Society for
Mathematical
Biology

Check for
updates

Analysis of the Geographic Transmission Differences of COVID-19 in China Caused by Population Movement and Population Density

Yi Hu[1] · Kaifa Wang[1] · Wendi Wang[1]

Abstract

The coronavirus disease (COVID-19) has led to a global pandemic and caused huge healthy and economic losses. Non-pharmaceutical interventions, especially contact tracing and social distance restrictions, play a vital role in the control of COVID-19. Understanding the spatial impact is essential for designing such a control policy. Based on epidemic data of the confirmed cases after the Wuhan lockdown, we calculate the invasive reproduction numbers of COVID-19 in the different regions of China. Statistical analysis indicates a significant positive correlation between the reproduction numbers and the population input sizes from Wuhan, which indicates that the large-scale population movement contributed a lot to the geographic spread of COVID-19 in China. Moreover, there is a significant positive correlation between reproduction numbers and local population densities, which shows that the higher population density intensifies the spread of disease. Considering that in the early stage, there were sequential imported cases that affected the estimation of reproduction numbers, we classify the imported cases and local cases through the information of epidemiological data and calculate the net invasive reproduction number to quantify the local spread of the epidemic. The results are applied to the design of border control policy on the basis of vaccination coverage.

Keywords Reproduction number · Population density · Population movement · Regional differences

✉ Wendi Wang
  wendi@swu.edu.cn

1 School of mathematics and statistics, Southwest University, Chongqing 400715, People's Republic of China

⚓ Springer

1 Introduction

In this ongoing epidemic of COVID-19, the implementation of non-pharmaceutical interventions (NPIs) can help to curb the spread of the epidemic, even if vaccines are available. NPIs include early case identification and isolation, banning public gatherings, personal contact restrictions, social alienation measures and personal protection actions. Many countries have implemented full or partial blockade measures to restrict gathering activities (Block et al. 2020; Ferguson et al. 2020; Gatto et al. 2020). Because Wuhan was the first city in China where the epidemic occurred, to prevent the further spatial spread of COVID-19 in China, Wuhan started to lockdown on January 23, 2020 (Tian et al. 2020). Afterward, all provinces started to implement prevention and control measures such as inter-city travel restrictions and encouraging the public to wash their hands, wear masks. China has implemented a series of public health and medical measures, such as building Fangcang Shelter Hospitals, reasonably allocated medical resources and achieved complete treatment for patients.

Data analysis and mathematical modeling can help to understand the spread of COVID-19 and provide important theoretical supports and decision-making basis for epidemic prevention and control. After the outbreak of COVID-19, compartment models are established (Zhao and Feng 2020; Huang et al. 2020; Roda et al. 2020; Song et al. 2020; Tang et al. 2020a, c; Yu et al. 2020; Zhang et al. 2020a) to estimate important parameters, such as the basic reproduction number (R_0), epidemic time, number of cases (Huo et al. 2021) and the final size. As an important indicator to characterize the spread of infectious diseases, R_0 is defined as the average number of new infections caused by a typical individual (Dietz 1974; Hethcote 2000; van den Driessche and Watmough 2002). Li et al. (2020) fit the initial case data in Wuhan and estimated that the basic reproduction number is 2.2 (1.4–3.9). Many studies have evaluated the control effect of NPIs on COVID-19 (Gatto et al. 2020; Ge et al. 2020; Jia et al. 2020; Luo et al. 2021; Tang et al. 2020d; Xue et al. 2021), such as the effect of building Fangcang Shelter Hospitals (He et al. 2021), wearing masks (Eikenberry et al. 2020). In terms of monitoring and evaluation research on prevention and control strategies, many articles have given valuable results (Fang et al. 2020; Huang et al. 2020). Moreover, Tang et al. (2020b) and Wang et al. (2020) analyzed the best time to resume work in Wuhan.

Population movement has a strong risk to diffuse the spread of the epidemic. Chen et al. (2020) studied the distribution of cases of the early epidemic in Wuhan and its relationship with population movement. Ye et al. (2020) analyzed the trend of the epidemic situation in Guangdong Province based on population movement. In addition, there is a strong positive correlation between the number of case and the number of population outflow from Zhang et al. (2020c) and Zhou et al. (2020).

To the best of our knowledge, most of researches focus on the establishment of a dynamic model in a fixed area and very few works have considered the spatial effects of COVID-19 transmission. Indeed, before the lockdown of Wuhan, approximately 5 million people flowed to other regions of China due to Chinese New Year (https://news.163.com/20/0126/22/F3ROV3FU0001899O.html) and caused the epidemic outbreaks in these sites. Notice that input numbers of migrants from Wuhan were different for different cities and provinces. Furthermore, the local population

densities (PD) vary greatly in different regions. These spatial differences may result in different epidemic risks, because each individual in a region with a higher PD probably has a larger social network with each other and sites near the epidemic center are easier to get more infectious inflows, which means more infectious contacts. Therefore, it is important to consider how the population migration and population density contribute to the outbreaks of COVID-19.

In the present paper, based on the reported epidemic data before the lockdown of Wuhan, we estimate the reproduction numbers of COVID-19 for the provinces and cities of China. Through correlation analysis and multiple linear regression, we reveal the relation between the reproduction numbers and population densities, and the relation between the reproduction numbers and population migrations. These results are helpful for the formulation of prevention and control measures.

The organization of the remaining paper is set as follows. Section 2 presents data collection and analysis. Section 3 introduces the main methods, including reproduction number and data smoothing method. In Sect. 4, we calculate the invasive reproduction numbers and show the correlation analysis results. We also analyze the spread of the local epidemic through the net invasive reproduction number. The paper ends with a brief discussion.

2 Data Collection and Analysis

The daily epidemic report data of 31 provinces (municipalities and autonomous regions) in China are obtained from the website of Chinese Center for Disease Control and Prevention (http://2019ncov.chinacdc.cn/2019-nCoV/global.html). The epidemic data of each city in Hubei Province are obtained from the Health Commission of Hubei Province (http://wjw.hubei.gov.cn/fbjd/dtyw/). The data for calculating population movement sizes come from Baidu Migration website (https://qianxi.baidu.com/2020/).

