medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Predicting the number of reported and unreported cases for the COVID-19 epidemics in China, South Korea, Italy, France, Germany and United Kingdom

Zhihua Liu[(a)], Pierre Magal[(b)]*and Glenn Webb[(c)]

[(a)] *School of Mathematical Sciences, Beijing Normal University, Beijing 100875, People's Republic of China*

[(b)] *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France. CNRS, IMB, UMR 5251, F-33400 Talence, France.*

[(c)] *Mathematics Department, Vanderbilt University, Nashville, TN, USA*

April 9, 2020

1 Summary

Background The novel coronavirus (SARS-CoV-2) is currently causing concern in the medical, epidemiological and mathematical communities as the virus is rapidly spreading around the world. Internationally, there are more than 1 200 000 cases detected and confirmed in the world on April 6. The asymptomatic and mild symptomatic cases are just going to be really crucial for us to understand what is driving this epidemic to transmit rapidly. Combining a mathematical model of severe (SARS-CoV-2) transmission with data from China, South Korea, Italy, France, Germany and United Kingdom, we provide the epidemic predictions of the number of reported and unreported cases for the SARS-CoV-2 epidemics and evaluate the effectiveness of control measures for each country.

Methods We combined a mathematical model with data on cumulative confirmed cases from China, South Korea, Italy, France, Germany and United Kingdom to provide the epidemic predictions and evaluate the effectiveness of control measures. We divide infectious individuals into asymptomatic and symptomatic infectious individuals. The symptomatic infectious phase is also divided into reported (severe symptoms) and unreported (mild symptoms) cases. In fact, there exists a period for the cumulative number of reported cases to grow (approximately) exponentially in the early phase of virus transmission which is around the implementation of the national prevention and control measures. We firstly combine the date of the implementation of the measures with the daily and cumulative data of the reported confirmed cases to find the most consistent period for the cumulative number of reported cases to grow approximately exponentially with the formula $\chi_1 \exp(\chi_2 t) - \chi_3$, thus we can determine the parameters χ_1, χ_2, χ_3 in this formula and then determine the parameters and initial conditions for our model by using this formula and the plausible biological parameters for SARS-CoV-2 based on current evidence. We then provide the epidemic predictions, evaluate the effectiveness of control measures by simulations of our model.

Findings Based on the simulations using multiple groups of parameters (d_1, d_2, N), here $[d_1, d_2]$ is the consistent period for the cumulative number of reported cases to grow approximately exponentially with the formula $\chi_1 \exp(\chi_2 t) - \chi_3$ and N is the date at which public intervention measures became effective, we found that the ranges of the turning point, the final size of reported and unreported cases are respectively Feb.6 − 7, 67 000 − 69 000 and 45 000 − 46 000 for China, Feb.29−Mar.1, 9 000 − 9 400 and 2 250 − 2 350 for South Korea, Mar.24 − 26, 156 000 − 177 000, and 234 000 − 265 000 for Italy, Mar.30−Apr.9, 104 000 − 212 000, and 177 000 − 318 000 for France, Mar.30−Apr.20, 141 000 − 912 000, and 197 000 − 1 369 000 for Germany, Apr.1−May12, 140 000 − 473 000, and 210 000 − 709 000 for United

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

*Corresponding author.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Kingdom. Our prediction relies on the cumulative data of the reported confirmed cases. As more data become available, the ranges become smaller and smaller, that means the prediction becomes better and better. It is evident that our estimates and simulations have shown good correspondence with the distribution of the cumulative data available of the reported confirmed cases for each country and in particularly, the curves plotted by using different parameter groups (d_1, d_2, N) for reported and unreported cases tend to be consistent in China and South Korea (see (e) in Figures 2-3). For Italy, France, Germany and United Kingdom, the prediction can be updated to higher accuracy with on-going day by day reported case data (see Figures 4-7).

Interpretation We used the plausible biological parameters f, ν, η for SARS-CoV-2 based on current evidence which might be refined as more comprehensive data become available. Our prediction also relies on the cumulative data of the reported confirmed cases. Using multiple groups of parameters (d_1, d_2, N), we have attempted to make the best possible prediction using the available data. We found that with more cumulative data available, the curves plotted by using different parameter groups (d_1, d_2, N) for reported and unreported cases will be closer and closer, and finally tend to be consistent. This shows that when we have no enough cumulative data available, we need to use all possible parameter groups to predict the range of turning point, the final size of reported and unreported cases. When we have enough cumulative data, for example, when we get the data after the turning point, we only need to use any one of these parameter groups to get a prediction with high accuracy.

Funding NSFC (Grant No. 11871007), NSFC and CNRS (Grant No. 11811530272) and the Fundamental Research Funds for the Central Universities.

2 Introduction

As coronavirus outbreaks surge worldwide, more and more facts [8] show that many new patients which are asymptomatic or have only mild symptoms can transmit the virus. Research [9] traced COVID-19 infections which resulted from a business meeting in Germany attended by someone infected but who showed no symptoms at the time and found that four people were ultimately infected from that single contact. It has also been confirmed that asymptomatic transmission occurs in [1] which echoes the report in [9]. A German research team [11] showed that some new crown pneumonia patients had higher viral levels in the throat swabs during the early stage of the disease, that is, when the symptoms were mild. It is reported in [7] that 13 evacuees from Wuhan, China on chartered flights were infected, of whom 4, or 31%, never developed symptoms. The estimated asymptomatic proportion in [6] is at 17.9% which overlaps with a derived estimate of 31% from data of Japanese citizens evacuated from Wuhan [7]. A team in China [10] suggests that by February 18, there were 37,400 people with the virus in Wuhan whom authorities didn't know about. The asymptomatic and mild symptomatic cases were missed because authorities aren't doing enough testing, or 'preclinical cases' in which people are incubating the virus but would not be ill enough to seek medical help, would probably slip past screening methods such as temperature checks. The asymptomatic and mild symptomatic cases are just going to be really crucial for us to understand the rapid transmission.

