medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Forecasting the Worldwide Spread of COVID-19 based on Logistic Model and SEIR Model

Xiang Zhou[1], Na Hong[2], Yingying Ma[2], Jie He[2], Huizhen Jiang[3], Chun Liu[2], Guangliang Shan[3], Longxiang Su[1], Weiguo Zhu[4,5], Yun Long[1]

[1]Department of Critical Care Medicine, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

[2]Digital China Health Technologies Co. Ltd., Beijing 100080, China

[3]Department of Epidemiology and Biostatistics, Institute of Basic Medicine Sciences, Chinese Academy of Medical Sciences (CAMS) & School of Basic Medicine, Peking Union Medical College, Beijing, 100730, China.

[4]Department of General Internal Medicine, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

[5]Department of Information Management, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

ABSTRACT

Background: As the outbreak of coronavirus disease 2019 (COVID-19), a sudden case increase in late February 2020 in global attracted deep concern. Italy, South Korea, Iran, France, Germany, Spain, the U.S. and Japan are probable the most severe countries. Collecting epidemiological data and predicting epidemic trends are important to develop and measure public intervention strategies. Epidemic predictions results yield by different mathematical models are out of line, therefore, we sought to compare different models and their prediction results, so as to generate objective conclusions.

Methods: We used the number of cases reported from January 23 to March 20, 2020 to estimate possible spread size and peak time of COVID-19, especially in 8 high risk countries. Logistic growth model, basic SEIR model and adjusted SEIR model were adopted for predicting. Considering different model inputs may infer different model outputs, we implemented three model predictions with three scenarios of epidemic development.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results: When contrasting all 8 countries short-term prediction results and peak predictions, the difference between the models was relatively large. The logistic growth model estimated a smaller epidemic size than the basic SERI model, however, once we added parameters which considered the effects of public health interventions and control measures, the adjusted SERI model results demonstrated a considerably rapid decelerate of the epidemic development. Our results demonstrated contact rate, quarantine scale, quarantine initiate time and length are important factors to control the epidemic size and length.

Conclusions: We demonstrated a comparative assessment of the predictions of COVID-19 outbreak of 8 high risk countries using multiple methods. By forecasting epidemic size and peak time as well as simulating the effects of public health interventions, the intent of this paper is to help understand the transmission dynamics of COVID-19 and recommend operation suggestions to slow down the epidemic. It is suggested that quickly detecting cases, enough quarantine implementation and public self-protection behaviors are critical to slow down the epidemic.

Keywords: COVID-19, epidemic predictions, SERI model, logistic growth model, public intervention strategies

Introduction

A novel coronavirus is a major public health event.. Although the governments had implemented various measures to protect their cities or countries, such like traffic restrictions, quarantine requirement for travelers, contact tracing, etc., a large-scale global movement of population have already caused rapid spread of the disease, resulting in an epidemic throughout the world. In February 2020, the World Health Organization (WHO) named the disease COVID-19, which stands for coronavirus disease 2019[1]. The virus that causes COVID-19 is named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The SARS-CoV-2 virus spreads mainly from person-to-person through between people who are in close contact with one another, and through respiratory droplets produced when an infected person coughs or sneezes[2]. As the epidemic came under control in China, global spread of the COVID-19 had caused a surge in Asia, Europe, Middle East and North American. As March 11, 2020, with the global risk continuously increased, there were already more than 118,000 cases in 114 countries and 4,291 people have lost their lives, WHO characterized COVID-19 as a pandemic[3].

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

With growing number of cases have occurred in more than 150 countries and regions, modeling the COVID-19's transmission dynamics and estimating its development are crucial to provide decision supports for public health departments and healthcare policy makers. Mathematical models are widely used in evaluating epidemic transmissions, forecasting the trend of disease spread, and providing optimal intervention strategies and control measures. Considerable recent studies have contended to estimate COVID-19's scale and severity, several mathematical models and predicting approaches have attempted to estimate the transmission of COVID-19[4-8]. Majority of the researches estimated the basic reproductive number R_0, a key parameter to evaluate the potential of COVID-19 transmission. However, different models often yield different conclusions in terms of differences in model structure and input parameters. It is imperative and critical to improve early predictive and warning capability for the pandemic.

In face of this new infectious disease and its complicated features with many unknown factors, single model estimations may infer biased results, therefore, we tried to make overall rigorous estimations by comparing different model results. To achieve an objective estimation, we investigated and implemented two most common approaches and one extended approach: the logistic model, the susceptible-exposed-infected-removed (SEIR) model and the adjusted SEIR model. Countries had different start time and level of interventions and measures to reduce risk of domestic secondary infections of COVID-19. We compared the models that have taken these effects into account or not, predicted the spread of the epidemic, and tried to compare different recommendations from 3 different models.