Notice that the Spring Festival travel season began on January 10, 2020, and Wuhan started to lock down at 10 a.m. on January 23, 2020. We select the time period of data from January 10 to January 24, 2020, a total of 15 days. The population flows among the regions except for Wuhan are ignored because the dispersal individuals were susceptible and had little influence on disease transmissions and local population densities. Thus, we focus only on the emigration flows of Wuhan. The specific formula for computing the population input size (PIS) of each site (province or large city), coming from Wuhan during this period, is given by

$$\text{PIS} = \sum_i \text{IP}_i \times \frac{\text{index}_i}{\sum_i \text{index}_i} \times N_{\text{total}}.$$

Here, IP_i represents the inflow proportion from Wuhan on the ith day, index_i represents the migration index in Wuhan on the ith day, reflects the scale of migration population, which can be horizontally compared between cities, $i = 1, \ldots, 15$, N_{total} represents the total number of migrants from Wuhan, which is 5 million (https://news.163.com/20/0126/22/F3ROV3FU0001899O.html).

Table 1 The PIS from Wuhan and PD at provincial level in China

Region	PIS	PD	Region	PIS	PD
Anhui	113,921	451	Jiangsu	73,842	751
Beijing	44,544	1313	Jiangxi	106,634	278
Chongqing	63,249	376	Jilin	8461	144
Fujian	45,271	325	Liaoning	16,595	295
Gansu	17,501	62	Ningxia	4228	104
Guangdong	95,756	631	Shaanxi	36,249	188
Guangxi	40,099	207	Shandong	55,151	636
Guizhou	28,261	204	Shanghai	33,914	3823
Hainan	18,667	264	Shanxi	29,857	237
Hebei	47,177	400	Sichuan	62,244	172
Heilongjiang	14,010	80	Tianjin	7641	1304
Henan	283,921	575	Xinjiang	10,440	15
Hunan	174,886	326	Yunnan	26,756	123
Inner Mongolia	8964	21	Zhejiang	54,315	564

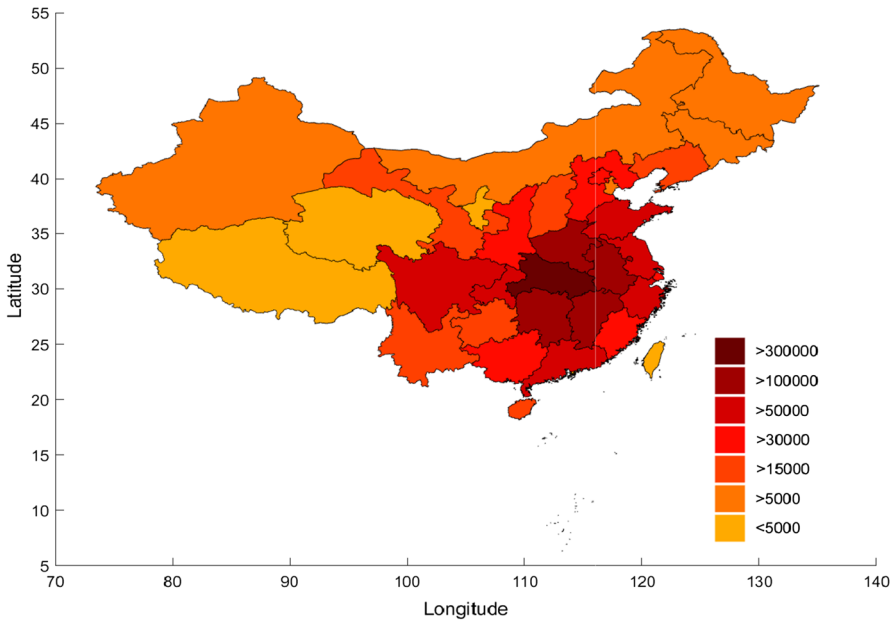PIS, population input size, person; PD, population density, person/km^2

Fig. 1 Geographic distribution of Wuhan's population outflow sizes from January 10 to January 24, 2020 (Color figure online)

Table 2 The PIS from Wuhan and PD at prefecture-level cities in Hubei Province

Region	PIS	PD	Region	PIS	PD
Enshi	66,405	141	Suizhou	110,313	230
Ezhou	140,353	676	Tianmen	72,295	485
Huanggang	459,650	364	Xiangyang	137,432	286
Huangshi	132,239	539	Xianning	176,204	258
Jinmen	114,832	234	Xiantao	102,915	449
Jinzhou	228,290	396	Xiaogan	479,663	552
Qianjiang	39,730	482	Yichang	100,065	197
Shiyan	66,902	144			

PIS, population input size, person; PD, population density, person/km^2

The numbers of PIS at provincial level are shown in Table 1. Figure 1 shows the geographic distribution of population movements. Outside Hubei Province, there were large differences in the size of population movements. Henan, Hunan and Anhui, which are the neighboring provinces of Hubei, ranked the top three in PIS, while Ningxia, Qinghai and Tibet, which are far away from Wuhan, were in the bottom of PIS rank. In addition, in order to describe the geographic distribution of population mobility on a smaller scale, we further focus on the cities in Hubei Province. Before the lockdown of Wuhan, nearly 70% of the total migrant population in Wuhan flowed into the prefecture-level cities in Hubei Province every day, and the corresponding PISs at prefecture-level cities in Hubei Province are shown in Table 2. The three prefecture-level cities with the largest inflows were Xiaogan, Huanggang and Jingzhou, whereas Shiyan City, Enshi Prefecture, Qianjiang City and Shennongjia Forest Region had relatively small inflows.

