Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Identification of the high-risk residence communities and possible risk factors of COVID-19 in Wuhan, China

Xiaojing Guo, Xinyue Zhou, Fengshi Tian, Hui Zhang

A R T I C L E   I N F O

Keywords:
COVID-19
Residence community
Risk factor
Correlation analysis

A B S T R A C T

The coronavirus disease 2019 (COVID-19) has become a public health emergency of international concern. It is important to identify high-risk residence communities and the risk factors for decision making on targeted prevention and control measures. In this paper, the number of confirmed and suspected cases of COVID-19 in the residence communities in Wuhan, China was collected together with the characteristic variables of the residence communities and the distances between the residence communities and nearby crowded places. The correlation analysis was conducted between the number of confirmed cases and the characteristic/distance variables. Concerning the characteristic variables, there are significant positive correlations between the number of COVID-19 confirmed cases and the construction area, covered area, total number of houses, total number of buildings, volume ratio, property charge, and number of second-hand houses in the residence communities in Wuhan, while minor or no correlation is observed for the average price of houses, construction year, greening ratio, or number of sold houses. Concerning the distance variables, there are significant negative correlations between the number of confirmed cases and the distances from the residence communities to the nearest universities, business clusters, and railway stations, while minor or no correlation is observed for the Huanan Seafood Wholesale Market, kindergartens, primary schools, middle schools, shopping malls, cinemas, subway stations, bus stops, inter-city bus stations, airport, general hospitals, or appointed hospitals for COVID-19 pandemic. Therefore, the residence communities which are newly-built, where the volume ratio or property charge is high or the construction area, covered area, or total number of houses, buildings, second-hand houses, or sold houses is large, or which are close to universities, business clusters, subway stations, or railway stations are the high-risk ones where strict measures should be taken. This study provides the authorities with a valuable reference for precise disease prevention and control on the residence community level in similar cities in the world.

1. Introduction

The coronavirus disease 2019 (COVID-19) has spread over the world and developed into a public health emergency of international concern (PHEIC) on January 30, 2020 [1,2]. Some findings have proved the person-to-person transmission of the novel coronavirus [3] and its infectivity during the incubation period [4], which is a big challenge for disease prevention and control [5,6]. It greatly affects the social and economic development as well as people’s normal work and life [7,8]. For example, self-isolation, social distancing, and travel restrictions have caused unemployment and panic-buying of food has increased food demand [9]. The prevention measures of the pandemic have also impacted the wellbeing of disadvantaged groups [10]. Although the transmission of COVID-19 has been effectively suppressed in China, there are still a small number of imported cases from abroad and local cases emerging, implying the risk of a rebound [11]. Therefore, the containment of the epidemic and the resumption of normal work and life should be carried on simultaneously, which requires targeted prevention and control measures on a fine scale.

Wuhan, the capital city of Hubei Province, China, had suffered a lot from the threat of COVID-19 since December 2019, which has drawn much attention from the researchers [12–16]. The relationship between the epidemic data and the socioeconomic factors may be discovered from the early transmission of COVID-19 in Wuhan when the effects of the confounders such as countermeasures can be ignored. Therefore, we conducted a correlation analysis between the number of COVID-19 confirmed cases by February 14, 2020 in the residence communities in Wuhan and the characteristic data of these residence communities as well as the distances from the residence communities to some kinds of

© 2019 The authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
crowded places. In this way, the socioeconomic factors including the characteristic variables and distance variables that may be correlated with the number of confirmed cases can be found out and the residence communities with high risk can be identified accordingly. This study would provide valuable guidance for the authorities to prioritize targeted prevention and control measures in high-risk residence communities, making interventions accurate to smaller units and minimizing the impact on the normal work and life in low-risk areas. These suggestions could be applied to the cities with similar situation in the world.

