Open access • Posted Content • DOI:10.1101/2020.02.21.20026468

Comparative Analysis of Early Dynamic Trends in Novel Coronavirus Outbreak: A Modeling Framework — Source link

Lin, Huazhen, Liu, Wei, Gao, Hong, Nie, Jinyu, Fan, Qiao

Institutions: Southwestern University of Finance and Economics

Published on: 25 Feb 2020 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Coronavirus and Outbreak

Comparative Analysis of Early Dynamic Trends in Novel Coronavirus Outbreak: A Modeling Framework

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 coronavirus disease (COVID-19) represents a significant public health threat globally. Here we describe efforts to compare epidemic growth, size and peaking time for countries in Asia, Europe, North America, South America and Australia in the early epidemic phase.

Methods Using the time series of cases reported from January 20, 2020 to February 13, 2020 and transportation data from December 1, 2019 to January 23, 2020 we have built a novel time-varying growth model to predict the epidemic trend in China. We extended our method, using cases reported from January 26, 2020 - or the date of the earliest case reported, to April 9, 2020 to predict future epidemic trend and size in 41 countries. We estimated the impact of control measures on the epidemic trend.

Results Our time-varying growth model yielded high concordance in the predicted epidemic size and trend with the observed figures in China. Among the other 41 countries, the peak time has been observed in 28 countries before or around April 9, 2020; the peak date and epidemic size were highly consistent with our estimates. We predicted the remaining countries would peak in April or May 2020, except India in July and Pakistan in August. The epidemic trajectory would reach the plateau in May or June for the majority of countries in the current wave. Countries that could emerge to be new epidemic centers are India, Pakistan, Brazil, Mexico, and Russia with a prediction of 10^5 cases for these countries. The effective reproduction

Email addresses: linhz@swufe.edu.cn (Huazhen Lin), qiao.fan@duke-nus.edu.sg (Qiao Fan)

number R_t displayed a downward trend with time across countries, revealing the impact of the intervention remeasures i.e. social distancing. R_t remained the highest in the UK (median 2.62) and the US (median 2.19) in the fourth week after the epidemic onset.

Conclusions New epidemic centers are expected to continue to emerge across the whole world. Greater challenges such as those in the healthcare system would be faced by developing countries in hotspots. A domestic approach to curb the pandemic must align with joint international efforts to effectively control the spread of COVID-19. Our model promotes a reliable transmissibility characterization and epidemic forecasting using the incidence of cases in the early epidemic phase.

Keywords: Coronavirus, COVID-19, SARS-CoV-2, transmissibility, outbreaks, reproduction number R, generalized growth modeling

---

1. Introduction

In early December 2019, a novel severe acute respiratory syndrome coronavirus, SARS-COV-2, emerged into the human population in Wuhan, China[1–7]. The first coronavirus disease 2019 (COVID-19) case outside China was reported on January 13, 2020 in Thailand. In just several weeks, the local transmission started rapidly in a broad array of countries in Asia, Europe, North America, South America and Australia, with the emergence of new epicenters of spread, such as the US, Italy, Spain, and France (https://www.who.int). The rapid spreading of SARS-COV-2 has led to a major global public health threat. The World Health Organization (WHO) declared the spread of COVID-19 a pandemic on March 11, 2020.

Early epidemic forecasts consisting of the likely trajectory of an unfolding outbreak can help guide the type and intensity of interventions[8–10]. The vast majority of these approaches considered early exponential growth dynamics, an assumption that could lead to substantial overestimation of epidemic size and peak timing. To enhance the ability to forecast epidemics, it is crucial to characterize the shape of epidemic growth and accurately assess early trends of sub-exponential growth phenomenon[11–13]. Currently, the global case count continues to rise, but there is a limited understanding of the extent of the outbreak and epidemic growth profile, particularly for the new emerging epicenters.

In this study, we attempt to assess and compare the extent of the outbreak across countries, draw preliminary conclusions about the impact of control measures, and characterize real-time effective reproduction number R_t. Our model, without making explicit assumptions about the epidemic growth profile, is a generalizable framework to estimate the early dynamic trends of COVID-19 from the incidences of cases. We estimated the epidemic trend and size in China using the cases reported from January 20, 2020 to February 13, 2020, and in other 41 countries using data from January 26, 2020, - or the date of the earliest case reported, to April 9, 2020.

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 29 provinces (25 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. All cases 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 both the percentages and volumes of migration from 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 Supplementary 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 Supplementary Figure 1, we observed that Guangdong 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 Huanggang 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 Guangdong, Henan, Huanggang and Xiaogan have the largest number of estimated primary infected cases imported from Wuhan on January 20, 2020.

We obtained the number of COVID-19 confirmed cases in 41 countries using data from January 23, 2020, or the date of the earliest case reported, to April 9, 2020 from the offical websites of WHO at https://www.who.int. We included 41 countries with at least confirmed cases on April 9 2020 in five continents: Asia, Europe, North America, South America, and Australia.

2.2. Modelling the transmissibility of COVID-19

We introduce the main notation here. All times are calendar times, measured in days since the start of 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,
W_{kt}	underline number of infected individuals who are infectious,
m	duration of infectious period (day),

where the subscript k represents country, 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. It is obvious that the average cases in province k diagnosed at day t is proportional to the scale of infectious cases on day $t-1$, $W_{k,t-1}$. We then assume a Poisson distribution for the new cases diagnosed in province k at day t with mean $\gamma_{kt} W_{k,t-1}$, that is

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

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 $\eta_t = \eta_t(a)$.

The new cases diagnosed at day t may be not reported fully. That is,

$E(dY_{kt}) = pdW_{kt}$ and $p < 1$. The estimation for p need more information except that we have. Fortunately, simple mathematical derivation shows that the value of p may influence the prediction of the absolute epidemic size but will not affect the trend of the epdimic, for example, the reproduction number, the duration and the peak time of the epidemic and relative epidemic size, in which we are interested. Hence, we suppose $p = 1$ in the paper. Since

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

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[14, 15].

2.2.2. Modeling for Hubei, Wuhan and the other countries

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,

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

The modeling and the likelihood function for Wuhan and countries, including Singapore, South Korea, Japan, Italy, German, Span, France and Iran, are similar with those for 28 provinces except that α_k is directly estimated by the diagnosed cases in Wuhan and other 41 countries respectively.

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

The equation $dW_{kt} = \gamma_{kt} W_{k,t-1}$ implies that, 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[16, 17].