The PD of each province or city comes from the 2019 Statistical Yearbook (National Bureau of statistics 2020), which lists the population numbers of the regions at the end of 2018 and the areas of the administrative divisions of China. The results are shown in Tables 1 and 2, where the PD is the ratio of population numbers to area in the unit of person/km^2.

3 Method

3.1 Lotka–Euler equation

Our objective is to use the data of daily reported cases to estimate the reproduction numbers of COVID-19 in the regions of China. As usual, we consider the case that the daily counts of new cases in the initial phase increase exponentially at a rate r. Let $b(t)$ be the total number of infectious inputs and local new infections at time t. Under the assumption that the infection process is stationary, we have

$$b(t) = b(t-a)e^{ra} \tag{1}$$

for $a > 0$.

If $F(a)$ is the infection rate and $S(a)$ is the proportion of infectious individuals at time a since local infection or arrival time for an infectious immigrant, then $F(a)S(a)$ is the expected rate of secondary cases at time a. It follows from Wallinga and Lipsitch (2007) that

$$b(t) = \int_0^\infty b(t-a)F(a)S(a)\mathrm{d}a.$$

In view of (1), we get the Lotka–Euler renewal equation

$$1 = \int_0^\infty e^{-ra}F(a)S(a)\mathrm{d}a. \tag{2}$$

Let us define the invasive reproduction number R_0 by

$$R_0 = \int_0^\infty F(a)S(a)\mathrm{d}a.$$

The invasive reproduction number is an appropriate quantity to measure the transmission risk as the number of new cases increases exponentially with a rate $r > 0$ when $R_0 > 1$, and decline exponentially with a rate $r < 0$ when $R_0 < 1$. In addition, it includes the influences of continuous infectious inputs, whereas the classical basic reproduction number considers only the local new transmissions. More importantly, it can be estimated by direct fitting the data of daily reported cases. Indeed, the rate $F(a)S(a)$ can be normalized to a distribution $g(a)$:

$$g(a) = \frac{F(a)S(a)}{\int_0^\infty F(a)S(a)\mathrm{d}a} = \frac{F(a)S(a)}{R_0}.$$

This is a distribution of serial interval, which is defined as the mean duration between the time of infection of a secondary infector and the time of infection of its primary infector (Wallinga and Lipsitch 2007). Substituting this expression into (2), we get

$$\frac{1}{R_0} = \int_0^\infty e^{-ra}g(a)\mathrm{d}a.$$

Note that the right-hand side of this equation is the Laplace transform of function $g(a)$. If $M(z) = \int_0^\infty e^{za}g(a)\mathrm{d}a$, which is the moment generating function of the distribution $g(a)$, it follows from Wallinga and Lipsitch (2007) that the invasive reproduction number R_0 can be computed by

$$R_0 = \frac{1}{M(-r)}. \tag{3}$$

3.2 Distribution of Serial Interval

The estimation of the reproduction number under a given growth rate depends on the specific distribution of the serial intervals (Wallinga and Lipsitch 2007). The gamma distribution is rich in shape and is very effective in fitting the data of serial interval of COVID-19 (Li et al. 2020; Zhang et al. 2020b). Thus, we fix the distribution of serial interval as a gamma distribution. The probability density function is

$$g(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

where α is the shape parameter, and β is the scale parameter. Exponential distribution and chi-square distribution are special cases of the gamma distribution. The corresponding moment generating function of gamma distribution is

$$M(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{(t-\beta)x} \mathrm{d}x. \tag{4}$$

For $t < \beta$, by the transformation $v = (\beta - t)x$ we obtain

$$M(t) = \frac{\beta^\alpha}{(\beta - t)^\alpha \Gamma(\alpha)} \int_0^\infty v^{\alpha-1} e^{-v} dv = \left(\frac{\beta}{\beta - t} \right)^\alpha. \tag{5}$$

Fixing $t = -r$ in (5), one obtains the formula of the invasive reproduction number

$$R_0 = \frac{1}{M(-r)} = \left(\frac{\beta + r}{\beta} \right)^\alpha. \tag{6}$$

3.3 Exponential Growth Rate

We estimate the exponential growth rate r from the time series of daily confirmed cases (t_i, x_i), $i = 0, \ldots, n-1$. For the region like Guangdong Province, the epidemic data in the early stage are well fitted by an exponential curve. Figure 2 shows the fitting graph of the confirmed cases of Guangdong Province, where January 22 of 2020, the date of the first reported case in Guangdong Province, is fixed as the initial date for fitting.

For cities or provinces with a small number of cases, the fitting effect of exponential growth is poor. This may result from the stochastic effects such as the errors of data reports, the variation of infection progressions and environmental perturbations. In order to filter the stochastic fluctuations, we ask the techniques of Earn et al. (2020) and Ma et al. (2013) where a logistic model is first used to simulate the increase in cumulative cases and the modified daily cases are extracted from the cumulative cases. More specifically, we assume that $c(t)$ (the total number of cases at t) satisfies
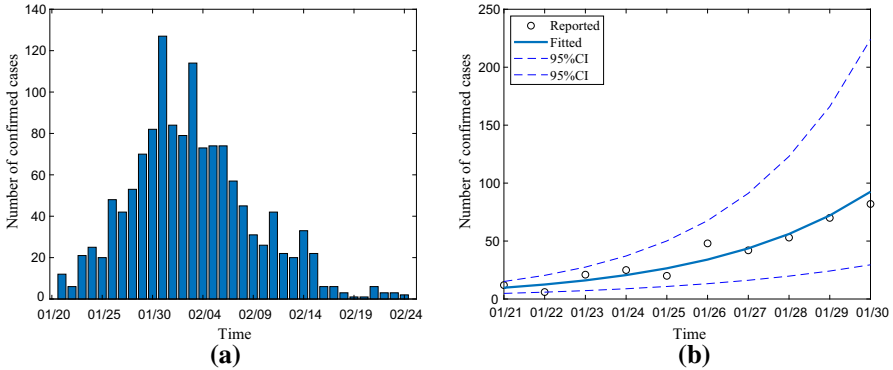
Fig. 2 a Daily confirmed data in Guangdong Province and **b** fitting graph of exponential growth rate of Guangdong Province (Color figure online)
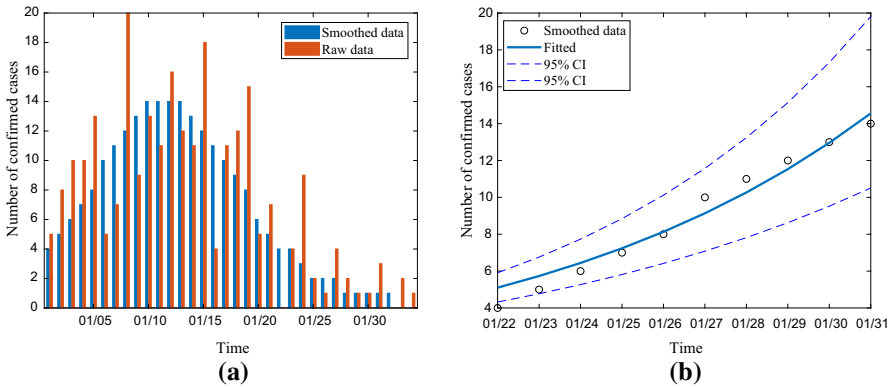

