

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

Huazhen Lin<sup>a</sup>, Wei Liu<sup>a</sup>, Hong Gao<sup>a</sup>, Jinyu Nie<sup>a</sup>, Qiao Fan<sup>b</sup>

<sup>a</sup>*Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China*

<sup>b</sup>*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](mailto:linhz@swufe.edu.cn) (Huazhen Lin)*

*Preprint submitted to Elsevier*

*February 22, 2020*

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 1. Introduction

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

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

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

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

## 41 **2. Methods**

### 42 *2.1. Sources of Data*

43 We obtained the number of COVID-19 confirmed cases of time series data  
44 between January 20, 2020 to February 13, 2020 in China from the official  
45 websites of the National Health Commission of China and Provincial Health  
46 Committees (<http://www.nhc.gov.cn>). The data of cases for each of the  
47 28 provinces (24 provinces plus 4 municipalities including Beijing, Shanghai,  
48 Chongqing and Tianjin) at 23 time points were included, as well as for Wuhan  
49 and major cities in Hubei. The start date of January 20 was chosen because  
50 the official diagnostic protocol released by WHO on January 17 allows the  
51 new COVID-19 cases to be diagnosed accurately and rapidly. The end date  
52 of February 13 was chosen is because the diagnosis criteria for Wuhan and  
53 Hubei province were changed to include those identified by clinic diagnosis  
54 since February 13. All cases in our study were laboratory-confirmed with the  
55 detection of viral nucleic acid following the case definition by the National  
56 Health Commission of China.

57 Wuhan is connected to other cities in China via high-speed railway, high-  
58 way, and airplane flights. Population mobility statistics to estimate the ex-  
59 posed sizes in cities outside Wuhan were based on transport-related databases  
60 below: 1) Railway and airline travel data: the daily numbers of outbound  
61 high-speed trains from Wuhan with corresponding travelling hours were ob-  
62 tained from the high-speed rail network (<http://shike.gaotie.cn>) from Decem-  
63 ber 1, 2019 to January 23, 2020., and similarly daily numbers of outbound  
64 flight and hours for air transport were obtained from the Citytrip network  
65 (<https://www.ctrip.com>) from December 1, 2019 to January 23, 2020. We  
66 calculated daily travelling hours which equal to the product of the outbound  
67 trip counts and the travelling hours for rail and air transport respectively  
68 from Wuhan to each major city. For a given province, we summarized the  
69 total number of travelling hours across all cities in that province. 2) Highway  
70 mileage data: we collected highway mileage data from bus station networks at  
71 <https://www.qichezhan.cn>. It contains the highway mileage from Wuhan to  
72 16 cities in the Hubei Province. 3) Migration data: we obtained population  
73 migration data from the Baidu Migration Map (<http://qianxi.baidu.com>)  
74 which includes the rate of migration among the population leaving Wuhan  
75 to other cities and provinces from January 1 to 28, 2020. Total travelling  
76 hours for rail and air flight, and migration scales are plotted by the province  
77 in Figure 1. Accumulated time on trains, on airplanes, highway mileage and  
78 population migration scales were used to model the underlying epidemic sizes  
79 in the provinces or cities outside Wuhan at the time 0 of this study which is  
80 on January 20, 2020.

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

## 89 *2.2. Modelling the transmissibility of COVID-19*

90 We introduce the main notation here. All times are calendar times, mea-  
91 sured in days since the beginning of 20 January, 2020, which was the start of

