) and \(j\), respectively, and \(\bar{Y}\) is the average of the number of newly confirmed cases in the entire region. A value of zero indicates that there is no spatial autocorrelation in the data. A positive Moran’s I value indicates the clustering of similar values, whereas a negative Moran’s I value indicates the clustering of dissimilar values. The larger the absolute Moran’s I value, the stronger the spatial autocorrelation.

In this study, the number of cases is skewed and thus the spatial dependency may not be properly captured. Therefore, to adjust for the skewness, we used logarithmic transformation of the newly confirmed cases instead of the number of cases itself. Because we had many zeros in the dataset, we added 0.01 to the data for log transformation.

Similar to that in a previous study (Meng et al., 2005), we used six different types of neighborhoods. In Model 1, two provinces are considered adjacent if they share a border. In Model 2, we used the distance between two provinces. In this case, we determined the centroid for each province using the gCentroid function in the rgeos package of the statistical software R. The distance between two provinces is defined as the Euclidean distance between the centroids of these provinces. The extent to which the two provinces are adjacent is defined as the inverse of the distance. In Models 1 and 2, spatial adjacency is 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 are the key foci. Thus, Models 3 and 4 consider population and population density. We ranked the population (population density) for each province. A province is defined as adjacent to both the previous and the following ranked provinces. Thus, the first-ranked and last-ranked provinces have only one adjacent neighbor. In terms of medical care resources, Models 5 and 6 consider the number of doctors and hospital or medical center beds. The definition of an adjacent neighbor is the same as that in Models 3 and 4. We
used Moran’s I function in the ape package of the statistical software R. The significance level of Moran’s I test is 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 is the sum of all cases in mainland China. As shown in Figure 4, the number of cases has increased almost exponentially. To prevent an exponential spread over mainland China, it is important to detect the spatial spread in the early stages.

Because COVID-19 spread from Hubei Province, the epicenter of the outbreak, we investigated the number of new confirmed cases in the provinces neighboring Hubei. We selected the provinces of Hunan, Sichuan, and Tianjin as representative areas of first-order, second-order, and third-order neighboring 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 Tianjin. From January 22, the number of newly confirmed cases in Hunan clearly increased. The infection first increased in Hubei and then in the first-order neighboring provinces, such as Hunan, and the second-order neighboring provinces, such as Sichuan. The infection finally spread to the third-order neighboring provinces, including Tianjin. This supports the fact that COVID-19 spread spatially and that investigation of spatial dependency is very essential.

In this study, we examined whether a spatial association existed in the cases of COVID-19. We used Moran’s I statistic, a measure of spatial association, for the number of confirmed cases with different types of neighborhoods. 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 the x-axis in Models 1–5, except for the first few days. Since approximately January 22, the number of new confirmed cases showed significant spatial dependency in Models 1 and 2. The maximum value of Moran’s I statistic in Models 1 and 2 was 0.4598 and 0.0841, respectively. The farther the statistic is from zero, the stronger the spatial dependency. Therefore, the numbers 0.4598 and 0.0841 are significant, with $P$-values < 0.05. For population-related neighborhoods, both Models 3 and 4 showed a spatial clustering tendency since January 22, except for two and four days, respectively. Among the days with significant spatial dependency, the maximum value was 0.6991 and 0.7336 in Models 3 and 4, respectively.
Models 3 and 4 also had significant $P$-values < 0.05. For medical-care-based neighborhoods, Model 5 showed a spatial association since January 23. However, no spatial association existed in Model 6. Since January 23, the average $P$-value was 0.0129 and 0.6638 in Models 5 and 6, respectively, which shows 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 neighboring Hubei; especially, the first-order neighboring provinces showed an increased number of confirmed cases after January 22. Then, the second-order and third-order provinces showed a steeply increasing number of cases from January 23 and January 24, respectively. This shows the spread of COVID-19. Eventually, the impact spread to all provinces in mainland China.

We investigated the spatial dependency through Moran’s I with different types of spatial connections. Except for the number-of-hospital-bed-based neighborhood, a spatial clustering tendency was observed in every neighborhood type from approximately January 22. The regions connected by express trains to Wuhan, such as Shenzhen, Shanghai, and Beijing, had 5, 2, and 2 confirmed cases, respectively on January 21 (Zhao et al., 2020). This supports the idea that COVID-19 spread via the traffic network. Our results show that the spatial association of infections was detected on January 22. On January 23, the Chinese government closed off Wuhan City to prevent the spread of COVID-19. Our findings could link with such government policy. The results of our evaluation using geographical and distance-based neighborhoods 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 neighborhoods also showed a strong spatial association. More 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, we 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. To the best of our knowledge, this study is the first to investigate the virus’s spatial spread to various types of neighborhoods in mainland China.

