A Model to Predict COVID-19 Epidemics with Applications to South Korea, Italy, and Spain

By Zhihua Liu, Pierre Magal, Ousmane Seydi, and Glenn Webb

Our team has developed several differential equations models of COVID-19 epidemics [1-3] that use early reported case data from around the world to predict the future number of cases. These models incorporate three important elements of COVID-19: (1) the number of asymptomatic infectious individuals (with very mild or no symptoms), (2) the number of symptomatic reported infectious individuals (with severe symptoms), and (3) the number of symptomatic unreported infectious individuals (with less severe symptoms). They also decompose COVID-19 epidemics into three phases:

- Phase I, during which *the number of cumulative reported cases increases linearly each day*
- Phase II, *during which the number of cumulative reported cases increases exponentially each day*
- Phase III, *during which the number of daily reported cases decreases each day*.

The transitions between phases are generally difficult to determine, but one can estimate them from reported cases data as time progresses.

Country	χ_1	χ_2	t_0	t_1	μ	N	S_0	f	τ_0	$\mathcal{R}_0$
South Korea	0.758	0.287	Feb. 1	Feb. 22	0.6	Feb 27	51,700,000	0.8	1.05×10^{-8}	4.01
Italy	63.7	0.135	Feb. 10	Mar. 12	0.095	Mar. 16	60,500,000	0.4	4.16×10^{-9}	2.57
Spain	433.3	0.194	Feb. 3	Mar. 13	0.125	Mar. 20	46,700,000	0.4	7.11×10^{-9}	3.39

Figure 2. *We obtain the parameters χ_1, χ_2 by fitting $\chi_1 \exp(\chi_2 t) - 1.0$ to the cumulative reported cases data between the dates $[t_1, t_2]$ for each country: (1) t_1 = February 22 to t_2 = March 1 for South Korea, (2) t_1 = March 12 to t_2 = March 21 for Italy, and (3) t_1 = March 13 to t_2 = March 21 for Spain. The values I_0, U_0, τ_0, t_0, and $\mathcal{R}_0$ are obtained via equations (3)-(6). The parameters $\nu = 1/6$, $\eta = 1/7$, $\alpha = 1/1$, $\chi_3 = 1.0$, and $R_0 = 1.0$ for all three countries.*

The rates between phases are generally difficult to determine, but one can estimate them from reported cases data as time progresses.

Our model consists of the following differential equations and initial conditions:

$$\begin{cases} S'(t) = -\tau(t)S(t)[I(t)+U(t)], \\ \quad S(t_0) = S_0, \\ E'(t) = \tau(t)S(t)[I(t)+U(t)] - \alpha E(t), \\ \quad E(t_0) = E_0 \\ I'(t) = \alpha E(t) - \nu I(t), \quad I(t_0) = I_0, \\ R'(t) = \nu_1 I(t) - \eta R(t), \quad R(t_0) = R_0, \\ U'(t) = \nu_2 I(t) - \eta U(t), \quad U(t_0) = U_0. \end{cases} \tag{1}$$

Here, $t \geq t_0$ is time in days, t_0 is the beginning date of the epidemic, $S(t)$ is the number of individuals susceptible to infection at time t, $E(t)$ is the number of asymptomatic noninfectious (exposed or latent infected) individuals at time t, $I(t)$ is the number of asymptomatic but infectious individuals at time t, $R(t)$ is the number of reported symptomatic infectious individuals at time t, and $U(t)$ is the number of unreported symptomatic infectious individuals at time t.