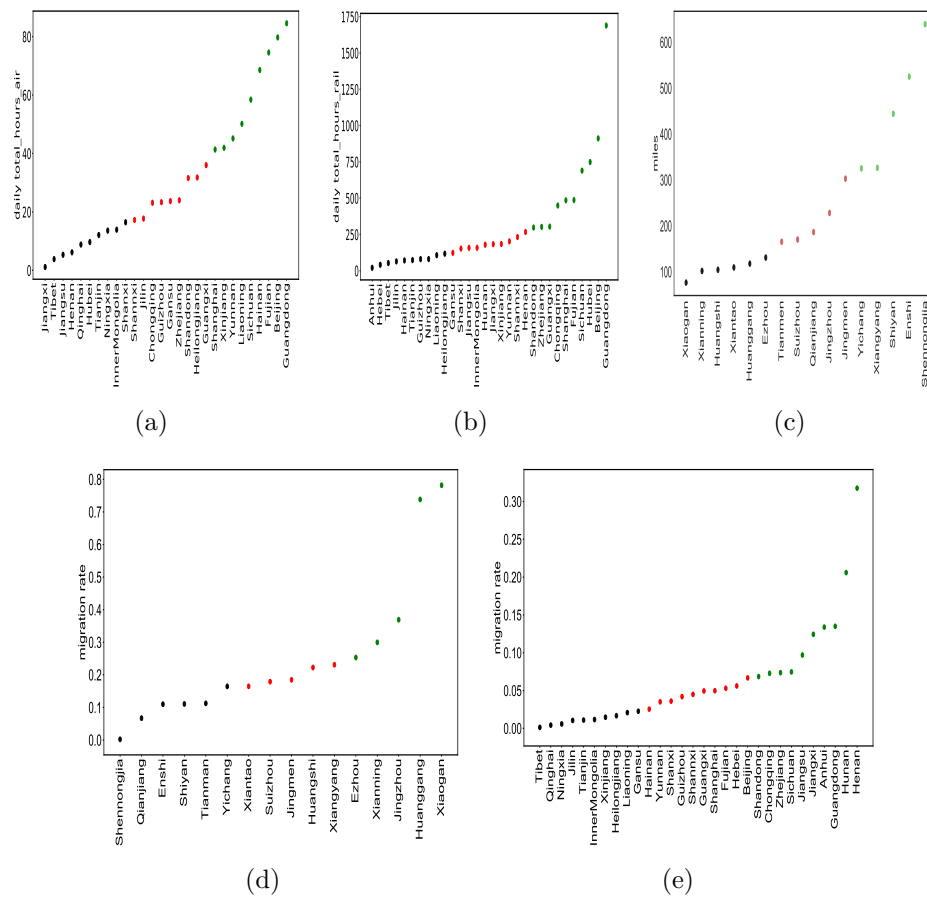


Figure 1: (a) Daily average total traveling hours through airplane from Wuhan; (b) daily total traveling hours through railway from Wuhan; (c) highway mileage from Wuhan to the cities in Hubei; (d) and (e): migration rate from Wuhan during January 1 to 28, 2020.

92 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,
$\alpha_k$	number of underlying primary infected cases on January 20, 2020,
$W_{kt}$	underline number of infected individuals who are infectious,
$\gamma_{kt}$	transmission rate defined by $dW_{kt}/W_{k,t-1}$ ,
$\eta_{kt}$	ratio of transmission rate defined by $\gamma_{kt}/\gamma_{k,t-1}$ ,
$m$	duration of infectious period (day),

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

### 101 2.2.1. Modeling for 28 provinces

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

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

105 where  $\beta = (\beta_1, \beta_2, \beta_3)'$  are estimated by the observed  $Y_{kt}$  in provinces  $k =$   
 106  $1, \dots, K$  and  $t = 1, \dots, T$ .

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

$$dY_{kt} = Y_{kt} - Y_{k,t-1} \sim \text{Poisson}(\gamma_{kt}W_{kt}), \quad (2)$$

112 where ‘ $\sim$ ’ means ‘distributed as’.

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

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

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

120 Finally, we notice that

$$dW_{kt} = \gamma_{kt}W_{k,t-1}, \quad W_{kt} = W_{k,t-1} + dW_{kt}. \quad (4)$$

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

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

125 Denote  $\gamma_1 = (\gamma_{11}, \dots, \gamma_{K1})'$  and all of the parameters by  $\delta = (\gamma'_1, a', \beta)'$ .  
 126 Taken (1), (2), (3) and (5) together, the loglikelihood function was

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

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

130 *2.2.2. Modeling for Hubei and Wuhan*

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

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

135 The modeling and the likelihood function for Wuhan are similar with  
136 those for 28 provinces except that  $\alpha_k$  is directly estimated by the diagnosed  
137 cases in Wuhan.

138 *2.2.3. The calculation of the time-dependent reproduction number  $R_t$*

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

145 In addition, by Bettencourt and Ribeiro (2008)[16], we also calculate the  
146 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  
147  $\lambda_{k,t} = \gamma_{kt} W_{kt}$ .

148 *2.2.4. Predication of potential turning point in NCPI outbreak*

149 With the estimated parameters by maximizing the loglikelihood (6), we  
150 can estimate and predict the average new cases  $dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1} (\gamma_{kj} + 1) \alpha_k$ ,  
151 then the cumulative cases  $\tilde{W}_{kt} = \prod_{j=0}^t (\gamma_{kj} + 1) \alpha_k$ . Based on the cumulative  
152 cases, we can predict the turning point in the COVID-19 outbreak. In the  
153 paper, we defined the turning point to be the day when the number of the  
154 cumulative cases reached a plateau, which satisfying  $|\partial f(v)/\partial v| \leq c_0$ , where  
155  $f(v) = \frac{\partial \tilde{W}_{kt}/\partial t|_{t=v}}{\partial \tilde{W}_{kt}/\partial t|_{t=v-1}}$  and  $c_0$  is a prespecified small number. We take  $c_0 =$   
156  $2e - 03$  through the analysis.

157 **3. Results**

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



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

### 166 3.1. Results from 28 provinces

167 With the form of  $\eta(t) = a_0 + a_1(t - t_1)_- + a_2(t - t_2)_-$  with  $t_1$  being January  
168 23 and  $t_2$  being January 27 and  $m = 6$ , we estimate  $\delta = (\gamma'_1, a', \beta)'$  based on  
169  $t = 1, \dots, T$  of 28 provinces using the method displayed in Section 2. Table  
170 1 displays the resulting estimators for  $\beta$ . Rail transportation and migration  
171 from Wuhan had significant effects on number of infectious on January, 20 so  
172 the epidemiological scale ( $p$ -value = 0.015 and  $p$ -value = 0.02, respectively),  
173 but not for air transportation ( $p$ -value = 1.0).