In previous works [3, 4, 5], our team has developed differential equations models of COVID-19 epidemics. Our goal was to predict forward in time the future number of cases from early reported case data in regions throughout the world. Our models incorporate the following important elements of COVID-19 epidemics: (1) the number of asymptomatic infectious individuals (with no or very mild symptoms), (2) the number of symptomatic reported infectious individuals (with severe symptoms) and (3) the number of symptomatic unreported infectious individuals (with mild symptoms). With our model, we can show the prediction of the final size of the asymptomatic infectious, reported (with severe symptoms) and unreported cases (with mild symptoms) which is an important epidemiological problem research teams around the world are trying to solve.

In an early phase of the epidemic, the reported case data grows exponentially, which corresponds to a constant transmission rate. We assume that government measures and public awareness cause this early constant transmission rate to change to a time dependent exponentially decreasing rate. We identify the early constant transmission rate using a method developed in [3]. We then identify the time dependent exponentially decreasing transmission rate from reported case data, and project forward the time-line of the epidemic course. With this time dependent transmission rate, the effectiveness of control measures for each country could be evaluated.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Our model is applicable to COVID-19 epidemics in any region with reported case data, which can be updated to higher accuracy with on-going day by day reported case data.

3 Research in context

Evidence before this study
We searched PubMed, BioRxiv, and MedRxiv for articles published in English, Chinese, French, using the search terms "2019-nCoV", "novel coronavirus", "COVID-19", "SARS-CoV-2", "asymptomatic", "mild symptoms", AND "unreported case" with no time restrictions. We found several estimates of asymptomatic, mild symptoms, and unreported case. However, we obtained no estimations for turning point, the number of asymptomatic, the number of mild symptomatic, and the number of unreported case by combining a mathematical model with a phenomenological model and data.

Added value of this study Our study gives the estimations of the turning point, the number of asymptomatic, the number of mild symptoms, and the number unreported case by using a mathematical model we present (which includes susceptible, asymptomatic infectious, reported symptomatic infectious, unreported symptomatic infectious individuals), together with a phenomenological model and data. Our model also incorporates government and social distancing measures, through the time-dependent transmission rate. Based on the values of f, μ, N in Table 2, it is evident that the strong government measures such as isolation, quarantine, and public closings, should start as early as possible, and should be as strong as possible. After a first outbreak of SARS-CoV-2, South Korea came back to linearly growing phase now. It seems that this epidemic of SARS-CoV-2 is not easy to eradicate. If the strong government measures are reduced too early or too extensively, the epidemic may return to a new exponential growth phase and then outbreaks again.

Implications of all the available evidence We see that estimation of the number of asymptomatic infectious and unreported cases has major importance in understanding the severity of this epidemic. Strong measures are needed to curb mild and asymptomatic cases that are fueling the pandemic. Based on the study, for China and South Korea, the major distancing measures can not be reduced too early or too extensively, otherwise the epidemic may return to new exponential growth phase, for Italy, France, Germany and United Kingdom, very strong testing and isolation measures should start as early as possible, and should be as strong as possible.

4 Methods

Model
To provide the epidemic predictions, evaluate the effectiveness of control measures, we fit the following transmission dynamic model to the cumulative confirmed case data in China, South Korea, Italy, France, Germany and United Kingdom:

$$\begin{cases} S'(t) = -\tau(t)S(t)[I(t)+U(t)], \\ I'(t) = \tau(t)S(t)[I(t)+U(t)] - \nu I(t), \\ R'(t) = \nu_1 I(t) - \eta R(t), \\ U'(t) = \nu_2 I(t) - \eta U(t), \end{cases} \tag{4.1}$$

4.1

with the initial data

$$S(t_0) = S_0 > 0,\ I(t_0) = I_0 > 0,\ R(t_0) = 0 \text{ and } U(t_0) = U_0 \geq 0. \tag{4.2}$$

4.2

Here $t \geq t_0$ is time in days, t_0 is the beginning date in the model of the epidemic, $S(t)$ is the number of individuals susceptible to infection at time t, $I(t)$ is the number of asymptomatic infectious individuals at time t, $R(t)$ is the number of reported sever symptomatic infectious individuals at time t, and $U(t)$ is the number of unreported mild symptomatic infectious individuals at time t.

The transmission rate at time t is $\tau(t)$. Asymptomatic infectious individuals $I(t)$ are infectious for an average period of $1/\nu$ days. Reported symptomatic individuals $R(t)$ are infectious for an average period of $1/\eta$ days, as are unreported symptomatic individuals $U(t)$. We assume that reported symptomatic

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

infectious individuals $R(t)$ are reported and isolated immediately, and cause no further infections. The asymptomatic individuals $I(t)$ can also be viewed as having a low-level symptomatic state. All infections are acquired from either $I(t)$ or $U(t)$ individuals. f is the fraction of asymptomatic infectious that become reported symptomatic infectious. The rate asymptomatic infectious become reported symptomatic is $\nu_1 = f\,\nu$, the rate asymptomatic infectious become unreported symptomatic is $\nu_2 = (1-f)\,\nu$, where $\nu_1 + \nu_2 = \nu$. The cumulative number of reported cases at time t is given by the formula

$$CR(t) = \nu_1 \int_{t_0}^{t} I(\sigma)d\sigma, \text{ for } t \geq t_0, \tag{4.3}$$

4.3

and the cumulative number of unreported at time t is given by the formula

$$CU(t) = \nu_2 \int_{t_0}^{t} I(\sigma)d\sigma, \text{ for } t \geq t_0. \tag{4.4}$$

4.4

The interpretation of the parameters and initial conditions of the model are given in Table 1 and a flow diagram of the model is given in Figure 1.