Methods

We collected the epidemic situation of COVID-19 in eight high risk countries that distributed in 3 continents, and compared the perdition results with the logistic model and the SEIR model with different parameter setting scenarios. Data is from the Coronavirus COVID-19 Global Cases published by Center for Systems Science and Engineering (CSSE) of Johns Hopkins University[9]. We used the existing reported data from January 23 to March 20 2020 for observing, performing parameter estimation, and forecasting COVID-19 dynamics in different countries/regions.

Logistic growth model and parameters estimate

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In Scenario 1, we have an assumption that the epidemic trend obeys logistic growth curve. We use logistic model to predict the disease trends. The logistic model's essence is that curve fitting and its prediction results highly depend on the historical data. It has been often used in epidemics dynamics prediction in previous studies[4,10,11]. Mathematically, the logistic model describes dynamic evolution of infected individuals being controlled by the growth rate and population capacity. According to the following ordinary differential equation (a), we will get logistic function (b), the model describes dynamic evolution of the reported number of confirmed cases P being controlled by the growth rate r, and the initial value of P_0 is the confirmed number of cases when T=0. The maximum case volume in the environment is K, which is the limit that can be reached by increasing to the final value of P(t), and r is the growth rate. We used the least squares method to fit the logistic growth function, and then to predict the number of future confirmed cases. Since the case numbers reported at very early stage are usually inaccuracy or missing, the initiate date of the model was set as the day since the 100th confirmed case was reached.

$$\frac{dP}{dt} = rP(1 - \frac{P}{K}) \qquad \text{(a)}$$

$$P(t) = \frac{KP_0e^{rt}}{K+P_0(e^{rt}-1)} \qquad \text{(b)}$$

For evaluating the logistic model's capability to predict infectious disease like COVID-19, we've fit the logistic curve every 10 days since the day as the 100th confirmed case was reached and each time we make 7 days predictions and use reported data for evaluation. For the experiment analysis, each time of perdition, we got different results errors, as listed in Appendix A. With the number of confirmed cases increases, the predicted following 7 days infections and future peak size and peak time are constantly changing. In addition, the shape of the curve will probably change due to exogenous effects like new burst of infection, control measures implementation, public behaviors and so on.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SEIR model and parameters estimate

Based on the epidemiological characteristics of COVID-19 infection, the SEIR model is more commonly adopted to study the dynamic of this disease. SEIR is a deterministic metapopulation transmission model, it simulates each individual in the population as a separate compartment, with the assumption that each individual in the same compartment has the same characteristics. By plugging in different setting of parameters, however, the models yield different results and we compared their results to observe patterns of the COVID-19 spread under two different scenarios, the basic SEIR (Scenario 2: without any interventions and measures) and the adjusted SEIR (Scenario 3: with strict interventions and measures).

In Scenario 2, we only used basic SEIR model, and the population is divided into four classes: susceptible(S), exposed (E), infectious (I) and removed(R). The essence of SEIR model is a system of ordinary differential equations about time. The disease trend it predicts only depends on parameters and the start time. The model is measured by below equation, and the entire population was initially susceptible with the assumption that all people have no immunity against COVID-19, the initiate number of cases were collected from the reported data. For the reason that reliable data are still scarce during the early days of a new outbreak, the initiate date of the model was set as the day since the 100th confirmed case was reached for each country, which indicates different initiate dates of 8 observed countries.

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = -\frac{\beta SI}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

where S is the number of susceptible population, E is the number of exposed population, I is the number of infected population, R is the number of the recovery or death, N is the number of the whole population, $\beta = k * b$ is the product of the people exposed to by infected population k and

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the probability of transmission b. $\gamma = 1/D$, is the average rate of recovery or death in infected populations, where D is the average duration of the infection, σ is the transition rate of the exposed individual develops into infected.