2. Literature review

Numerous studies have used infectious disease dynamics models to estimate the progression of COVID-19 under some assumptions, based on which valuable insights were provided for the authorities to take effective prevention and control measures at the macro level [17–23]. A comprehensive systematic review investigated the effects of physical distancing, face masks, and eye protection on the transmission of COVID-19, which provided specific instructions on these simple and common intervention methods [24]. A mathematical model was used to assess the effects of case isolation and contact tracing on the control of COVID-19 outbreaks [25]. Besides, a lot of papers have studied the impact of some natural factors on the severity of COVID-19 such as temperature [26–29], relative humidity [26,28,30], absolute humidity [26], wind [30], and nitrogen dioxide (NO₂) levels [31]. Since the epidemic of infectious diseases like COVID-19 is influenced not only by meteorological factors but also by socioeconomic factors, the evaluation of socioeconomic factors is also important [32]. Furthermore, making precise prevention and control strategies depends on the analyses of socioeconomic influencing factors to assess the spatial risk of COVID-19 and identify high-risk locations on a fine scale. The spatial transmission of COVID-19 and its association with several socioeconomic factors were also investigated [33–36]. The impacts of several socioeconomic factors on the COVID-19 transmission in China were examined [37]. The potential risk zones were identified based on nine socioeconomic variables in Chinese megacities [38]. We would like to take a step further to study the transmission pattern on the residence community level, which contributes to identifying high-risk residence communities and discovering possible risk sources.

3. Methods

3.1. Data collection and processing

3.1.1. Epidemic data of residence communities in Wuhan

Wuhan is the capital city of Hubei Province and one of the megacities in China. It acts as an important industrial center, science and education center, and comprehensive transportation hub in China. During the COVID-19 outbreak, the epidemic situation in Wuhan was issued by Wuhan Municipal Health Commission every day. Many communities in Wuhan have reported the cumulative number of confirmed cases and suspected cases of COVID-19 in the communities. These epidemic data in the residence communities in Wuhan by February 14, 2020 were collected for analysis. The suspected cases are likely to be converted into confirmed cases, while the possibility of conversion is not a constant value [15,39]. Therefore, four epidemic variables corresponding to four different conversion possibilities were calculated from the number of confirmed cases and suspected cases in a residence community. The four epidemic variables were N₀, N₄₀, N₇₀, and N₁₀₀ corresponding to the assumed conversion possibilities of 0, 40%, 70%, and 100% respectively. With N_con firm denoting the number of confirmed cases in a residence community and N_suspect denoting the number of suspected cases in a residence community, the four epidemic variables can be calculated by

\[ N_0 = N_{\text{confirm}} \]  
\[ N_{40} = N_{\text{confirm}} + 40\% \times N_{\text{suspect}} \]  
\[ N_{70} = N_{\text{confirm}} + 70\% \times N_{\text{suspect}} \]  
\[ N_{100} = N_{\text{confirm}} + N_{\text{suspect}} \]  

These four residence community epidemic variables were used for further analysis in this paper.

3.1.2. Characteristic data of residence communities in Wuhan

With the epidemic data of the residence communities in Wuhan, we searched for the detailed characteristic data of each residence community so that the relationship between the number of confirmed (and suspected) cases and the characteristic data of the corresponding residence community could be explored. The detailed residence community information was from Fang.com, a website for purchasing, selling, and renting houses in China [40]. This online platform provides much information about the residence communities in Wuhan including names, addresses, developers, and comments. A total of 11 characteristic variables were collected from Fang.com which were the average price of houses, construction year, construction area, covered area, total number of houses, total number of buildings, greening ratio, volume ratio, property charge, number of second-hand houses, and number of sold houses. The explanations of these variables are as follows.

- Average price of houses: the average price of the houses in the residence community.
- Construction year: the year when the residence community was completed.
- Construction area: the sum of the horizontal area of each floor of all the buildings in the residence community.
- Covered area: the horizontal area of the land used by the residence community.
- Total number of houses: the total number of houses in the residence community.
- Total number of buildings: the total number of buildings in the residence community.
- Greening ratio: the ratio of green land area to total land area in the residence community.
- Volume ratio: the ratio of total building area to land area in the residence community.
- Property charge: an amount of money paid by the property owner or user to the property management agency for the services related to residents’ life in the residence community such as the repair and renovation of utilities as well as the maintenance of hygiene and security.
- Number of second-hand houses: the total number of second-hand houses in the residence community.
- Number of sold houses: the total number of sold houses in the residence community.