2.2.4. Predication of potential peak time and turning point in COVID-19 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 new cases and cumulative cases, we can predict the peak time and the turning point in the COVID-19 outbreak. In the paper, we defined the peak time to be the day at which the incidence cases began to decline, and 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

3.1. Fitting a generalized growth model using a time-series data

We fit the time-varying generalized growth model using the early stage outbreak series data for Wuhan, Hubei province, China, and other 41 countries with at least 2,000 confirmed cases on the date of April 9 2020. The optimal fitted model for each country or region was chosen based on the prediction error of the lowest values (Supplementary Figures 2 to 4). The corresponding empirical distributions of the parameters are shown in Table 1. The parameter m, the estimated mean infectious duration, ranged from 4 to 13 days, with the highest value of 13 for the epidemic in Italy. The short duration of 4 days was estimated for South Korea. The value of m reflects the duration of infectious period, which in practice could be intervened by control measures such as early diagnosis and isolation. Another parameter $\eta(t)$ quantifies how rapidly the growth rate $\gamma(t)$ at time t changed; the functional form of $\eta(t)$ was estimated by pointwise estimation using data till t for each model respectively. The growth rates declined most rapidly in China and Thailand (median η: 0.86 to 0.89), and most slowly in India, Mexico, Singapore and Sweden (median η: 0.98 to 0.99).

3.2. Predictive performance of model fitting

Using the calibrated model based on m and the form of $\eta(t)$, we estimated the real-time growth rates based on the number of confirmed cases at $t = 1, \ldots, T$ by maximizing the likelihood function displayed in the Method section. The underlying growth process was lower than the constant exponential growth rate as medium $\eta(t)$ was consistently estimated at less than 1(Supplementary figure 3), similar to findings of sub-exponential growth dynamics for epidemics of influenza, Ebola etc with the deceleration growth factor below 1 [18, 19].

For China, we used case incidence data from January 20 to February 13 as the epidemic reporting period to fit the model for trajectory predictions. The observed values fall within 95% CI of the prediction band in general, suggesting a good model fitting (Figure 1 and Supplementary Figures 8-9). We predicted that the epidemics would fade out around February 19 to 24 across the 28 provinces, about 4 weeks after the intervention starting from January 20; this is in concordance with the actual timing in which no more new cases were observed at the end of February (http://www.nhc.gov.cn). For 28 provinces in China, the estimated size was at 8553 to 9460 and 11,000 to 12,600 for Hubei province; both 95% CI estimates cover the observed figures (Table 2). For Wuhan, due to the changes in the diagnosis criteria, the predicted number of cases was not comparable after February 13. We estimated the epidemic would fade out around the end of February, which is close to the final date in which no more new cases were observed in Wuhan.

Within each province, the observed final epidemic size was within the 95% CIs of prediction except 5 provinces, yet the prediction was still within 10% +/- flanking the upper or lower bound (Supplementary Table 1). Given the intervention measures are similar across provinces in China besides Hubei, it is not surprising that we observed the trend that the provinces with a higher estimated risk of imported cases alpha from Wuhan had an increased epidemic size. However, there were some exceptions. For example, Beijing ($\alpha = 28; 95\% CI : 7 - 49$) and Shanghai ($\alpha = 16; 4 - 28$) were in the high-risk group but the final infection numbers were 286 and 261. On the contrary, Heilongjiang was in the low-risk group ($\alpha = 4 : 1 - 6$) whereas the final infections were at 252. This implies the stronger intervention locally or compliance in densely populated municipalities (Beijing or Shanghai), than that in Heilongjiang, the north-east region far away from Wuhan. Overall our time-varying growth model provides a good fitting using time-series confirmed cases for the epidemic of COVID-19 in China.

3.3. Projection of the epidemic and final size in 41 countries

We projected the future growth trajectory underlying the outbreak using the daily confirmed case data from January 23, 2020 - or the date of the earliest case reported - to April 9, 2020 in 42 countries across Asia, Europe, North America, South America and Australia. Our prediction is based on the estimated parameters assuming the sustainability of intervention measures. The time-varying growth model fitting to the time-series data performed well as shown by the observed cases generally falling in the 95% CI of the prediction bound in each plot (Figure 1 and Supplementary Figures 8-12), except for South Korea in the very early phase. The predicted cases were higher than the observed number of infections from February 8 to 25, likely due to a proportion of infections being greatly understated - the hiding cases from the religious group.

We predicted the epidemic size and duration of the outbreak with the assumption of current model fitting parameters (Table 2). In Asia, the epidemic will continue until May in Philippine, Malaysia and Iran, and July in Indonesia. Strikingly, the epidemic would have the longest outbreak spread until February or March 2021 in Indian and Pakistan, with the final infections at more than a million. For Japan and Singapore, we were not able to do predictions on the epidemic sizes and duration because the estimated time-varying growth rates kept on increasing at the end of the study period – which is concordant to the big daily jump in cases observed in these two countries recently; thus incidence case data in a longer period are required for valid predictions using our model. In Europe, the epidemic will not fade out until May or June for the majority of countries. However the epidemic may last until August (Russia and Sweden). In North America, the epidemic will continue until May (Canada, Dominican Republic), June (US) and October (Mexico), and the infection are estimated to be around 1 million in the US, and 10^5 in Mexico. In South America, the epidemic will continue until May (Chile) and June (Brazil, Colombia and Panama). Among them, Brazil has the largest estimated infections with more than 10^5.

3.4. Estimation of peak time

We estimated the peak time, the day at which the incidence cases began to decline based on the estimated daily cases $\mathbf{d}W_{kt}$ (see Methods). Among 41 countries, the peak time has been observed in 28 countries before or around April 9, 2020 (Table 3 & Supplementary Figures 13-15). The estimations in the above countries are concordant to the observed period during which incidence cases declined. For other countries, under the current parameters with continuity of intervention measures, we estimated the epidemic would peak in April or May 2020 for the majority of countries, except for India (July) and Pakistan (August). The predicted numbers of accumulated cases at the peak point were also consistent with the observed figures. Based on the predicted number of infections, we estimated the maximum number of ventilators for the peak requirement. For example, at least 500,000 ventilators need to be prepared for the epidemic peak in India, $\sim 100,000$ for Pakistan. Considering the death or recovery rate, the number is regarded as the upper bound in practice.

3.5. Impact of the reduction of the infectious period

Shortening the time from the infection onset (symptomatic or asymptomatic) to isolation, termed as 'infectious period' in this paper, is vital as it will reduce transmission. The control strategies could rely on social distancing, earlier isolation of cases and population-based testing to identify the presymptomatic or asymptomatic cases. Here we evaluated the impact of shortening the infectious period on the epidemics, showings the epidemic curve for each country, assumed to be 1 or 2 days shorter in the duration of infectivity infectious period (Supplementary Figures 13-15). A reduction of the infectious period by 1 or 2 days would have a negligible effect on the peak time, as the estimation was essentially the same. The reduction of the infectious period by 1 or 2 days would lead to a significantly flatted epidemic curve across all the countries. For example, the 2-days decrease in the infectious period would result in a significant reduction of the final epidemic size in India (31.1%), Pakistan (39.3%), Russia (22.6%), Brazil (26.8%), and Mexico(18.0%).