Fig. 3 a Daily confirmed data and smoothed data in Guangxi Province and **b** fitting graph of exponential growth rate of Guangxi Province (Color figure online)

a Logistic equation:

$$\frac{\mathrm{d}c(t)}{\mathrm{d}t} = r_c c(t) \left[1 - \frac{c(t)}{K} \right], \tag{7}$$

where r_c is the intrinsic growth rate of the epidemic and K represents the final scale of the epidemic. After that, we fix $x(t_i) = c(t_i) - c(t_{i-1})$ and then fit the filtered data by an exponential curve. Take Guangxi Province as an example, the actual data fluctuate largely because the sizes of new cases are small, as shown in Fig. 3a. Through the filtration, the random fluctuation of data is eliminated to get a better exponential fit, as shown in Fig. 3b.

3.4 Statistical Analysis

Shapiro–Wilk test is used to test the normality of the PD, PIS and reproduction number. If variables obey the normal distribution, then Pearson correlation coefficient is selected. Otherwise, Spearman correlation coefficient is selected. Multiple linear regression is selected to explore the relationships among R_0, PD and PIS. Results are considered as statistically significant when the p value is less than 0.05.

4 Results

In the epidemic spread, a considerable part of the cases were imported at the initial phase and the cases due to local transmissions occurred over time. Notice that the confirmed data do not distinguish the imported cases from the local cases. Therefore, we define the invasive reproduction number on the basis of confirmed data as R_0, and the reproduction number from the local case data as the net invasive reproduction number (R_0^{net}), which describes the local transmission risk of the epidemic after the arrival of imported cases.

4.1 The Invasive Reproduction Number

First, we use the data of confirmed cases in the early stage of COVID-19 spread from January 20, 2020, to February 20, 2020, to fit the exponential growth rate of each region. Second, based on the data of 425 cases in Wuhan, we previously used the maximum likelihood method to estimate the gamma distribution of serial interval, and our results in Hu et al. (2020) are in good agreement with Li et al. (2020) and Zhang et al. (2020b). Taking the estimated parameters $\alpha = 5.3183$, $\beta = 0.6010$ in Hu et al. (2020), by (6) we obtain the calculation formula for invasive reproduction number:

$$R_0 = \left(1 + \frac{r}{0.601}\right)^{5.3183}. \tag{8}$$

Table 3 shows the results of invasive reproduction numbers at provincial level in China. Similarly, we also obtain the reproduction numbers at prefecture-level cities in Hubei Province, which are shown in Table 4. On the whole, the reproduction numbers in the early stage of the epidemic transmissions in the regions (cities) are between 2.0 and 6.6. Therefore, there was a high risk of epidemic outbreak.

In order to reveal how population movement and local population size influence the risk of COVID-19 outbreaks, we conduct the correlation analysis between invasive reproduction number and PIS, and the correlation analysis between invasive reproduction number and PD. After the test of normality, we find that the Pearson correlation is selected for the first one, and the Spearman correlation is appropriate for the second case. These correlation analyses show that the invasive reproduction number of a province (city) in China (except Hubei Province) has a significant correlation with the PIS (the correlation coefficient $r = 0.628$, $P < 0.0001$) and has also a significant correlation with the local PD ($r = 0.650$, $P < 0.0001$). Furthermore, in Hubei

Table 3 Estimations of invasive reproduction numbers and their 95% confidence interval (CI) at provincial level in China

Region	R_0	95% CI	Region	R_0	95% CI
Anhui	4.16	(3.65, 4.73)	Jiangxi*	3.65	(3.02, 4.38)
Beijing	3.69	(3.29, 4.14)	Liaoning	3.39	(1.84, 4.01)
Fujian	2.90	(2.32, 3.60)	Inner Mongolia	2.54	(1.88, 3.38)
Gansu	2.65	(2.21, 3.16)	Ningxia	2.46	(2.23, 2.69)
Guangdong*	4.80	(4.25, 5.40)	Shandong*	3.66	(3.27, 4.08)
Guangxi	2.57	(2.24, 2.93)	Shanxi	3.76	(3.16, 4.44)
Guizhou	3.16	(2.67, 3.72)	Shaanxi	3.36	(2.74, 4.09)
Hainan	2.85	(2.55, 3.17)	Shanghai*	3.91	(3.39, 4.50)
Hebei	3.42	(3.13, 3.73)	Sichuan*	2.95	(2.67, 3.25)
Henan*	4.53	(3.98, 5.13)	Tianjin	3.24	(3.01, 3.47)
Heilongjiang*	3.55	(3.06, 4.11)	Xinjiang	2.58	(2.39, 2.80)
Hunan*	4.11	(3.55, 4.76)	Yunnan*	3.40	(2.72, 4.23)
Jilin	3.38	(2.55, 4.41)	Zhejiang*	4.10	(3.24, 5.14)
Jiangsu*	3.56	(3.17, 3.99)	Chongqing	2.70	(2.36, 3.08)
Hubei	5.81	(5.76, 5.86)			

