Title: **Intrinsic growth rules of patients infected, dead and cured with 2019 novel coronavirus in mainland China**

Chuanliang Han[1△], Yimeng Liu[2,3,4△], Saini Yang[2,3,4*]

[1]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China

[2]Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China

[3]State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China

[4]Academy of Disaster Reduction and Emergency Management Ministry of Emergency Management and Ministry of Education, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

[△]C.L.H and Y.M.L contributed equally to this work

***Correspondence authors:**

Saini Yang,

Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education & State Key Laboratory of Earth Surface Processes and Resource Ecology & Academy of Disaster Reduction and Emergency Management Ministry of Emergency Management and Ministry of Education, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

Email: yangsaini@bnu.edu.cn

Conflict of Interest: The authors declare no competing conflicts of interest.

Acknowledgements: This study is sponsored by the National Key Research and Development Program of China (2018YFC1508903), the National Natural Science Foundation of China (41621061) and the support of International Center for Collaborative Research on Disaster Risk Reduction (ICCR-DRR).

Abstract

Background

An outbreak of a novel coronavirus (SARS-CoV-2)-infected pneumonia (COVID-19) was first occurred in Wuhan, China, in December 2019 and then spread rapidly to other regions. It has been declared that at least one confirmed case infected by SARS-CoV-2 was found in each province of China by late January 2020.

Methods

We collected the time series data of the cumulative number of confirmed infected, dead, and cured cases from the health commissions in 31 provinces in mainland China. A simple descriptive model in a logistic form was formulated to infer the intrinsic epidemic rules of COVID-19. Furthermore, we compared the intrinsic epidemic rules of COVID-19 in Hubei with that of severe acute respiratory syndrome (SARS) in Beijing, which was obtained from the Ministry of Public Health of China in 2003.

Results

The feature of data of three population groups in all the provinces could be well captured (R^2>0.95) by the proposed descriptive model with slight variation among different provinces. By comparing fitted parameters, we found that the fitted curve of infected case is the earliest to be saturated and has the lowest semi-saturation period in both COVID-19 and SARS. While all three population groups of SARS is later to saturate and has longer semi-saturation period than that of COVID-19.

Conclusions

Our model is robust and stable to depict the intrinsic growth rule for the cumulative number of confirmed infected, dead, and cured cases in 30 provinces, which could depict the epidemic situation in mainland China very well. Despite the virus caused SARS (SARS-CoV) and the virus caused COVID-19 (SARS-CoV-2) are homologous, their epidemiologic characteristics are quite different. The duration of the outbreak would be shorter for COVID-19.

Introduction

In December 2019, a novel coronavirus-infected pneumonia causing human infection was first identified in Wuhan, a city of Hubei province located in central China with a population of 11 million people (Zhu et al., 2020; Hui et al., 2020). The coronavirus was firstly named as 2019-nCoV on January 12, 2020 and may have a probable bat origin (Zhou et al., 2020). Human to human transmission through 2019-nCoV has been confirmed among closed contacts since the middle of December 2019 (Wang et al. 2020; Li et al., 2020). As of February 12, 2020, infections of 2019-nCoV have been identified in all Chinese provincial administrative regions and in dozens of countries around the world (Munster et al., 2020; Holshue et al, 2020; Ki et al., 2020; Shigemura et al., 2020; Nshiera et al., 2020). So far, the coronavirus has been renamed as COVID-19 on February 12, 2020 officially. Different from a previous famous beta coronavirus in 2003, severe acute respiratory syndrome coronavirus (SARS-CoV) (Ksiazek et al., 2003; Kuiken et al., 2003; Drosten et al., 2003) that has a mortality rates of 10% (WHO), that of COVID-19 is 3.25% in Hubei province, and 0.71% in rest of mainland China (from Chinese Center for Disease Control and Prevention，CCDC) up to February 19, 2020.

Here, we used a descriptive model to fit data on the 2019-nCoV cases (infection, death, cure) from January 20, 2020 to February 19, 2020, aiming to capture the epidemiologic characteristics and dynamics of NCIP in 31 provinces in Mainland China. Furthermore, we compared these characteristics with SARS in 2003 using the same model structure in order to validate this model and understand the differences among these two diseases.

