Trends in Transmissibility of 2019 Novel Coronavirus-infected Pneumonia in Wuhan and 29 Provinces in China

Huazhen Lin[a], Wei Liu[a], Hong Gao[a], Jinyu Nie[a], Qiao Fan[b]

[a] *Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China*
[b] *Centre for Quantitative Medicine, Program in Health Services & Systems Research, Duke-NUS Medical School, Singapore*

Abstract

Background The 2019 novel coronavirus infected pneumonia (COVID-19) represents a significant public health threat. The COVID-19 emerged in December 2009 in Wuhan, China and rapidly spread to other regions and countries. The variation in transmission patterns and disease spread in regard to time or among different locations, partially reflecting the public health intervention effects, remains to be quantified. As most transmissibility-related epidemic parameters are unknown, we sought, with minimal assumptions, to estimate real-time transmissibility and forecast new cases using dynamic modelling.

Methods Using the cases reported from the National Health Commission of China and transportation data, including the total number of travelling hours through railway, airplane, and car outbound from Wuhan, we have built a time-series model to estimate real-time underlying transmission rates of newly generated cases sequentially from January 20, 2020 to Feb 13, 2020 in Wuhan, Hubei province and other 28 provinces in China. We quantified the instantaneous transmission rate and relative reproduction number (R_t) of COVID-19, and evaluated whether public health intervention affected the case transmissibility in each province. Based on the current estimates, we have predicted the trend of disease spread with a high level of certainty.

Findings We estimated that R_t declined from the range of 4 to 5 towards

Email address: `linhz@swufe.edu.cn` (Huazhen Lin)

1 and remained below unity, while there was an initial growth followed by a decline in a shorter period in Hubei and other provinces. The ratio of transmission rates decreased dramatically from January 23 to 27 likely due to the rigorous public health intervention implemented by the government beginning on January 23, 2020. The mean duration of the infectious period was 6 to 9 days. We have predicted the trend of infection sizes which became stable in provinces around February 19 to 24, 2020, and the date of containment would be one-week later in Wuhan.

Interpretation Public health interventions implemented at both the social and personal levels are effective in preventing outbreaks of COVID-19 in Wuhan and other provinces. Model prediction results suggested that COVID-19 will be contained around the end of February 2020 in China.

Keywords: Coronavirus, COVID-19, transmissibility, dynamic reproduction number R, statistical modelling, pneumonia outbreak

---

1. Introduction

In early December 2019, a novel coronavirus, SARS-COV-2, emerged into the human population in Wuhan, the pandemic center, China[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. The number of SARS-COV-2 infected pneumonia (COVID-19) cases has increased rapidly since then in Wuhan. On January 19, 2020, the first case reported in another province was a person who traveled from Wuhan. From January 23, 2020, a series of substantial public health interventions including travel bans into and from Wuhan, social isolation, quarantine and wearing face masks, have been implemented in Wuhan and subsequently other cities in China such as all 9 localities have successively launched the first-level response to major public health emergencies. As of February 20 (at the time of writing), there have been 75,567 confirmed cases and the death toll stands at 2239 (*http://www.nhc.gov.cn*), 2-fold greater than what was reported ten days ago, and greatly surpasses the infections and tolls from severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak in 2003.

For early assessment of the epidemic potential of COVID-19 outbreak and potential effects of government and public health intervention, it is essential to quantify the epidemiological parameters in real-time from a time series data. Using the earlier phase of outbreak data, the basic reproduction number R_0 has been estimated to be somewhere between 2.2 to 4.0 in

Wuhan[2, 3, 11, 12, 13, 14]. In practice, it becomes crucial to monitor quantitative changes in transmission rates and the effective reproduction number R_t over time to reveal the impacts of control measures[15, 16, 17, 18, 19]. The transmissibility depends on the biological properties of the coronavirus, as well as the contact patterns which can be intervened at the national or social levels in populations. The dynamic changes in transmissibility of COVID-19 in Wuhan and across provinces in China remain unknown. We hypothesized that there could be a significant reduction of transmissibility with time which is in accordance with the public health interventions in Wuhan and other provinces.

Here we provide a real-time model-based analysis to estimate trends in transmissibility of COVID-19 from January 20 to February 13 in Wuhan, Hubei and other 28 provinces in China and forecast the turning point to reach a potential outbreak plateau based on the surveillance case counts as well as transportation and population immigration data from Wuhan to other cities. Our method, without relying on an assumption of epidemiological parameters for disease progression which is absent for the novel pathogen in our study, is a flexible and generalizable approach to directly estimate the distribution of the transmissibility trend.

2. Methods

2.1. Sources of Data

We obtained the number of COVID-19 confirmed cases of time series data between January 20, 2020 to February 13, 2020 in China from the official websites of the National Health Commission of China and Provincial Health Committees (http://www.nhc.gov.cn). The data of cases for each of the 28 provinces (24 provinces plus 4 municipalities including Beijing, Shanghai, Chongqing and Tianjin) at 23 time points were included, as well as for Wuhan and major cities in Hubei. The start date of January 20 was chosen because the official diagnostic protocol released by WHO on January 17 allows the new COVID-19 cases to be diagnosed accurately and rapidly. The end date of February 13 was chosen is because the diagnosis criteria for Wuhan and Hubei province were changed to include those identified by clinic diagnosis since February 13. All cases in our study were laboratory-confirmed with the detection of viral nucleic acid following the case definition by the National Health Commission of China.