The province with * means its R_0 is computed by raw data, the others are calculated by filtered data

Table 4 Estimations of invasive reproduction numbers and their 95% CI at prefecture-level cities in Hubei Province

Region	R_0	95% CI	Region	R_0	95% CI
Enshi	2.29	(2.04, 2.57)	Suizhou*	4.03	(3.28, 4.91)
Ezhou*	4.29	(4.01, 4.59)	Tianmen	3.62	(3.33, 3.93)
Huanggang*	3.65	(3.13, 4.23)	Xiangyang	3.49	(2.69, 4.47)
Huangshi*	3.62	(2.95, 4.41)	Xianning	2.32	(2.18, 2.46)
Jinmen	3.58	(3.31, 3.86)	Xiantao	3.42	(2.91, 4.01)
Jinzhou*	4.54	(4.20, 4.89)	Xiaogan*	4.73	(4.13, 5.30)
Qianjiang	3.02	(2.84, 3.20)	Yichang*	3.69	(3.08, 4.40)
Shiyan*	3.26	(2.90, 3.65)	Wuhan*	6.26	(5.99, 6.54)

The city with * means its R_0 is computed by raw data, the others are calculated by filtered data

Province, there are a significant correlation between the reproduction number and PIS ($r = 0.595$, $P = 0.019$), and a significant correlation between the reproduction number and local PD ($r = 0.516$, $P = 0.049$).

Though the above-mentioned correlations are statistically significant, the scatter plots indicate that the linear correlations are weak, which are shown in Fig. 4. This should be the case because a variety of control measures were adopted to reduce social activities in China from the beginning. In addition, people were urged to wear surgical masks for protection. These NPIs decreased the correlations.
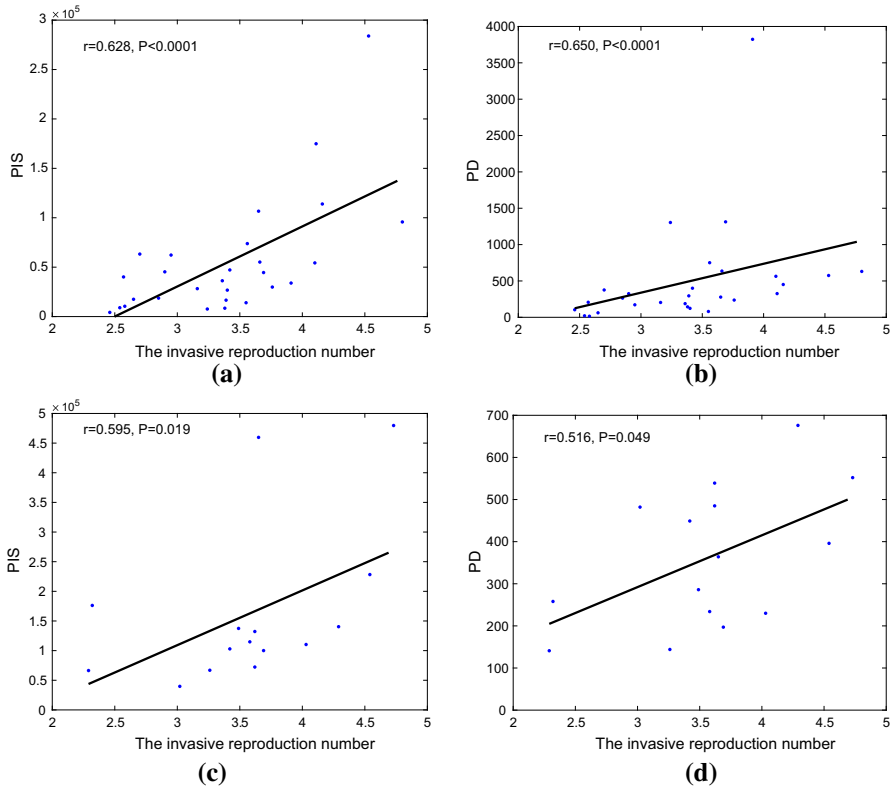
Fig. 4 Scatter plots and correlation coefficients of the invasive reproduction numbers versus PIS or PD. **a** Scatter plot and correlation coefficient of invasive reproduction numbers versus PIS at provincial level in China; **b** scatter plot and correlation coefficient of the invasive reproduction numbers versus PD at provincial level in China; **c** scatter plot and correlation coefficient of the invasive reproduction numbers versus PIS at prefecture-level cities in Hubei Province; and **d** scatter plot and correlation coefficient of the invasive reproduction numbers versus PD at prefecture-level cities in Hubei Province. *PIS* population input size, *PD* population density (Color figure online)

With the development of epidemics, the vaccination program (Kaur and Gupta 2020) and the border reopening measures (Nali et al. 2021) are implemented, which cause a variety of influences on PD and PIS. First, the PD of susceptible population becomes lower because the vaccination of COVID-19 provides a barrier for those vaccinated individuals. In addition, the border reopened policy results in the increase in PIS. In order to figure out how much the PD and PIS contribute to the disease transmission risk, we fit the values of R_0 by the PD and PIS. Notice that the disease transmissions in one site are attributed to local transmission and infectious contacts with imported individuals. We select

$$R_0 = a * \text{PD} + b * \text{PIS} * \text{PD} \tag{9}$$

Table 5 Parameter estimations of regression model

Model	Unstandardized coefficients		Standardized coefficients	t	P
	B	Standard error			
PD	1.632×10^{-3}	6.310×10^{-4}	0.344	2.588	0.013
PIS*PD	2.498×10^{-8}	6.964×10^{-9}	0.477	3.587	0.001

PIS population input size, *PD* population density
* $F = 24.897$, $R^2 = 0.548$

to fit the data, where a and b are constants to measure the intensities of local transmission and the transmission with imported individuals, respectively. Using the data in Tables 1, 2, 3 and 4 and multiple linear regression, we obtain the regression model, in which R_0 is the dependent variable, and the two indicators (PD and PIS) are explaining variables. Table 5 shows the results of the parameter estimation of the regression model, including regression coefficients, standard errors, standardized coefficients and the results of hypothesis test.