Method

Sources of data

We obtained the time series data of 2019-nCoV cases from the National Health Commission of China, and the provincial Health Commission of 31 provincial administrative regions in Mainland China (January 20 to February 19, 2020). The data used in this study include the number of confirmed 2019-nCoV cases, the cumulative number of suspected, dead, and cured cases. All cases were laboratory confirmed following the case definition by national health commission of China (from CCDC website). The basic test procedure has been described in detail in previous work (Huang et al., 2020; Li et al., 2020). We only chose the cases that are tested positive by real-time RT-PCR with specific primers and probes to SARS-CoV-2 both at the corresponding provincial CDC and the National Institute for Viral Disease Control at CCDC.

The severe acute respiratory syndrome (SARS) in 2003 was another fatal coronavirus epidemics over the last two decades, we collected the data (the cumulative number of suspected, dead, and cured cases) of SARS during March 1 to August 16, 2003, in Beijing from the National Health and Family Planning Commission of PRC.

Epidemic curve modeling

We assumed that the population of each region was susceptible to COVID-19, the number of susceptible people per unit time was proportional to the number of uninfected persons, and isolation or closure of the city could affect the spread effect of infected to susceptible people to a certain extent. Previous studies have showed that daily cumulative number of patients could be explained by logistic function (Huang et al., 2003; Wang et al., 2003). Based on the above assumption and previous studies, we modeled the epidemic information of virus as logistic form (Eq. (1))

$$\begin{cases} \frac{dI(t)}{dt} = \alpha I(t)(1 - \frac{I(t)}{N_1}) \\ I(0) = I_0 \end{cases} \quad (1)$$

where N_1 is maximum number of cumulative infections, $I(t)$ is cumulative number of patients at time t, α is the incidence growth rate, and I_0 is the number of infected cases at the initial time.

Analytical solution of Eq. (1) could be written in the form of Eq. (2)

$$I(t) = \frac{N_1}{1+C_1 e^{-\alpha N_1 t}} \quad (2)$$

$$C_1 = (\frac{N_1}{I_0} - 1)e^{\alpha N_1 t_0} \quad (3)$$

Similarly, we could write the analytical solutions for the dead and cured cases by Eq. (4) and Eq. (6).

$$D(t) = \frac{N_2}{1+C_2 e^{-\beta N_2 t}} \quad (4)$$

$$C_2 = (\frac{N_2}{D_0} - 1)e^{\beta N_2 t_0} \quad (5)$$

$$R(t) = \frac{N_3}{1+C_3 e^{-\gamma N_3 t}} \quad (6)$$

$$C_3 = (\frac{N_3}{R_0} - 1)e^{\gamma N_3 t_0} \quad (7)$$

Where β, γ is the growth rate of dead cases and the growth rate of cured cases, respectively. N_2 and N_3 is maximum number of cumulative dead and cured cases. D_0 and R_0 is the number of dead and cured cases at initial time.

Here we simplified Eqs. (2),(4),(6) and unified them as Eq. (8)

$$N(t) = \frac{A}{1+e^{-k(t-t_0)}} \quad (8)$$

where $N(t)$ is the general form of $I(t)$, $D(t)$ or $R(t)$, A denotes N_1, N_2, *or* N_3, and k represents α, β, or γ, indicating the intrinsic rule for growth of cases. t_0 represents the semi-saturation period, illustrating the inflection point of the sigmoid curve.

The data processing and modeling were performed on MATLAB (The MathWorks) with custom scripts. The nonlinear least square (NLS) algorithm was adopted for data fitting and parameter estimation. We used the MATLAB function “nlinfit” to

minimize the sum of squared differences between the data points and the fitted values. Eq. (8) was used in this work as the descriptive model to depict the intrinsic growth rule for infected, dead and cured cases.