Wuhan is connected to other cities in China via high-speed railway, highway, and airplane flights. Population mobility statistics to estimate the exposed sizes in cities outside Wuhan were based on transport-related databases below: 1) Railway and airline travel data: the daily numbers of outbound high-speed trains from Wuhan with corresponding travelling hours were obtained from the high-speed rail network (http://shike.gaotie.cn) from December 1, 2019 to January 23, 2020., and similarly daily numbers of outbound flight and hours for air transport were obtained from the Citytrip network (https://www.ctrip.com) from December 1, 2019 to January 23, 2020. We calculated daily travelling hours which equal to the product of the outbound trip counts and the travelling hours for rail and air transport respectively from Wuhan to each major city. For a given province, we summarized the total number of travelling hours across all cities in that province. 2) Highway mileage data: we collected highway mileage data from bus station networks at https://www.qichezhan.cn. It contains the highway mileage from Wuhan to 16 cities in the Hubei Province. 3) Migration data: we obtained population migration data from the Baidu Migration Map (http://qianxi.baidu.com) which includes the rate of migration among the population leaving Wuhan to other cities and provinces from January 1 to 28, 2020. Total travelling hours for rail and air flight, and migration scales are plotted by the province in Figure 1. Accumulated time on trains, on airplanes, highway mileage and population migration scales were used to model the underlying epidemic sizes in the provinces or cities outside Wuhan at the time 0 of this study which is on January 20, 2020.

From Figure 1, we observed that Guandong has the largest traveling hours through railway and airplane outbound from Wuhan among the provinces. Also, the largest population has immigrated from Wuhan to Henan. In Hubei province, Cities of Huanggan and Xiaogan are the closest to Wuhan in terms of mileage and the scale of migration. These simple observations are consistent with our result that Guandong, Henan, Huanggan and Xiaogan have the largest number of estimated primary infected cases imported from Wuhan on January 20, 2020.

2.2. Modelling the transmissibility of COVID-19

We introduce the main notation here. All times are calendar times, measured in days since the begining of 20 January, 2020, which was the start of

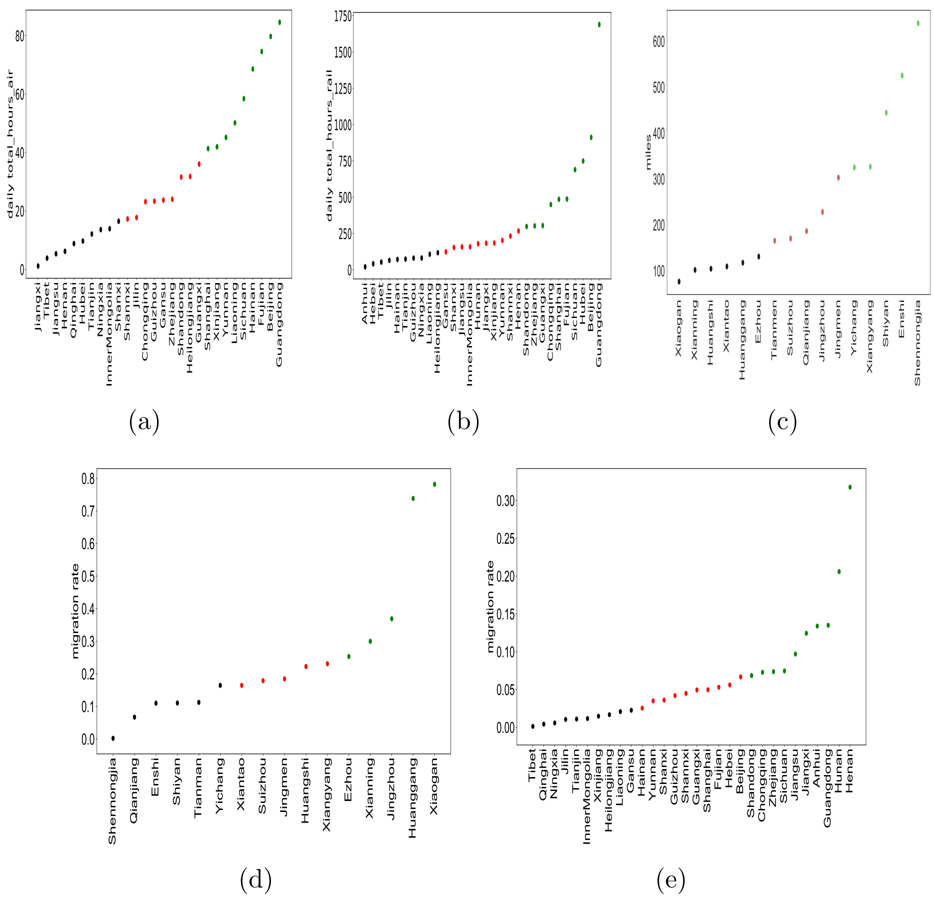


Figure 1: (a) Daily average total traveling hours through airplane from Wuhan; (b) daily total traveling hours through railway from Wuhan; (c) highway mileage from Wuhan to the cities in Hubei; (d) and (e): migration rate from Wuhan during January 1 to 28, 2020.

the epidemic.

Y_{kt}	number of accumulated diagnosed case till day t,
TR_k	daily traveling hours on trains from Wuhan,
FL_k	daily traveling hours on airplane from Wuhan,
RM_k	highway mileage from Wuhan,
MI_k	volumes of migration from Wuhan from January 1 to 28, 2020,
α_k	number of underlying primary infected cases on January 20, 2020,
W_{kt}	underline number of infected individuals who are infectious,
γ_{kt}	transmission rate defined by $dW_{kt}/W_{k,t-1}$,
η_{kt}	ratio of transmission rate defined by $\gamma_{kt}/\gamma_{k,t-1}$,
m	duration of infectious period (day),

where the subscript k represents province or city k, the subscript t represents day t. TR_k and FL_k are constructed based on the two reasons. One is that the longer people stay on the train or plane, the more likely he(she) is to get infected. Another is that the infection happens in local area, hence the number of trains or planes has more information than the population taking trains or planes. In addition, in Hubei province, most people left Wuhan by cars, we use RM_k as one of measurement for the spatial distance between city k and Wuhan.

2.2.1. Modeling for 28 provinces

First, we build an index α_k to represent the baseline infected cases in province k on 20 January, 2020. Particularly, we will use TR_k, FL_k and MI_k to measure the relationship between provice k and Wuhan. We suppose

$$\alpha_k = \beta_1 \times TR_k + \beta_2 \times FL_k + \beta_3 \times MI_k, \text{ for province } k, \tag{1}$$

where $\beta = (\beta_1, \beta_2, \beta_3)'$ are estimated by the observed Y_{kt} in provinces $k = 1, \cdots, K$ and $t = 1, \cdots, T$.