In Scenario 3, considering the contribution to the epidemic dynamics by public health interventions factors, such as the government locked down the cities, took measures to track and quarantine people who have close contact with confirmed cases, advocated citizens to keep social distances and wash hands frequently, etc., we further used an adjusted SEIR model for COVID-19 estimation. This model considered the contact rate and quarantined proportion of the COVID-19 transmission, and divided the population into seven classes: susceptible, exposed, infectious, removed, quarantined susceptible, quarantined exposed and quarantined infected. A fraction of the susceptible population was quarantined and identified as Sq and a fraction of the exposed population was isolated and identified as Eq. We provide their detailed equation and parameter definitions as follows[8]:

$$\begin{cases} \frac{dS}{dt} = -[c\beta + cq(1-\beta)]S(I+\theta E) + \lambda S_q, \\ \frac{dE}{dt} = c\beta(1-q)S(I+\theta E) - \sigma E, \\ \frac{dI}{dt} = \sigma E - (\delta_I + \alpha + \gamma_I)I, \\ \frac{dS_q}{dt} = (1-\beta)cqS(I+\theta E) - \lambda S_q, \\ \frac{dE_q}{dt} = \beta cqS(I+\theta E) - \delta_q E_q, \\ \frac{dH}{dt} = \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H, \\ \frac{dR}{dt} = \gamma_I I + \gamma_H H \end{cases}$$

Countries all reduced their people's social contact although by implementing multiple different control measures, we give a preliminary estimation for the contact rate and quarantined proportions as model parameters. In addition, as the quarantine and control measures take effects after a period of cases increasing, usually it has surged into a considerable number, therefore, the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

initiate dates of 8 observed countries used in the adjusted model were different, they were defined as the dates countries governments declared strict interventions and control measures for a large-scale of population. We assumed that the contact rate decreased since the government has implemented strict control measures. In terms of our previous study and related studies, the contact rate c is below 8 after a large-scale intervention initiated. The initial populations of each country are acquired from published data[12], and the initial infected, recovered population were set based on the reported data[9]. Model parameters are estimated on the basis of fitting reported data from the initiate date, the probability of transmission per contact of each country were estimated using early stages data of each country based on Monte Carlo simulation. We assumed that the median incubation period was 5-6 days (ranging from 0-14 days) based on the WHO report[13], the initial quarantine proportion of uninfected susceptible were 60%~80% of the population (under the assumption strict quarantine policy are large-scale executed), with the quarantine period set as 28 days, the mortality rate is derived from reported data of each country. Based on the above assumptions, we implemented the adjusted SEIR model to get the minimal estimation.

Spreading potential evaluation

To disclose the epidemic growth potential of each country, we used two parameters to assess. First, the growth rate of logistic curves, the greater it is, the faster the curve grows. Second, the basic reproduction number R_0. We used the next generation matrix to derive a formula for the basic reproduction number, as follows:

$$R_0 = \frac{\beta}{\gamma} = \frac{k * b}{1/D}$$

where β is the product of the population exposed to infected population(k) and the probability of transmission (b). γ=1/D is the average rate of recovery or death in infected populations, where D is the average duration of the infection, in terms of related studies for the COVID-19 infection assumption[7], we set D=7 to get the R_0 estimation.

Results

COVID-19 Epidemic Progressing

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We summarized the epidemic curves of the eight observed countries, to help observe the global trends and disclose spread pattern of different countries. Cases raised rapidly in these countries, including Italy, South Korea, Iran, France, etc. The epidemic curves are shown in Figure 1, the initiate date of the curves was set as the earliest day since the 100th confirmed case was reached for 8 countries, which is February 20, 2020 (South Korea, other countries initiate dates could be found at Appendix B). As the epidemic curve follows the rule of rising, peaking, and then decline, however, in our observed period, all the 8 countries are in their speedy rising stages but not yet reach their peak and decline stages. Among these countries, numbers of confirmed patients in Italy are largest. After a small burst, as a result of a series of emergency prevention and control measures, the number of confirmed diagnoses in South Korea has increased slowly. Iran, Germany, Spain and the U.S. have around 20000 confirmed patients as of March 20, and there are no trends of slowdown. Iran has more mild patients, so cure rates are higher. Confirmed number in Japan is cumulative 963 as of March 20, and cured number is 191 which has the smallest number of cases of these 8 countries.

In Figure 2, the death tolls of Italy due to COVID-19 reached 4032 as of March 20 and the mortality rate of Italy reached 8.57%. Iran's early negligence also caused a very high mortality rate 40% at its early stage. Some countries also appear high mortality rates, like the United States and Japan. However, South Korea and Germany controlled the mortality rate at around 1%, which is a positive indication.

Model Predictions

When contrasting results from three prediction models, we achieved quite different results for COVID-19 development, as shown in Table 1 (As of March 20, 2020) and Table 2. For the reason that different models built on different theory and assumption, their output measurement were varied, cumulative number is for logistic model and active number were measured by SEIR models. The results disclosed the differences of three mathematical models, and further disclose the prediction differences without/with consideration of interventions.