The time-dependent transmission rate parameter is $\tau(t)$. Newly-infected noninfectious asymptomatic individuals $E(t)$ incubate for an average period of $1/\alpha$ days. Asymptomatic infectious individuals $I(t)$ are infectious for an average period of $1/\nu$ days. Reported symptomatic infectious individuals $R(t)$ are infectious for an average period of $1/\eta$ days, as are unreported symptomatic infectious individuals $U(t)$. We assume that reported symptomatic infectious individuals $R(t)$ are reported and isolated immediately, and cause no further infections. One can also view the asymptomatic individuals $I(t)$ as having a low-level symptomatic state. All infections are acquired from either $I(t)$ or $U(t)$ infectious individuals. The fraction f of asymptomatic infectious cases becomes reported symptomatic infectious, and the fraction $1-f$ becomes unreported symptomatic infectious. The rate at which asymptomatic infectious cases become reported symptomatic is $\nu_1 = f\nu$ and the rate at which asymptomatic infectious cases become unreported symptomatic is $\nu_2 = (1-f)\nu$, where $\nu_1 + \nu_2 = \nu$.

The cumulative number of reported cases $CR(t)$ at time t is

$$CR(t) = \nu_1 \int_{t_0}^{t} I(\sigma)d\sigma, \quad t \geq t_0,$$

the cumulative number of unreported cases $CR(t)$ at time t is

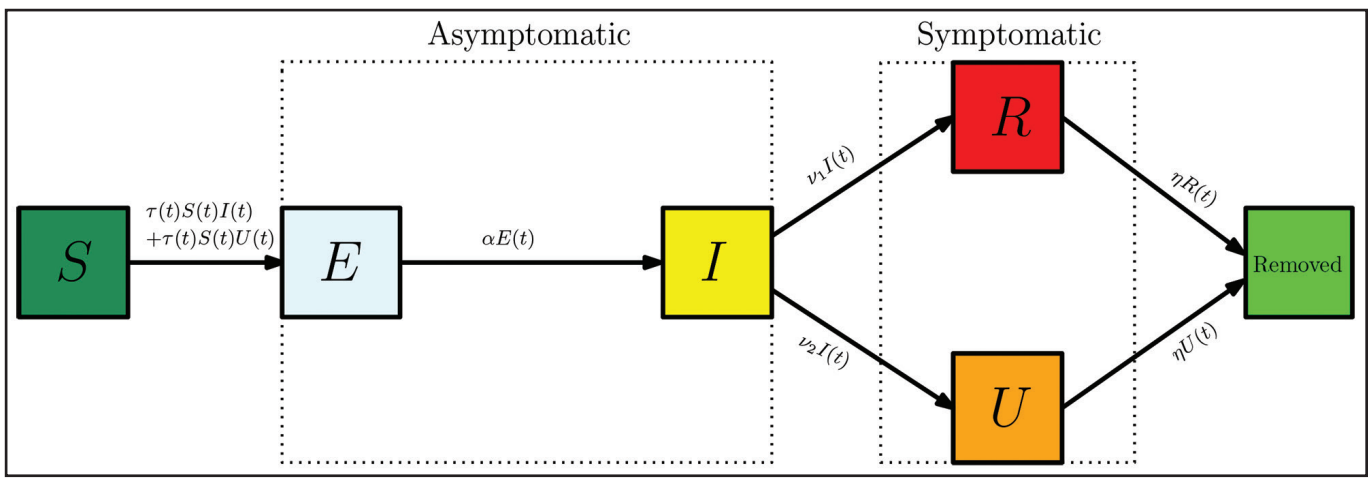

Figure 1. *Compartments and flow chart of the model.*

$$CU(t) = \nu_2 \int_{t_0}^{t} I(s)ds, \quad t \geq t_0,$$

and the daily number of reported cases $DR(t)$ at time t is

$$DR'(t) = \nu_1 I(t) - DR(t), \quad t \geq t_0, \quad DR(t_0) = DR_0.$$

Figure 1 depicts a flow diagram of the model.

Parameters

The fraction f of total reported symptomatic infectious cases is unknown and varies from region to region. We assume that $\eta = 1/7$, which means that the average period of infectiousness of both unreported and reported symptomatic infectious individuals is seven days. We also assume that $\nu = 1/6$, which means that the average period of infectiousness of asymptomatic infectious individuals is six days. Finally, we assume that $\alpha = 1$, which means that the average period of exposed individuals is one day. We can modify these values as further epidemiological information becomes known; at present, they are consistent with accepted values.