We consider a scenario where PIS varies from 1 to 10^5 and PD varies from 1 to 1000. Then numerical computations of invasive reproduction number on the basis of (9) and Table 5 yield Fig. 5a where the green plane corresponds to $R_0 = 1$. This figure indicates that suitable sizes of PIS and PD can drive the reproduction number below unity to control the epidemic disease. To demonstrate this, we let the vaccination ratio be α and assume that the vaccine is 100% effective. Then the susceptible PD is $(1 - \alpha) * PD$ and (9) become

$$R_0 = a * (1 - \alpha) * \text{PD} + b * (1 - \alpha) * \text{PIS} * \text{PD}. \tag{10}$$

If PD $= 1000$ and $\alpha = 0.6$, we get a relation between R_0 and PIS, which is shown in Fig. 5b. From this function, we see that $R_0 < 1$ when PIS $< 23,170$. This means that the daily input number should be less than $23,170/15 \approx 1544$ to control the epidemic. In other words, the control strategy should be selected from the green region in Fig. 5b.

Let us fix $R_0 = 1$ in (10). Then the blue surface in Fig. 5c shows the surface of critical vaccination ratio as PIS and PD vary, that is, if the vaccination ratio is above this surface, the epidemic is controlled because $R_0 < 1$. In order to consider border reopening scenario under vaccination, we fix PD $= 800$ and $R_0 = 1$. Then we get a curve in the surface to describe the critical relation between the vaccination ratio and PIS, which is marked by red color in Fig. 5c. To be clearer, we present Fig. 5d to show the curve in the PIS-α-plane, in which the green region means that the epidemic is controlled. It reveals that only when vaccination ratio exceeds a certain critical value (here is 23.4% in this scenario, see blue pentagram in Fig. 5d), can the border be reopened, and a higher vaccination ratio leads to the more openness.
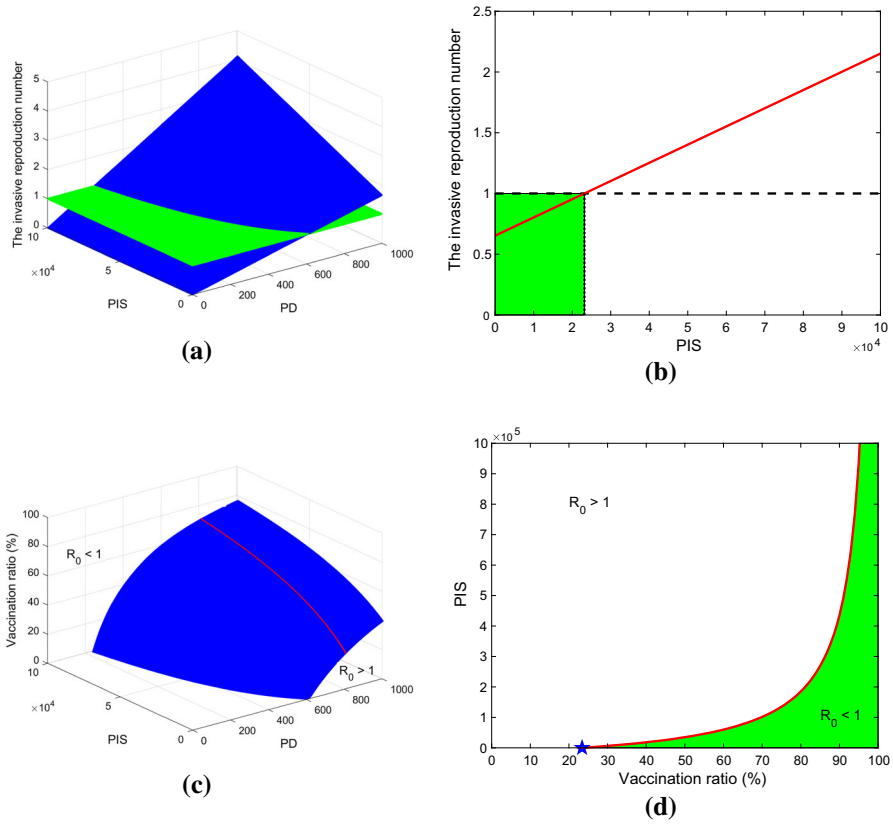
Fig. 5 **a** Numerical simulation of (9), **b** relation between PIS and R_0 when $\alpha = 0.6$ and PD $= 1000$, **c** surface of vaccination ratio versus PIS and PD, which is defined by (10) in the case of $R_0 = 1$ and **d** graph of border reopening region for (α, PIS) when $R_0 = 1$, PD $= 800$ (Color figure online)

4.2 The Net Invasive Reproduction Number

Since there were sequential inputs of imported cases in the early stage of COVID-19 spread, which affect the estimation of invasive reproduction numbers, we classify the imported cases and local cases through the information of epidemiological data and calculate the net invasive reproduction number to quantify the local spread of the epidemic. The cities or provinces with the full epidemiological information include Shanghai, Yunnan, Guizhou, Shaanxi, Beijing, Liaoning, Sichuan and Henan. The daily reported epidemiological data of these regions are collected through the websites of the Health Commissions or from the literature (Cheng et al. 2020b; Liu et al. 2020). The infected persons are divided as imported cases and local cases. An imported case is defined as a patient who had a history of living in Hubei Province within 14 days before the onset of illness. Otherwise, it is classified as a local case (Cheng et al. 2020a).