Results

As of Feb 19, 2020, laboratory-conformed COVID-19 infections have been reported in each of the 31 Chinese provincial administrative region. So far, there have been a total of 75002 confirmed cases of 2019-nCoV infections in mainland China, including 2118 deaths and 16157 cases cured (From CCDC). Hubei province contributed a majority of cases to this statistical numbers (Fig 1A). Based on the limited infected case data in the early and middle phase of this new coronavirus outbreak, we firstly established a descriptive model with three parameters for each of the 31 provinces in mainland China by (Fig 1B & Fig 2B, See Methods). Fitting goodness of models was expressed by the R-square, which ranged from 0.95 to 0.999, indicating that the models we established fitted the data well. The panel B of Figure 1 illustrates examples of time series data of infected cases and corresponding fitting curves in four provinces (provinces of Xinjiang, Beijing, Guangdong, Zhejiang and Heilongjiang). It is clear that different province varied in patterns of increase of infected cases.

---------- Insert Figure 1 near here ----------

In our model, k is an important parameter to depict the intrinsic growth rule of cumulative number of infected patients (Eq. (8)). Based on our assumption, the growth pattern was supposed to follow a sigmoid form, and the parameter k here denotes how rapid the number of infected cases get to saturate. If k is large, the S form of the curve would be sharp, and the curve will rise up very quickly to be saturated to the maximum number. While if k is only slightly larger than 0, the curve would show a slow and smooth increase from 0 to the maximum.

To further illustrate the spatial distribution of the infected cases' growth pattern, we exerted further analyses of the fitted infected case curves of 30 provinces. We excluded Tibet in this analysis because the number of confirmed infected case did not

increase since Jan 29, 2020. And the only one confirmed case in Tibet has been cured on Feb 12. The spatial distribution of parameter k is shown in Fig 2A, illustrating provinces varied in the increase patterns of cumulative infected cases. Fitted curves of provinces such as Jilin and Qinghai has larger value of k, which means that the number of infections peaked earlier in these regions than in other provinces. By contrast, provinces like Hainan, Hebei, Ningxia, has smaller value of k than other provinces, suggesting the infections of these provinces would continue increasing when infections of other provinces reach saturation (Fig 2B). Fig 2C shows the spatial distribution of predicted maximum number of infected cases (parameter A) in each province of mainland China. Our model shows that the maximum number of infected COVID-19 cases in mainland China would be 81498 (95% confidence interval: [79099, 83896]). Among all the provinces in mainland China, the Hubei province is predicted to have largest number of infected COVID-19 cases, which would be 68056 (95% confidence interval: [65679,70432]). Despite Tibet is predicted to have only one infected case, the model predicted Qinghai province would have 19 cases (95% confidence interval: [18, 19]), which is the fewest compared with other provinces in mainland China.

---------- Insert Figure 2 near here ----------

Moreover, to compare the intrinsic growth rule of infected cases, dead cases and cured cases between COVID-19 and SARS, we applied the descriptive model to fit the data of SARS in Beijing 2003 (Fig 3A). Since the time series data of SARS was full phase of the outbreak, the parameters estimated by our model was solid (R-square>0.95). We found the k value of fitted curves of infected SARS cases in Beijing was 0.18, which was much lower than that of Hubei province of COVID-19 (k = 0.25) (Fig 3 ABC). Furthermore, we found that all 30 provinces' k value were significantly larger than 0.18 (Fig 3D). We also noted the t_0 of fitted curves of infected SARS cases in

Beijing was larger than that of fitted curves of infected COVID-19 cases in each province in mainland China (Fig 3E). Moreover, the increasing rate (k) was negatively correlated (Pearson, r=-0.47, p=0.0072) to semi-saturation period (Fig 3F). We concluded that at the provincial level, the COVID-19 spread more quickly than the SARS did. The infections of COVID-19 increased more rapid and would reach saturation earlier than SARS according to our model.

---------- Insert Figure 3 near here ----------

The outbreak of SARS in Beijing was the most serious among all the provincial administrative regions, and the outbreak of COVID-19 was most serious in Hubei province. Here we used our descriptive model to fit the dead and cured cases' curves for SARS in Beijing and COVID-19 in Hubei province (Fig 4 A-C, F-H, R-square≥ 0.95), and compared the growth features of three types cases of SARS and COVID-19 by two parameters, k and t_0(Fig 4 DEIJ). We found that k values of the three types cases of SARS were lower than that of COVID-19 (Fig 4 DE), and semi-saturation period for them in SARS was higher than COVID-19's, which suggested the duration of COVID-19 would be shorter than that of SARS. For both SARS and COVID-19, their infected curves have the lowest semi-saturation period (Fig 4 EJ), while cured curves have the highest which is in keeping with the common sense that the cure is always later than infection and death. For both SARS in Beijing and COVID-19 in Hubei, the k values of death case curves were lowest compared with curves of infected cases and cured cases. For both SARS and COVID-19, the k value of infected case curve was larger than that of cured curve.