3.6. Estimation of real-time effective reproduction number R_t

We quantified the real-time reproduction number R_t (see Method section)[17, 20, 21], based on the estimation parameters of growth rates γ and the infectious period m. The R_t exhibited a declining trend with time and variability in the estimates across countries (Table 4). The uncertainty of Rt was largest in the first week and gradually became smaller with time (Table 4). In the first week, 28 countries displayed R_t below 5 while 14 exhibited large R_t above 5. Italy and Spain displayed the highest R_t at 12.36, and 9.26 respectively. R_t estimations become closer ranging from 0.95 (Singapore) to 5.54 (Spain) in the second week. During the first month of the epidemic period, all countries displayed a declined trend towards 1 in the epidemic period with varying deceleration rates, revealing the impacts of the intervention strategies. In the fourth week, 14 countries exhibited R_t below 1, the highest in UK (2.62) and the US (2.19). For Singapore and Japan, although both displayed the declining trend in the first month, there is an inclining trend starting from April, implying the intervention could not be effective in the late stage (Supplementary Figure 16). R_t displayed the fastest deceleration in China with the most pronounced changes during the third to fourth weeks; this may reveal the significant impact of intervention measures implemented since January 20, 2020 in the first week.

Table 1: Summary of fitting parameters for the selected model in 10 countries in 2020.

	Epidemic study period	m	Medium η_t (95%)	form of η_t
Asia				
China	01/20– 02/13	6	0.89(0.87, 0.91)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Hubei Province, China	01/20– 02/13	9	0.86(0.84, 0.88)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Wuhan, China	01/20– 02/13	6	0.88(0.86, 0.90)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
India	03/04–04/09	9	0.99(0.98, 1.00)	$a_0 + a_1(t-t_1)_-$
Indonesia	03/08–04/09	10	0.97(0.96, 0.98)	$a_0 + a_1(t-t_1)_-$
Iran	03/19–04/09	8	0.96(0.96, 0.98)	$a_0 + a_1(t-t_1)_-$
Japan	02/14–04/09	8	0.97(0.96, 0.98)	$a_0 + a_1(t-t_1)_-$
South Korea	01/24– 04/09	4	0.93(0.92, 0.94)	$a_0 + a_1(t-t_1)I(t_1 \leqslant t \leqslant t_2)$ $+a_1(t_2-t_1)I(t > t_2)$
Malaysia	02/27-04/09	11	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Pakistan	03/09-04/09	6	0.92(0.89, 0.95)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Singapore	01/23-04/09	5	0.99(0.98, 1.00)	$a_0 + a_1(t-t_1)_-$
Thailand	03/09-04/09	10	0.89(0.88, 0.90)	a_0
Israel	03/03–04/09	7	0.92(0.90, 0.94)	a_0
Saudi Arabia	03/08-04/09	10	0.96(0.94, 0.98)	a_0
Turkey	03/15-04/09	6	0.90(0.89, 0.91)	$a_0 + a_1(t-t_1)_-$
North America				
US	02/24-04/09	6	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_-$
Canada	02/27–04/09	8	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Dominican Republic	03/19-04/09	11	0.92(0.88, 0.96)	$a_0 + a_1(t-t_1)_-$
Mexico	03/13-04/09	11	0.98(0.96, 1.00)	a_0
South America				
Brazil	03/09–04/09	8	0.97(0.95, 0.99)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Chile	03/10–04/09	7	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Colombia	03/10–04/09	8	0.96(0.94, 0.98)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Panama	03/10-04/09	10	0.95(0.93, 0.97)	$a_0 + a_1(t-t_1)_-$
Europe				
Austria	02/29–04/09	10	0.90(0.88, 0.92)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Belgium	03/01–04/09	10	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Czechia	03/03–04/09	7	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_-$
Finland	03/05–04/09	8	0.97(0.96, 0.98)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
France	02/27–04/09	6	0.96(0.94, 0.98)	$a_0 + a_1(t-t_1)_-$
Germany	02/26–04/09	7	0.94(0.93, 0.95)	a_0
Ireland	03/08–04/09	7	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_-$
Italy	02/21–04/09	13	0.93(0.92, 0.94)	$a_0 + a_1(t-t_1)_-$
Luxembourg	03/10–04/09	6	0.90(0.89, 0.91)	$a_0 + a_1(t-t_1)_-$
Netherlands	02/28–04/09	7	0.94(0.93, 0.95)	$a_0 + a_1(t-t_1)_-$
Norway	02/28–04/09	6	0.93(0.92, 0.94)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Philippines	03/05-04/09	6	0.93(0.91, 0.95)	a_0
Poland	03/07-04/09	9	0.95(0.94, 0.96)	a_0
Portugal	03/03-04/09	11	0.92(0.91, 0.93)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$
Romania	03/04-04/09	7	0.95(0.94, 0.96)	a_0
Russia	03/11-04/09	9	0.97(0.96, 0.98)	a_0
Spain	02/25-04/09	9	0.93(0.92, 0.94)	a_0
Sweden	02/29-04/09	12	0.98(0.97, 0.99)	a_0
Switzerland	02/29-04/09	10	0.93(0.92, 0.94)	a_0
United Kingdom	02/26-04/09	11	0.95(0.94, 0.96)	a_0
Australia				
Australia	02/29–04/09	8	0.92(0.91, 0.93)	$a_0 + a_1(t-t_1)_- + a_2(t-t_2)_-$

Table 2: Estimation of median growth rate, final epidemic size and ending date for each country