These variables reflect the characteristics of a residence community from different aspects and may be linked with the epidemic data of the residence community.

3.1.3. Location data of places with a crowded population in Wuhan

The person-to-person transmission of COVID-19 implies that the number of confirmed and suspected cases in a residence community may be associated with the distances between the residence community and the places with a crowded population. In this paper, 15 kinds of crowded places in Wuhan were collected and categorized into five types, i.e., the Huanan Seafood Wholesale Market, schools (kindergartens, primary schools, middle schools, and universities), places for shopping and
entertainment (business clusters, shopping malls, and cinemas), places for transportation (subway stations, bus stops, inter-city bus stations, railway stations, and an airport), and hospitals (general hospitals and appointed hospitals for COVID-19). The data sources of the addresses of these places are shown in Table 1. Considering that the number of appointed hospitals for COVID-19 increased with time as the infectious disease developed, the appointed hospitals for COVID-19 on February 1, 2020 and on February 14, 2020 were collected.

The addresses of the residence communities and the places with a crowded population in Wuhan were transformed into the longitude and latitude data on the Baidu map to determine their locations. The spatial distribution of the residence communities and crowded places we studied in this paper is shown in Fig. 1.

**Table 1**

Data sources of the addresses of residence communities and places with a crowded population.

| Places                     | Data sources of addresses                        |
|----------------------------|--------------------------------------------------|
| Residence communities      | https://wuhan.esf.fang.com/housing/              |
| Kindergartens              | http://wh.bendibao.com/wangdian/                |
| Primary schools            | http://wh.bendibao.com/wangdian/                |
| Middle schools             | http://wh.bendibao.com/wangdian/                |
| Universities               | https://www.dxsbb.com/news/1679.html             |
| Business clusters          | https://zhidao.baidu.com/                       |
| Shopping malls             | http://wh.bendibao.com/wangdian/                |
| Cinemas                    | http://wh.bendibao.com/wangdian/                |
| Subway stations            | https://download.csdn.net/                      |
| Bus stops                  | https://download.csdn.net/                      |
| Inter-city bus stations    | http://wh.bendibao.com/wangdian/                |
| Railway stations           | http://wh.bendibao.com/wangdian/                |
| Airport                    | http://wh.bendibao.com/wangdian/                |
| General hospitals          | http://wh.bendibao.com/wangdian/                |
| Appointed hospitals        | http://wjw.wuhan.gov.cn/ztzl_28/fl/ftzgg/      |

3.1.4. Calculation of distances between residence communities and crowded places

The distance between two places can be calculated from their longitude and latitude. The equation is

\[ \text{Dis} = R \times \arccos(\sin \text{Lat}_1 \times \sin \text{Lat}_2 + \cos \text{Lat}_1 \times \cos \text{Lat}_2 \times \cos (\text{Lng}_1 - \text{Lng}_2)) \]  

where Dis (m) is the distance between two places and \( R = 6,371,000 \) (m) is the radius of the earth. Besides, \( \text{Lng}_1 \) and \( \text{Lat}_1 \) are the longitude and latitude of the first place and \( \text{Lng}_2 \) and \( \text{Lat}_2 \) are the longitude and latitude of the second place. Taking the kindergartens as an example, the distance between a residence community and the nearest kindergarten can be calculated by

\[ \text{Dis}_{\min} = \min(\text{Dis}_{1}, \text{Dis}_{2}, \ldots, \text{Dis}_{i}, \ldots, \text{Dis}_{n}) \]  

where \( \text{Dis}_i \) is the distance between the residence community and the \( i \)th kindergarten and \( n \) is the total number of kindergartens. The calculation of distances by Eqs. (5) and (6) was conducted on all the residence communities and all the place types.