Table 6 The net invasive reproduction numbers and their 95% CI

Region	R_0^{net}	95% CI	Region	R_0^{net}	95% CI
Shanghai	4.48	(3.65, 5.45)	Sichuan	2.23	(2.07, 2.44)
Yunnan	1.65	(1.53, 1.79)	Beijing	2.78	(2.42, 3.19)
Guizhou	3.23	(2.53, 4.08)	Liaoning	2.99	(2.45, 3.61)
Shaanxi	3.12	(2.57, 3.76)	Henan	3.39	(2.80, 4.08)

Based on the data of local cases and (8), the net invasive reproduction numbers are shown in Table 6. Compared with Table 3, the net invasive reproduction number in most areas is less than the total invasive reproduction number. This means that the reproduction number decreases with the removal of imported cases. Hence, the total invasive reproduction number could cause an overestimation of local disease transmission risk. Especially, this is the case for a city or a province with the strict prevention measures or the lower PD, such as Southwest China, Beijing and Liaoning.

However, an attention should be paid to Shanghai's net invasive reproduction number $R_0^{net} = 4.48$ (95%CI 3.65−5.45), which is larger than the invasive reproduction number and indicates the risk of local spread of the epidemic in Shanghai is relatively higher. This weird paradox may be caused by the following reasons. In early 2020, similar to other provinces and cities, the outbreak of COVID-19 in Shanghai originated from the imported cases in Wuhan. Approximatively, the spread of disease could be roughly divided into two stages: imported case transmission and local case transmission. As an international large city, Shanghai had the ability to quickly identify and isolate these imported cases through information means to block their transmission. This reduced the invasive reproduction number in the period of imported case transmission. However, for the stage of local transmission, the local asymptomatic individuals and local exposed individuals contributed much more than the imported members, but they could not be isolated as rapidly as the imported individuals because there was no massive nucleic acid test during that time. This, together with the highest population density in China, intensified the net reproduction number.

Furthermore, the R_0^{net} in Yunnan and Henan dropped significantly because of geographic reasons or control measures. In fact, Yunnan is far from Wuhan, the center of the epidemic, and has a relatively small PD. Henan took the lead in adopting prevention and control measures in the early stage of the epidemic because it neighbors Wuhan, which had a greater advantage in preventing the spread of COVID-19.

5 Discussion

In this paper, we have calculated the invasive reproduction number in each province (large city) of China, studied the relationship between the reproduction numbers and PIS, and the relations between the reproduction number and local PD. These results reveal how the population migration and PD contribute to the outbreaks of COVID-19 in China, which are important for guiding the policy of NPIs. First, since there

is a significant correlation between PD and reproduction number, it is crucial to take more stringent prevention measures, such as reduction of vaccination hesitancy (Razai et al. 2021), home isolation and community control measures, for the region with the higher PD. Second, since the PIS intensifies the reproduction number, it is important to conduct the lockdown policy to curb the spread of COVID-19, and it is necessary to strengthen import and export quarantine when the border is reopened. Moreover, we have calculated the net invasive reproduction numbers based on local cases and analyzed the characteristics of local transmission of COVID-19, which indicates that the total invasive reproduction number in most cases, could cause an overestimation of local disease transmission risk.

Though this paper has obtained some results in the assessment of the transmission capacity of COVID-19, there are some limitations. First, the lack of understanding of COVID-19 in the early stage and the limited detection capabilities have led to large fluctuations in the number of daily reports. And there are asymptomatic infections and incubation cases, which affects the estimation of reproduction number. Second, the Lotka–Euler equation is used to calculate the reproduction number for the phase of early epidemic growth. It is interesting to consider how the geographic differences affect the outcomes of COVID-19 spread in the later stage when a variety of intervention measures are used. Third, the epidemic of COVID-19 is still prevalent all over the world. A variety of mutant viruses have occurred, such as α-virus in British (Davies et al. 2021), β-virus in South Africa (Tegally et al. 2020), γ-virus in Brazil (Faria et al. 2021), δ-virus in India (Campbell et al. 2021) and the latest Omicron-virus (Karim and Karim 2021). Various researches show that these variants have the ability to weaken the efficacy of vaccination. Therefore, it might be an attractable study to investigate how the mutations influence the vaccination and further impact the reproduction numbers. We leave these as future research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Baidu migration. https://qianxi.baidu.com/2020/

Block P, Hoffman M, Raabe IJ et al (2020) Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nat Hum Behav 4:588–596

Campbell F, Archer B, Laurenson-Schafer H et al (2021) Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill 26(24):2100509

Chen Z, Zhang Q, Lu Y et al (2020) Distribution of the 2019-nCoV epidemic and correlation with population emigration from Wuhan, China. Chin Med J 133(9):1044–1055

Cheng C, Chen SY, Geng J et al (2020a) Preliminary analysis on COVID-19 case spectrum and spread intensity in different provinces in China except Hubei province. Chin J Epidemiol 41(10):1601–1605

Cheng XW, Zhou LJ, Huang T et al (2020b) Epidemiological characteristics of the novel coronavirus pneumonia in Sichuan province. J Prev Med Inf 36(8):947–958

Chinese Center for Disease Control and Prevention. Distribution of novel coronavirus pneumonia. http://2019ncov.chinacdc.cn/2019-nCoV/global.html

Davies NG, Abbott S, Barnard RC et al (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science 372(6538):eabg3055

Dietz K (1974) Transmission and control of arbovirus diseases. In: Proceedings of the Society for Industrial and Applied Mathematics. Epidemiology: Philadelphia, pp 104–121

Earn D, Ma J, Poinar H et al (2020) Acceleration of plague outbreaks in the second pandemic. Proc Natl Acad Sci 117(44):27703–27711

Eikenberry SE, Mancuso M, Iboi E et al (2020) To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect Dis Model 5:293–308

Fang LH, Hou JW, Lai JJ et al (2020) Mathematical modeling of COVID-19 spreading dynamics based on a real megapolis map: an elementary study of computational simulations and intervention strategies. Acta Math Appl Sin 43(02):241–259