		Ending date		Final epidemic size	
Region	γ_t (95% CI)	Prediction	Observation	Prediction	Observation
Asia					
China	0.107(0.097, 0.117)	2020/02/22	2020/02/22	$8553 \sim 9460$	8666
Hubei Province, China	0.177(0.151, 0.203)	2020/02/25	2020/02/25	$1.10e4 \sim 1.26e4$	1.14e4
Wuhan, China	0.197(0.039, 0.355)	2020/02/29	2020/03/02	-	-
India	0.215(0.175, 0.254)	2021/02/03	-	$> 1e6$	-
Indonesia	0.162(0.138, 0.185)	2020/07/22	-	$18648 \sim 83899$	-
Iran	0.141(0.130, 0.151)	2020/05/25	-	$1.01e5 \sim 1.32e5$	-
Japan	0.099(0.088, 0.110)	-	-	-	-
South Korea	0.150(0.129, 0.171)	2020/04/14	-	$8469 \sim 12449$	-
Malaysia	0.174(0.147, 0.200)	2020/05/11	-	$4994 \sim 6271$	-
Israel	0.323(0.261, 0.386)	2020/05/10	-	$9636 \sim 17046$	-
Pakistan	0.211(0.180, 0.243)	2021/03/04	-	$> 1e6$	-
Philippines	0.277(0.189, 0.364)	2020/05/31	-	$6778 \sim 10201$	-
Saudi Arabia	0.216(0.188, 0.243)	2020/06/23	-	$9302 \sim 21866$	-
Singapore	0.089(0.079, 0.098)	-	-	-	-
Thailand	0.187(0.172, 0.202)	2020/04/27	-	$2328 \sim 2956$	-
Turkey	0.314(0.297, 0.331)	2020/05/29	-	$91290 \sim 123019$	-
North America					
US	0.368(0.336, 0.400)	2020/06/04	-	$984100 \sim 1216034$	-
Canada	0.312(0.247, 0.378)	2020/05/12	-	$43866 \sim 105344$	-
Dominican Republic	0.150(0.109, 0.191)	2020/05/30	-	$7751 \sim 11364$	-
Mexico	0.167(0.144, 0.190)	2020/10/16	-	$> 1e5$	-
South America					
Chile	0.227(0.190, 0.264)	2020/05/23	-	$11536 \sim 15342$	-
Brazil	0.212(0.175, 0.249)	2020/06/04	-	$> 1e5$	-
Colombia	0.187(0.160, 0.213)	2020/06/13	-	$17692 \sim 105180$	-
Panama	0.199(0.171, 0.226)	2020/06/20	-	$7250 \sim 15462$	-
Europe					
Austria	0.246(0.215, 0.277)	2020/04/26	-	$12769 \sim 15887$	-
Belgium	0.241(0.210, 0.272)	2020/05/09	-	$44728 \sim 53322$	-
Czechia	0.216(0.192, 0.239)	2020/05/13	-	$6441 \sim 9609$	-
Finland	0.144(0.115, 0.172)	2020/06/09	-	$9263 \sim 23582$	-
France	0.221(0.179, 0.263)	2020/06/15	-	$> 3e5$	-
Germany	0.249(0.223, 0.275)	2020/05/21	-	$1.43e5 \sim 1.88e5$	-
Ireland	0.211(0.181, 0.241)	2020/05/19	-	$9414 \sim 15730$	-
Italy	0.180(0.169, 0.192)	2020/05/07	-	$1.58e5 \sim 1.81e5$	-
Luxembourg	0.150(0.130, 0.170)	2020/04/29	-	$2933 \sim 3907$	-
Netherlands	0.207(0.193, 0.220)	2020/06/07	-	$31582 \sim 40301$	-
Norway	0.177(0.153, 0.202)	2020/05/07	-	$5351 \sim 9636$	-
Poland	0.209(0.191, 0.228)	2020/06/22	-	$13693 \sim 28531$	-
Portugal	0.290(0.250, 0.329)	2020/05/12	-	$16759 \sim 22456$	-
Romania	0.241(0.198, 0.285)	2020/06/20	-	$5822 \sim 22280$	-
Russia	0.268(0.235, 0.300)	2020/08/06	-	$> 3e5$	-
Spain	0.274(0.238, 0.310)	2020/05/18	-	$184932 \sim 219819$	-
Sweden	0.123(0.103, 0.143)	2020/08/31	-	$> 8e4$	-
Switzerland	0.191(0.170, 0.212)	2020/05/14	-	$25285 \sim 32773$	-
United Kingdom	0.283(0.242, 0.325)	2020/06/11	-	$92514 \sim 312237$	-
Australia					
Australia	0.266(0.233, 0.298)	2020/04/16	-	$4932 \sim 6969$	-

The ending date is the day when the number of the cumulative cases reach a plateau (see Methods). For Japan and Singapore, the predicted trend does not reach a plateau because of increasing growth rate at last time points (Supplementary Figure 3(h) & 4(h))

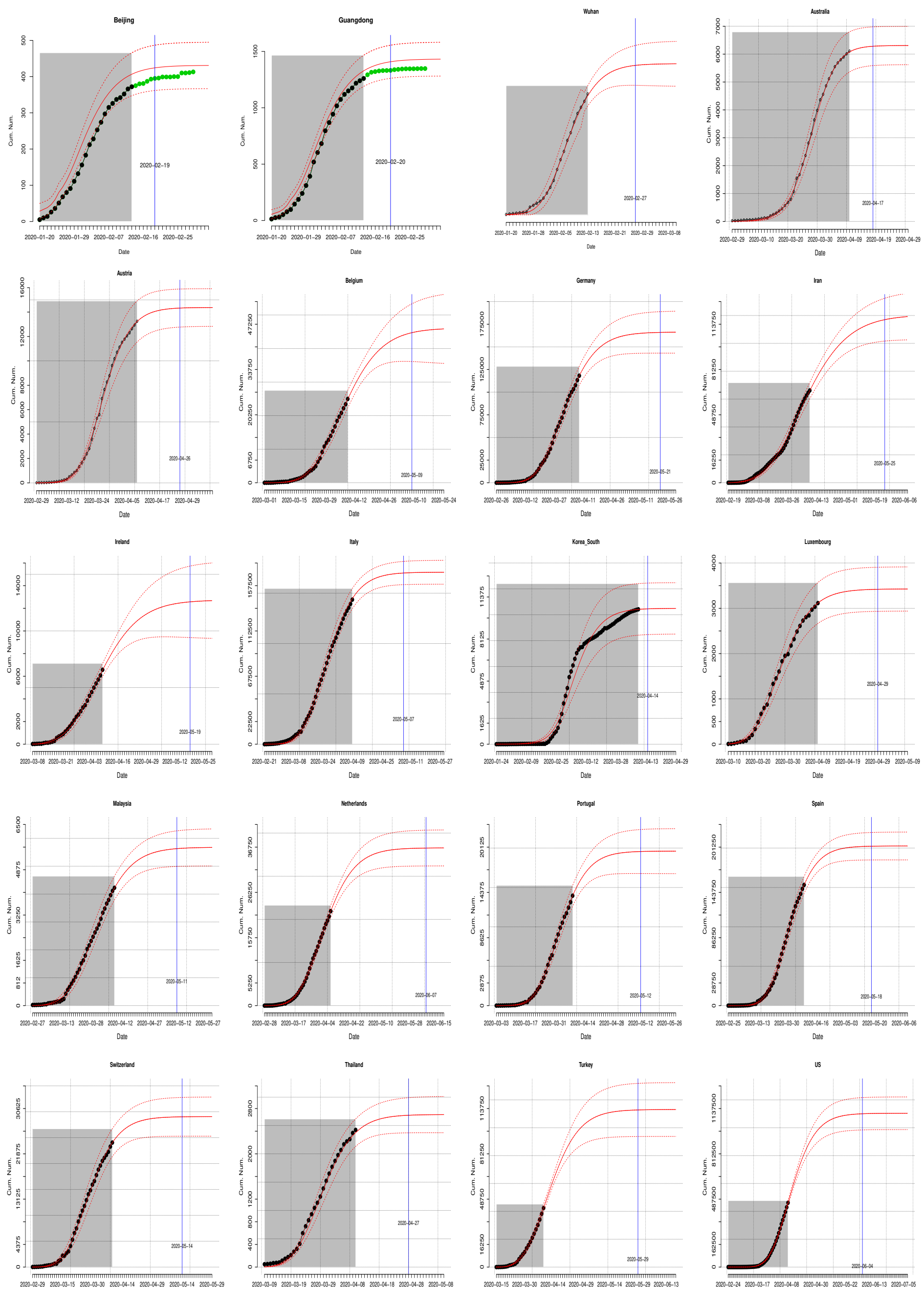

Figure 1: The estimated (red-solid) and observed (black-dotted) cumulative number of infectious over time t, as well as 95% CI (red-dashed) of the estimators. In addition, the green dots in the plot for Beijing and Guangdong in China refer to the number of reported cases which were not included for the model fitting.