A COVID-19 epidemic transitions from phase I to phase II at time $t_1 > t_0$. Before t_1, the cumulative number of reported cases increases linearly each day. After t_1, the cumulative number of reported cases increases exponentially each day. We estimate the value of t_1 from data pertaining to the cumulative reported cases. We then fit an exponentially growing curve $CR(t)$ to the cumulative reported cases data in an estimated time interval $[t_1, t_2]$ according to the following formula:

$$CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3, \quad t_1 \leq t \leq t_2. \tag{2}$$

We typically set $\chi_3 = 1$ but allow for other values. The initial value S_0 corresponds to the population of the reported case data's region. The other initial conditions are

$$I_0 = \frac{\chi_2 \chi_3}{f(\nu_1 + \nu_2)}, \quad E_0 = \frac{\chi_2 + \nu}{\alpha} I_0, \quad U_0 = \frac{\nu_2}{\chi_2 + \eta} I_0. \tag{3}$$

Furthermore, the value of t_0 (when $R(t_0) = CR(t_0) = 0$) for starting time t_0 of the epidemic is given by

$$CR(t_0) = 0 \Leftrightarrow \chi_1 \exp(\chi_2 t_0) - \chi_3 = 0 \Rightarrow t_0 = \frac{1}{\chi_2}\left(\ln(\chi_3) - \ln(\chi_1)\right). \tag{4}$$

Additionally,

$$\tau_0 = \frac{(\chi_2 + \alpha)E_0}{S_0[I_0 + U_0]} = \frac{(\chi_2 + \nu)(\chi_2 + \alpha)(\chi_2 + \eta)}{\alpha S_0(\chi_2 + \eta + \nu_2)}, \tag{5}$$

and the basic reproductive number is given by

$$R_0 = \frac{(\chi_2 + \nu)(\chi_2 + \alpha)(\chi_2 + \eta)}{\nu\alpha(\chi_2 + \eta + \nu_2)} \left(1 + \frac{(1-f)\nu}{\eta}\right). \tag{6}$$

We derive these formulas for I_0, E_0, U_0, t_0, τ_0, and $\mathcal{R}_0$ in [1]; their values connect the phase II reported cases data to the parameterisation and initialisation of our differential equations model.

During phase II of the epidemic, $\tau(t) \equiv \tau_0$ is constant. When strong government measures such as isolation, quarantine, and public closings are implemented, phase III begins. The timing of the implementation of these measures—and their subsequent impact on disease transmission—is complex. We use an exponentially decreasing time-dependent transmission rate $\tau(t)$ in phase III to incorporate these effects. The formula for $\tau(t)$, which has phase III beginning on day N, is

$$\begin{cases} \tau(t) = \tau_0, \ 0 \leq t \leq N, \\ \tau(t) = \tau_0 \exp\left(-\mu(t-N)\right), \ N < t. \end{cases} \tag{7}$$

We choose the date N and intensity μ of the public measures so that the cumulative reported cases in the epidemic's numerical simulation align with the cumulative reported case data at an identified date after day N. In this way, we can project forward the time path of the epidemic after the government-imposed public measures take effect.

Applications

We apply our model to the COVID-19 epidemics in South Korea,[1] Italy,[2] and Spain.[3] Figure 2 provides the parameters for these three countries.

COVID-19 Epidemic in South Korea: We divide the epidemic in South Korea into four stages (see Figure 3):

(1) Before February 22: Phase I.

(2) February 22 to March 1: Phase II.

(3) March 2 to March 8: Phase III. The South Korean government implemented extensive testing, isolation, contact tracing of confirmed cases, and quarantine policies after February 20, which took effect in daily reports after March 2.

