Supplementary Information for

Estimating the reproductive number R_0 of SARS-CoV-2 in the United States and eight European countries and implications for vaccination

Ruian Ke*, Ethan Romero-Severson, Steven Sanche, Nick Hengartner

*Correspondences to: Ruian Ke

Email: rke@lanl.gov

This PDF file includes:

Supplementary text
SI references
Supplementary figure S1
Supplementary tables S1 to S2
SI references

Supplementary Text

1. Data Collection

We extracted data from https://github.com/CSSEGISandData/COVID-19 (John Hopkins Center for Systems Science and Engineering). The date of data access and extraction is March 31, 2020. The data consists of time series of cumulative case confirmations and deaths by country. We used data for the following countries: France (FR), Italy (IT), Spain (SP), Germany (GR), Belgium (BE), Switzerland (SW), Netherlands (NT), United Kingdom (UK) and the US (US). We calculated daily case confirmation and death incidence from cumulative counts.

A few entries in the collected data show signs of bulk reporting. The following procedure was performed to aggregate data where bulk reporting was suspected. It is applied to both case confirmation and death incidence data:

- If an increase of more than 1000% was observed between two consecutive dates, the data was aggregated over the two-day period and accordingly, incidence is considered to be over the two days.
- If the number of reported cases was zero on a date, the data was aggregated over the two-day period.
- If two consecutive days had zero incidence, the data was aggregated over the three-day period.

2. Mathematical model

Standard SEIR model for COVID-19

We first construct a basic susceptible (S)- exposed (E) – infected (I) – recovered (R) model for COVID-19 and then extend upon the basic model. The ordinary differential equations are:

$$\frac{dS}{dt} = -\beta \frac{S}{N} I$$
$$\frac{dE}{dt} = \beta \frac{S}{N} I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

where N is the total number of population, β is the infectivity, $1/k$ is the latent period, i.e. from infection to onset of infectiousness, $1/\gamma$ is the infectious period. Note that for simplicitly, we assumed that R compartment include both recovered and dead individuals in this basic model and γ is the overall rate of individuals leaving I compartment.

We are interested in the dynamics of early exponential growth. We make the common assumption of a constant susceptible population (S(t)=N) during this period, and get a reduced EIR model:

$$\frac{dE}{dt} = \beta I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

The long-term solution of the EIR model is driven by a single exponential whose rate is determined by the dominant eigenvalue, λ, of the Jacobian matrix of the EIR model. The exponential growth rate r, is thus the same as the dominant eigenvalue, λ:

$$r = \lambda = \frac{\sqrt{(k+\gamma)^2 + 4k(\beta - \gamma)} - (k+\gamma)}{2}$$

We can calculate β for a specific value of the exponential growth rate r, as:

$$\beta = \frac{1}{k} r^2 + \frac{k+\gamma}{k} r + \gamma$$

If we let $I^*(t)$ be the total number of infected individuals, and define that $I^*(t) = E(t) + I(t) = I_0^* e^{rt}$, we get the following expressions for $E(t)$ and $I(t)$:

$$I(t) = \frac{k+r}{k+r+\beta} I_0^* e^{rt}$$
$$E(t) = \frac{\beta}{k+r+\beta} I_0^* e^{rt}$$

Thus, we calculate the true daily incidence predicted by the model as:

$$\Omega(t) = \int_{t-1}^{t} \beta I(s)\, ds = \int_{t-1}^{t} \beta \frac{k+r}{k+r+\beta} I_0^* e^{rs}\, ds = \frac{\beta(k+r)}{r(k+r+\beta)} I_0^* \left(e^{rt} - e^{r(t-1)}\right)$$

Extending the model to consider case confirmation

We extend the EIR model to consider case confirmation. In this model, we consider that among newly infected individuals at time *t*, βI(t), a fraction, $\theta(t)$, of them will be tested at a later time. We use an Erlang distribution with a shape parameter m and a mean duration of 1/g for the period between infection and case confirmation. The ODE model is then:

$$\frac{dE}{dt} = \beta I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$
$$\frac{dC_0}{dt} = \theta(t)\beta I - mgC_0$$
$$\cdots$$
$$\frac{dC_i}{dt} = mg(C_{i-1} - C_i)$$
$$\cdots$$
$$\frac{dC_m}{dt} = mgC_{m-1}$$

where i=1,…,m-1, C_0 and C_i correspond to the infected individuals who is tested before their infection is confirmed, and C_m is the cumulative number of confirmed cases.