Symbol	Interpretation	Method
t_0	Date the epidemic started	fitted
S_0	Number of susceptible at time t_0	fixed
I_0	Number of asymptomatic infectious at time t_0	fitted
U_0	Number of unreported symptomatic infectious at time t_0	fitted
$\tau(t)$	Transmission rate at time t	fitted
N	Date at which public intervention measures became effective	fitted
μ	Intensity of the public intervention measures	fitted
$1/\nu$	Average time during which asymptomatic infectious are asymptomatic	fixed
f	Fraction of asymptomatic infectious that become reported symptomatic infectious	fixed
$\nu_1 = f\,\nu$	Rate at which asymptomatic infectious become reported symptomatic	fitted
$\nu_2 = (1-f)\,\nu$	Rate at which asymptomatic infectious become unreported symptomatic	fitted
$1/\eta$	Average time symptomatic infectious have symptoms	fixed

Table1

Table 1: *Parameters and initial conditions of the model.*

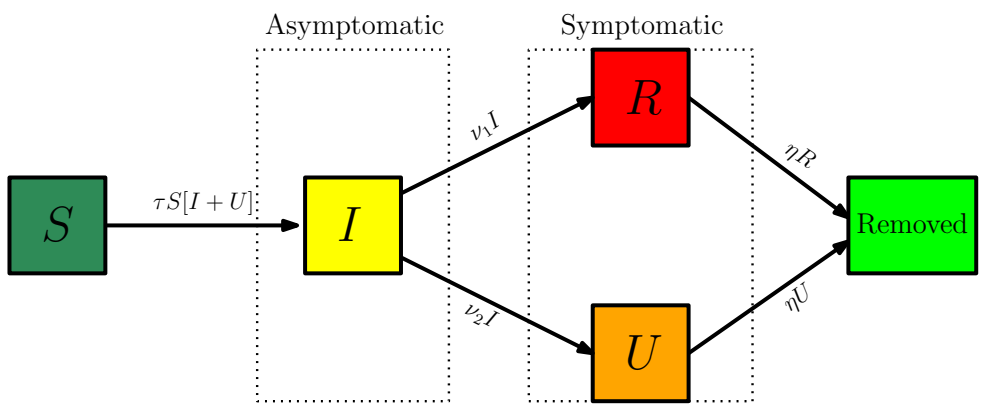

fig1

Figure 1: *Compartments and flow chart of the model.*

Method to estimate the parameters

We assume that $(100 \times f)\,\%$ of symptomatic infectious cases go reported. The actual value of f is unknown and varies from country to country. We assume $\eta = 1/7$, which means that the average period of infectiousness of both reported and unreported symptomatic infectious individuals is 7 days. We assume $\nu = 1/7$, which means that the average period of infectiousness of asymptomatic infectious individuals is 7 days. These values can be modified as further epidemiological information becomes known.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

After a period of linear growth, the epidemic of COVID-19 starts to enter into a second phase where the cumulative number of reported cases $CR(t)$ grows approximately exponentially. We assume that during this phase cumulative number of reported cases is described by the following phenomenological model:

$$CR(t) = \chi_1 \exp(\chi_2\, t) - \chi_3,\ t \geq t_0. \tag{4.5}$$

4.5

We fix the value $\chi_3 = 1$. The values of χ_1 and χ_2 are fitted to cumulative reported case data in the second phase when it is recognized that $CR(t)$ is growing exponentially (i.e. we use an exponential fit $\chi_1 \exp(\chi_2\, t)$ to fit the data $CR(t)+1$). We assume the initial value S_0, corresponds to the population of the region of the reported case data. The value of the susceptible population $S(t)$ is assumed to be only slightly changed by removal of the number of people infected in the beginning of the second phase. The following formulas for I_0, U_0, t_0, τ_0, and $\mathcal{R}_0$ were derived in [3]. Their numerical values are identified by using (4.5) from the exponential growth phase of the epidemic. The other initial conditions are

$$I_0 = \frac{\chi_2 \chi_3}{f\nu},\ \ U_0 = \left(\frac{\nu_2}{\eta + \chi_2}\right) I_0. \tag{4.6}$$

4.6

The transmission rate $\tau(t)$ is assumed to be constant when the number of reported infectious symptomatic cases starts growth exponentially fast:

$$\tau_0 = \left(\frac{\chi_2 + \nu_1 + \nu_2}{S_0}\right)\left(\frac{\eta + \chi_2}{\nu_2 + \eta + \chi_2}\right). \tag{4.7}$$

4.7

The model starting time of the epidemic is

$$t_0 = \frac{1}{\chi_2}\Big(\log(\chi_3) - \log(\chi_1)\Big). \tag{4.8}$$

4.8

The value of the basic reproductive number is

$$\mathcal{R}_0 = \left(\frac{\tau_0 S_0}{\nu_1 + \nu_2}\right)\left(1 + \frac{\nu_2}{\eta}\right). \tag{4.9}$$

4.9

When strong government measures such as isolation, quarantine, and public closings are implemented, the last phase of the epidemic begins. The actual effects of these measures are complex, and we use an exponential decrease for a time-dependent decreasing transmission rate $\tau(t)$ in the third phase to incorporate these effects. The formula for $\tau(t)$ is

$$\begin{cases} \tau(t) = \tau_0,\ 0 \leq t \leq N, \\ \tau(t) = \tau_0 \exp\left(-\mu\,(t - N)\right),\ N < t. \end{cases} \tag{4.10}$$

4.10

The date N and the value μ are chosen so that the cumulative reported cases in the numerical simulation of the epidemic aligns with the cumulative reported case data after day N when the public measures take effect. In this way we are able to project forward the time-path of the epidemic after the government imposed public restrictions take effect.