174 Figure 2(a) displays the transmission rates  $\gamma_{k1}$  on January 20, 2020 by  
175 province and Figure 2(c) displays the ratio of transmission rate  $\eta_t$  over day  $t$ .  
176 Since  $\gamma_{kt} = \gamma_{k1} \prod_{j=2}^t \eta_j$  and the ratio of transmission rate  $\eta_t$  are the same over  
177 28 provinces, the initial transmission rates  $\gamma_{k1}$  can be used to compare the  
178 strength of the prevention and control for the NCP epidemic in 28 provinces.  
179 The transmission rates  $\gamma_{k1}$  varied greatly by province, from 0.58 to 0.98 for  
180 Beijing and Heilongjiang, respectively.

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

185 Those with the highest imported cases  $\alpha_k$  did not necessarily exhibit the  
186 highest transmission rate  $\gamma_{k1}$ . For instance, Beijing and Sichuan had the  
187 highest underlying cases but low transmission rates (0.58 and 0.65 in Fig-  
188 ure 2(a)); on the contrary, Heilongjiang and Jilin had the lowest underlying  
189 cases but high infection rates (0.83 – 0.98 in Figure 2(a)). Interestingly, this  
190 might be in accordance with the effects implemented in each province; Bei-  
191 jing and Sichuan are known for their substantial and immediate public health  
192 intervention starting on January 20, 2020.

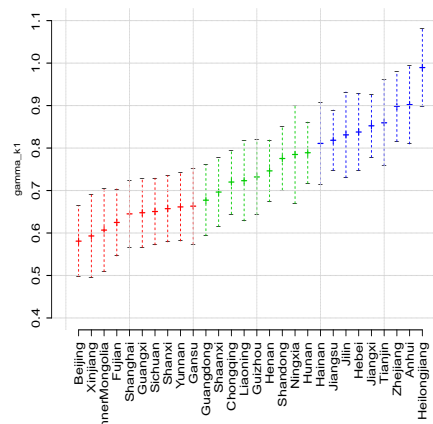
193 Figure 2(d) displays the  $R_t$  and 95% confidence intervals (CI). The re-  
194 production number increased to 2.15 at January 26 and declined to 1 on  
195 February 1, then gradually decreased to 0.26 at February 13. We also cal-

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

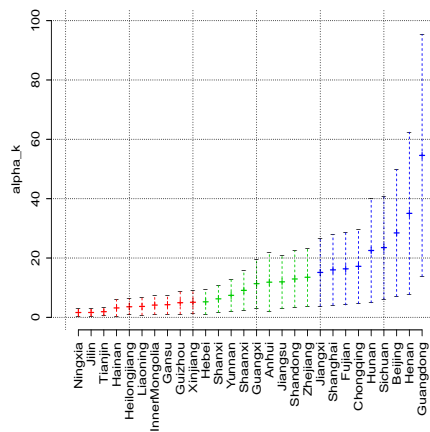
	28 provinces			16 cities in Hubei		
	train	airplane	migration	train	highway	migration
Est.	42.628	1e-05	28.740	1e-05	7.781	20.908
SD	17.498	0.853	12.321	0.106	3.719	5.967
<i>p</i> -value	0.015	1.000	0.0197	0.999	0.036	4e-04

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

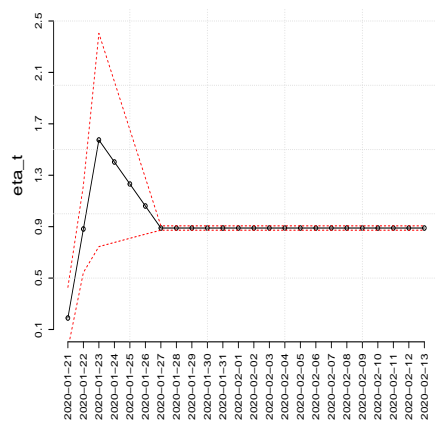
198 We further predicted the number of accumulated cases until March for  
 199 each province using the optimal model we chose above (Figure 3 and Supple-  
 200 mentary Figures 2 and 3). For almost all provinces, observed values perfectly  
 201 fall within 95% CI of the prediction band, suggesting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



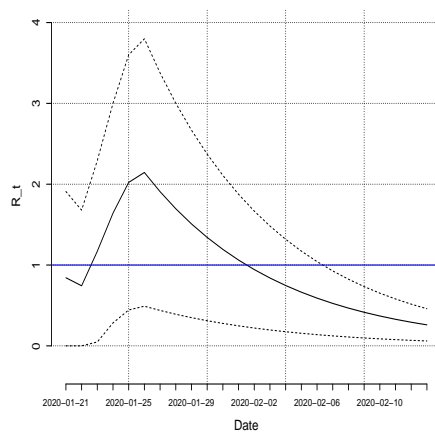
(a)



(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.