In Scenario 1, based on logistic model, we predicted the epidemic trends of 8 countries. Table 1 and Appendix B lists the detail results of prediction and growth trajectories of 8 countries. For instance, the model placed the peak time of Italy as 50 days after its initiate date February 23 2020, with a maximum number of infected individuals of about 97442. Infection of US was unable to predict based on logistic model using reported data, because it is exponentially

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

growing which is hard to fit logistic curve, which reflected that the US is in high spreading stages of transmission.

In Scenario 2, the basic SEIR results showed that the confirmed cases will take 9-22 months (275-650 days) to reach the peak and most of the population would eventually be infected in a long period if there are no any control measures. Appendix C lists the detail results of basic SEIR model predictions of different countries. In addition, the active number of cases at the peak time will reach to around 10-20% of these countries' population, overloading the healthcare system, which is the worst scenario, as shown in Table 2.

In Scenario 3, the adjusted SEIR model results show that under strict control measures, the active cases will reach the peak in 16-32 days (from the early of April to the middle of April of 2020) after the initiate day when cases reached 100th. Appendix D shows the detail results of adjusted SEIR model predictions of 8 countries. Japan and South Korea took isolation measures when the spreading was still in the early stage of transmission, its peak value was low and the cumulative number of infected people was relatively small. Especially in Japan, our predicted peak value is less than 1560, which shows that the spreading has been well controlled in the early stage. However, the transmission in European countries and the United States are in the phase of outbreak as the date of our data collected. Adjusted SEIR model predicted that the peak value of Spain, Italy, Germany, France and other countries are between 10420 to 85750.

According to our parameter estimation methods, we disclosed the dynamics of the number of cases. The Scenario 3 estimation is based on strict quarantine assumption, the results displayed that implementing the control measures would decrease the epidemic peak significantly and bring forward the epidemic peak. It supports that the contact rate is an important factor that reflect the effects of control measures, with the formulation and implementation of extreme epidemic prevention measures to reduce the rate of contact, the epidemic size and peak would reduce. However, the epidemic still shows a long tail after peek, our study found that longer quarantine time of susceptible would reduce the long tail after peek.

Table 1. Short-term epidemic predictions of 8 countries

Countries / Models		Italy	Iran	South Korea	Germany	France	US	Spain	Japan
Logistic Model	7 days predicts (cumulative)	75406	22319	8139	32594	24250	272880	26692	1281

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(Scenario 1)									
Basic SEIR (Scenario 2)	7 days predicts (active)	22620	27840	42480	51820	71830	79120	12040	1235
Adjusted SEIR (Scenario 3)	7 days predicts (active)	70120	34040	10210	67640	24950	34810	58020	1540

Table 2. Epidemic peak predictions of 8 countries

Countries / Models		Italy	Iran	South Korea	Germany	France	US	Spain	Japan
Logistic Model (Scenario 1)	Days to Peak	50	40	25	49	45	N/A	32	65
	Peak value (cumulative)	97442	23414	8139	235902	30076	N/A	33890	1375
Basic SEIR (Scenario 2)	Days to Peak	370	275	360	275	300	340	410	650
	Peak value (*10^4, active)	790	1050	550	1750	750	620	465	615
Adjusted SEIR (Scenario 3)	Days to Peak	23	23	32	18	17	25	16	32
	Peak value (active)	71950	36240	10420	85750	36980	41850	61420	1560

Spreading Potentials

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The growth rates of the logistic curves are also listed as a spreading potential index for comparing 8 countries situations in Table 3. Our results suggest that these countries are all with high risk of virus rapid transmissions except for Japan, in which the spreading shows slowing down. On the basis of evidence from previous transmission dynamics studies, the documented COVID-19 basic reproductive number (R_0) range from 2.0 to 4.9[14-17]. Our estimated R_0 range is between 1.687~3.864 (with the assumption of 7 days mean infection period in terms of reported COVID-19 studies).

Table 3. Spreading potentials of 8 countries

Countries / Index	Italy	Iran	South Korea	Germany	France	US	Spain	Japan
growth rate	0.205	0.216	0.340	0.263	0.25	0.348	0.330	0.122
R_0(D =7)	3.094	3.465	3.663	3.878	3.269	3.472	3.864	1.687

Discussions

According to the COVID-19 data tracking, it suggested that these 8 countries have delayed the phase of preventing the epidemic and entered the outbreak phase of epidemic with the indeed community spread cases. Mortality rates analysis disclosed that undetected transmission events may have occurred in some countries. Seven of the eight countries have R0 levels above 3, which deserves our attention.