---------- Insert Figure 4 near here ----------

Discussion

We modeled the epidemiologic characteristics of 2019-nCoV for 31 provinces in mainland China based on the data from Jan 20 to Feb 19 during the outbreak. Furthermore, we compared these characteristics with SARS in 2003 using the same model framework.

Compared with dynamical models such as SIR (Susceptible – Infectious - Recovered) or SEIR (Susceptible – Exposed - Infectious - Recovered), our model is relatively simple but robust. We used a differential equation in logistic form to describe the dynamics of the infected, dead and healed cases caused by COVID-19 respectively (See Methods). The increasing of the population no matter in which type of case (infection, death or recovery) has multiple reasons. When the outbreak of COVID-19 originated in Wuhan, it did not cause too much concern at the beginning so that it leads to many infections and starts to export potential exposed cases to other provinces. But as the successive closure of the cities in Hubei province, and isolation of the community in all the provinces in mainland China, the virus spread tendency was weakened. In our model, we also considered these measures taken by the government. We rewrote Eq. (1) to Eq. (9).

$$\frac{dI(t)}{dt} = \alpha I(t) - rI^2(t) \tag{9}$$

where α is growth rate of the number of infections; *I(t)* is cumulative number of infections at time *t*; r could be understood as a parameter to reflect the effect of measures aiming at controlling the virus spread. Here we have decomposed $I(t)$ into two parts. The first part $\alpha I(t)$ could represent the rapid infections growth in the early phase of the outbreak without any measures for epidemic control. The second part $rI^2(t)$ illustrates the slower infections growth after taking effective isolation measures for epidemic control. For death cases, this idea still holds. But things are different for cured case. In contrast, at the early phase of the outbreak, the number of cured cases increase slowly. But once there are effective treatments, the number of

cured cases will rise up quickly. The growth of cured cases would again be slow since the number of patients is limited. All these facts were considered in our highly summarized model (Eq. (8)) in a macroscopic scale. All data sets have a R-square greater than 0.95 when fitted by our framework, so that we could further discuss some intrinsic mechanisms for the growth rules of infection, death and recovery through the parameters estimated from the model (See Methods).

The value of k from Eq. (8) is a critical parameter. k represents how rapid the number of infected cases get to the saturation. We found the variability of k values among fitted curves of infected COVID-19 cases in 30 provinces (Fig 2 AB) is large (Ranging from 0.19 to 0.38). It is clear to show the real situation in each province and this guides us to be more careful against the epidemic. Further we compared this characteristics between COVID-19 and SARS in 2003 in typical provinces (Fig 3 AB). We noted that k values of fitted COVID-19 infection curves in 30 provinces of mainland China (exclude Tibet) are larger than that of SARS in Beijing 2003 (Fig 3 C), which suggests the time cost of this epidemic will be shorter compared with SARS in 2003. This result may related to several aspects. Firstly, it may due to the basic spread property of the virus. Besides, the isolation policy of all the provinces in mainland China was more strictly enforced than it was in 2003. Moreover, the improved clinical skills today may also helpful for shortening the duration of epidemic.

Furthermore, we analyzed semi-saturation period (Fig 4 EK) (t_0 in Eq. (8)) and make comparison in three types of cases (infection, death and recovery) (Fig 4 A-C G-I). Our model could fit them very well (R-square $\geq (R - s)$. k values for the fitted curves of the three types of cases caused by COVID-19 were larger than that of SARS, while t_0 values were the opposite. This again suggests that the epidemic of COVID-19 will be quicker to pass away than SARS. For both SARS and COVID-19, fitted

curves of infections have lowest semi-saturation period (Fig 4 EK), while fitted curves of cured cases have the highest semi-saturation period, which follows the common sense that the cure was always later than infection and death since it costs time to get on the research of effect plans to cure the patients. For both SARS and COVID-19, the *k* value of infected case curve was larger than that of cured curve, while that of death case is the lowest.