3.2. Correlation analysis between epidemic data and residence community variables

The correlation analysis is a statistical tool that studies the relationship between two variables [41–43]. In this paper, we performed a correlation analysis between the epidemic data in residence communities and the residence community variables in Wuhan to find the possible association between them. There were 27 residence community variables in total, including 11 characteristic variables and 16 distance variables. First, the correlation analysis was conducted between the number of confirmed cases and the characteristic variables which were the average price of houses, construction year, construction area, covered area,
4 Results

4.1 Descriptive statistics of epidemic data and residence community variables

The average values and standard deviations of the epidemic data and the residence community variables in Wuhan are shown in Table 2. It can be seen from the last column of Table 2 that the number of Wuhan residence communities we researched is 383. There are some characteristic variables the number of which is less than 383 due to data missing. The average value of the number of confirmed cases is 6, indicating that 6 residents are confirmed COVID-19 per residence community on average. Similarly, the average number of suspected cases per residence community is 5. The average value for kindergartens is 582 (m), indicating that the distance between a residence community and the nearest kindergarten is 582 (m) on average. The average values and standard deviations of other distance variables can be understood in the same way.

4.2 Correlation between epidemic data and characteristic variables in residence communities

The correlation analysis was conducted between the epidemic data and the characteristic variables in the residence communities in Wuhan to find out the possible relationship between the number of confirmed cases of COVID-19 and the descriptive characteristics of the residence communities. The four epidemic variables \(N_0\), \(N_{40}\), \(N_{70}\), and \(N_{100}\) calculated by Eqs. (11)–(4) represent the number of COVID-19 confirmed cases in a residence community assuming that the probabilities of suspected cases transforming into confirmed cases are 0, 40%, 70%, and 100%, respectively. The Pearson correlation coefficients and significance between the number of COVID-19 confirmed cases \(N_0\), \(N_{40}\), \(N_{70}\), and \(N_{100}\) and the 11 residence community characteristic variables are shown in Table 3. The scatterplots of \(N_{100}\) and the characteristic variables are presented in Fig. 2.
It can be seen from Table 3 that all the correlation coefficients are positive. Significant positive correlations are shown between the number of COVID-19 confirmed cases and the construction year, construction area, covered area, total number of houses, total number of buildings, volume ratio, property charge, number of second-hand houses, and number of sold houses in the residence communities in Wuhan. More specifically, \( N_0 \), \( N_{40} \), \( N_{70} \), and \( N_{100} \) positively correlate with the construction area, covered area, total number of houses, total number of buildings, volume ratio, property charge, and number of second-hand houses among which the positive correlations between the number of COVID-19 confirmed cases and the construction area, covered area, total number of houses, total number of buildings, and number of second-hand houses are significant at the 0.01 level for all the four assumed conversion possibilities. As for the construction year, the positive correlations are significant at the 0.01 level for \( N_{40} \) and \( N_{100} \) and at the 0.05 level for \( N_{40} \). Besides, \( N_{70} \) and \( N_{100} \) positively correlate with the number of sold houses, correlations both significant at the 0.05 level. No significant correlation is observed between the number of COVID-19 confirmed cases and the average price of houses or greening ratio.

4.3. Correlation between epidemic data and distance variables in residence communities

Similarly, we performed correlation analysis between the epidemic variables \( N_0 \), \( N_{40} \), \( N_{70} \), and \( N_{100} \) and the distance variables for the residence communities in Wuhan to find out whether the number of COVID-19 confirmed cases in a residence community was related to the distance between the residence community and the nearest crowded places, and what kinds of crowded places had stronger correlations. The Pearson correlation coefficients and significance between the number of COVID-19 confirmed cases \( N_0 \), \( N_{40} \), \( N_{70} \), and \( N_{100} \) and the 16 distance variables are presented in Table 4. The scatterplots of \( N_0 \) and the distance variables are presented in Fig. 3.

From Table 4, it can be seen that there are significant negative correlations between the residence community epidemic variables and the distances from the residence communities to the nearest universities, business clusters, subway stations, and railway stations, which indicates that the number of COVID-19 confirmed cases in a residence community is likely to increase if there are universities, business clusters, subway stations, or railway stations nearby. To be more concrete, \( N_0 \), \( N_{40} \), \( N_{70} \), and \( N_{100} \) negatively correlate with the distances between the residence communities and the nearest universities, business clusters, and railway stations and correlations are significant at the 0.01 level for universities and business clusters and at the 0.05 level for railway stations. Besides, \( N_0 \) negatively correlates with the distances between the residence communities and the nearest subway stations which is significant at the 0.05 level. No significant correlation is observed between the number of COVID-19 confirmed cases and the other distance variables, i.e., the distances between the residence communities and the Huanan Seafood Wholesale Market as well as the nearest kindergartens, primary schools, middle schools, shopping malls, cinemas, bus stops, inter-city bus stations, airport, general hospitals, and appointed hospitals for COVID-19 in Wuhan.