Table 3: The predicted or observed peak date and the corresponding cumulated size among the different countries

	Peak date		# Maximum cases at the peak		# Ventilator needed
	Prediction	Observation	Prediction	Observation	for the peak
Asia					
India	2020/07/18	-	$> 1e6$	-	$> 5e5$
Indonesia	2020/04/24	-	7804	-	$390 \sim 780$
Iran	2020/04/03	2020/03/28-04/05	51615	47226	$2581 \sim 5162$
Israel	2020/04/01	2020/03/24-04/08	5929	5717	$296 \sim 593$
SKorea	2020/03/04	2020/02/25-03/10	4435	4693	$222 \sim 444$
Japan	-	-	-	-	-
Malaysia	2020/03/29	2020/03/24-04/02	2469	2382	$123 \sim 247$
Pakistan	2020/08/05	-	1.20e6	-	$6.01e4 \sim 1.20e5$
Philippines	2020/04/04	2020/04/03-04/06	2914	3254	$146 \sim 291$
Saudi Arabia	2020/05/13	-	26803	-	$1340 \sim 2680$
Singapore	-	-	-	-	-
Thailand	2020/03/28	2020/03/22-03/29	1222	987	$61 \sim 122$
Turkey	2020/04/14	2020/04/11-04/16	65276	63144	$3264 \sim 6528$
North America					
US	2020/04/09	2020/04/04-04/16	459761	489999	$22988 \sim 45976$
Canada	2020/04/08	2020/04/02-04/09	19208	15836	$960 \sim 1921$
Dominican Republic	2020/04/16	-	3512	-	$176 \sim 351$
Mexico	2020/05/23	-	60307	-	$3015 \sim 6031$
South America					
Chile	2020/04/07	2020/04/03-04/16	5071	6212	$254 \sim 507$
Brazil	2020/05/22	-	3.167927e+05	-	$15839 \sim 31679$
Colombia	2020/04/30	-	8173	-	$409 \sim 817$
Panama	2020/04/12	2020/04/09-04/12	3083	2872	$154 \sim 308$
Europe					
Austria	2020/03/26	2020/03/23-03/29	6641	6710	$332 \sim 664$
Belgium	2020/04/05	2020/04/02-04/07	19368	18875	$968 \sim 1937$
Czechia	2020/04/01	2020/03/28-04/05	3556	3586	$178 \sim 356$
Finland	2020/04/19	-	4242	-	$212 \sim 424$
France	2020/04/14	2020/04/10-04/16	1.60e5	1.21e5	$8022 \sim 16043$
Germany	2020/03/31	2020/03/24-04/08	72246	74873	$3612 \sim 7225$
Ireland	2020/04/06	2020/04/09-04/11	5421	7864	$271 \sim 542$
Italy	2020/03/25	2020/03/16-03/31	7.54e4	6.68e4	$3771 \sim 7541$
Luxembourg	2020/03/27	2020/03/21-03/31	1573	1434	$79 \sim 157$
Netherlands	2020/04/03	2020/04/02-04/10	15710	18862	$786 \sim 1571$
Norway	2020/03/23	2020/03/22-03/27	2659	3012	$133 \sim 266$
Poland	2020/04/12	2020/04/05-04/11	6796	5207	$340 \sim 680$
Portugal	2020/04/02	2020/03/31-04/10	8884	11196	$444 \sim 888$
Romania	2020/04/11	2020/04/09-04/16	5943	6424	$297 \sim 594$
Russia	2020/05/11	-	141110	-	$7056 \sim 14111$
Spain	2020/03/30	2020/03/25-04/01	87412	76795	$4371 \sim 8741$
Sweden	2020/05/09	-	37389	-	$1869 \sim 3739$
Switzerland	2020/03/27	2020/03/23-03/30	12516	12392	$626 \sim 1252$
United Kingdom	2020/04/13	2020/04/05-04/14	86715	70811	$4336 \sim 8672$
Australia					
Australia	2020/03/28	2020/03/25-03/30	3482	3384	$174 \sim 348$

* Peak date has been observed. Due to the fluctuation of the daily reported cases, the observed peak interval was listed in stead of the exact peak date.
The number of maximum cases at the peak is the accumulated number of cases at the peak period.
The number of ventilator for the peak requirement is calculated based on the predicted infection. We assume $5 - 10\%$ of the infected patients need the ventilator[22].