The daily number of reported cases from the model can be obtained by computing the solution of the following equation:

$$DR'(t) = \nu_1\, I(t) - DR(t),\ \text{for } t \geq t_0 \text{ and } DR(t_0) = DR_0. \tag{4.11}$$

4.11

A major challenge for the predictions from the reported cases data, is to determine the date interval for the second phase when the number of reported infectious symptomatic cases grows exponentially and the date N (at which public intervention measures became effective), which are key elements of our model, and they depend strongly on the implementation of social distancing measures. If these measures are implemented gradually, then the difficulty is increased. Usually, the measures government implemented took effect in daily reported cases after some days.

Another difficulty in applying our model is how to fix the value of the parameter f. A decreased value of f corresponds to a greater final size of the epidemic. The value of f is unknown, but information about the level of testing relates to the value of f. The increased testing can increase the value of f. Mortality can also be used as a reference to estimate the value of f. High mortality indicates high unreported ratio. In fact, from the values of f, N and μ, we can also obtain some information of the actual effects of these measures of testing, quarantining and isolation implemented by the governments in these countries.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 principle of our method is the following. By using an exponential best fit method we obtain a best fit of (4.5) to the data over a time $[d_1, d_2]$ and we derive the parameters χ_1 and χ_2. We fix $f = 0.4, 0.6$ or 0.8, $\nu = 1/7$ and $\eta = 1/7$. The values of I_0 U_0, τ_0, and t_0 are obtained by using (4.6)-(4.8). Next we fix N to some values and we obtain μ by trying to get the best to the data.

The uncertainty in our prediction is due to the fact that several sets of parameters (d_1, d_2, N) may give a good fit to the data. As a consequence, at the early stage of the epidemics (in particular before the turning point) the outcome of our method can be very different from one set of parameters (d_1, d_2, N) to another. We try to solve this uncertainty problem by using several choice of the period $[d_1, d_2]$ to fit an exponential growth of the data to determine χ_1 and χ_2 and using several choice for N (the parameter $\chi_3 = 1$ being fixed). We vary the time interval $[d_1, d_2]$ during which we use the data to obtain χ_1 and χ_2 by using an exponential fit. In the simulations below we vary the first day d_1, the last day d_2, N (date at which public intervention measures became effective) such that all possible sets of parameters (d_1, d_2, N) will be considered. For each (d_1, d_2, N) we evaluate μ to obtain the best fit of the model to the data. We use the mean absolute deviation as the distance to data to evaluate the best fit to the data. We obtain a large number of best fit depending on (d_1, d_2, N, f) and we plot the smallest mean absolute deviation $\mathrm{MAD}_{\min}$. Then we plot all the best fit with mean absolute deviation between $\mathrm{MAD}_{\min}$ and $\mathrm{MAD}_{\min} + 5$.

Remark 4.1 *The number* 5 *chosen in* $\mathrm{MAD}_{\min} + 5$ *is questionable. We use this value for all the simulations since it gives sufficiently many runs that are fitting very well the data and which gives later a sufficiently large deviation.*

5 Results

Combining a mathematical model with multiple groups of parameters (d_1, d_2, N), we project the future number of cases, both reported and unreported for China, South Korea, Italy, France, Germany and United Kingdom. Here $[d_1, d_2]$ is the period to fit an exponential growth of the data to determine χ_1 and χ_2, and N is the date at which public intervention measures became effective. For each (d_1, d_2, N) we evaluate μ to obtain the best fit of the model to the data. We use the mean absolute deviation as the distance to data to evaluate the best fit to the data. We obtain a large number of best fit depending on (d_1, d_2, N) and we plot the one with the smallest mean absolute deviation $\mathrm{MAD}_{\min}$ and all the best fit with mean absolute deviation in between $\mathrm{MAD}_{\min}$ and $\mathrm{MAD}_{\min} + 5$. Thus we could summarize the parameters giving the best fit to the data so far. In Table 2 and show the range of the turning point, the final size for both reported and unreported cases, the maximum number of the daily reported cases in Table 3 for each country.

Country	χ_1	χ_2	χ_3	t_0	μ	N	S_0	f	τ_0
China	2.05	0.28	1	Dec. 30	0.16	Jan. 28	1.40005×10^9	0.6	2.66×10^{-10}
South Korea	0.27	0.32	1	Feb. 4	0.47	Feb. 26	51.47×10^6	0.8	3.18×10^{-10}
Italy	4.78	0.19	1	Jan. 24	0.06	Mar. 8	60.48×10^6	0.4	1.93×10^{-10}
France	0.36	0.21	1	Feb. 3	0.04	Mar. 10	66.99×10^6	0.4	2.08×10^{-10}
Germany	140.64	0.24	1	Feb. 9	0.05	Mar. 12	82.79×10^6	0.4	2.28×10^{-10}
	126.95	0.25	1	Feb. 10	0.04	Mar. 11	82.79×10^6	0.8	2.64×10^{-10}
United Kingdom	40.18	0.22	1	Feb. 13	0.04	Mar. 17	66.44×10^6	0.4	2.1451×10^{-10}

Table2

Table 2: *The parameters* χ_1, χ_2, χ_3 *are estimated by using an exponential best fitting for* $\chi_1 \exp(\chi_2 t)$ *to* $CR(t) + \chi_3$ *where* $CR(t)$ *corresponds to the cumulative data for each country during: (1) Jan.* $19 - 31$ *for mainland China; (2) Feb.* $20 - 25$ *for South Korea; (3) Feb.* $29-$*Mar.* 11 *for Italy; (4) Mar.* $7 - 15$ *for France; (5) Mar.* $16 - 10$ *(when* $f = 0.4$*) and Mar.* $7 - 16$ *(when* $f = 0.8$*) for Germany; (6) Mar.* $6 - 13$ *for United Kingdom. The parameters* $\nu = 1/7$ *and* $\eta = 1/7$*. The values of* I_0 U_0*,* τ_0*, and* t_0 *are obtained by using (4.6) to (4.8).*