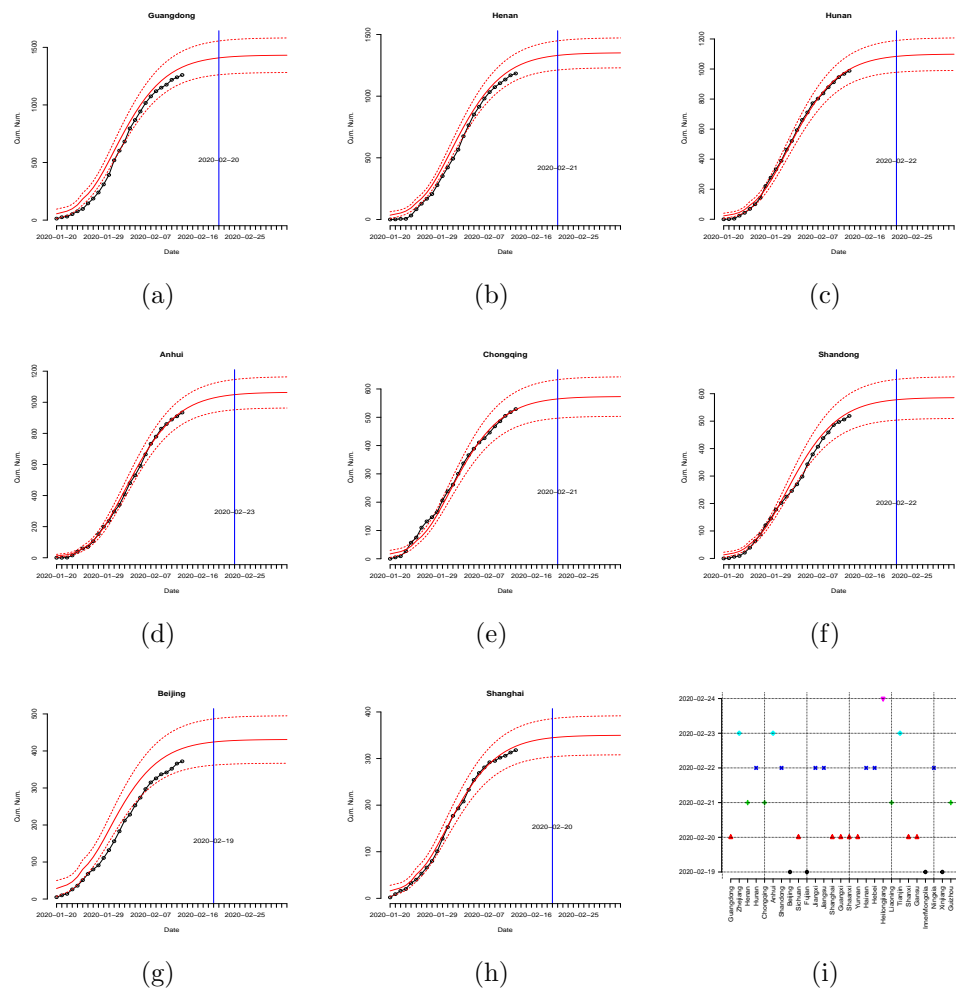


Figure 3: (a)-(h): The estimated (red-solid) and observed (black-dotted) cumulative number of infectious over day  $t$ , as well as 95% CI (red-dashed) for typical provinces. (i): The day reaching plateau by provinces.

208 *3.2. Results from the cities in Hubei province*

209 With the form of  $\eta(t) = a_0 + a_1(t - t_1)_- + a_2(t - t_2)_-$  with  $t_1$  being  
 210 January 23 and  $t_2$  being January 28 and  $m = 9$ , we estimate  $\delta = (\gamma'_1, a', \beta')$   
 211 based on  $t = 1, \dots, T$  of 16 cities in Hubei using the method displayed in  
 212 Section 2. The duration of the infectious period  $m$  was 9 days, 3 days longer  
 213 than that estimated in other provinces; this could be due to the delay in the  
 214 diagnosis/hospital admission of infected cases in cities in Hubei.

215 Table 1 displays the resulting estimators for  $\beta$ . Highway transportation  
216 and migration from Wuhan have significant effects on number of infectious  
217 on January, 20 so the epidemiological scale ( $p$ -value = 0.036 and  $p$ -value  
218 =  $4e - 04$ , respectively), but not for rail transportation ( $p$ -value = 1.0). This  
219 is different from the conclusion for 28 provinces where rail transportation is  
220 significant. This may be attributed to that the cities in Hubei is close to  
221 Wuhan and the popular transportation leaving Wuhan is by cars.

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

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

231 Again, those with higher imported cases  $\alpha_k$  did not necessarily exhibit  
232 higher transmission rate  $\gamma_{k1}$ . For instance, although Suizhou, Huangshi and  
233 Ezhou had lower underlying cases (6.06 – 7.50) but the highest transmission  
234 rate (0.74 – 0.80); on the contrary, Enshi had higher underlying cases (9.12)  
235 but lower transmission rates (0.48 in Figure 4(a)).