We are interested in new confirmed cases during a day, and thus it is reasonable to assume $\theta(t)$ does not change (at θ_t) in one day period. With this assumption, we solve the ODE above, and get:

$$C_i(t) = \theta_t \frac{(mg)^i}{(mg+r)^{i+1}} \frac{\beta(k+r)}{k+r+\beta} I_0^* e^{rt}$$

for i=1,…,m-1.

The daily new confirmed case count, $\Psi(t)$, is calculated as

$$\Psi(t) = \int_{t-1}^{t} mgC_{m-1}(s)\, ds = \theta_t \left(\frac{mg}{mg+r}\right)^m \frac{\beta(k+r)}{r(k+r+\beta)} I_0^* \left(e^{rt} - e^{r(t-1)}\right)$$
$$= \theta_t \left(\frac{mg}{mg+r}\right)^m \Omega(t)$$

Extending the EIR model to explicitly consider death

Now, we extend the EIR model to explicitly consider death of infected individuals. Let X be the case fatality ratio, and again, we assume an Erlang distribution for the period between infection to death. We get the following model:

$$\frac{dE}{dt} = \beta I - kE$$
$$\frac{dI}{dt} = kE - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$
$$\frac{dI_{d,1}}{dt} = X\beta I - ndI_{d,1}$$
$$\cdots$$
$$\frac{dI_{d,n}}{dt} = nd(I_{d,n-1} - I_{d,n})$$
$$\frac{dD}{dt} = ndI_{d,k}$$

where n and $1/d$ are the shape parameter and the mean duration of the Erlang distribution for the period from infection to death.

Solving the ODEs above, we get:

$$I_{d,i}(t) = \frac{(nd)^{i-1}}{(r+nd)^i} X \frac{\beta(k+r)}{k+r+\beta} I_0^* e^{rt}, \quad i = 1..n$$
$$D(t) = \left(\frac{nd}{r+nd}\right)^n X \frac{\beta(k+r)}{r(k+r+\beta)} I_0^* e^{rt}$$

Then, the daily death count, $\Phi(t)$, is:

$$\Phi(t) = \int_{t-1}^{t} ndI_{d,n}\, ds = \left(\frac{nd}{r+nd}\right)^n X \frac{\beta(k+r)}{r(k+r+\beta)} I_0^* \left(e^{rt} - e^{r(t-1)}\right)$$

$$= \left(\frac{nd}{r+nd}\right)^n X\ \Omega(t)$$

3. Choice of parameter values and ranges

The latent period, 1/k
We set 1/k=3 days as a baseline. Previously, we and others estimated the incubation period, i.e. the duration between infection and symptom onset to be between 4-6 days (1-3). It is possible that infectiousness onset starts 1-2 days before symptom onset (4). Therefore, we set the latent period at 3 days with a range of variation between 3-4 days.

Distribution of duration from infection to case confirmation
We assumed an Erlang distribution for the duration. We set the mean of the duration to be 12 days, i.e. g=1/12 /day. We varied this mean duration between 10-14 days in the sensitivity analysis. The shape parameter is set to 2 (m=2).

Distribution of duration from infection to death
We assumed an Erlang distribution for the duration with a mean of 1/d and a shape parameter n. We and others showed that the mean time from symptom onset to death is between 16.5 and 18.5 days (3, 5, 6). Assuming an incubation period between 4-5 days (1-3), we set the mean duration of infection to death tobe 1/d=21.5 days. We varied 1/d between 20.5 and 23.5 days in sensitivity analysis. We set n=4 according to our previous estimate (between 3 and 5) (3).

Infection fatality ratio, X
We set X=0.01 as baseline with ranges between 0.04-0.015. This point estimates and range are set according to two previous studies (6, 7).

SI References:

1. Guan WJ, *et al.* (2020) Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*.
2. Lauer SA, *et al.* (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 172(9):577-582.
3. Sanche S, *et al.* (2020) The novel coronavirus, SARA-nCoV-2, is highly contagious and more infectious than initially estimated. *Emerging Infectious Diseases* (accepted).
4. Zou L, *et al.* (2020) SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 382(12):1177-1179.
5. Zhou F, *et al.* (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395(10229):1054-1062.
6. Verity R, *et al.* (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infetious Diseases*.
7. Wu JT, *et al.* (2020) Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 26(4):506-510.