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Country	Range of turning point	Range of daily reported	Range of reported	Range of unreported
China	Feb. 6 − 7	3 250 − 3 450	67 000 − 69 000	45 000 − 46 000
South Korea	Feb. 29−Mar. 1	520 − 680	9 000 − 9 400	2 250 − 2 350
Italy	Mar 24 − 26	5 380 − 5 700	156 000 − 177 000	234 000 − 265 000
France	Mar. 30−Apr. 9	3 560 − 5 521	104 000 − 212 000	177 000 − 318 000
Germany ($f = 0.4$)	Mar. 30−Apr. 20	4 800 − 21 120	141 000 − 912 000	197 000 − 1 369 000
Germany ($f = 0.8$)	Mar. 29−Apr. 16	4 490 − 16 510	124 000 − 674 000	31 000 − 169 000
United Kingdom	Apr. 1−May 12	3 000 − 38 100	140 000 − 473 000	210 000 − 709 000

Table3

Table 3: *The range of the turning point, the maximum value of the daily reported number of cases, the final size of cumulative reported and unreported cases, for China, South Korea, Italy, France, Germany, United Kingdom.*

6 Discussion

In the case of China and South Korea, according to our model, the peak of the epidemic occurred approximately on February 6 and February 29, respectively from Figures 2-3 and the daily number of cases reaches a maximum of approximately 3500 and 700 cases respectively, near the turning point. We see that our model agrees very well the data for China and South Korea. Compared to China and South Korea, the public interventions in Italy, France, Germany and United Kingdom were relatively late. The peak of the epidemic occurs in Italy around March 24, and the peak of the maximum daily number of cases in our simulation is approximately 5 500, which agrees well with the daily reported cases data for Italy. For France, Germany and United Kingdom, the number of daily reported cases are still rising. Our simulations captures these increasing values, but the advance later for both countries requires more data available.

Based on our estimated results in Table 3 and all the Figures, we found that the curves plotted by using all possible groups of parameters (d_1, d_2, N) for cumulative reported and unreported case and the daily reported case in China and South Korea become closer and closer, and finally tend to be consistent as more and more data of cumulative reported case is used (see Figures 2-3). For Italy, we can see from Figure 4 that these curves are very close to each other since we have the data after the turning point. But for France, Germany and United Kingdom, these curves could not tend to be consistent with the data available now. This shows that when we have enough cumulative data available, for example, when we get the data after the turning point, we could use only one of these parameter groups to get the prediction with high accuracy. When we have a few cumulative data available, we need to use all possible parameter groups to predict a range for the turning point, the final size of cumulative reported and unreported case and the daily reported case. It is evident that we used plausible biological parameters f, ν, η for SARS-CoV-2 based on current evidence which might be refined as more comprehensive data become available. Our prediction also relies on the cumulative reported data. The prediction becomes better and better as more data become available. Using multiple groups of parameters (d_1, d_2, N), we have attempted to make the best possible prediction using the available data. As more data for particularly Italy, France, Germany and United Kingdom become available, it will be possible to refine these estimates.

Our model incorporates social distancing measures through the time dependent transmission rate $\tau(t)$. It is evident that these measures should start as early as possible, and should be as strong as possible. The consequences of late public interventions may have severe consequences for the epidemic outcome. The example of South Korea shows that a background level of daily cases may persist for an extended time which also means that South Korea came back to linearly growth phase. If the strong measures are reduced too early or too extensively, the epidemic may return to new exponential growth phase.

There are many crucial epidemiological problems which research teams are racing to understand, for example, how to estimate the last day for COVID-19 outbreak and the proportion of people with mild or no symptoms who could be spreading the pathogen. Recently we estimated the last day for COVID-19 outbreak in mainland China and present the probability distribution of the extinction date of the epidemics combining our model (4.1) with the stochastic simulations. In fact, with our model (4.1), we could also predict the proportion of asymptomatic or mild symptomatic infectious which we will focus on in the future work.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