[1] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_South_Korea

[2] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_Italy

[3] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_Spain

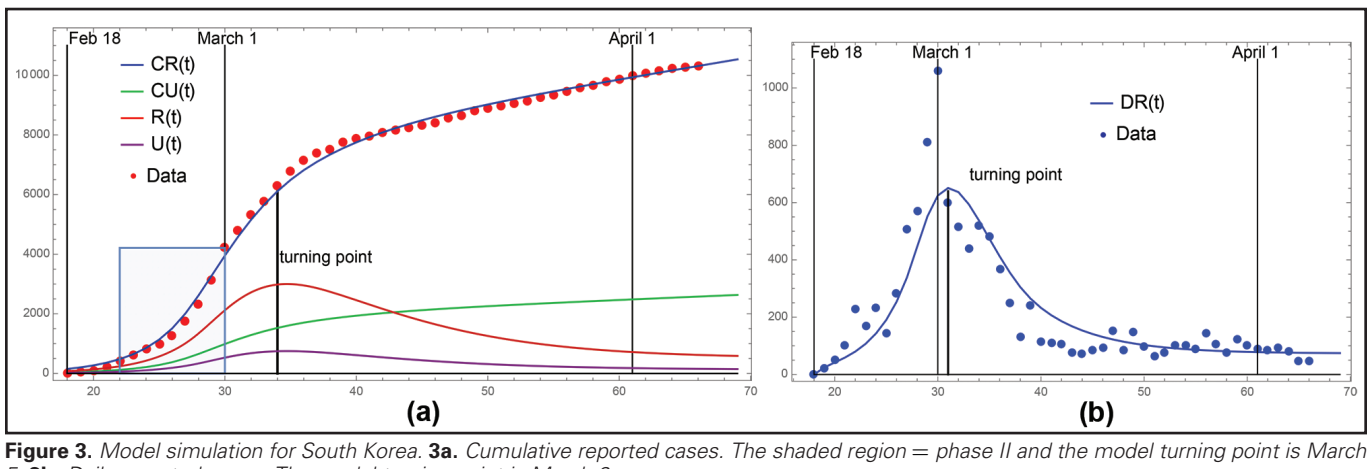

Figure 3. *Model simulation for South Korea.* **3a.** *Cumulative reported cases. The shaded region = phase II and the model turning point is March 5.* **3b.** *Daily reported cases. The model turning point is March 2.*

(4) After March 8: The daily reported cases remained approximately the same each day, and the cumulative reported cases increased linearly. This stage corresponds to a new phase I, with a low-level background generation of reported cases each day.

To account for this new phase, we modify model (1) by replacing $\tau(t)$ with a novel transmission function $\tau(t, S(t), I(t), U(t))$ that depends on t, $S(t)$, $I(t)$, $U(t)$, as follows:

$$\begin{cases} \tau(t,S(t),I(t),U(t)) = \tau_0, \ \ t_0 \le t \le 27; \\ \tau(t,S(t),I(t),U(t)) = \tau_0 \exp \\ \quad \left(-0.6\left(t-27\right)\right), \ \ 27 < t \le 37; \\ \tau(t,S(t),I(t),U(t)) = 23.0\tau_0 \exp \\ \quad \left(-0.6\left(37-27\right)\right) \\ \quad \left(\dfrac{S(37)[(I(37)+U(37)]}{(S(t)[(I(t)+U(t)]}\right), \ \ 37 < t. \end{cases} \tag{8}$$

We select the value 23.0 to match the slope of the linear increasing cumulative reported cases data after day 37. The equations and initial values remain the same, except for this novel τ function. The formulas in (8) connect the new phase I to the transmission rate in the model equations and the model outputs of $E(t)$, $I(t)$, $U(t)$, $R(t)$, $CU(t)$, $CR(t)$, $DR(t)$. One can apply the form of (8) to other examples that transition from phase III to a new phase I, corresponding to a linearly-increasing growth rate of cumulative reported cases. This new phase I can further transition to yet another phase I with a slower linearly increasing growth rate.

COVID-19 Epidemic in Italy: We divide the epidemic in Italy into three stages (see Figure 4):

(1) Before March 12: Phase I.

(2) March 12 to March 21: Phase II.

(3) After March 24: Phase III. Beginning March 1, the Italian government implemented extensive public regional lockdown measures, which were extended to all of Italy on March 10. These measures began reducing the number of reported daily cases approximately two weeks later.