5. Discussion

Wuhan, the capital city of Hubei Province and one of the megacities in China, has been confronted with stresses from the transmission of COVID-19 since December 2019 [39]. The epidemic of infectious diseases like COVID-19 is impacted not only by biological factors but also by socioeconomic factors. It is important to analyze the relationship between the epidemic data in the residence communities and the socioeco-
Table 4
Pearson correlation coefficients between the number of COVID-19 confirmed cases and the distance variables in residence communities.

| Distance variables                        | $N_0$ | $N_{40}$ | $N_{70}$ | $N_{100}$ |
|------------------------------------------|-------|---------|---------|---------|
| Huanan Seafood Wholesale Market          | −0.048| −0.068  | −0.077  | −0.083  |
| Kindergartens                            | −0.056| −0.059  | −0.059  | −0.059  |
| Primary schools                          | −0.074| −0.074  | −0.072  | −0.070  |
| Middle schools                           | −0.016| −0.016  | −0.016  | −0.015  |
| Universities                             | −0.156∗∗| −0.163∗∗| −0.162∗∗| −0.159∗∗|
| Business clusters                        | −0.157∗∗| −0.155∗∗| −0.148∗∗| −0.141∗∗|
| Shopping malls                           | −0.096| −0.095  | −0.091  | −0.087  |
| Cinemas                                  | −0.087| −0.084  | −0.080  | −0.076  |
| Subway stations                          | −0.100∗| −0.097  | −0.093  | −0.088  |
| Bus stops                                | 0.005 | 0.006   | 0.012   | 0.016   |
| Inter-city bus stations                  | −0.017| −0.042  | −0.054  | −0.063  |
| Railway stations                         | −0.115∗| −0.124  | −0.125∗| −0.124  |
| Airport                                  | 0.052 | 0.033   | 0.021   | 0.011   |
| General hospitals                        | −0.043| −0.051  | −0.053  | −0.055  |
| Appointed hospitals (on February 1)      | −0.076| −0.078  | −0.077  | −0.075  |
| Appointed hospitals (on February 14)     | −0.062| −0.048  | −0.038  | −0.030  |

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Fig. 3. Scatter plots of $N_0$ and the distance variables in residence communities.
omic factors to identify high-risk residence communities and find possible risk sources, based on which targeted containment measures can be made and implemented to inhibit the transmission of COVID-19. This paper performed a correlation analysis between the number of COVID-19 confirmed cases in the residence communities in Wuhan and the characteristic and distance variables of the residence communities. Among them, the characteristic variables refer to the detailed information of the residence communities including the average price of houses, construction year, construction area, covered area, total number of houses, total number of buildings, greening ratio, volume ratio, property charge, number of second-hand houses, and number of sold houses. The distance variables refer to the shortest distances from a residence community to the crowded places including the Huanan Seafood Wholesale Market, schools (kindergartens, primary schools, middle schools, and universities), places for shopping and entertainment (business clusters, shopping malls, and cinemas), places for transportation (subway stations, bus stops, inter-city bus stations, railway stations, and an airport), and hospitals (general hospitals and appointed hospitals for COVID-19). Besides, there are four epidemic variables \( N_0, N_{Ai}, N_{Bi}, \) and \( N_{i0} \) denoting the number of COVID-19 confirmed cases in the residence communities corresponding to the assumed possibilities of suspected cases converting into confirmed cases being 0%, 40%, 70%, and 100% respectively. The results show that the number of COVID-19 confirmed cases in the residence communities positively correlates with 9 of the 11 characteristic variables and negatively correlates with 4 of the 16 distance variables. Besides, the correlation significance is consistent within \( N_0, N_{Ai}, N_{Bi}, \) and \( N_{i0} \), for most characteristic and distance variables, indicating that the possibility of suspected cases converting into confirmed cases hardly affects the results.