Table 4: Estimated Medium R_t (95%CI) in the subsequent week

		First week	Second week	Third week	Fourth week
Asia	China	1.64(0.98, 2.30)	1.34(1.21, 1.48)	0.59(0.55, 0.63)	0.26(0.24, 0.28)
	Hubei Province, China	3.11(2.58, 3.64)	2.05(1.86, 2.23)	0.73(0.68, 0.77)	0.26(0.23, 0.29)
	Wuhan, China	2.50(1.56, 3.44)	1.51(0.83, 2.19)	0.64(0.40, 0.88)	0.27(0.17, 0.37)
	India	2.34(1.31, 3.38)	1.23(0.83, 1.63)	1.12(0.81, 1.42)	1.02(0.71, 1.32)
	Indonesia	4.49(3.66, 5.31)	1.77(1.30, 2.23)	1.24(0.98, 1.51)	0.99(0.87, 1.11)
	Iran	3.56(2.39, 4.73)	2.01(1.78, 2.23)	1.37(1.21, 1.53)	1.05(0.94, 1.16)
	Israel	7.70(3.86, 11.55)	4.26(2.79, 5.73)	2.36(1.88, 2.83)	1.30(1.16, 1.45)
	Japan	4.36(3.21, 5.50)	1.89(1.55, 2.22)	1.02(0.90, 1.14)	0.68(0.60, 0.76)
	SKorea	1.52(0.99, 2.06)	1.42(0.99, 1.84)	1.32(0.98, 1.65)	1.22(0.96, 1.49)
	Malaysia	5.77(4.81, 6.72)	3.39(2.77, 4.02)	2.19(1.66, 2.72)	1.46(1.18, 1.74)
	Pakistan	4.90(3.58, 6.22)	2.32(1.82, 2.83)	1.02(0.86, 1.18)	0.71(0.60, 0.82)
	Philippines	4.18(1.20, 7.17)	2.59(1.31, 3.87)	1.60(1.12, 2.09)	0.99(0.83, 1.16)
	Saudi Arabia	2.91(1.55, 4.27)	2.14(1.53, 2.75)	1.57(1.33, 1.80)	1.15(0.97, 1.33)
	Singapore	1.38(1.05, 1.72)	0.95(0.77, 1.14)	0.70(0.59, 0.81)	0.54(0.47, 0.61)
	Thailand	6.03(4.55, 7.50)	2.99(2.60, 3.37)	1.48(1.38, 1.58)	0.73(0.63, 0.84)
	Turkey	5.53(4.10, 6.95)	1.97(1.85, 2.10)	1.08(0.95, 1.21)	-
North America	US	3.03(2.53, 3.54)	3.69(3.13, 4.26)	3.24(2.68, 3.79)	2.19(1.91, 2.47)
	Canada	2.92(2.50, 3.34)	2.25(1.60, 2.91)	2.45(1.78, 3.13)	2.00(1.58, 2.41)
	Dominican Republic	3.91(2.87, 4.94)	1.16(0.52, 1.80)	0.84(0.60, 1.08)	-
	Mexico	2.78(1.80, 3.76)	2.45(2.00, 2.90)	2.16(1.99, 2.32)	1.90(1.68, 2.11)
South America	Chile	5.29(4.03, 6.54)	2.33(1.78, 2.89)	1.31(1.09, 1.53)	0.85(0.76, 0.95)
	Brazil	6.91(4.55, 9.27)	4.36(3.30, 5.41)	1.84(1.56, 2.12)	1.54(1.38, 1.71)
	Colombia	5.37(3.82, 6.92)	2.55(1.90, 3.19)	1.34(1.13, 1.55)	1.07(0.90, 1.24)
	Panama	4.89(2.67, 7.11)	2.03(1.51, 2.55)	1.41(1.07, 1.76)	1.00(0.70, 1.30)
Europe	Austria	4.91(3.63, 6.19)	3.91(2.80, 5.01)	2.75(2.14, 3.36)	1.38(1.17, 1.58)
	Belgium	6.61(4.95, 8.28)	4.27(3.36, 5.19)	2.67(2.26, 3.08)	1.76(1.59, 1.92)
	Czechia	4.07(2.81, 5.33)	2.33(2.07, 2.59)	1.47(1.28, 1.67)	0.93(0.76, 1.09)
	Finland	7.76(5.23, 10.28)	2.50(1.23, 3.78)	1.38(0.80, 1.97)	1.10(0.72, 1.48)
	France	4.29(3.39, 5.19)	1.80(1.06, 2.54)	1.26(0.83, 1.69)	0.95(0.71, 1.18)
	Germany	5.52(4.16, 6.88)	3.48(2.83, 4.13)	2.19(1.91, 2.48)	1.38(1.27, 1.49)
	Ireland	4.08(2.88, 5.27)	1.77(1.36, 2.18)	1.15(0.92, 1.38)	0.76(0.63, 0.90)
	Italy	12.36(10.99, 13.74)	5.14(4.52, 5.76)	3.11(2.77, 3.46)	1.91(1.71, 2.10)
	Luxembourg	4.05(2.48, 5.63)	1.71(1.47, 1.94)	0.82(0.70, 0.94)	0.40(0.33, 0.47)
	Netherlands	4.39(3.01, 5.77)	2.54(2.38, 2.70)	1.68(1.57, 1.79)	1.13(1.04, 1.22)
	Norway	3.89(3.09, 4.70)	2.43(2.09, 2.77)	1.38(1.09, 1.66)	0.75(0.62, 0.89)
	Poland	3.45(2.85, 4.05)	2.46(2.15, 2.77)	1.75(1.61, 1.90)	1.25(1.17, 1.33)
	Portugal	5.87(4.44, 7.30)	4.30(3.38, 5.22)	3.24(2.74, 3.74)	1.83(1.68, 1.99)
	Romania	3.42(2.06, 4.78)	2.40(1.73, 3.08)	1.69(1.38, 1.99)	1.19(1.06, 1.31)
	Russia	3.41(2.61, 4.22)	2.71(2.28, 3.14)	2.15(1.97, 2.33)	1.70(1.62, 1.79)
	Spain	9.26(6.73, 11.79)	5.54(4.34, 6.73)	3.31(2.78, 3.84)	1.98(1.76, 2.20)
	Sweden	2.39(1.64, 3.14)	2.07(1.56, 2.58)	1.79(1.47, 2.12)	1.55(1.35, 1.75)
	Switzerland	6.39(4.64, 8.15)	3.77(3.05, 4.49)	2.22(1.95, 2.50)	1.31(1.18, 1.44)
	United Kingdom	7.46(5.18, 9.74)	5.26(4.00, 6.52)	3.71(3.06, 4.36)	2.62(2.30, 2.93)
Australia	Australia	3.72(3.41, 4.03)	2.48(2.16, 2.79)	2.16(1.90, 2.41)	1.43(1.30, 1.57)

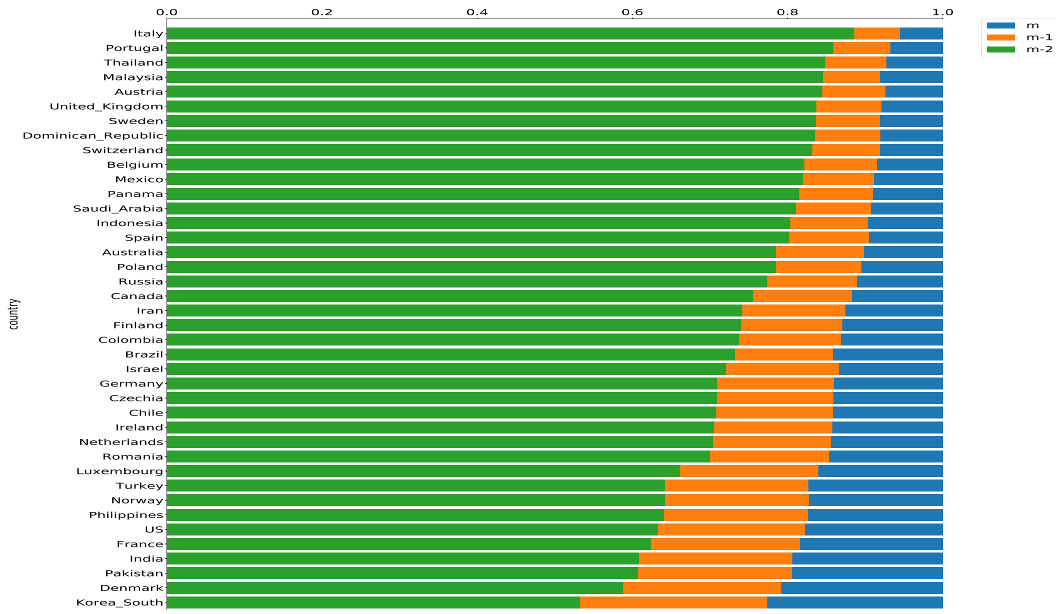

Figure 2: The proportion of epidemic size when the duration of infectious period m is reduced by 1 and 2. The reduction of infectious duration refers to shorten the time from the infection onset (symptomatic or asymptomatic) to isolation by intervention measures.