References

Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003; 348: 1967–76.

Holshue ML et al. (2020) First Case of 2019 Novel Coronavirus in the United States. N Engl J Med.

Huang C et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet.

HUANG De-sheng, GUAN Peng, ZHOU Bao-sen. Fitness of morbidity and discussion of epidemic characteristics of SARS based on logistic models[J]. Chinese Journal of Public Health, 2003, 19(6): 1-2.

Hui DS, E IA, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, McHugh TD, Memish ZA, Drosten C, Zumla A, Petersen E (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 91:264-266.

Ki M, nCo VT (2020) Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Republic of Korea. Epidemiol Health:e2020007.

Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003; 348: 1953–66.

Kuiken T, Fouchier RAM, Schutten M, et al. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet 2003; 362: 263–70.

Li Q et al. (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med.

Shigemura J, Ursano RJ, Morganstein JC, Kurosawa M, Benedek DM (2020) Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: mental health consequences and target populations. Psychiatry Clin Neurosci.

Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Wang Y, Liu X, The compound Logistic model used to describe epidemic situation dynamics of SARS in Beijing, JOURNAL OF CHINA JILIANG UNIVERSITY,

2005

WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. Dec 31, 2003. https://www. who.int/csr/sars/country/table2004_04_21/en/ (accessedJan 19, 2020).

Zhao Y, Li Y, Wang X, Peng Z (2020) Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA.

Zhou P et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I, Research T (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med.

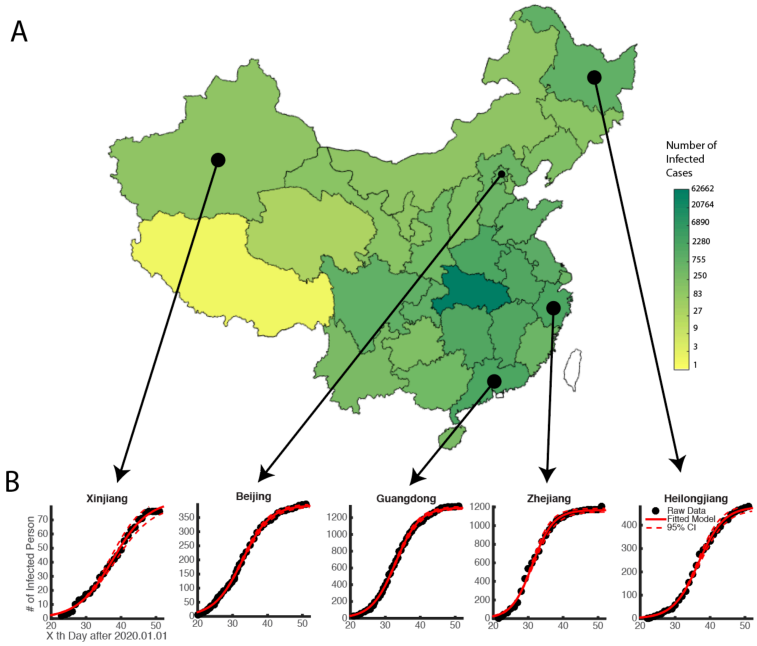

Figure 1 Spatial distribution of the number of COVID-19 infected cases in mainland China.

Panel A is the spatial distribution of the number of COVID-19 infected cases of provinces in mainland China.

Panel B takes five provinces (Xinjiang, Beijing, Guangdong, Zhejiang and Heilongjiang) as example to show the time series of infected cases and corresponding fitted curve. The horizontal axis is the x_{th} day after Jan 1, 2020. And the vertical axis denotes the number of infected cases in corresponding province. The black dots in each panel are raw data of the cumulative infected cases. The red line is the fitted curve by our descriptive model. The dashed red lines show the 95% confidence interval of the fitted curves.