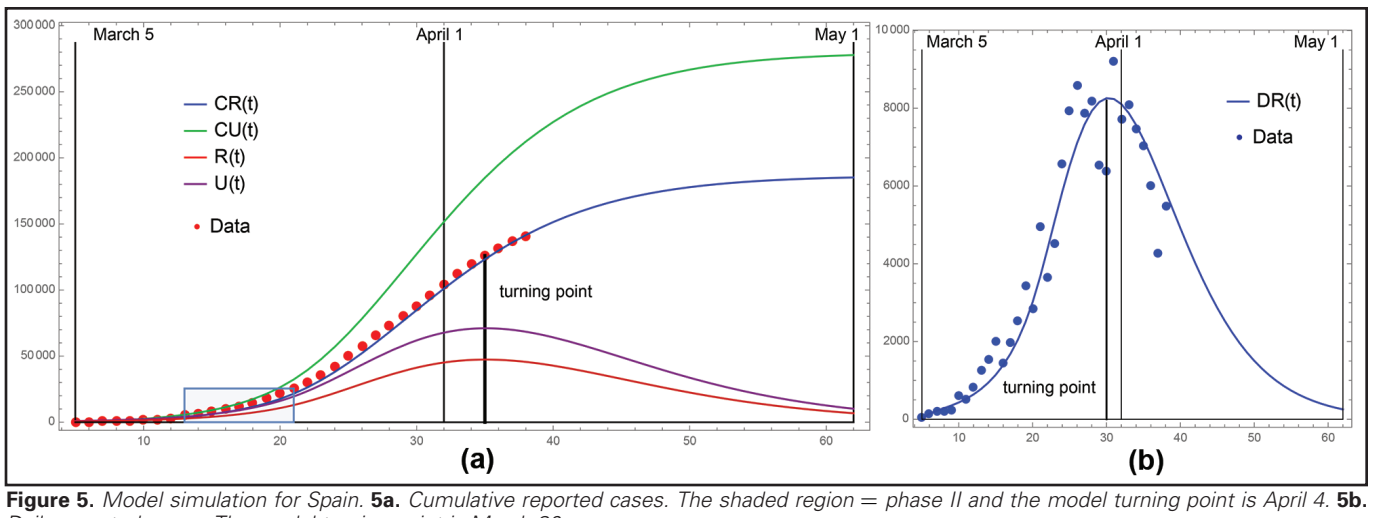

Figure 5. *Model simulation for Spain.* **5a.** *Cumulative reported cases. The shaded region = phase II and the model turning point is April 4.* **5b.** *Daily reported cases. The model turning point is March 30.*

COVID-19 Epidemic in Spain: We divide the epidemic in Spain into three stages (see Figure 5):

(1) Before March 13: Phase I.

(2) March 13 to March 21: Phase II.

(3) After March 28: Phase III. The Spanish government implemented partial shutdown measures on March 13 and imposed a general state of alarm for all of Spain on March 14. These measures began reducing the number of reported daily cases approximately two weeks later.

Concluding Thoughts

We have applied a new method [1-3] to predict a COVID-19 epidemic's evolution in a particular geographical region, based on reported case data from that region. Our model focuses on unreported cases, asymptomatic infectious cases, and the epidemic evolution's division through a succession of phases. Our method can be predictive when the epidemic is growing exponentially in phase II. Specifically, we demonstrate a technique to identify the exponentially increasing rate of cumulative reported cases in phase II [3]. When public measures to ameliorate the epidemic begin in phase II, we model these measures with a time-dependent exponentially decreasing transmission rate. These mitigations result in phase III: a subsequent reduction in daily reported cases. We determine the transition from phase II to phase III—which may require more than a week—in the model simulations.

The epidemic has attenuated in South Korea because of major measures that encourage social distancing. These measures involve surveillance, extensive testing, and isolation and contact tracing for reported and suspected cases. However, the cumulative number of reported cases in South Korea has not flattened; instead, it is growing linearly at a low rate. The epidemics in Italy and Spain have evidently passed the turning point, according to data about the daily reported cases. The cumulative reported cases may not flatten but instead continue growing linearly at a low rate, as in South Korea.