So far, we are not sure the key epidemiological parameters that affected spread and persistence. We hence make assumptions as least as possible. With the notations defined above, it is obvious that the average new cases in province k at day t is $\gamma_{kt} W_{kt}$. We then assume a Poisson distribution for the new cases diagnosed in province k at day t with mean $\gamma_{kt} W_{kt}$, that is

$$dY_{kt} = Y_{kt} - Y_{k,t-1} \sim Possion(\gamma_{kt} W_{kt}), \tag{2}$$

where '$\sim$' means 'distributed as'.

Under the unified leadership of the central goverment, we suppose the trend of γ_{kt} over day t is the same for 28 provinces, that is, $\eta_{kt} = \eta_t$ is independent of k so that

$$\gamma_{kt} = \eta_t \times \gamma_{k,t-1}. \tag{3}$$

To avoid strong assumptions about the evolution of the epidemic, we allow η_t to be arbitray function of t. We determine the functional form of η_t by pointwise estimating η_t and checking the resulting pattern over t. Denote the resulting functional form for η_t by $\eta_t = \eta_t(a)$.

Finally, we notice that

$$dW_{kt} = \gamma_{kt} W_{k,t-1}, \; W_{kt} = W_{k,t-1} + dW_{kt}. \tag{4}$$

With the chain calculation, we have $dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1}(\gamma_{kj}+1)W_{k0}$ and $W_{kt} = \prod_{j=0}^{t-1}(\gamma_{kj}+1)W_{k0}$, where $W_{k0} = \alpha_k$ and $\gamma_{k0} = 0$. In practice, the infected patients will be isolated and removed from the infectious source. With the notation m of duration of infectious period, we hence have

$$W_{kt} = \prod_{j=0}^{t-1}(\gamma_{kj}+1)\alpha_k - I(t>m)\prod_{j=0}^{t-(m+1)}(\gamma_{kj}+1)\alpha_k. \tag{5}$$

Denote $\gamma_1 = (\gamma_{11}, \cdots, \gamma_{K1})'$ and all of the parameters by $\delta = (\gamma_1', a', \beta')'$. Taken (1), (2), (3) and (5) together, the loglikelihood function was

$$\begin{aligned} L(\delta) &= \sum_{k=1}^{K}\sum_{t=1}^{T}\{dY_{kt}\log(\lambda_{kt}) - \lambda_{kt}\} + C \\ &= \sum_{k=1}^{K}\sum_{t=1}^{T}\Big(dY_{kt}\log[\gamma_{kt}X_k^{\mathrm{T}}\beta\left\{\Pi_{i=1}^{t-1}(1+\gamma_{ki}) - I(t>m)\Pi_{i=1}^{t-m}(1+\gamma_{ki})\right\}] \\ &- \gamma_{kt}X_k^{\mathrm{T}}\beta\left\{\Pi_{i=1}^{t-1}(1+\gamma_{ki}) - I(t>m)\Pi_{i=1}^{t-m}(1+\gamma_{ki})\right\}\Big) + C, \end{aligned} \tag{6}$$

where C is a constant independent of δ, m is determined by minimizing the prediction error. The confidence intervals were obtained based on 200 bootstrap resampling[20, 21].

2.2.2. Modeling for Hubei and Wuhan

The modeling and the loglikelihood function for Hubei are similar with those for 28 provinces except that FL_k is replaced by RM_k and provinces are replaced by cities, because there are not flights between the cities in Hubei and Wuhan, and the most people leave Wuhan by cars or buses. Specifically,

$$\alpha_k = \beta_1 \times TR_k + \beta_2 \times RM_k + \beta_3 \times MI_k, \text{ for city } k \text{ in Hubei.} \quad (7)$$

The modeling and the likelihood function for Wuhan are similar with those for 28 provinces except that α_k is directly estimated by the diagnosed cases in Wuhan.

2.2.3. The calculation of the time-dependent reproduction number R_t

When $W_{k,t-1} = 1$, we have $\gamma_{kt} = dW_{kt}$. Hence γ_{kt} is the average number of new infections created by an infectious individual in one day, then $\phi_t = \sum_{k=1}^{K} \gamma_{kt}/K$ is the corresponding average number across provinces or cities. Since one infectious individual can make infection for m days, an infectious individual then can lead to $R_t = m\phi_t$ new infections, which indeed is the time-dependent reproduction number.

In addition, by Bettencourt and Ribeiro (2008)[16], we also calculate the time-dependent reproduction number by $\tilde{R}_t = m \log\left\{\frac{\sum_k \lambda_{k,t+1}}{\sum_k \lambda_{k,t}}\right\} + 1$, where $\lambda_{k,t} = \gamma_{kt} W_{kt}$.

2.2.4. Predication of potential turning point in NCPI outbreak

With the estimated parameters by maximizing the loglikehood (6), we can estimate and predict the average new cases $dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1}(\gamma_{kj} + 1)\alpha_k$, then the cumulative cases $\tilde{W}_{kt} = \prod_{j=0}^{t}(\gamma_{kj} + 1)\alpha_k$. Based on the cumulative cases, we can predict the turning point in the COVID-19 outbreak. In the paper, we defined the turning point to be the day when the number of the cumulative cases reached a plateau, which satisfying $|\partial f(v)/\partial v| \leqslant c_0$, where $f(v) = \frac{\partial \tilde{W}_{kt}/\partial t|t=v}{\partial \tilde{W}_{kt}/\partial t|t=v-1}$ and c_0 is a prespecified small number. We take $c_0 = 2e - 03$ through the analysis.

3. Results

We used a series of values of m, ranging from 3 to 23 to fit the model. The optimal model with $m = 6$ for 28 provinces and Wuhan and $m = 9$ for

Hubei were chosen based on the lowest prediction errors in Supplementary Figures 1(a), 4(a) and 6(a), and $\eta(t) = a_0 + a_1(t - t_1)_- + a_2(t - t_2)_-$ by Supplementary Figures 1(b), 4(b) and 6(b) for all of 28 provinces, Hubei and Wuhan but with different t_1 and t_2, where $t_- = \min(t, 0)$. The functional form of $\eta(t)$ is obtained by pointwise estimating η based on the data till t for $t = 3, \cdots, T$.