4. Discussions

Here we predicted that the COVID-19 pandemic will persistently spread over many countries and will last on average 5 to 6 months since January 2020 for the current wave. We estimated, given no further intervention measures, that India would emerge to be a new epidemic center, as well as Pakistan, Brazil, Russia and Mexico. Effective intervention measures by reducing the infectious period would result in an 50% reduction in the final epidemic sizes. R_t had a declining trend in almost all countries, revealing the impact of the intervention measures i.e. social distancing.

Our real-time estimation framework can yield a reliable prediction of the final epidemic size using the data from the early phase of the epidemic. We evaluated the performance of our model using incident case data from January 20 to February 13, 2020 in China. We predicted the epidemic would fade out on February 19 to 24 across 28 provinces in China with a one-week lag in Wuhan; the epidemic size was estimated at 8,500 to 9,500, and 11,000 to 12,600 in 28 provinces and Hubei respectively. The outbreak size and epidemic duration estimated are found to be highly consistent with the observed figures (https://www.who.int). The prediction of the final epidemic size based on the model that assumes early exponential growth could tend to overestimate the epidemic size, which has been shown in the previous studies[10, 23–28].

We estimated that the most affected countries in the next wave of the pandemic will be India, Pakistan, Brazil, Mexico as well as Russia. All these countries are currently undergoing regional or national lockdown except Mexico. Given that thousands of ventilators are needed for the peak around May to July, it is immediately important for these countries to prepare the ICU beds. In Europe, most countries have peaked in the curve, and the final outbreak size was estimated to exceed 10^5 in Italy, France, UK, Germany and Spain. For the United States, the estimated final epidemic size will surpass 1 million, peaking around July 2020. A recent report calculated that 81% of British and the United States populations would be infected, with approximately a half and 2.2 million deaths respectively[29]. The high value is based on the assumption of R_0 at 2.4, while our results demonstrate that real-time R_t decelerated under interventions. Thus, their estimates are likely to be the higher bound of the true value. In Asia, besides China, South Korea and Iran, Thailand and Malaysia have peaked in March. Indonesia and Saudi Arabia would peak in April. Singapore and Japan recently implemented lockdown and more data will be needed for the epidemic prediction. Given the case-fatality rate at 2.3%[6], we estimated around 23,000 deaths in India and Pakistan, and around 2,300 in Mexico and Brazil, and 6,900 in Russia.

This is the first study to compare country-specific temporal R_t. It is natural to expect a declining trend of R_t to 1 owing to stochastic effects for epidemics governed by subexponential growth[18, 30], while a faster decline may suggest a larger impact of intervention measures or behavior changes31. A majority of countries displayed R_t below 5 in the first week of the epidemic period. Several countries exhibited large R_t, ie. Italy or Spain, suggesting a rapid increase of cases at the beginning or a large variation in the under-reporting rates in the early epidemic phase. For Wuhan, R_t declined from the median 2.5 to 0.3 within a month, which is compatible with the R_0 estimation in a range of 2.2 to 3.6 in Wuhan[2, 10, 12, 13, 16, 23, 31], and below 1 after interventions[32]. All countries displayed a declined trend towards 1 in the epidemic period with varying deceleration rates, revealing the variation in the impact of the intervention strategies. For further research on the intervention effectiveness, individual data, as well as a series of intervention measures for each country, may be required to quantify the effects in detail.

Our study showed that by shortening the infectious period by two days, we could effectively reduce the final epidemic size by up to 50%. What strategies can effectively reduce the infectious period? Usually, it can be done - similar to the other coronaviruses - by early detection and isolation of symptomatic patients and tracing of close contacts. However, the presymptomatic transmission has been reported in many countries such as China, the US, and Singapore[33–36], etc. The existence of presymptomatic or asymptomatic transmission would present difficult challenges for disease control, which underscores the importance of social distancing. For instance, at the onset of the epidemic, the spread has been well controlled in Singapore, accounted for by a range of intervention measures, such as contact tracing and quarantining, that were instituted from January 23-the onset of the first case in Singapore[37]. However, the recent increase in cases in Singapore, initially starting from the imported cases followed by the outbreak in dormitories of immigration workers, led to its circuit period from April 7 to May 4. Quarantine measures for the infected regions or at the national scale, either a complete lockdown as in China or partial lockdown such as in Europe, are effective in flattening the curve[38]. Besides the quarantine measures, South Korea was also successful in suppressing the outbreak, attributed to the rapid measures to perform large-scale diagnostic testing for the public for case isolation[39].

During the early phase of the epidemic, we forecasted the epidemic size as a function of the time-varying growth model, similar but more flexible to the previous approaches to model sub-exponential growth dynamics[18–21, 40], without relying on any epidemiological parameter assumptions. The traditional transmission epidemiological model is defined through a susceptible, exposure, infectious, removed (SEIR) scheme, which requires epidemiological parameters from detailed case studies[2, 10, 31, 41]. Furthermore, given the variation of the transmissibility, mostly due to the intervention strategies implemented and behavioral changes in the population, it is desirable to quantify the dynamics of R_t over time[20, 42–44]. Several studies attempted to forecast the number of epidemics for COVID-19 in China using the constant growth rate[27, 45, 46], which is not the case in our model of nonmonotonic behavior of the growth rate. Our model, compared to previous studies, has greater flexibility in a data-mining manner to fit and predict future trajectories.

In summary, we compared the epidemic trajectory, characterized dynamic R_t in 42 countries, and predicted that the new epidemic centers will continue to emerge in the next wave. Meanwhile, we highlighted the importance of various effective interventions in flattening the epidemic curve, such as social distancing, to shorten the infectious period. By carefully characterizing the shape of the epidemic growth phase, we believe our study represents a significant step in modeling real-time transmission and providing accurate forecasts of epidemics.

Our study has several limitations. Firstly, the projection of the temporal trend of an outbreak using the early-stage dataset could be dramatically influenced by the changes in the intervention strategies later on. For example, using data up to March 15, our previous analysis overestimated the epidemic size in U.S as the social distance measures (i.e. quarantine) were implemented starting from March 21. The performance of models used in this work will be continuously improved with data coming in from an ongoing outbreak, thus real-time estimates of key epidemiological parameters can be available before the epidemic fully ends. Secondly, the number of infections estimated might not be comparable across the countries; for example, the number of infections in Germany is not comparable to the number of infections in Italy or China, as the latter did not perform large-scale population-based testing and thus the cases could be more severe. Lastly, the analyses are highly reliant on the reporting criteria and quality of the data. The under-reporting of infection is likely a common scenario in the majority of countries. For example, a recent study showed that the reported number of confirmed positive cases was 50-85-fold lower than the actual number of infections in 3330 people in Santa Clara County, US[47]. A more realistic and comprehensive analysis could be performed that includes accurate epidemic data and information. In current imperfect situation, our model could still be used for more advanced analyses, including estimations of the epidemic size, peak points and dynamic R_t.