The high mortality was reported due to the main reason of undetected transmission is probably that some of the cases are asymptomatic or lacking of testing kits. Since our Scenario 2 and Scenario 3 are separately based on maximal assumption (without any interventions) and minimal assumption (with strict interventions). In Scenario 3, under the assumption that infected people are promptly tested and counted, the adjusted model suggested the peak time is about to reach after strictly implementing large-scale two rounds of 14 days quarantine. The decrease of epidemic size of Japan and South Korea showed positive evidences as shown in Table 1. However, considering that each country has different culture and healthcare situations and the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

implementation of the policies and control measures are in different levels, the actual situation of those countries should be between Scenario 2 and Scenario 3. It is worth pointing out, compared with no interventions, if governments take strict control measures to reduce the movement of population, implement prompt diagnosis and isolation, the peak time will reach and the peak size would greatly decrease to a relatively low level in about 30 days and it is best triggered in early stage of the epidemic; however, it will need to be maintained several months until a vaccine becomes available. Conversely, whenever the epidemic is not end, there is still a possibility of future outbreaks if governments loosen the interventions with people returning back to close social distances.

The results of spreading potentials show that R_0 of COVID-19 is high, which is similar with SARS viruses with R_0 values between 2.0 and 5.0[18], higher of influenza viruses H1N1 with R_0 values between 1.2 and 3.7[19], and Ebola viruses with R_0 values between 1.34 and 3.65[20]. The COVID-19 is a highly contagious human-to-human transmission disease. The R_0 is expected to decrease substantially compared to values at the early stage after the government implemented control measures, however, no matter what kind of policy each country executes, each policy has its limitation to defend COVID-19 and the sustained transmission chains will occur until there is vaccine or virus disappears with season or population immunity[21]. Therefore, detecting all transmission events are the most key issue of the COVID-19 control in current stage, because any undetected case in a local area could begin a new epidemic chain of transmission, in addition, the public should take adequate protective measures against the transmission of COVID-19.

From the view of mathematical models, the SEIR model is designed for infectious disease estimation, however, the logistic growth model is designed to fit the curve development. Logistic model may fit the existing data better when comparing with SEIR model since it is trained from the existing data, but it cannot be accurately judged and incorporate infectious characters. Therefore, we regarded that it is better for near term predictions. On the other hand, SEIR model introduces more variables and factors by considering the interaction and association of multiple groups of people, it is more reasonable than the logistic model as it follows the rules of infectious disease development, but the predict results vary greatly when considering different interventions and settings.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The study has some limitations. The mathematical models allow to quickly incorporate multiple inputs to yield prediction results. However, this process involves making assumptions about uncertain factors, for example, it is difficult to exactly determine the extend people follow the local government's quarantine policies or measures and public behaviors such like washing hands, taking masks, social distancing, etc. That may affect the actual contact rate and the subsequent development of the epidemic. The models also lack enough data to estimate quarantine proportions of a certain population. Actually, the epidemic evolvement is quite complicated and our study has only taken account of a few factors. In addition, a lack of testing kits means many cases have not been tested in some countries and without robust testing, official number of cases is incomplete. When working with incomplete data, a small error in one factor can have outsize effect.

In conclusion, our study demonstrated that reducing contact rate are key measures to control the spread of disease at early stage, and execute enough quarantine time would decrease the case scale after peek. Therefore, implementing strong containment policies during the early spreading stages of COVID-19 and flattening the peak to avoid overloading the healthcare system should be listed as the main action of these high-risk countries/regions. After the strict quarantine period, governments still need to raise the public awareness of precaution and taking self-protecting behaviors to make the epidemic under control as scattered events.

References

1 Organization, W. H. *Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020*, <https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020 (> (2020).

2 Prevention, C. o. D. C. a. *Coronavirus Disease 2019 (COVID-19)*, <https://www.cdc.gov/coronavirus/2019-ncov/prepare/transmission.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fabout%2Ftransmission.html> (

3 *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*, <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4 Hermanowicz, S. W. Forecasting the Wuhan coronavirus (2019-nCoV) epidemics using a simple (simplistic) model. *medRxiv* (2020).

5 Liu, T. *et al*. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). (2020).

6 Imai, N. *et al*. Report 2: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China. *Imperial College London* (2020).

7 Yang, Z. *et al*. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* (2020).

8 Hong, N. *et al*. Evaluating the secondary transmission pattern and epidemic prediction of the COVID-19 in metropolitan areas of China. *medRxiv* (2020).

9 Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30120-1 (2020).

10 Wu, K., Darcet, D., Wang, Q. & Sornette, D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. *arXiv preprint arXiv:2003.05681* (2020).

11 Pell, B., Kuang, Y., Viboud, C. & Chowell, G. Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics* **22**, 62-70 (2018).

12 *Countries in the world by population (2020)*, <https://www.worldometers.info/world-population/population-by-country/> (2020).