3.1. Results from 28 provinces

With the form of $\eta(t) = a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$ with t_1 being January 23 and t_2 being January 27 and $m = 6$, we estimate $\delta = (\gamma_1', a', \beta')'$ based on $t = 1, \cdots, T$ of 28 provinces using the method displayed in Section 2. Table 1 displays the resulting estimators for β. Rail transportation and migration from Wuhan had significant effects on number of infectious on January, 20 so the epidemiological scale (p-value = 0.015 and p-value = 0.02, respectively), but not for air transportation (p-value = 1.0).

Figure 2(a) displays the transmission rates γ_{k1} on January 20, 2020 by province and Figure 2(c) displays the ratio of transmission rate η_t over day t. Since $\gamma_{kt} = \gamma_{k1} \prod_{j=2}^{t} \eta_j$ and the ratio of transmission rate η_t are the same over 28 provinces, the initial transmission rates γ_{k1} can be used to compare the strength of the prevention and control for the NCP epidemic in 28 provinces. The transmission rates γ_{k1} varied greatly by province, from 0.58 to 0.98 for Beijing and Heilongjiang, respectively.

Figure 2(b) displays the underlying infected cases α_k on January 20, 2020. The top five provinces with the highest number of cases were in Guangdong, Henan, Beijing, Sichuan and Hunan respectively, and the lowest numbers were in Ningxia, Jilin, Hainan and Heilongjiang.

Those with the highest imported cases α_k did not necessarily exhibit the highest transmission rate γ_{k1}. For instance, Beijing and Sichuan had the highest underlying cases but low transmisison rates (0.58 and 0.65 in Figure 2(a)); on the contrary, Heilongjiang and Jilin had the lowest underlying cases but high infection rates ($0.83 - 0.98$ in Figure 2(a)). Interestingly, this might be in accordance with the effects implemented in each province; Beijing and Sichuan are known for their substantial and immediate public health intervention starting on January 20, 2020.

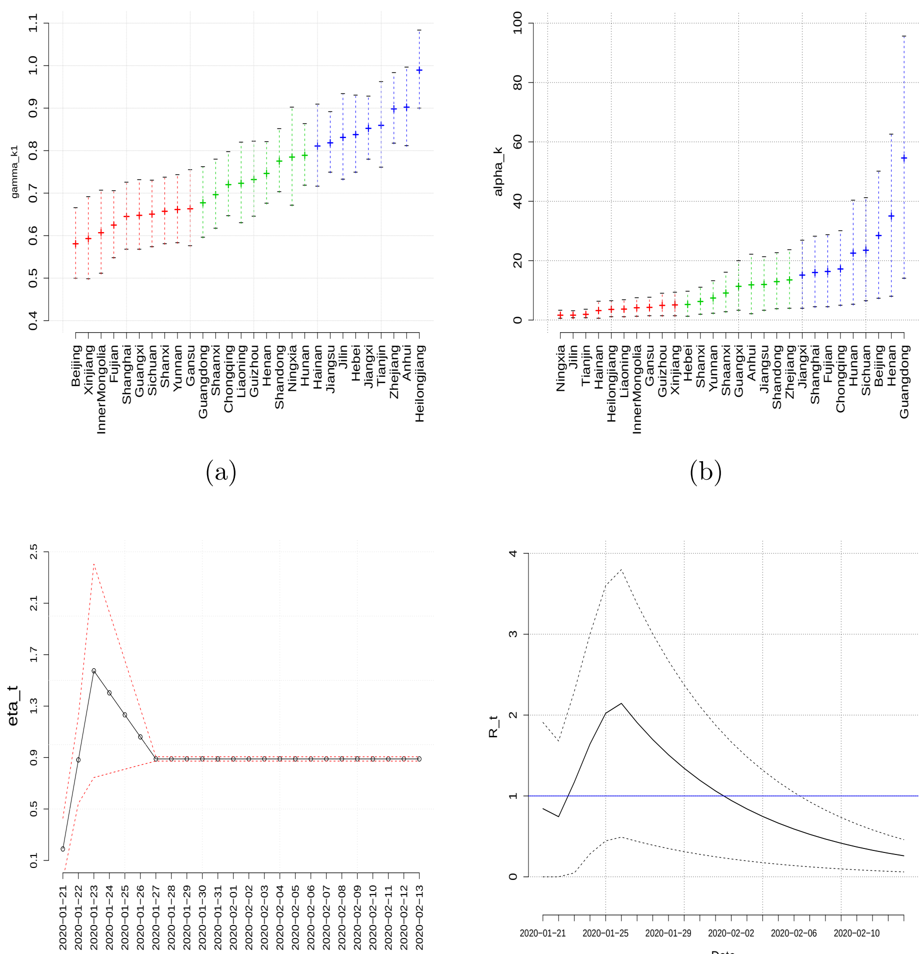
Figure 2(d) displays the R_t and 95% confidence intervals (CI). The reproduction number increased to 2.15 at January 26 and declined to 1 on February 1, then graduately decreased to 0.26 at February 13. We also cal-

Table 1: Effects of transport modes and population immigration to estimate infection cases at baseline

	28 provinces			16 cities in Hubei		
	train	airplane	migration	train	highway	migration
Est.	42.628	1e-05	28.740	1e-05	7.781	20.908
SD	17.498	0.853	12.321	0.106	3.719	5.967
p-value	0.015	1.000	0.0197	0.999	0.036	4e-04

culate $\tilde{R}_t$ from Bettencourt and Ribeiro (2008)[16]. The shape of R_t and $\tilde{R}_t$ are concordant, whereas $\tilde{R}_t$ dramatically fluctuated in the begining.

We further predicted the number of accumulated cases until March for each province using the optimal model we chose above (Figure 3 and Supplementary Figures 2 and 3). For almost all provinces, observed values perfectly fall within 95% CI of the prediction band, suggesting a good fitting, except a few cases. For example, the predicted numbers were somewhat larger than the observed values in Bejing, which could be due to an over-estimation of the underlying case at the baseline. Given no changes in the current control measures, we observed that the predicted numbers reached a plateau from February 19 to 24 in all provinces (Figure 3(i)), marked by a blue line with the date of turning point labeled in each plot.