In terms of the characteristic variables, there are significant positive correlations between the number of COVID-19 confirmed cases and the construction year, construction area, covered area, total number of houses, total number of buildings, volume ratio, property charge, number of second-hand houses, and number of sold houses in the residence communities in Wuhan, while there is no significant correlation between the number of COVID-19 confirmed cases and the average price of houses or greening ratio. The results would provide useful references for decision-makers to evaluate the risk level of residence communities and identify the high-risk residence communities where the infectious disease is more likely to spread. They may be explained from several aspects. First, the positive correlation between the number of confirmed cases and the construction year indicates that the newly built residence communities may have a higher degree of epidemic severity than the old residence communities. Perhaps the reasons are that the neighborhood residence committees take a more active part in residence community management and pay more attention on epidemic prevention and control in old residence communities. Besides, the residents in old residence communities are usually more familiar with each other and their relationships are closer. Therefore, they are more willing to cooperate with the neighborhood residence committees and follow the epidemic-related regulations, which is conducive to the implementation of containment strategies. Second, the positive correlation between the number of confirmed cases and the construction area, covered area, total number of houses, or total number of buildings is probably because that broader area or more houses or buildings imply more residents in a residence community, which increases the possibility of the person-to-person transmission of COVID-19. Third, the positive correlation between the number of confirmed cases and the volume ratio may be because that higher volume ratio suggests higher population density and poorer ventilation [44]. Fourth, it seems that the property management agency does not greatly contribute to the inhibition of COVID-19 spread considering the positive correlation between the number of confirmed cases and the property charge. So we cautiously suggest that the property management agency collaborate with the neighborhood residence committee to better serve the residence community residents especially in the case of disease transmission. Last, the positive correlation between the number of confirmed cases and the number of second-hand houses or sold houses may be due to the larger floating population in the residence communities with more second-hand houses or sold houses considering that the strangers to Wuhan tend to stay in the second-hand houses temporarily. The floating population is exposed to various people when traveling, which increases the risk of infection. Moreover, the person density is usually higher in the second-hand houses because it is common for several tenants to live together to share the rent, which may also lead to a higher infection rate.

In terms of the distance variables, the results of each type of crowded places are presented and analyzed respectively as follows, which helps to identify the crowded places with high risk and find possible infection sources.

As for the Huanan Seafood Wholesale Market, there is no significant correlation between the number of COVID-19 confirmed cases in the residence communities and the distances from the residence communities to the Huanan Seafood Wholesale Market. It may be owed to the closure of the market since January 1, 2020. Perhaps further investigation is needed on the presence of the virus on the market and the people who have contacted with the virus in a visit to the market.

As for schools, the number of confirmed cases negatively correlates with the distances between the residence communities and the nearest universities, while no significant correlation is observed between the number of confirmed cases and the shortest distances from the residence communities to the kindergartens, primary schools, or middle schools. There may be several reasons for the findings. First, the kindergartens, primary schools, and middle schools have entered winter vacation since January 15, 2020. Second, the floating population is large in universities compared with that in kindergartens, primary schools, or middle schools, considering that university students are from various provinces and cities and they need to move between the dormitories, canteens, and classrooms or even go outside the universities for all kinds of activities. Third, some universities are open to the public so that the nearby residents can exercise in the university stadiums, which also increases the contact rate among people.

As for places for shopping and entertainment, the correlation is negative between the number of confirmed cases and the distances from the residence communities to the nearest business clusters, while the correlation is not significant for shopping malls and cinemas. Perhaps it is because that business clusters combine the places for shopping, entertainment, education, work, industry, and so on, leading to a large floating population with complexity. By contrast, the composition of shopping malls and cinemas is relatively simple.