Although we conducted this study in the early stage of the COVID-19 outbreak to determine whether there was a spatial association, our study has a few limitations. First, we used the reported dataset for the daily number of newly confirmed cases in the 31 provinces of China. This does not include the number of suspected cases, so it is difficult to understand the spatio-temporal transmission of COVID-19. However, it is important to investigate the spatial and temporal characteristics of the COVID-19 outbreak at an early-stage. Second, we considered only six types of neighborhoods, but other types of neighborhoods not covered in this study, such as the urban-rural relationship, might also be significant (Meng et al., 2005). Third, we only investigated spatial spread in mainland China. As infections have also occurred in other countries, investigating the spatial spread of COVID-19 worldwide 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, we will be able to determine 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 analyzed during the current study are available from the websites https://ncov.dxy.cn/ncovh5/view/pneumonia and http://statista.com.

Competing interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the research fund of the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07047712) and by the Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG18C0088).

Authors’ contributions

J.C. designed the study; D.K. and H.C. contributed to data acquisition; D.K., H.C., and J.C. carried out the statistical analysis; D.K., H.C., J.H.K., and J.C. drafted the manuscript. All authors contributed to the interpretation of data and revision of the manuscript. All authors read and approved the final manuscript.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thanks to all of the persons who were struggling in healthcare fields to overcome the COVID-19 outbreak. This study were performed under the research project named ‘Research and Development on Integrated Surveillance System for Early Warning of Infectious Diseases (RISEWIDs).’
References

1. 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. 2017;12(7):e0181215; https://doi.org/10.1371/journal.pone.0181215.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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; https://doi.org/10.1016/j.tmaid.2020.101568.

8. Statista. http://statista.com, 2020 (accessed 21 February 2020).
Figure 1. Map of the cumulative cases of COVID-19 in mainland China.
Figure 2. Map of the population (left panel) and population density (right panel) in mainland China.
Figure 3. Map of the number of doctors (left panel) and hospital beds (right panel) in mainland China.
Figure 4. Time series plot of the number of newly confirmed COVID-19 cases in mainland China.
Figure 5. Plots of the incidence in Hubei (upper panel) and in provinces neighboring Hubei (lower panel).
Figure 6. Plots of Moran's I statistic and $P$-values.
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                              | 126824            | 5.19                                          |
| Beijing       | 297                       | 2154                | 1312.53                             | 99807             | 5.74                                          |
| Chongqing     | 411                       | 3102                | 376.46                              | 76379             | 7.1                                           |
| Fujian        | 224                       | 3941                | 318.08                              | 91110             | 4.88                                          |
| Gansu         | 67                        | 2637                | 61.93                               | 59560             | 6.17                                          |
| Guangdong     | 1018                      | 11346               | 631.39                              | 276361            | 4.56                                          |
| Guangxi       | 172                       | 4926                | 207.32                              | 105979            | 5.2                                           |
| Guizhou       | 75                        | 3600                | 204.31                              | 81475             | 6.82                                          |
| Hainan        | 111                       | 934                 | 264.19                              | 22289             | 4.8                                           |
| Hebei         | 171                       | 7556                | 400.21                              | 211387            | 5.58                                          |
| Heilongjiang  | 263                       | 3773                | 82.96                               | 89489             | 6.63                                          |
| Henan         | 914                       | 9605                | 575.15                              | 235649            | 6.34                                          |
| Hubei         | 22112                     | 5917                | 318.29                              | 152040            | 6.65                                          |
| Hunan         | 772                       | 6899                | 325.73                              | 180882            | 6.99                                          |
| Jiangsu       | 408                       | 8051                | 784.7                               | 233263            | 6.11                                          |
| Jiangxi       | 661                       | 4648                | 278.49                              | 87304             | 5.37                                          |
| Jilin         | 65                        | 2704                | 144.29                              | 77108             | 6.18                                          |
| Liaoning      | 94                        | 4359                | 293.73                              | 120431            | 7.21                                          |
| Inner Mongolia| 50                        | 2534                | 21.42                               | 73563             | 6.27                                          |
| Ningxia       | 43                        | 688                 | 103.61                              | 19435             | 5.96                                          |
| Qinghai       | 18                        | 603                 | 8.35                                | 16153             | 6.49                                          |
| Shaanxi       | 184                       | 3864                | 187.76                              | 99036             | 6.57                                          |
| Province   | Population | Income | Price per Capita | GDP  | Growth Rate |
|------------|------------|--------|------------------|------|-------------|
| Shandong   | 379        | 10047  | 639.53           | 290416 | 6.06        |
| Shanghai   | 269        | 2424   | 3823.04          | 71580 | 5.74        |
| Shanxi     | 96         | 3718   | 237.27           | 99490 | 5.6         |
| Sichuan    | 344        | 8341   | 171.59           | 204956 | 7.18        |
| Tianjin    | 81         | 1560   | 1309.05          | 43105 | 4.37        |
| Xinjiang   | 39         | 2487   | 14.94            | 63312 | 7.19        |
| Tibet      | 1          | 344    | 2.8              | 8322  | 4.88        |
| Yunnan     | 135        | 4830   | 122.56           | 99669 | 6.03        |
| Zhejiang   | 1006       | 5737   | 563.56           | 190782 | 5.79        |