(a) (b)

(c) (d)

Figure 2: The estimation and 95% CI. (a) the estimated transmission rate on January 20 in each province; (b) the estimated underlying infected cases on January 20 in each province; (c) the estimated ratio of transmission rate over time; (d) Sequential estimation of reproduction number from the daily COVID-19 cases.

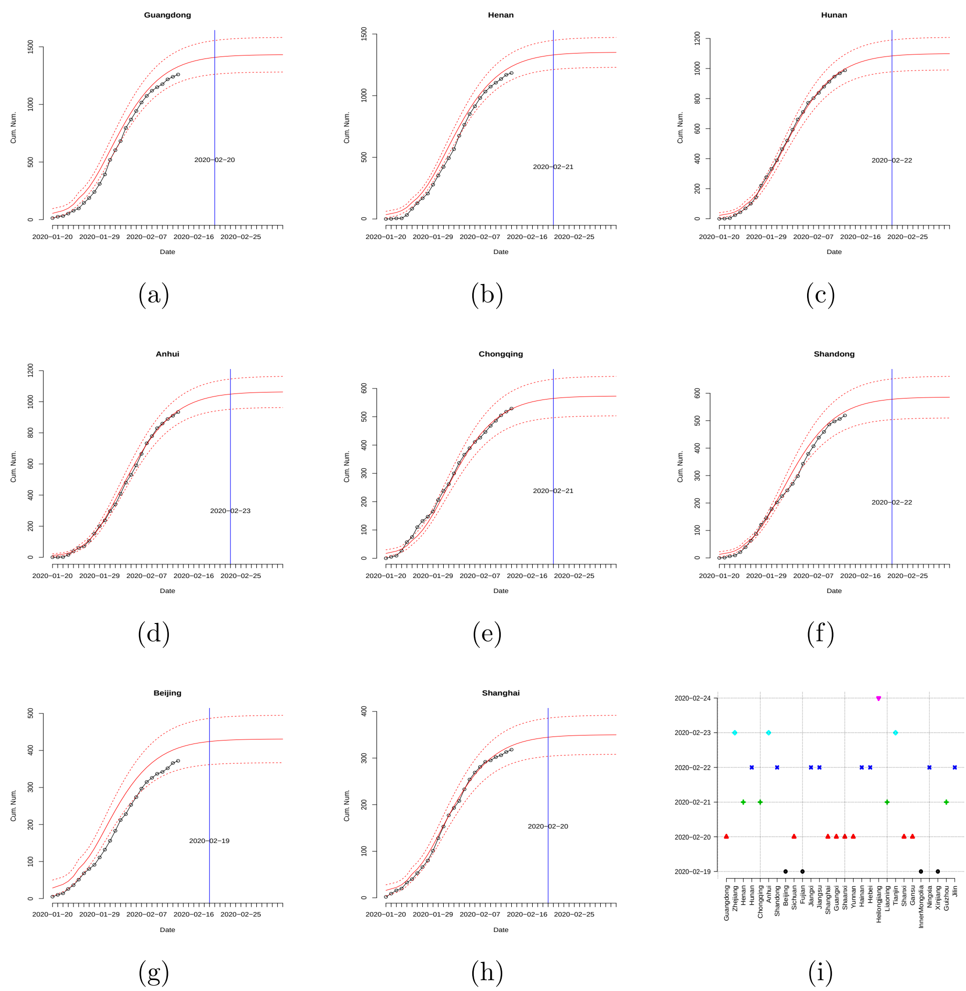


Figure 3: (a)-(h): The estimated (red-solid) and observed (black-dotted) cumulative number of infectious over day t, as well as 95% CI (red-dashed) for typical provinces. (i): The day reaching plateau by provinces.

3.2. *Results from the cities in Hubei province*

With the form of $\eta(t) = a_0 + a_1(t - t_1)_- + a_2(t - t_2)_-$ with t_1 being January 23 and t_2 being January 28 and $m = 9$, we estimate $\delta = (\gamma_1', a', \beta')'$ based on $t = 1, \cdots, T$ of 16 cities in Hubei using the method displayed in Section 2. The duration of the infectious period m was 9 days, 3 days longer than that estimated in other provinces; this could be due to the delay in the diagnosis/hospital admission of infected cases in cities in Hubei.

Table 1 displays the resulting estimators for β. Highway transportation and migration from Wuhan have significant effects on number of infectious on January, 20 so the epidemiological scale (p-value $= 0.036$ and p-value $= 4e-04$, respectively), but not for rail transportation (p-value $= 1.0$). This is different from the conclusion for 28 provinces where rail transportation is significant. This may be attributed to that the cities in Hubei is close to Wuhan and the popular transportation leaving Wuhan is by cars.

Figure 4(a) displays the transmission rates γ_{k1} on January 20, 2020 by city and Figure 4(c) displays the ratio of transmission rate η_t over day t. The transmission rates γ_{k1} varied by city, from 0.45 to 0.80 for Enshi and Suizhou, respectively, except for Shennongjia which is a scenic area. The top two cities with the highest transmission rates were in Suizhou and Huangshi, and the lowest transmission rates were in Shennongjia and Enshi.

Figure 4(b) displays the underlying infected cases α_k on January 20, 2020. The top two cities with the highest number of cases were in Xiaogan and Huanggang, and the lowest numbers were in Qianjiang and Tianmen.

Again, those with higher imported cases α_k did not necessarily exhibit higher transmission rate γ_{k1}. For instance, although Suizhou, Huangshi and Ezhou had lower underlying cases $(6.06 - 7.50)$ but the highest transmission rate $(0.74 - 0.80)$; on the contrary, Enshi had higher underlying cases (9.12) but lower transimission rates (0.48 in Figure 4(a)).

Figure 4(d) displays the R_t and its 95% CI. The reproduction number decreased dramatically to 0.97 on January 22, 2020, then increased to 3.70 on January 27, 2020 and decreased to 1 on February 5.