Supplementary Figure

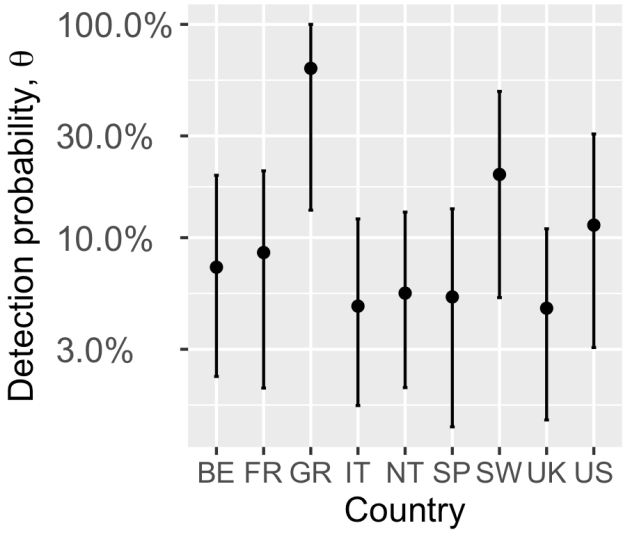

Figure S1. Estimates and confidence interval ranges of and the detection probability $\boldsymbol{\theta}$ in each country.

Supplementary Tables

Table S1. Sensitivity analysis showing the estimated growth rate, *r*, and detection probability, θ, are robust to choices of number of data points used for inference. The 2nd and 3rd columns show the period where case and death counts were collected for inference in the main text, and the estimated values of r and θ using these data are shown in the 'Main text' columns. The estimated values of r and θ using 15, 13 and 10 days of data points are also listed.

Country	Dates of case count data*	Dates of death count date	Growth rate, *r* (/day)				Detection probability, θ			
			Main text	15 days	13 days	10 days	Main text	15 days	13 days	10 days
Belgium	Mar. 6 – 18	Mar. 19 – Mar. 31	0.18	0.19	0.18	0.17	0.07	0.07	0.06	0.06
France	Feb. 29 – Mar. 17	Mar. 10 – Mar. 26	0.22	0.22	0.23	0.22	0.09	0.09	0.08	0.07
Germany	Mar. 1 – 22	Mar. 17 – Mar. 31	0.23	0.23	0.24	0.22	0.62	0.62	0.55	0.45
Italy	Feb. 23 – Mar. 11	Feb. 28 – Mar. 15	0.22	0.23	0.23	0.22	0.05	0.05	0.03	0.04
Netherlands	Mar. 6 – 23	Mar. 15 – 31	0.18	0.18	0.18	0.17	0.05	0.06	0.04	0.05
Spain	Mar. 2 – 14	Mar. 10 – 25	0.3	0.29	0.32	0.34	0.05	0.05	0.04	0.04
Switzerland	Mar. 5 – 20	Mar. 17 – 31	0.18	0.19	0.18	0.19	0.2	0.22	0.18	0.18
United Kingdom	Mar. 5 – 24	Mar. 15 – 31	0.2	0.2	0.21	0.19	0.05	0.05	0.03	0.04
United States	Mar. 3 – 15	Mar. 8 – 23	0.28	0.28	0.28	0.27	0.11	0.12	0.08	0.10

* The end date is set to the date of lock-down or resitriction of movement as reported in the Coronavirus Government Response Tracker (https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker).

Table S2. Comparison of models for the detection probability function, θ, using AIC scores. The three models correspond to the three models listed in the Methods in the main text. Model 1, i.e. constant θ, is the best model for all 9 countries considered (bold AIC scores).

Country	Model 1 Constant	Model 2 Hill-type function	Model 3 Linear function
Belgium	221.7	227.7	227.7
France	338.7	343.8	344.7
Germany	402.3	407.6	408.3
Italy	344.7	350.7	350.7
Netherlands	264.6	270.6	270.6
Spain	280.4	279.6	286.4
Switzerland	260.2	262.2	266.2
UK	327.5	332.5	333.5
US	234	239.5	240