Our model incorporates government and social distancing measures through the time-dependent transmission rate τ. These measures should begin as early as possible and be as strong as possible. If such efforts cause the epidemic to substantially subside, South Korea indicates that a background level of daily cases may persist for an extended time. If countries reduce major distancing measures too early or too extensively, the epidemic can enter a new phase II and see another exponential increase in cumulative cases. Control of COVID-19 epidemics is possible, as evidenced by the situation in South Korea. The future of COVID-19 and its human toll is currently uncertain, and we hope that mathematical models will be of use.

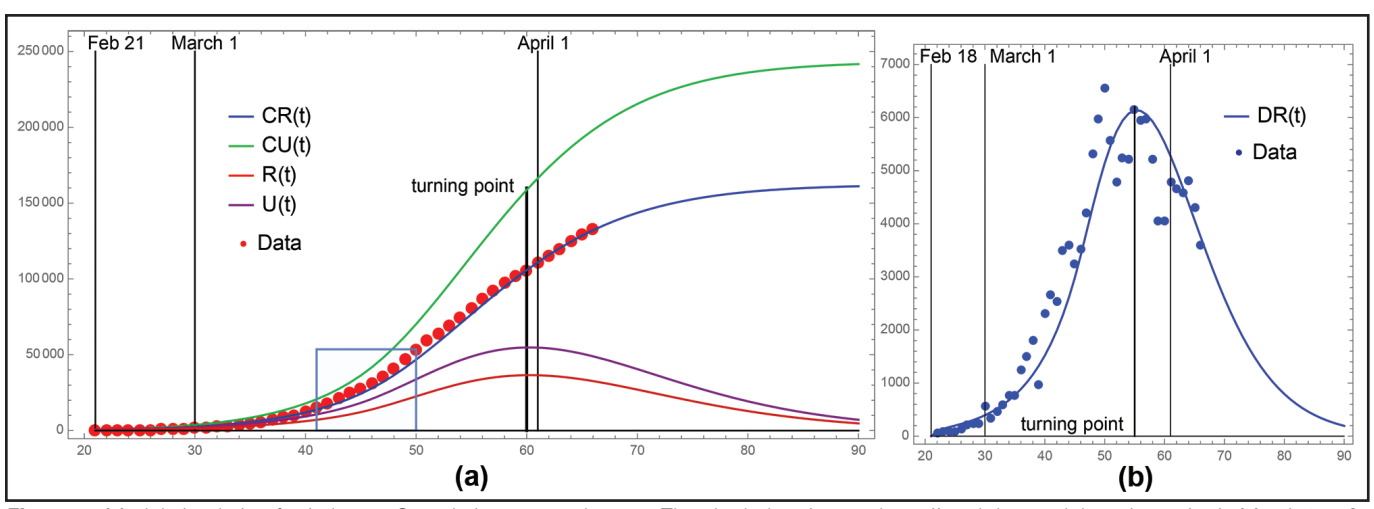

Figure 4. *Model simulation for Italy.* **4a.** *Cumulative reported cases. The shaded region = phase II and the model turning point is March 31.* **4b.** *Daily reported cases. The model turning point is March 26.*

The figures in this article were provided by the authors.

References

[1] Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). A COVID-19 epidemic model with latency period. To appear.

[2] Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). Predicting the cumulative number of cases for the COVID- 19 epidemic in China from early data. Preprint, *medRxiv*.

[3] Liu, Z., Magal, P, Seydi, O., & Webb, G. (2020). Understanding unreported cases in the 2019-nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions. *MPDI Biol., 9*(3), 50.

Zhihua Liu is a professor in the School of Mathematical Sciences at Beijing Normal University in China. Pierre Magal is a professor of mathematics at the University of Bordeaux in France. Ousmane Seydi is an assistant professor at Ecole Polytechnique de Thiès in Senegal. Glenn Webb is a professor of mathematics at Vanderbilt University.