As the diagnosis criteria for Wuhan and Hubei province expanded to include those identified by clinic diagnosis since February 13, the reported cases increased dramatically since then. As our model was initially built up using the laboratory-confirmed cases, the predicted cases were not presented after February 13 (Supplementary Figure 5). However, the estimated trend, without changes in other parameters, was still informative to predict the turning point. We estimate the growth curve to reach the plateau from February 22 to 26, 2020 (Figure 4(e)).

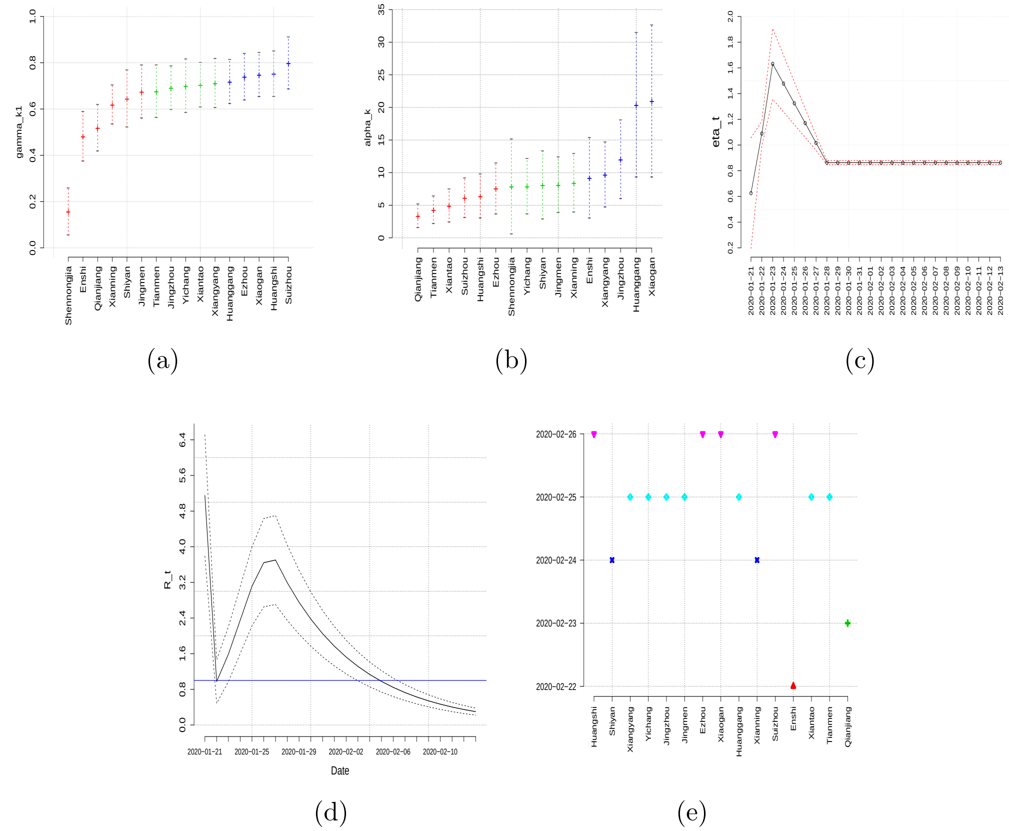
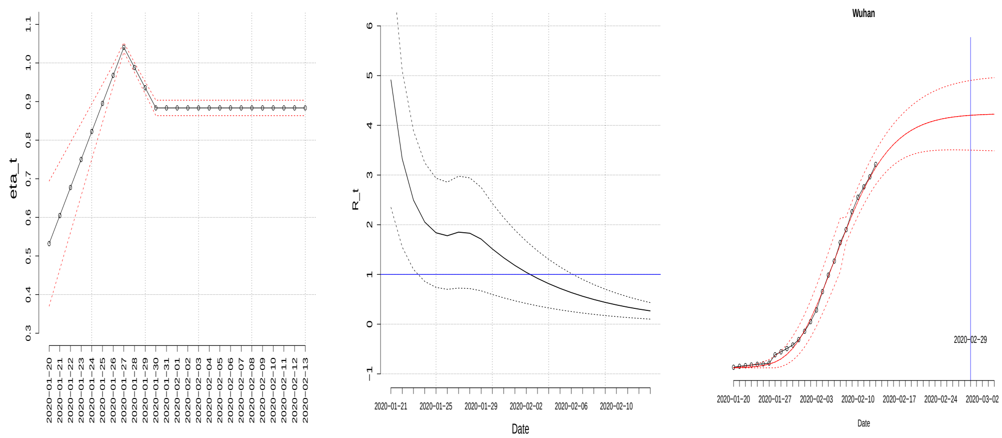


Figure 4: The estimation and 95% CI. (a) the estimated transmission rate on January 20 by cities in Hubei; (b) the estimated underlying infected cases on January 20 by cities in Hubei; (c) the estimated ratio of transmission rate over time in Hubei; (d) Sequential estimation of reproduction number from the daily COVID-19 cases in Hubei. (e) The day reaching plateau by cities in Hubei.

3.3. Results from WuHan city

With the form of $\eta(t) = a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$ with t_1 being January 27 and t_2 being January 30 and $m = 6$, we estimated the parameters based on $t = 1, \cdots, T$ in Wuhan using the method displayed in Section 2.

Figures 5(a) and 5(b) displays the estimators of η_t and R_t, as well as their 95% CI, respectively, for Wuhan. The reproduction number R_t decreased greatly from 4.9 to 2 in 3 days, remained stable till January 28, and then continuously decreased to 1 at February 2. In our prediction model, the observed cases were within the prediction band before February 13. Due to the changes of diagnosis criteria, the predicted cases were not comparable after February 13. However, as mentioned above, the estimated trend was still informative to predict the turning point. Thus we estimate that the date of containment would be around February 29, 2020 in Wuhan (Figure 5(c)).