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] 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.

[4] 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.

[5] 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.

[6] Z. Wu, J. M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention, Jama.

[7] K. Sun, J. Chen, C. Viboud, Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study, The Lancet Digital Health.

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

[9] M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, P.-Y. Boëlle, E. D'Ortenzio, Y. Yazdanpanah, S. P. Eholie, M. Altmann, et al., Preparedness and vulnerability of african countries against importations of covid-19: a modelling study, The Lancet.

[10] 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.

[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] 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.

[13] J. Riou, C. L. Althaus, Pattern of early human-to-human transmission of wuhan 2019 novel coronavirus (2019-ncov), december 2019 to january 2020, Eurosurveillance 25 (4).

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

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

[16] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, M. H. Wang, 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 92 (2020) 214 – 217. doi:https://doi.org/10.1016/j.ijid.2020.01.050. URL http://www.sciencedirect.com/science/article/pii/S1201971220300539

[17] S. Zhao, S. S. Musa, H. Fu, D. He, J. Qin, Simple framework for real-time forecast in a data-limited situation: the zika virus (zikv) outbreaks in brazil from 2015 to 2016 as an example, Parasites & vectors 12 (1) (2019) 344.

[18] G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, Mathematical models to characterize early epidemic growth: A review, Physics of life reviews 18 (2016) 66–97.

[19] C. Viboud, L. Simonsen, G. Chowell, A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks, Epidemics 15 (2016) 27–37.

[20] 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.

[21] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, other, Transmission dynamics and control of severe acute respiratory syndrome, Science.

[22] W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui, et al., Clinical characteristics of coronavirus disease 2019 in china, New England Journal of Medicine.

[23] H. Wang, Z. Wang, Y. Dong, R. Chang, C. Xu, X. Yu, S. Zhang, L. Tsamlag, M. Shang, J. Huang, et al., Phase-adjusted estimation of the number of coronavirus disease 2019 cases in wuhan, china, Cell Discovery 6 (1) (2020) 1–8.

[24] M. Shen, Z. Peng, Y. Xiao, L. Zhang, Modelling the epidemic trend of the 2019 novel coronavirus outbreak in china, bioRxiv.

[25] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, C. P. Jewell, Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions, medRxiv.

[26] M. A. Al-qaness, A. A. Ewees, H. Fan, M. Abd El Aziz, Optimization method for forecasting confirmed cases of covid-19 in china, Journal of Clinical Medicine 9 (3) (2020) 674.

[27] N. Imai, I. Dorigatti, A. Cori, C. Donnelly, S. Riley, N. M. Ferguson, Report 2: Estimating the potential total number of novel coronavirus cases in wuhan city, china, Imperial College London.

[28] S. Zhao, S. S. Musa, Q. Lin, J. Ran, G. Yang, W. Wang, Y. Lou, L. Yang, D. Gao, D. He, et al., Estimating the unreported number of novel coronavirus (2019-ncov) cases in china in the first half of january

2020: a data-driven modelling analysis of the early outbreak, Journal of clinical medicine 9 (2) (2020) 388.

[29] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al., Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand.

[30] G. Chowell, C. Viboud, L. Simonsen, S. M. Moghadas, Characterizing the reproduction number of epidemics with early subexponential growth dynamics, Journal of The Royal Society Interface 13 (123) (2016) 20160659.

[31] T. Liu, J. Hu, M. Kang, L. Lin, H. Zhong, J. Xiao, G. He, T. Song, Q. Huang, Z. R. and, Transmission dynamics of 2019 novel coronavirus (2019-ncov).

[32] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, S. Flasche, Early dynamics of transmission and control of covid-19: a mathematical modelling study, The Lancet Infectious Diseasesdoi:https://doi.org/10.1016/S1473-3099(20)30144-4. URL http://www.sciencedirect.com/science/article/pii/S1473309920301444

[33] G. Qian, N. Yang, A. H. Y. Ma, L. Wang, G. Li, X. Chen, X. Chen, A covid-19 transmission within a family cluster by presymptomatic infectors in china, Clinical Infectious Diseases.

[34] W. E. Wei, Presymptomatic transmission of sars-cov-2—singapore, january 23–march 16, 2020, MMWR. Morbidity and Mortality Weekly Report 69.

[35] L. Tian, X. Li, F. Qi, Q.-Y. Tang, V. Tang, J. Liu, X. Cheng, X. Li, Y. Shi, H. Liu, et al., Pre-symptomatic transmission in the evolution of the covid-19 pandemic, arXiv preprint arXiv:2003.07353.

[36] A. Kimball, Asymptomatic and presymptomatic sars-cov-2 infections in residents of a long-term care skilled nursing facility—king county, washington, march 2020, MMWR. Morbidity and mortality weekly report 69.

[37] J. E. L. Wong, Y. S. Leo, C. C. Tan, COVID-19 in Singapore—Current Experience: Critical Global Issues That Require Attention and Action, JAMAarXiv:https://jamanetwork.com/journals/jama/articlepdf/2761890/jama_wong_2020_vp_200026.pdf, doi:10.1001/jama.2020.2467.
URL https://doi.org/10.1001/jama.2020.2467

[38] A. L. Phelan, R. Katz, L. O. Gostin, The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance, JAMA 323 (8) (2020) 709–710. arXiv:https://jamanetwork.com/journals/jama/articlepdf/2760500/jama_phelan_2020_vp_200008.pdf, doi:10.1001/jama.2020.1097.
URL https://doi.org/10.1001/jama.2020.1097

[39] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, How will country-based mitigation measures influence the course of the covid-19 epidemic?, The Lancet.

[40] 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.

[41] 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.

[42] 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.

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

[44] 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.

[45] S.-m. Jung, A. R. Akhmetzhanov, K. Hayashi, N. M. Linton, Y. Yang, B. Yuan, T. Kobayashi, R. Kinoshita, H. Nishiura, Real-time estimation of the risk of death from novel coronavirus (covid-19) infection: Inference using exported cases, Journal of clinical medicine 9 (2) (2020) 523.

[46] S. W. Hermanowicz, Forecasting the wuhan coronavirus (2019-ncov) epidemics using a simple (simplistic) model, medRxiv.

[47] E. Bendavid, et al., Covid-19 antibody seroprevalence in santa clara county, california, medRxiv, 2020.04.14.20062463 69.