13 Organization, W. H. *Coronavirus disease 2019 (COVID-19) Situation Report*, <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200219-sitrep-30-covid-19.pdf> (2020).

14 Zhao, S. *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* (2020).

15 Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, doi:10.1016/S0140-6736(20)30260-9. (2020).

16 Shen, M., Peng, Z., Xiao, Y. & Zhang, L. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. *bioRxiv* (2020).

17 Park, S. W. *et al*. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *medRxiv* (2020).

18 Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine* (2020).

19 Boni, M. F. *et al*. Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses. *BMC Med* **7**, 43, doi:10.1186/1741-7015-7-43 (2009).

20 House, T. Epidemiological dynamics of Ebola outbreaks. *Elife* **3**, e03908 (2014).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.26.20044289; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

21 Ferguson, N. M. *et al*. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *London: Imperial College COVID-19 Response Team, March* **16** (2020).

Author Contributions

Weiguo Zhu and Yun Long are corresponding authors. Xiang Zhou, and Na Hong are co-first authors. L. S, W. Z and Y. L take responsibility for the integrity of the work as a whole, from inception to published article; X. Z, N. H and L. S were responsible for study design and conception; J. H, Y. M and H. J were responsible for data modeling and analysis; G. Shan guided the analysis; W. Z and C. L interpreted the results; X. Z, N. H, and L. S drafted the manuscript. All authors revised the manuscript for important intellectual content.

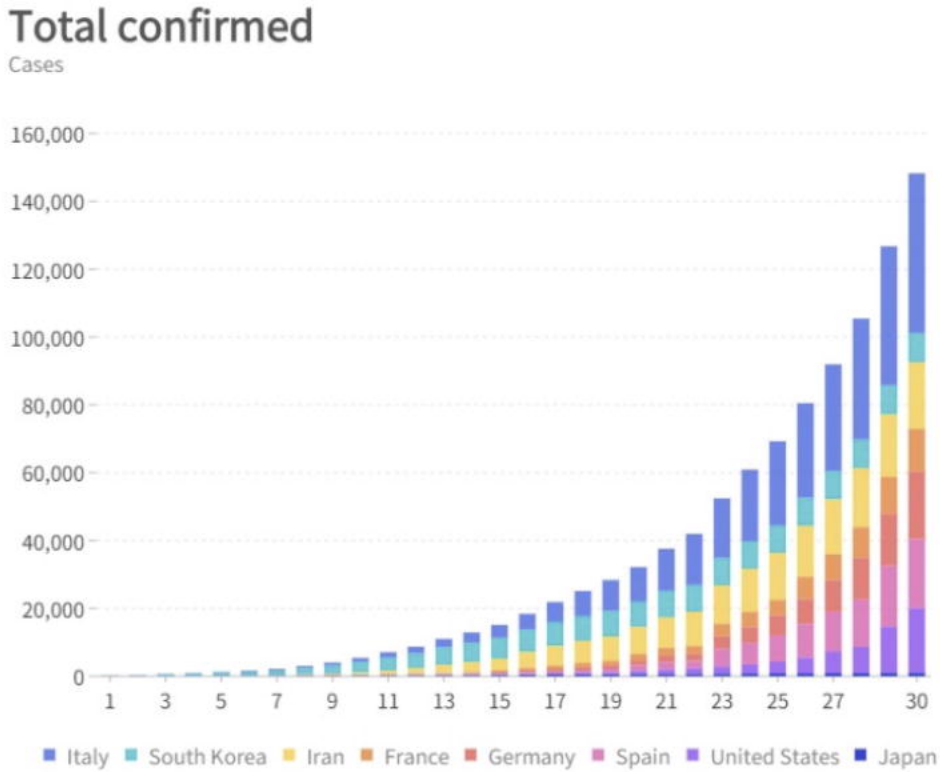
Declarations

The authors declare that there is no conflict of interest regarding the publication of this article.