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

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

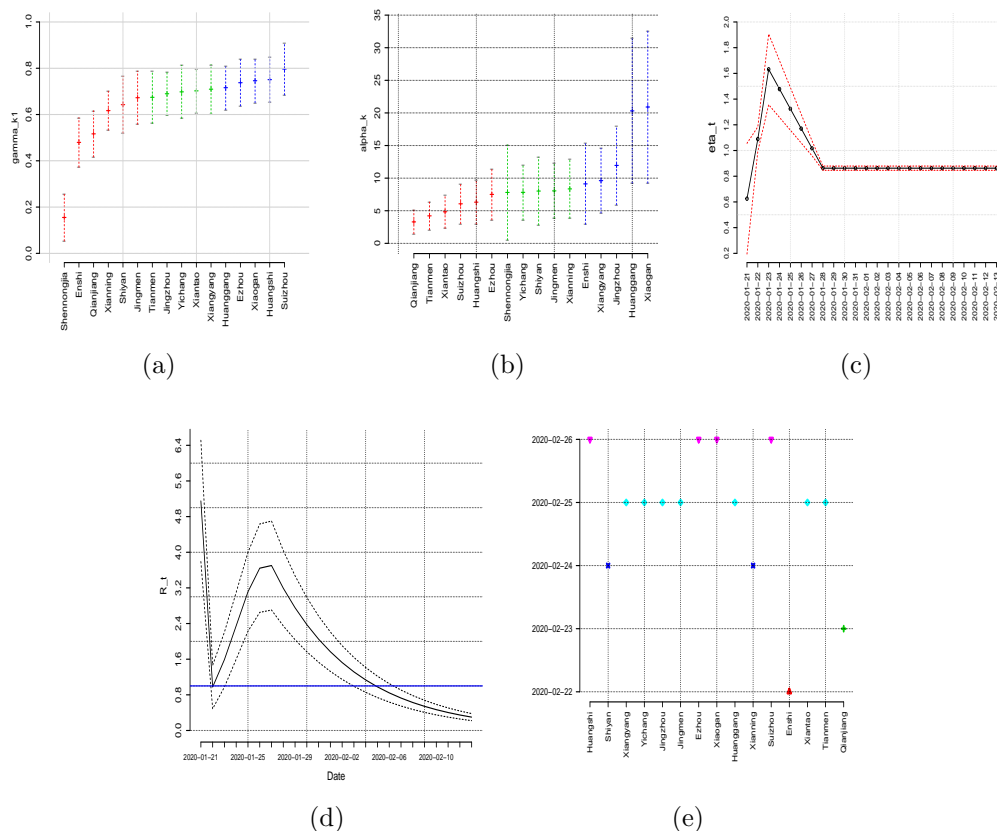


Figure 4: The estimation and 95% CI. (a) the estimated transmission rate on January 20 by cities in Hubei; (b) the estimated underlying infected cases on January 20 by cities in Hubei; (c) the estimated ratio of transmission rate over time in Hubei; (d) Sequential estimation of reproduction number from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

### 247 3.3. Results from WuHan city

248 With the form of  $\eta(t) = a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$  with  $t_1$  being January  
 249 27 and  $t_2$  being January 30 and  $m = 6$ , we estimated the parameters based  
 250 on  $t = 1, \dots, T$  in Wuhan using the method displayed in Section 2.

251 Figures 5(a) and 5(b) displays the estimators of  $\eta_t$  and  $R_t$ , as well as their  
 252 95% CI, respectively, for Wuhan. The reproduction number  $R_t$  decreased  
 253 greatly from 4.9 to 2 in 3 days, remained stable till January 28, and then  
 254 continuously decreased to 1 at February 2. In our prediction model, the  
 255 observed cases were within the prediction band before February 13. Due to

256 the changes of diagnosis criteria, the predicted cases were not comparable  
 257 after February 13. However, as mentioned above, the estimated trend was  
 258 still informative to predict the turning point. Thus we estimate that the date  
 259 of containment would be around February 29, 2020 in Wuhan (Figure 5(c)).