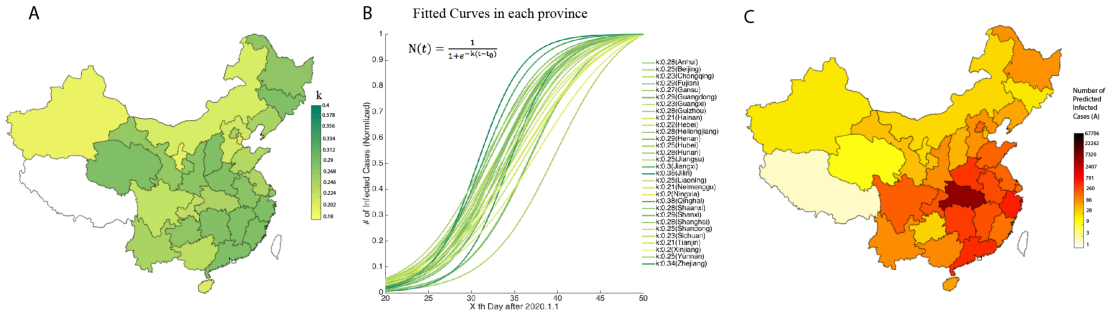

Figure 2 Spatial distribution of increasing rate (*k*) and the normalized fitted curves of each province of mainland China (exclude Tibet)

Panel A demonstrates the spatial distribution of parameter *k* of provinces in mainland China.

Panel B shows the fitted curves of infected cases in 30 provinces. The vertical axis is the number of infected cases (normalized by dividing the maximum number of infected cases in each province). The color of each province is consistent with Panel A.

Panel C is the spatial distribution of the predicted maximum number (parameter A in our model) of COVID-19 infected cases of provinces

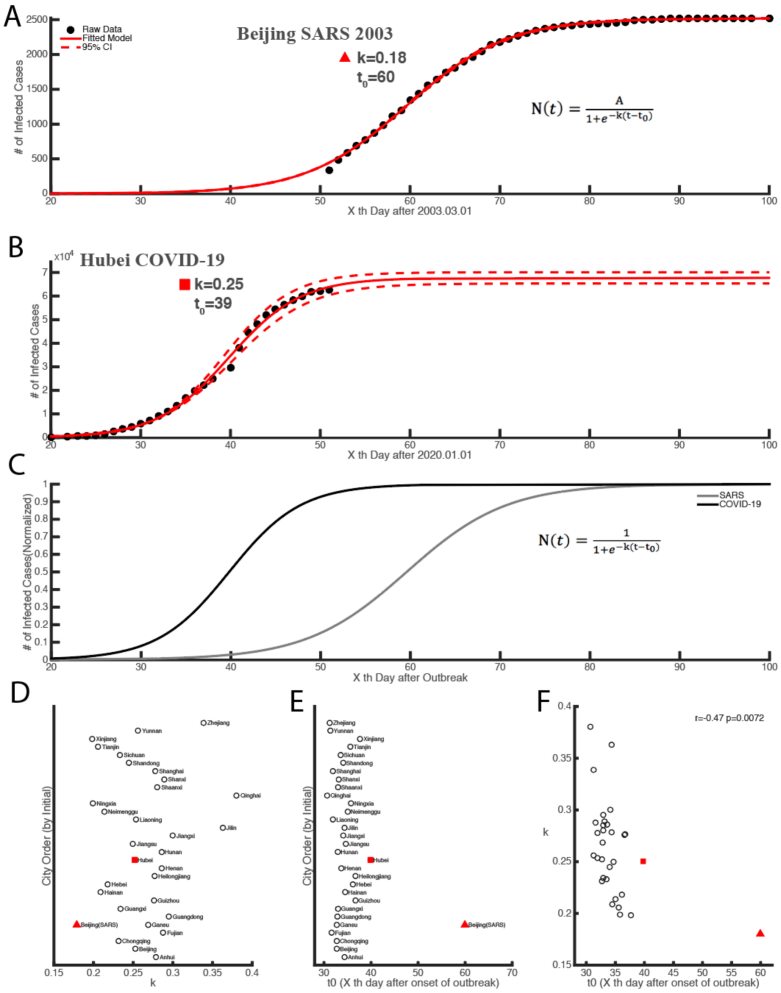

Figure 3 Comparison of intrinsic epidemic rules between COVID-19 in Wuhan and SARS in Beijing (2003)

Panel A shows the raw data and fitted curve of SARS cases in Beijing. The horizontal axis is the x_{th} day after March 1, 2003. The vertical axis denotes the cumulative number of SARS cases in Beijing. The black dots are the raw data of the SARS cases. The red line is the fitted curve by our descriptive model. The dashed red lines are the 95% confidence interval of the fitted curves.