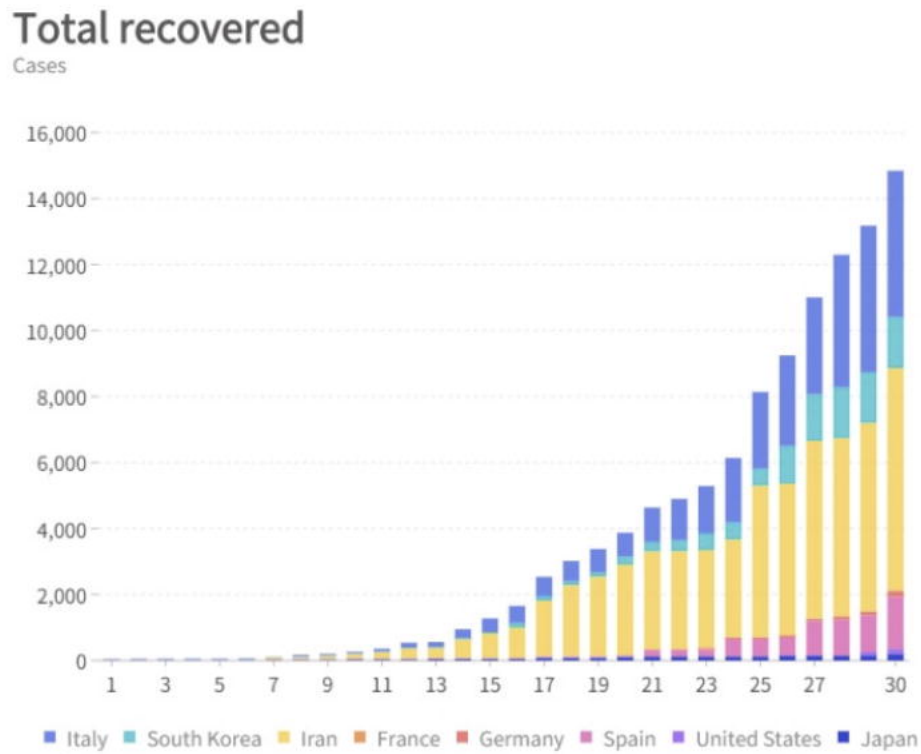
Legends

Figure 1: Epidemiological curves of 8 observed countries (February 20~March 20, 2020)

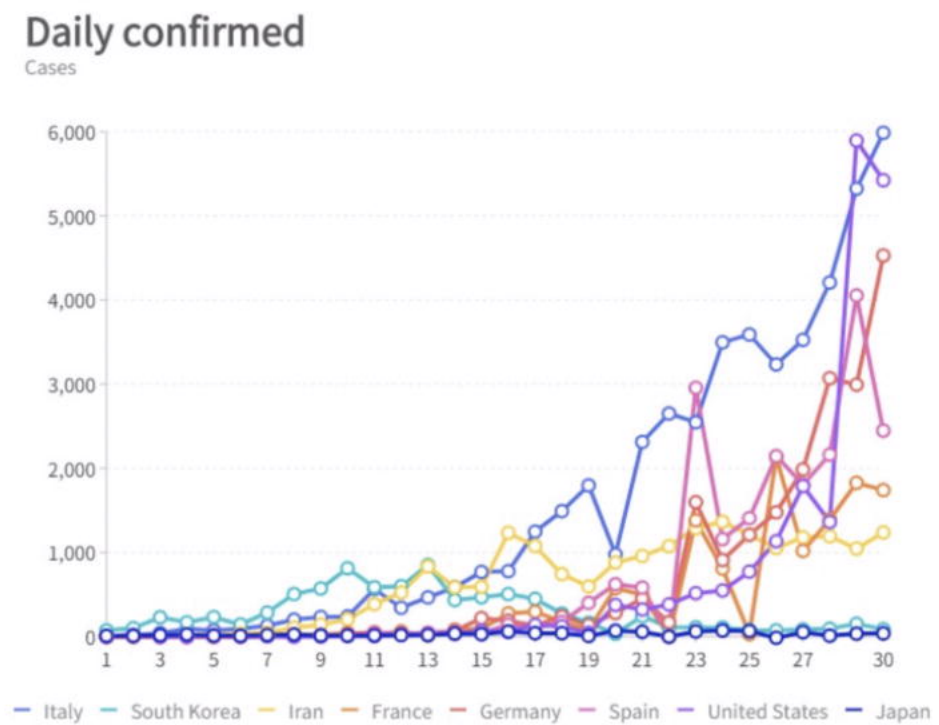
Figure 2: Mortality rate of 8 observed countries (February 20~March 20, 2020)