7 Supplementary

Predicting the number of cases for mainland China

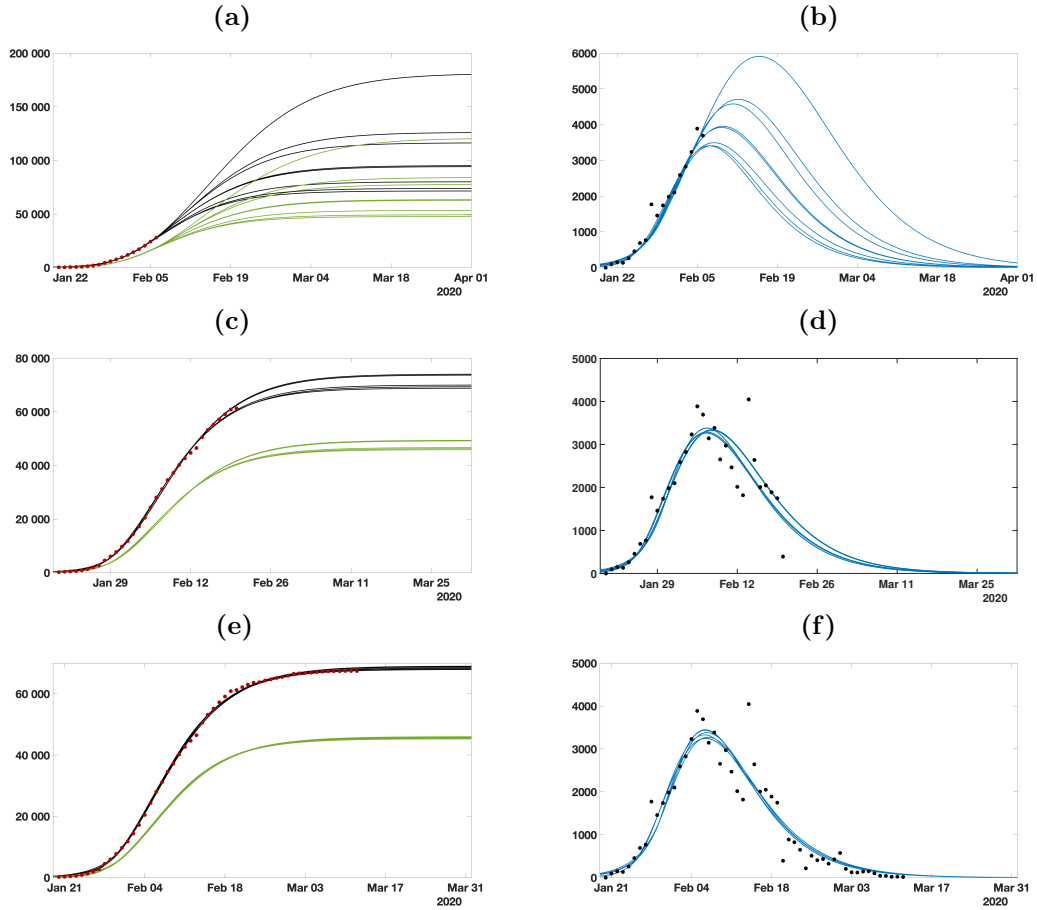

fig2

Figure 2: *On the left hand-side we plot the cumulative data (reds dots) and the best fits of the model $CR(t)$ (black curve) and $CU(t)$ (green curve). On the right hand-side we plot the daily data (black dots) with $DR(t)$ (blue curve). Figures (a) and (b) correspond the best fits of the model to the data for mainland China until February 6. Figures (c) and (d) correspond the best fits of the model to the data for mainland China until February 20. Figures (e) and (f) correspond the best fits of the model to the data for mainland China until March 12. In other words, in Figure (e) we use all the data, in Figure (c) we use the data excepted the 3 last weeks and in Figure (a) we use the data excepted the 5 last weeks. The parameters used are listed in Table 2 together with $f = 0.6$, $\nu = 1/7$ and $\eta = 1/7$.*

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Predicting the number of cases for South Korea

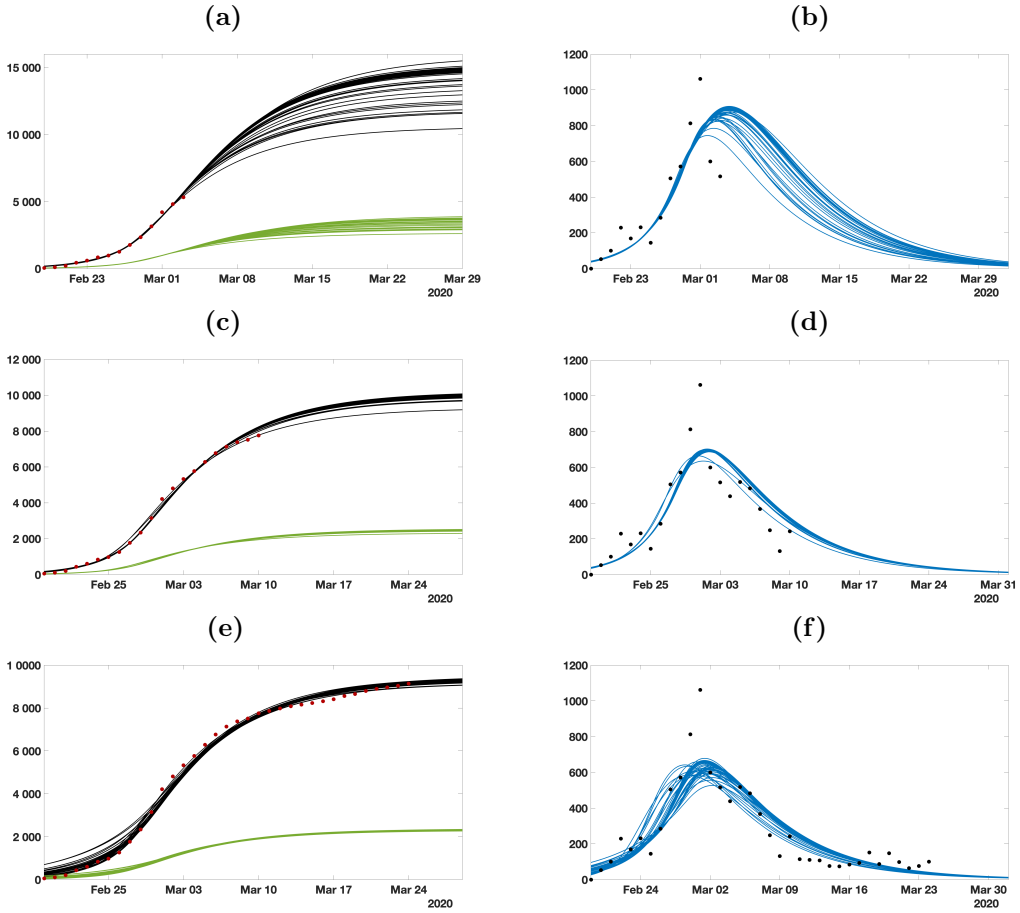

fig3

Figure 3: *On the left hand-side we plot the cumulative data (reds dots) and the best fits of the model $CR(t)$ (black curve) and $CU(t)$ (green curve). On the right hand-side we plot the daily data (black dots) with $DR(t)$ (blue curve). Figures (a) and (b) correspond the best fits of the model to the data for South Korea until March 3. Figures (c) and (d) correspond the best fits of the model to the data for South Korea until March 10. Figures (e) and (f) correspond the best fits of the model to the data for South Korea until March 24. In other words, in Figure (e) we use all the data, in Figure (c) we use the data excepted the 2 last weeks and in Figure (a) we use the data excepted the 3 last weeks. The parameters used are listed in Table 2 together with $f = 0.6$, $\nu = 1/7$ and $\eta = 1/7$.*