(a) (b) (c)

Figure 5: The estimation and 95% CI. (a) the estimated ratio of transmission rate over time in Wuhan; (b) Sequential estimation of reproduction number from the daily COVID-19 cases in Wuhan; (c) The estimated (red-solid) and observed (black-dotted) number of infectious over day t, as well as 95% CI (red-dashed) of the estimators for Wuhan.

4. Discussions

Here we described transmission dynamic patterns of the COVID-19 outbreak using time-serial data in Wuhan and 29 provinces in China from January 20 to February 13, 2020. The instantaneous R_t declined from the range of 4 to 5 towards 1, from January 21 to February 2 in Wuhan, while there was an initial growth followed by a decline in a shorter period in Hubei and other provinces. The ratio for transmission rates decreased dramatically from January 23 to 27, likely due to the rigorous public health intervention beginning on January 23, 2020. Model prediction results suggested that COVID-19 would be contained in provinces from February 19 to 24, 2020, and the date of containment would be one-week later in Wuhan.

The temporal distributions for R_t give a basis for assessing the evolution of transmissibility over time[15, 16, 22]. For Hubei and other provinces, there was an initial growth from January 21 to 27, 2020 with R_t up to 4 followed by a decline in a short period. The increase in R_t was in accordance with the increase in transmission, but at a slower rate. There was a dramatically declining trend in the ratio for the transmission rate from January 23 to 28. Since January 23, 2020, the Chinese government banned travels into and from Wuhan by air, rail and road access to halt the spread of cases from Wuhan to other cities. Also, almost all provinces initiated the highest level of public health emergency responses during January 23 and 25 including tracking and isolating close contacts with COVID-19 patients, self-quarantine, etc. We speculate that, based on the coincidence of a series of intervention effects, the decline in R_t could be due to a series of intervention measures to reduce the transmission rate implemented since January 23, 2020.

For Wuhan, the reproduction number R_t showed a trend of decline in general from 4.9 to 1 on February 2, 2020, although a slight shift upward on January 27 and 28. Different from other places, the outbreak in Wuhan could be in earlier January with sparse data. In our study, the initial R_t was estimated at 4.9 on January 21 and even higher at 8.13 before this date (data not shown), larger than the estimation in other places as expected. We also observed that the ratio for transmission rates decreased from January 28 in Wuhan, suggesting the intervention starting on January 23 might have lag effects on the transmission rates. Several studies have estimated the basic reproduction number R_0 in a range of 2.2 to 4.0 in Wuhan[2, 3, 11, 12], highly contagious compared to SARS[13, 23, 24, 25, 26, 27, 28]. One study[13] estimated R_t in 830 cases in Wuhan from December 24 and January 23 and found that R_t decreased from 8 to 2.52. The trend is similar to what we observed; however, with the interaction involved, we estimated R_t declined and approached the critical threshold at 1 on February 2, 2020.

The difference in transmission rates among different provinces highlights the importance of its contact patterns and control measures, given no required immunity obtained in the early outbreak phase[4, 18]. As of January 20, underlying cases imported from Wuhan were the highest in Guangdong, Henan, Beijing and Sichuan. Beijing and Sichuan had the lowest transmission rates (ranging from 0.58 to 0.65), and Heilongjiang, Jilin and Tianjin had the lowest underlying cases but high transmission rates (0.83 to 0.98). This could be related to the extent of efforts in each province to reduce the number of contacts in the transmission process, particularly challenging for those with the highest number of cases. It was reported a high sensitivity of the timing of implementation of control for SARS, a 1-week delay in implementation of control measures results in a 2.6-fold increase in mean epidemic size[16]. Thus the immediate stringent controls measures have critical impact to prevent the spread of diseases.

We forecasted that COVID-19 will be contained across provinces around February 19 to 24, and in Wuhan one week later. The model fitting to the data performed very well for all regions as shown by the observed cases generally falling in the 95% CI of the bound of prediction in the plots. Zhejiang and Guizhou provinces showed statistical anomalies with the largest predictor error (Supplementary Figure 3); the observed cases were either underestimated or overestimated. This could be due to the variation in the duration of the infectious period or the number of underlying case estimation at baseline as the transport data we have might not be completed, which warrant further investigation.

The estimation of the traditional SIR or SEIR models[29, 30, 31, 32, 33] requires information about the epidemiological parameters of disease progression, which is absent for the novel pathogen in our study. Studies, for example, used serial interval estimates based on previous estimations of SARS-Cov[2, 3, 12]. It would be desirable to estimate them from the model itself. The likelihood-based method modelling presented here can be used for this purpose in real-time estimations of epidemiological parameters, with minimal assumptions. We assumed secondary cases are generated according to a Poisson distribution[22, 34]. Our method relies on the inference of transmission among the underlying statistical distribution of the infected cases during the infectious period. It can be used to provide statistical expectations for new case predictions. Large time-series data permit us to estimate the rates on a daily basis. Thus, for the calculations in this study, we choose to use a minimal statistical approach for emerging infectious diseases which is more appropriate and generalizable.

Our study is not without limitations. First, our results are based on the diagnosis cases reported by the CDC in China. Underreporting is likely to occur which was not accounted for in our analysis. Among all cases reported until February 11, only 13.8% cases were reported prior January 20, 2020 and among them, 77.6% were among the Hubei province[35]. Since we are most interested in the temporal trend after January 20, the underreporting might not vary greatly with time since then. Second, as there is a series of interventions implemented during subsequent periods, we were not able to distinguish which intervention could be the main driver based on current data. Third, due to the changes in diagnosis, the diagnosed cases increased dramatically up to 14840 in Wuhan and Hubei from February 12, which makes our prediction challenging. Thus we did not provide the predicted cases for Wuhan and Hubei province after February 13, 2020.

In summary, we provide transmission dynamic patterns of the COVID-19 outbreak using time-serial data from January 20 to February 13, 2020 in China. The declining trend of R_t, as well as ratio for transmission rates, indicate the effects of public health intervention implemented by the government beginning on January 23, 2020. Model prediction results suggested that COVID-19 would be contained in provinces around 19 to 24 February, 2020, and the date of containment would be one-week later in Wuhan.

Declaration of interests

We declare no competing interests.