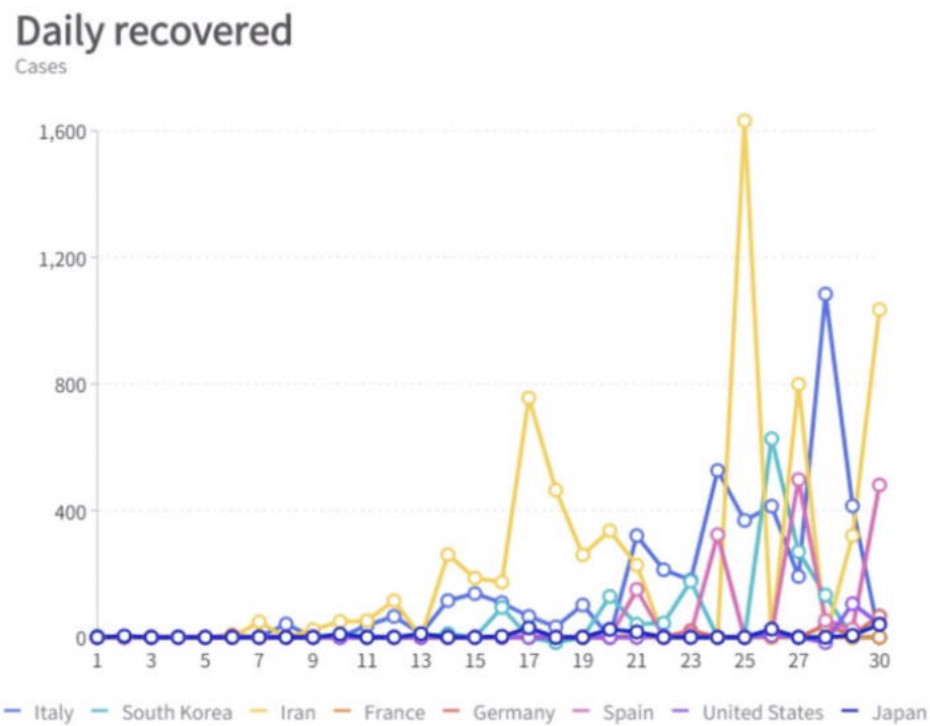
A. Cumulative confirmed cases timeline



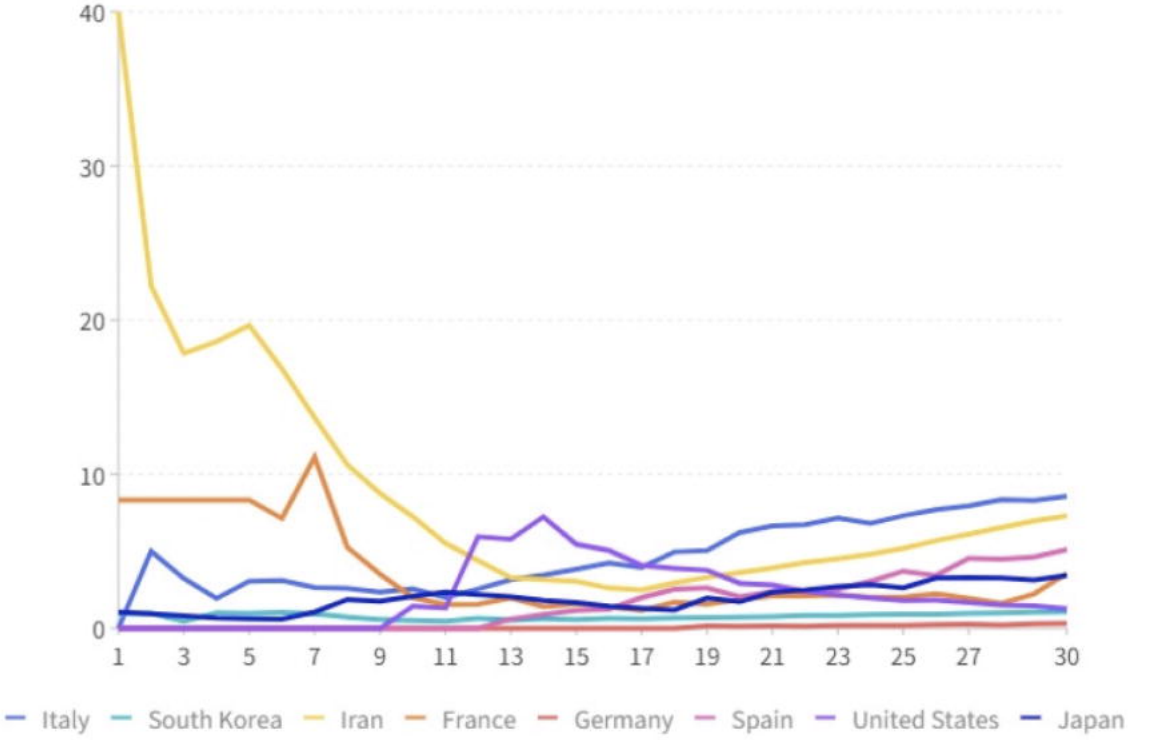
B. Cumulative confirmed cases timeline



C. Daily confirmed cases timeline



D. Daily confirmed cases timeline

Mortality rate

%