Panel B shows the raw data and fitted curve of COVID-19 cases in Hubei province. The horizontal axis is the x_{th} day after Jan 1, 2020. The vertical axis denotes the cumulative number of COVID-19 cases in Hubei. The black dots are raw data of the COVID-19 cases. The red line is the fitted curve by our descriptive model. The dashed red lines are the 95% confidence interval of the fitted curves.

Pannel C shows the fitted curves of infected SARS cases (gray curve) and COVID-19 (black curve), which are normalized by dividing the maximum number of infected cases in each province.

Panel D illustrates the k value (increasing rate of infected cases) of fitted curves of infected COVID-19 cases in each province in mainland China, where Hubei is represented by the solid red square (the same in D and E). k value of fitted curves of infected SARS cases in Beijing is represented by the solid red triangle (the same in D and E).

Panel E illustrates the t_0 (semi-saturation period) of fitted curves of infected COVID-19 cases in each province in mainland China, as well as t_0 of fitted curves of infected SARS cases in Beijing. The scatter plot of k and t_0 is shown in Panel F, indicating significant negative correlation between k and t_0.

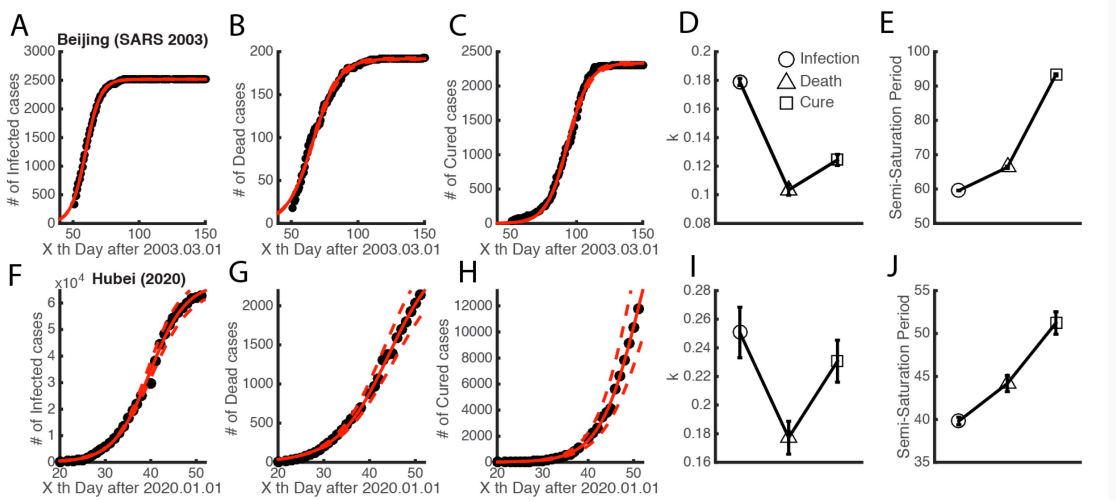

Figure 4 Intrinsic rules of infection, death and recovery of COVID-19 and SARS2003

Panel A-C show the raw data and fitted curve of SARS cases in Beijing. Their horizontal axises are the x_{th} day after Mar 1, 2003. The black dots in Panel A, B and C are raw data of the SARS infected cases, dead cases and cured cases, respectively. The red lines are the fitted curves by our descriptive model. The dashed red lines are the 95% confidence interval of the fitted curves.

Panel D and Panel E respectively shows the k value and the t_0 of fitted curves of three types cases caused by SARS (the circle, triangle and square denotes infected, dead and cured cases, respectively, and the same in Panel I and J). The error bars show the 95% confidence interval of the parameters.

Panel F-H show the raw data and fitted curve of COVID-19 cases in Hubei. Their horizontal axises are the x_{th} day after Jan 1, 2020. The black dots in Panel F, G and H are raw data of the COVID-19 infected cases, dead cases and cured cases, respectively. The red lines are the fitted curves by our descriptive model. The dashed red lines are the 95% confidence interval of the fitted curves.

Panel I and Panel J respectively shows the k value and the t_0 of fitted curves of three types cases caused by COVID-19, which is same with D and E.