As for places for transportation, we further divided them into two groups which were the places for transportation within the city (subway stations and bus stops) and the ones for transportation between cities (inter-city bus stations, railway stations, and an airport). Concerning the places for transportation within the city, there is a negative correlation between the number of confirmed cases and the shortest distances from the residence communities to the subway stations compared with the bus stops which do not have a significant correlation. The differences between the subway stations and the bus stops may explain the results. First, the subway stations are underground where the environment is relatively airtight with poor ventilation, while the bus stops are on the ground where the environment is open with better ventilation. Second, the departure interval in subway stations is much shorter than that in bus stops and the passenger capacity is much larger, leading to a larger floating population in subway stations. Concerning the places for transportation between cities, the number of confirmed cases negatively correlates with the shortest distances from the residence communities to the railway stations and does not correlate with the distances of the inter-city bus stations or the airport, which may be explained from several aspects. First, the passenger capacity is larger in railway stations than that in inter-city bus stations. Therefore, perhaps the population density is higher and the floating population is larger in the railway stations. Although there are also many passengers in the airport, its covered
area is broad, reducing the population density. Second, it can be inferred that long-distance land transportation may have a greater impact on the spread of COVID-19 since the trains usually travel across a few provinces while most of the inter-city buses travel across the cities within Hubei Province. Third, trains stop at several stations on the journey with many passengers getting on and off at each station, resulting in passengers being exposed to more people compared with inter-city buses and airports. Last, it can be seen from the average value and standard deviation of the distance variable of the airport in Table 2 that the airport is relatively far from the residence communities in the Wuhan city, having little impact on the residents. Nevertheless, perhaps the people who have used the airport should be considered.

As for hospitals, no significant correlation is observed between the number of confirmed cases and the distances from the residence communities to the nearest general hospitals or appointed hospitals for COVID-19 (neither on February 1 nor on February 14), reassuring the people living near to hospitals that there is not enough evidence to prove a higher risk of infection.

This study is subject to several limitations. First, not all of the residence communities in Wuhan had reported the number of confirmed and suspected COVID-19 cases by February 14, 2020, leading to data missing. Second, although we have proposed many possible explanations, the reasons for the correlation results need to be further analyzed, based on which targeted measures can be taken to diminish the influence of the risk factors. Third, the impact of other socioeconomic factors and the change of the number of cases with time should be considered once the data can be accessed. Nevertheless, this study discovers the correlations between the number of COVID-19 confirmed cases in residence communities and the socioeconomic factors, which is presented in Table 5. We investigated the closure date of the Huanan Seafood Wholesale Market and schools and analyzed its influence on the data of cases. The Huanan Seafood Wholesale Market was closed on January 1 blocking the spread of disease, which is consistent with the result that no significant correlation is observed between the number of COVID-19 confirmed cases in the residence communities and the distances from the residence communities to the Huanan Seafood Wholesale Market. The kindergartens, primary schools, and middle schools entered winter vacation on January 15 and were also closed, which may explain why there is no significant correlation between the number of confirmed cases and the shortest distances from the residence communities to the kindergartens, primary schools, or middle schools. It can be seen that the closure of crowded places is effective for the containment of the spread of COVID-19. This paper provides valuable references for the authorities to identify the high-risk residence communities and find possible risk sources, which is advantageous to the precise prevention and control of COVID-19 on the residence community level.

6. Conclusions

This paper performed correlation analysis between the number of COVID-19 confirmed cases in the residence communities in Wuhan, China and the socioeconomic factors including the residence community characteristic variables and the distances from the residence communities to the crowded places, considering that the factors significantly correlating with the number of confirmed cases may be associated with the risk level of residence communities. The results show that the residence communities which are newly-built, where the volume ratio or property charge is high or the construction area, covered area, or total number of houses, buildings, second-hand houses, or sold houses is large, or which are close to universities, business clusters, subway stations, or railway stations are at higher risk of the COVID-19 transmission. Also, the universities, business clusters, subway stations, and railway stations are the crowded places where COVID-19 is more likely to spread. The findings supply the authorities with useful guidance to identify high-risk residence communities and find possible risk sources, based on which targeted prevention and control strategies can be prioritized in the high-risk residence communities, benefiting precise interventions on the residence community level.

Declaration of Competing Interest

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.