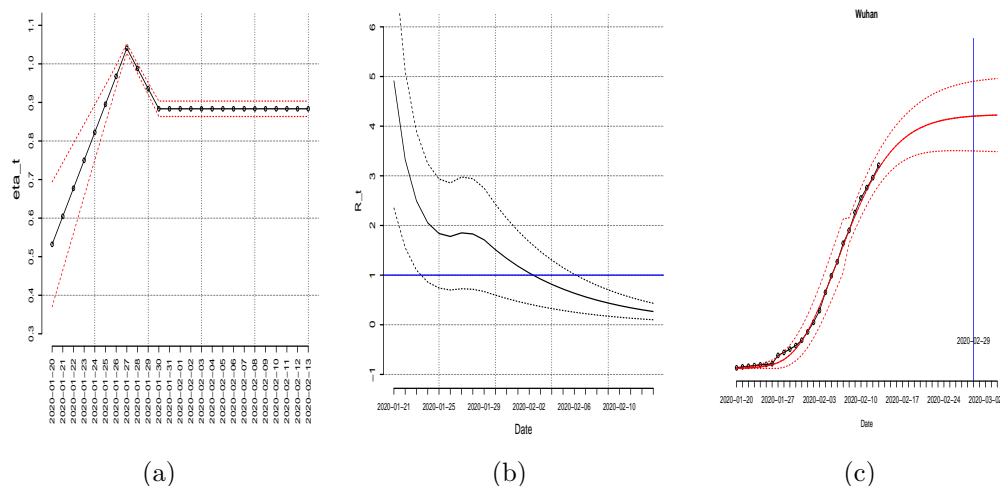


Figure 5: The estimation and 95% CI. (a) the estimated ratio of transmission rate over time in Wuhan; (b) Sequential estimation of reproduction number from the daily COVID-19 cases in Wuhan; (c) The estimated (red-solid) and observed (black-dotted) number of infectious over day  $t$ , as well as 95%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

#### 260 4. Discussions

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

271 The temporal distributions for  $R_t$  give a basis for assessing the evolution  
 272 of transmissibility over time[15, 16, 22]. For Hubei and other provinces, there

273 was an initial growth from January 21 to 27, 2020 with  $R_t$  up to 4 followed  
274 by a decline in a short period. The increase in  $R_t$  was in accordance with  
275 the increase in transmission, but at a slower rate. There was a dramatically  
276 declining trend in the ratio for the transmission rate from January 23 to 28.  
277 Since January 23, 2020, the Chinese government banned travels into and from  
278 Wuhan by air, rail and road access to halt the spread of cases from Wuhan  
279 to other cities. Also, almost all provinces initiated the highest level of public  
280 health emergency responses during January 23 and 25 including tracking and  
281 isolating close contacts with COVID-19 patients, self-quarantine, etc. We  
282 speculate that, based on the coincidence of a series of intervention effects,  
283 the decline in  $R_t$  could be due to a series of intervention measures to reduce  
284 the transmission rate implemented since January 23, 2020.

285 For Wuhan, the reproduction number  $R_t$  showed a trend of decline in  
286 general from 4.9 to 1 on February 2, 2020, although a slight shift upward  
287 on January 27 and 28. Different from other places, the outbreak in Wuhan  
288 could be in earlier January with sparse data. In our study, the initial  $R_t$   
289 was estimated at 4.9 on January 21 and even higher at 8.13 before this date  
290 (data not shown), larger than the estimation in other places as expected. We  
291 also observed that the ratio for transmission rates decreased from January  
292 28 in Wuhan, suggesting the intervention starting on January 23 might have  
293 lag effects on the transmission rates. Several studies have estimated the  
294 basic reproduction number  $R_0$  in a range of 2.2 to 4.0 in Wuhan[2, 3, 11, 12],  
295 highly contagious compared to SARS[13, 23, 24, 25, 26, 27, 28]. One study[13]  
296 estimated  $R_t$  in 830 cases in Wuhan from December 24 and January 23 and  
297 found that  $R_t$  decreased from 8 to 2.52. The trend is similar to what we  
298 observed; however, with the interaction involved, we estimated  $R_t$  declined  
299 and approached the critical threshold at 1 on February 2, 2020.

300 The difference in transmission rates among different provinces highlights  
301 the importance of its contact patterns and control measures, given no re-  
302 quired immunity obtained in the early outbreak phase[4, 18]. As of January  
303 20, underlying cases imported from Wuhan were the highest in Guangdong,  
304 Henan, Beijing and Sichuan. Beijing and Sichuan had the lowest transmis-  
305 sion rates (ranging from 0.58 to 0.65), and Heilongjiang, Jilin and Tianjin  
306 had the lowest underlying cases but high transmission rates (0.83 to 0.98).  
307 This could be related to the extent of efforts in each province to reduce the  
308 number of contacts in the transmission process, particularly challenging for  
309 those with the highest number of cases. It was reported a high sensitivity of  
310 the timing of implementation of control for SARS, a 1-week delay in imple-



311 mentation of control measures results in a 2.6-fold increase in mean epidemic  
312 size[16]. Thus the immediate stringent controls measures have critical impact  
313 to prevent the spread of diseases.

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

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

339 Our study is not without limitations. First, our results are based on the  
340 diagnosis cases reported by the CDC in China. Underreporting is likely to  
341 occur which was not accounted for in our analysis. Among all cases reported  
342 until February 11, only 13.8% cases were reported prior January 20, 2020  
343 and among them, 77.6% were among the Hubei province[35]. Since we are  
344 most interested in the temporal trend after January 20, the underreporting  
345 might not vary greatly with time since then. Second, as there is a series  
346 of interventions implemented during subsequent periods, we were not able  
347 to distinguish which intervention could be the main driver based on current  
348 data. Third, due to the changes in diagnosis, the diagnosed cases increased

349 dramatically up to 14840 in Wuhan and Hubei from February 12, which  
350 makes our prediction challenging. Thus we did not provide the predicted  
351 cases for Wuhan and Hubei province after February 13, 2020.