Faria NR, Mellan TA, Whittaker C et al (2021) Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 372(6544):815–821

Ferguson NM, Laydon D, Nedjati-Gilani G et al (2020) Report 9: impact of NPIs to reduce COVID-19 mortality and healthcare demand. https://doi.org/10.25561/77482

Gatto M, Bertuzzo E, Mari L et al (2020) Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. Proc Natl Acad Sci 117(19):10484–10491

Ge J, He DH, Lin ZG et al (2020) Four-tier response system and spatial propagation of COVID-19 in China by a network model. Math Biosci 330:108484

He Q, Xiao H, Li HM et al (2021) Practice in information technology support for Fangcang Shelter Hospital during COVID-19 epidemic in Wuhan, China. J Med Syst 45(4):830–841

Health Commission of Hubei Province. News. http://wjw.hubei.gov.cn/fbjd/dtyw/

Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42(4):599–653

Hu Y, Wang KF, Wang WD et al (2020) Analysis of transmissibility of COVID-19 and regional differences in disease control. Acta Math Appl Sin 43(2):227–237

Huang SZ, Peng ZH, Jin Z et al (2020) Studies of the strategies for controlling the COVID-19 epidemic in China: estimation of control efficacy and suggestions for policy makers (in Chinese). Sci Sin Math 50(06):885–898

Huo X, Chen J, Ruan S (2021) Estimating asymptomatic, undetected and total cases for the COVID-19 outbreak in Wuhan: a mathematical modeling study. BMC Infect Dis 21:476

Jia JS, Lu X, Yuan Y et al (2020) Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 582(7812):389–394

Karim SSA, Karim QA (2021) Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. Lancet 398(10317):11–17

Kaur SP, Gupta V (2020) COVID-19 vaccine: a comprehensive status report. Virus Res 288:198114

Li Q, Guan X, Wu P et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 382(13):1199–1207

Liu Y, Yang DY, Dong GP et al (2020) The spatio-temporal spread characteristics of 2019 novel coronavirus pneumonia and risk assessment based on population movement in Henan province: analysis of 1243 individual case reports. Econ Geogr 40(03):24–32

Luo XF, Feng SS, Yang JY et al (2021) Nonpharmaceutical interventions contribute to the control of COVID-19 in China based on a pairwise model. Infect Dis Model 6:643–663

Ma J, Dushoff J, Bolker BM et al (2013) Estimating initial epidemic growth rates. Bull Math Biol 76(1):245–260

Nali LHD, Salvador FS, Bonani GDSS et al (2021) Reopening borders: protocols for resuming travel during the COVID-19 pandemic. Clinics (Sao Paulo) 76:e2723

National Bureau of statistics (2020) China Statistical yearbook-2019. China Statistics Press, Beijing

Netease News. Mayor of Wuhan: more than 5 million people left Wuhan and 9 million people stayed. https://news.163.com/20/0126/22/F3ROV3FU0001899O.html

Razai MS, Chaudhry UAR, Doerholt K et al (2021) Covid-19 vaccination hesitancy. BMJ 373:n1138

Roda WC, Varughese MB, Han D et al (2020) Why is it difficult to accurately predict the COVID-19 epidemic? Infect Dis Model 5:271–281

Song PF, Lou Y, Zhu LP et al (2020) Multi-stage and multi-scale patch model and the case study of novel coronavirus. Acta Math Appl Sin 43(2):174–199

Tang B, Bragazzi NL, Li Q et al (2020a) An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infect Dis Model 5(1):248–255

Tang SY, Tang B, Bragazzi NL et al (2020b) Analysis of COVID-19 epidemic traced data and stochastic discrete transmission dynamic model (in Chinese). Sci Sin Math 50(8):1071–1086

Tang B, Wang X, Li Q et al (2020c) Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. J Clin Med 9(2):462–474

Tang B, Xia F, Tang SY et al (2020d) The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemic in the final phase of the current outbreak in China. Int J Infect Dis 96:636–647

Tegally H, Wilkinson E, Giovanetti M et al (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv. https://doi.org/10.1101/2020.12.21.20248640

Tian H, Liu Y, Li Y et al (2020) An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 368(6491):638–642

van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180:29–48

Wallinga J, Lipsitch M et al (2007) How generation intervals shape the relationship between growth rates and reproduction numbers. Proc R Soc B Biol Sci 274(1609):599–604

Wang X, Tang SY, Chen Y et al (2020) When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemic? A data-driven network modeling analysis (in Chinese). Sci Sin Math 50(07):969–978

Xue L, Jing SL, Sun W et al (2021) Evaluating the impact of the travel ban within mainland China on the epidemic of the COVID-19. Int J Infect Dis 107:278–283

Ye YY, Wang CJ, Zhang HO et al (2020) Spatio-temporal analysis of COVID-19 epidemic risk in Guangdong Province based on population migration. Acta Geogr Sin 75(11):243–256

Yu Z, Zhang G, Liu QZ et al (2020) The outbreak assessment and prediction of COVID-19 based on time-varying SIR model. J Univ Electron Sci Technol China 49(03):357–361

Zhang JP, Li Y, Yao MP et al (2020a) Analysis of the relationship between transmission of COVID-19 in Wuhan and soft quarantine intensity in susceptible population. Acta Math Appl Sin 43(02):20–31

Zhang J, Litvinova M, Wang W et al (2020b) Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect Dis 20(7):793–802

Zhang H, Shen HL, Xia L et al (2020c) Construction of COVID-19 epidemic prevention and control capacity based on big data perspective. Acta Math Appl Sin 43(02):468–481

Zhao H, Feng ZL (2020) Staggered release policies for COVID-19 control: costs and benefits of relaxing restrictions by age and risk. Math Biosci 326:108405

Zhou CH, Pei T, Du YY et al (2020) Big data analysis on COVID-19 epidemic and suggestions on regional prevention and control policy. Bull Chin Acad Sci 35(02):200–203