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Predicting the number of cases for Italy

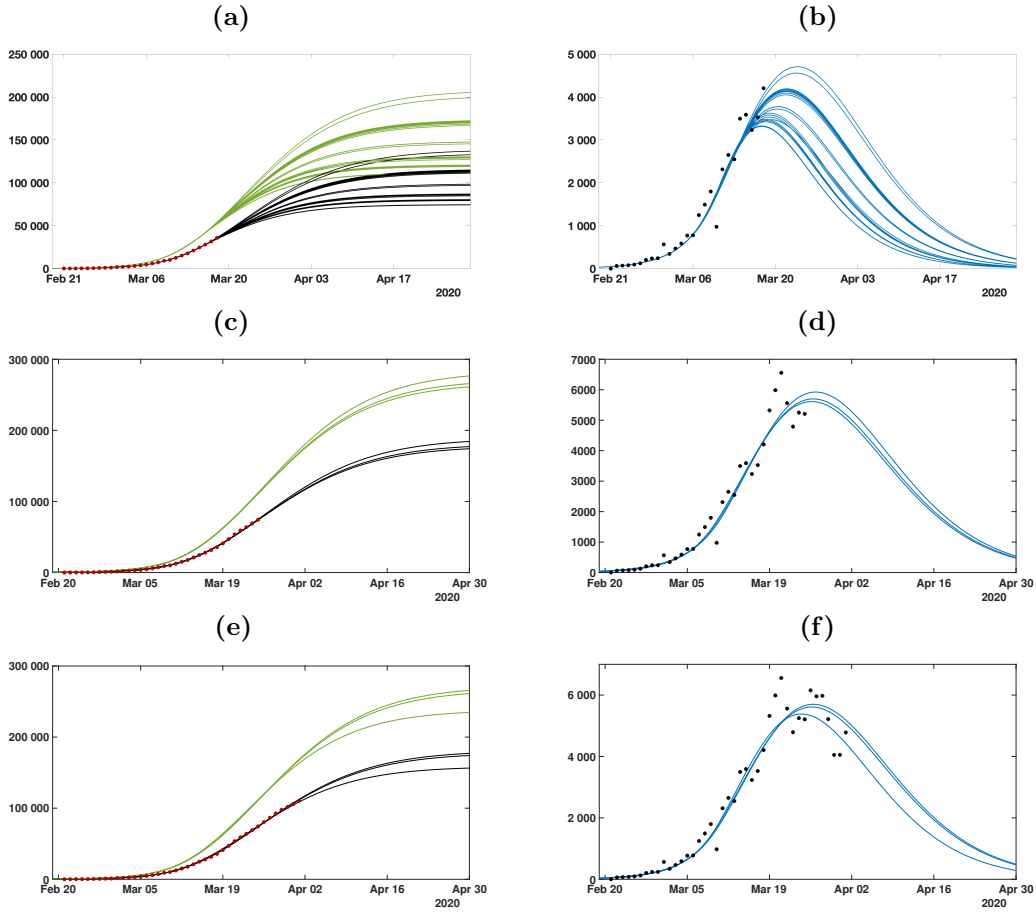

fig4

Figure 4: *On the left hand-side we plot the cumulative data (reds dots) and the best fits of the model $CR(t)$ (black curve) and $CU(t)$ (green curve). On the right hand-side we plot the daily data (black dots) with $DR(t)$ (blue curve). Figures (a) and (b) correspond the best fits of the model to the data for Italy until March 18. Figures (c) and (d) correspond the best fits of the model to the data for Italy until March 25. Figures (e) and (f) correspond the best fits of the model to the data for Italy until April 1. In other words, in Figure (e) we use all the data, in Figure (c) we use the data excepted the last weeks and in Figure (a) we use the data excepted the 2 last weeks. The parameters used are listed in Table 2 together with $f = 0.6$, $\nu = 1/7$ and $\eta = 1/7$.*

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Predicting the number of cases for France

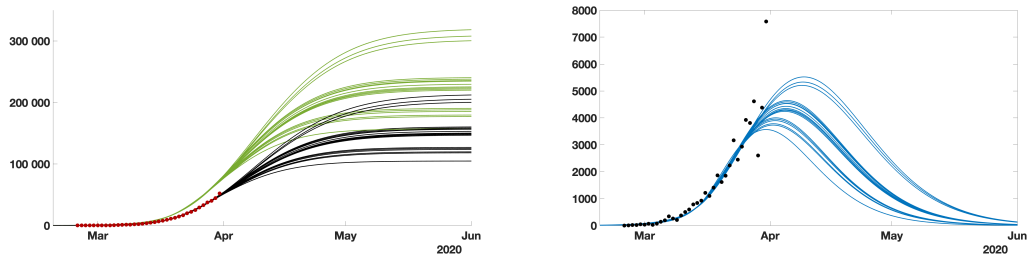

fig5

Figure 5: *On the left hand-side we plot the cumulative data (reds dots) for France until April 1 and the best fits of the model $CR(t)$ (black curve) and $CU(t)$ (green curve). On the right hand-side we plot the daily data (black dots) with $DR(t)$ (blue curve).*