352 In summary, we provide transmission dynamic patterns of the COVID-  
353 19 outbreak using time-serial data from January 20 to February 13, 2020 in  
354 China. The declining trend of  $R_t$ , as well as ratio for transmission rates,  
355 indicate the effects of public health intervention implemented by the gov-  
356 ernment beginning on January 23, 2020. Model prediction results suggested  
357 that COVID-19 would be contained in provinces around 19 to 24 February,  
358 2020, and the date of containment would be one-week later in Wuhan.

### 359 Declaration of interests

360 We declare no competing interests.

### 361 Acknowledgments

362 The research were partially supported by National Natural Science Foun-  
363 dation of China (Nos. 11931014 and 11829101) and Fundamental Research  
364 Funds for the Central Universities (No. JBK1806002) of China.

### 365 References

- 366 [1] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang,  
367 Z. Cheng, Y. Xiong, et al., Clinical characteristics of 138 hospitalized pa-  
368 tients with 2019 novel coronavirus–infected pneumonia in wuhan, china,  
369 Jama.
- 370 [2] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S.  
371 Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in  
372 wuhan, china, of novel coronavirus–infected pneumonia, New England  
373 Journal of Medicine.
- 374 [3] J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the poten-  
375 tial domestic and international spread of the 2019-ncov outbreak origi-  
376 nating in wuhan, china: a modelling study, The Lancet.
- 377 [4] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song,  
378 B. Huang, N. Zhu, et al., Genomic characterisation and epidemiology

- 379 of 2019 novel coronavirus: implications for virus origins and receptor  
380 binding, *The Lancet*.
- 381 [5] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R.  
382 Si, Y. Zhu, B. Li, C.-L. Huang, et al., A pneumonia outbreak associated  
383 with a new coronavirus of probable bat origin, *Nature* (2020) 1–4.
- 384 [6] D. K. Chu, Y. Pan, S. Cheng, K. P. Hui, P. Krishnan, Y. Liu, D. Y.  
385 Ng, C. K. Wan, P. Yang, Q. Wang, et al., Molecular diagnosis of a  
386 novel coronavirus (2019-ncov) causing an outbreak of pneumonia, *Clinical Chemistry*.  
387
- 388 [7] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang,  
389 Y. Liu, Y. Wei, et al., Epidemiological and clinical characteristics of 99  
390 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descrip-  
391 tive study, *The Lancet*.
- 392 [8] J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang,  
393 F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, et al., A familial cluster of  
394 pneumonia associated with the 2019 novel coronavirus indicating person-  
395 to-person transmission: a study of a family cluster, *The Lancet*.
- 396 [9] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-  
397 W. Tao, J.-H. Tian, Y.-Y. Pei, et al., A new coronavirus associated with  
398 human respiratory disease in china, *Nature* (2020) 1–8.
- 399 [10] D. S. Hui, E. I Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ip-  
400 polito, T. D. Mchugh, Z. A. Memish, C. Drosten, et al., The continuing  
401 2019-ncov epidemic threat of novel coronaviruses to global healththe  
402 latest 2019 novel coronavirus outbreak in wuhan, china, *International*  
403 *Journal of Infectious Diseases* 91 (2020) 264–266.
- 404 [11] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao,  
405 L. Yang, D. He, et al., Preliminary estimation of the basic reproduction  
406 number of novel coronavirus (2019-ncov) in china, from 2019 to 2020:  
407 A data-driven analysis in the early phase of the outbreak, *International*  
408 *Journal of Infectious Diseases*.
- 409 [12] Z. Zhu, J. Li, D. Gong, D. Wan, S. Chen, L. Guo, Y. Li, L. Sun,  
410 W. Liang, T. Song, et al., Time-varying transmission dynamics of novel  
411 coronavirus pneumonia in china.

- 412 [13] T. Liu, J. Hu, J. Xiao, G. He, M. Kang, Z. Rong, L. Lin, H. Zhong,  
413 Q. Huang, A. Deng, W. Zeng, X. Tan, S. Zeng, Z. Zhu, J. Li,  
414 D. Gong, D. Wan, S. Chen, L. Guo, Y. Li, L. Sun, W. Liang, T. Song,  
415 J. He, W. Ma, Time-varying transmission dynamics of novel coron-  
416 avirus pneumonia in china, bioRxiv: [https://www.biorxiv.org/  
417 content/early/2020/02/13/2020.01.25.919787.full.pdf](https://www.biorxiv.org/content/early/2020/02/13/2020.01.25.919787.full.pdf), doi:10.  
418 1101/2020.01.25.919787.  
419 URL [https://www.biorxiv.org/content/early/2020/02/13/2020.  
420 01.25.919787](https://www.biorxiv.org/content/early/2020/02/13/2020.01.25.919787)
- 421 [14] J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the poten-  
422 tial domestic and international spread of the 2019-ncov outbreak origi-  
423 nating in wuhan, china: a modelling study, *The Lancet*.
- 424 [15] Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive  
425 number of covid-19 is higher compared to sars coronavirus, *Journal of  
426 Travel Medicine*.
- 427 [16] L. M. Bettencourt, R. M. Ribeiro, Real time bayesian estimation of the  
428 epidemic potential of emerging infectious diseases, *PLoS One* 3 (5).
- 429 [17] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respi-  
430 ratory syndrome reveal similar impacts of control measures, *American  
431 Journal of epidemiology* 160 (6) (2004) 509–516.
- 432 [18] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, A. Vespignani,  
433 Measurability of the epidemic reproduction number in data-driven con-  
434 tact networks, *Proceedings of the National Academy of Sciences* 115 (50)  
435 (2018) 12680–12685.
- 436 [19] X.-S. Zhang, R. Pebody, A. Charlett, D. de Angelis, P. Birrell, H. Kang,  
437 M. Baguelin, Y. H. Choi, Estimating and modelling the transmissibility  
438 of middle east respiratory syndrome coronavirus during the 2015 out-  
439 break in the republic of korea, *Influenza and other respiratory viruses*  
440 11 (5) (2017) 434–444.
- 441 [20] B. Efron, Bootstrap confidence intervals for a class of parametric prob-  
442 lems, *Biometrika* 72 (1) (1985) 45–58.
- 443 [21] B. Efron, et al., Second thoughts on the bootstrap, *Statistical Science*  
444 18 (2) (2003) 135–140.