Acknowledgments

The research were partially supported by National Natural Science Foundation of China (Nos. 11931014 and 11829101) and Fundamental Research Funds for the Central Universities (No. JBK1806002) of China.

References

[1] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china, Jama.

[2] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, New England Journal of Medicine.

[3] J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study, The Lancet.

[4] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, et al., Genomic characterisation and epidemiology

of 2019 novel coronavirus: implications for virus origins and receptor binding, The Lancet.

[5] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature (2020) 1–4.

[6] D. K. Chu, Y. Pan, S. Cheng, K. P. Hui, P. Krishnan, Y. Liu, D. Y. Ng, C. K. Wan, P. Yang, Q. Wang, et al., Molecular diagnosis of a novel coronavirus (2019-ncov) causing an outbreak of pneumonia, Clinical Chemistry.

[7] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study, The Lancet.

[8] J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, The Lancet.

[9] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al., A new coronavirus associated with human respiratory disease in china, Nature (2020) 1–8.

[10] D. S. Hui, E. I Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drosten, et al., The continuing 2019-ncov epidemic threat of novel coronaviruses to global healththe latest 2019 novel coronavirus outbreak in wuhan, china, International Journal of Infectious Diseases 91 (2020) 264–266.

[11] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, et al., Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak, International Journal of Infectious Diseases.

[12] Z. Zhu, J. Li, D. Gong, D. Wan, S. Chen, L. Guo, Y. Li, L. Sun, W. Liang, T. Song, et al., Time-varying transmission dynamics of novel coronavirus pneumonia in china.

[13] T. Liu, J. Hu, J. Xiao, G. He, M. Kang, Z. Rong, L. Lin, H. Zhong, Q. Huang, A. Deng, W. Zeng, X. Tan, S. Zeng, Z. Zhu, J. Li, D. Gong, D. Wan, S. Chen, L. Guo, Y. Li, L. Sun, W. Liang, T. Song, J. He, W. Ma, Time-varying transmission dynamics of novel coronavirus pneumonia in china, bioRxivarXiv:https://www.biorxiv.org/content/early/2020/02/13/2020.01.25.919787.full.pdf, doi:10.1101/2020.01.25.919787.
URL https://www.biorxiv.org/content/early/2020/02/13/2020.01.25.919787

[14] J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study, The Lancet.

[15] Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of covid-19 is higher compared to sars coronavirus, Journal of Travel Medicine.

[16] L. M. Bettencourt, R. M. Ribeiro, Real time bayesian estimation of the epidemic potential of emerging infectious diseases, PLoS One 3 (5).

[17] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, American Journal of epidemiology 160 (6) (2004) 509–516.

[18] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, A. Vespignani, Measurability of the epidemic reproduction number in data-driven contact networks, Proceedings of the National Academy of Sciences 115 (50) (2018) 12680–12685.

[19] X.-S. Zhang, R. Pebody, A. Charlett, D. de Angelis, P. Birrell, H. Kang, M. Baguelin, Y. H. Choi, Estimating and modelling the transmissibility of middle east respiratory syndrome coronavirus during the 2015 outbreak in the republic of korea, Influenza and other respiratory viruses 11 (5) (2017) 434–444.

[20] B. Efron, Bootstrap confidence intervals for a class of parametric problems, Biometrika 72 (1) (1985) 45–58.

[21] B. Efron, et al., Second thoughts on the bootstrap, Statistical Science 18 (2) (2003) 135–140.

[22] L. Forsberg White, M. Pagano, A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic, Statistics in medicine 27 (16) (2008) 2999–3016.

[23] A. N. Desai, M. U. Kraemer, S. Bhatia, A. Cori, P. Nouvellet, M. Herringer, E. L. Cohn, M. Carrion, J. S. Brownstein, L. C. Madoff, et al., Real-time epidemic forecasting: Challenges and opportunities, Health security 17 (4) (2019) 268–275.

[24] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, et al., Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions, Science 300 (5627) (2003) 1961–1966.

[25] C. A. Donnelly, M. C. Fisher, C. Fraser, A. C. Ghani, S. Riley, N. M. Ferguson, R. M. Anderson, Epidemiological and genetic analysis of severe acute respiratory syndrome, The Lancet infectious diseases 4 (11) (2004) 672–683.

[26] C. Dye, N. Gay, Modeling the sars epidemic, Science 300 (5627) (2003) 1884–1885.

[27] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L.-M. Ho, T.-Q. Thach, P. Chau, et al., Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in hong kong, The Lancet 361 (9371) (2003) 1761–1766.

[28] J. S. M. Peiris, C.-M. Chu, V. C.-C. Cheng, K. Chan, I. Hung, L. L. Poon, K.-I. Law, B. Tang, T. Hon, C. Chan, et al., Clinical progression and viral load in a community outbreak of coronavirus-associated sars pneumonia: a prospective study, The Lancet 361 (9371) (2003) 1767–1772.

[29] S. V. Scarpino, G. Petri, On the predictability of infectious disease outbreaks, Nature communications 10 (1) (2019) 1–8.

[30] M. Y. Li, J. S. Muldowney, Global stability for the seir model in epidemiology, Mathematical biosciences 125 (2) (1995) 155–164.

[31] M. Y. Li, J. R. Graef, L. Wang, J. Karsai, Global dynamics of a seir model with varying total population size, Mathematical biosciences 160 (2) (1999) 191–213.

[32] M. Y. Li, H. L. Smith, L. Wang, Global dynamics of an seir epidemic model with vertical transmission, SIAM Journal on Applied Mathematics 62 (1) (2001) 58–69.

[33] P. E. Lekone, B. F. Finkenstädt, Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study, Biometrics 62 (4) (2006) 1170–1177.

[34] S. Cauchemez, P.-Y. Boëlle, C. A. Donnelly, N. M. Ferguson, G. Thomas, G. M. Leung, A. J. Hedley, R. M. Anderson, A.-J. Valleron, Real-time estimates in early detection of sars, Emerging infectious diseases 12 (1) (2006) 110.

[35] T. N. C. P. E. R. E. Team, The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19) in china, Chinese Journal of Epidemiology 41 (2) (2020) 145–151.