Predicting the number of cases for Germany

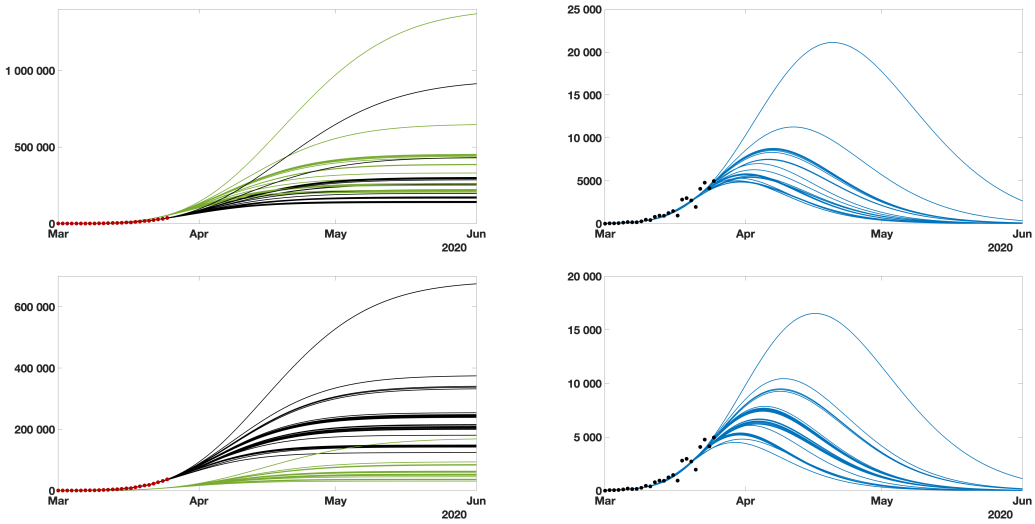

fig6

Figure 6: *We use $f = 0.4$ (top) and $f = 0.8$ (bottom). On the left hand-side we plot the cumulative data (reds dots) for Germany until April 2 and the best fits of the model $CR(t)$ (black curve) and $CU(t)$ (green curve). On the right hand-side we plot the daily data (black dots) with $DR(t)$ (blue curve).*

Predicting the number of cases for United Kingdom

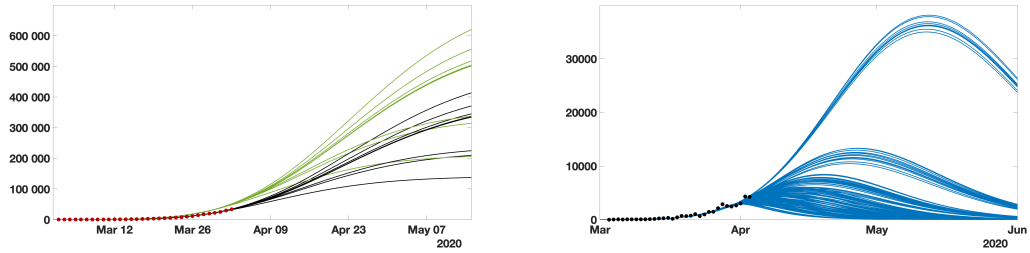

fig7

Figure 7: *On the left hand-side we plot the cumulative data (reds dots) for United Kingdom until April 2 and the best fits of the model $CR(t)$ (black curve) and $CU(t)$ (green curve). On the right hand-side we plot the daily data (black dots) with $DR(t)$ (blue curve).*

Declaration of interests

We declare no competing interests.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.09.20058974; this version posted April 10, 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 .

Keywords: corona virus, reported and unreported cases, isolation, quarantine, public closings; epidemic mathematical model

References

Guan [1] W. Guan, et al., Clinical Characteristics of Coronavirus Disease 2019 in China, *New England Journal of Medicine*, (2020). Published on February 28, 2020, PMID: 32109013. https://doi.org/10.1056/NEJMoa2002032.

Kucharski [2] A. Kucharski, T.W. Russell and C. Diamond, *et al.*, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet*, March 11, 2020. Published:March 11, 2020DOI: https://doi.org/10.1016/S1473-3099(20)30144-4

LMSW1 [3] Z. Liu, P. Magal, O. Seydi and G. Webb, Understanding unreported cases in the 2019-nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions, *MPDI Biology*, **9(3), 50** (2020). https://doi.org/10.3390/biology9030050

LMSW2 [4] Z. Liu, P. Magal, O. Seydi and G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data, *Mathematical Biosciences and Engineering* **17(4)** (2020), 3040-3051. https://doi.org/10.3934/mbe.2020172

LMSW3 [5] Z. Liu, P. Magal, O. Seydi and G. Webb, A COVID-19 epidemic model with latency period, *Infectious Disease Modelling (to appear)*

Chowell [6] K. Mizumoto, K. Kagaya, A. Zarebski and G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25(10)** (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180

Nishiura [7] H. Nishiura,et al., Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19), *International Journal of Infectious Diseases*, (2020). Published:March 13, https://doi.org/10.1016/j.ijid.2020.03.020.

Qiu [8] J. Qiu, Covert coronavirus infections could be seeding new outbreaks, *Nature*, (2020). https://www.nature.com/articles/d41586-020-00822-x

Rothe [9] C. Rothe, et al., Transmission of 2019-nCoV infection from an asymptomatic contact in Germany, *New England Journal of Medicine*, (2020). https://doi.org/10.1056/NEJMc2001468

Wang [10] C. Wang, et al., Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China, *medRxiv*. https://doi.org/10.1101/2020.03.03.20030593

Woelfel [11] R. Wölfel, et al., Virological assessment of hospitalized patients with COVID-2019, *Nature*, (2020). https://doi.org/10.1038/s41586-020-2196-x

Verity [12] R. Verity, L.C. Odell, I. Dorigatti, *et al.*, Estimates of the severity of COVID -19 disease, *medRxiv*, 2020. https://doi.org/10.1101/2020.03.09.20033357

Zou [13] L. Zou, et al., SARS-CoV-2 viral load in upper respiratory specimens of infected patients, *New England Journal of Medicine*, **382**(2020), 1177-1179. https://doi.org/10.1056/NEJMc2001737