- 445 [22] L. Forsberg White, M. Pagano, A likelihood-based method for real-time  
446 estimation of the serial interval and reproductive number of an epidemic,  
447 *Statistics in medicine* 27 (16) (2008) 2999–3016.
- 448 [23] A. N. Desai, M. U. Kraemer, S. Bhatia, A. Cori, P. Nouvellet, M. Her-  
449 ringer, E. L. Cohn, M. Carrion, J. S. Brownstein, L. C. Madoff, et al.,  
450 Real-time epidemic forecasting: Challenges and opportunities, *Health*  
451 *security* 17 (4) (2019) 268–275.
- 452 [24] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad,  
453 A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, et al.,  
454 Transmission dynamics of the etiological agent of sars in hong kong:  
455 impact of public health interventions, *Science* 300 (5627) (2003) 1961–  
456 1966.
- 457 [25] C. A. Donnelly, M. C. Fisher, C. Fraser, A. C. Ghani, S. Riley, N. M.  
458 Ferguson, R. M. Anderson, Epidemiological and genetic analysis of se-  
459 vere acute respiratory syndrome, *The Lancet infectious diseases* 4 (11)  
460 (2004) 672–683.
- 461 [26] C. Dye, N. Gay, Modeling the sars epidemic, *Science* 300 (5627) (2003)  
462 1884–1885.
- 463 [27] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser,  
464 S. Riley, L. J. Abu-Raddad, L.-M. Ho, T.-Q. Thach, P. Chau, et al.,  
465 Epidemiological determinants of spread of causal agent of severe acute  
466 respiratory syndrome in hong kong, *The Lancet* 361 (9371) (2003) 1761–  
467 1766.
- 468 [28] J. S. M. Peiris, C.-M. Chu, V. C.-C. Cheng, K. Chan, I. Hung, L. L.  
469 Poon, K.-I. Law, B. Tang, T. Hon, C. Chan, et al., Clinical progression  
470 and viral load in a community outbreak of coronavirus-associated sars  
471 pneumonia: a prospective study, *The Lancet* 361 (9371) (2003) 1767–  
472 1772.
- 473 [29] S. V. Scarpino, G. Petri, On the predictability of infectious disease out-  
474 breaks, *Nature communications* 10 (1) (2019) 1–8.
- 475 [30] M. Y. Li, J. S. Muldowney, Global stability for the seir model in epi-  
476 demiology, *Mathematical biosciences* 125 (2) (1995) 155–164.

- 477 [31] M. Y. Li, J. R. Graef, L. Wang, J. Karsai, Global dynamics of a  
478 seir model with varying total population size, *Mathematical biosciences*  
479 160 (2) (1999) 191–213.
- 480 [32] M. Y. Li, H. L. Smith, L. Wang, Global dynamics of an seir epidemic  
481 model with vertical transmission, *SIAM Journal on Applied Mathematics*  
482 62 (1) (2001) 58–69.
- 483 [33] P. E. Lekone, B. F. Finkenstädt, Statistical inference in a stochastic  
484 epidemic seir model with control intervention: Ebola as a case study,  
485 *Biometrics* 62 (4) (2006) 1170–1177.
- 486 [34] S. Cauchemez, P.-Y. Boëlle, C. A. Donnelly, N. M. Ferguson, G. Thomas,  
487 G. M. Leung, A. J. Hedley, R. M. Anderson, A.-J. Valleron, Real-time  
488 estimates in early detection of sars, *Emerging infectious diseases* 12 (1)  
489 (2006) 110.
- 490 [35] T. N. C. P. E. R. E. Team, The epidemiological characteristics of an  
491 outbreak of 2019 novel coronavirus diseases (covid-19) in china, *Chinese*  
492 *Journal of Epidemiology* 41 (2) (2020) 145–151.