

Universality in COVID-19 spread in view of the Gompertz function [†]

Akira Ohnishi¹, Yusuke Namekawa¹, and Tokuro Fukui¹

¹ *Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan*
* *E-mail: ohnishi@yukawa.kyoto-u.ac.jp*

.....
We demonstrate that universal scaling behavior is observed in the current coronavirus (COVID-19) spread in various countries. We analyze the numbers of infected people in selected eleven countries (Japan, USA, Russia, Brazil, China, Italy, Indonesia, Spain, South Korea, UK, and Sweden). By using the double exponential function called the Gompertz function, $f_G(x) = \exp(-e^{-x})$, the number of infected people is well described as $N(t) = N_0 f_G(\gamma(t - t_0))$, where N_0 , γ and t_0 are the final total number of infected people, the damping rate of the infection probability and the peak time of $dN(t)/dt$, respectively. The scaled data of infected people in most of the analyzed countries are found to collapse onto a common scaling function $f_G(x)$ with $x = \gamma(t - t_0)$ in the range of $f_G(x) \pm 0.05$. The recently proposed indicator so-called the K value, the increasing rate of infected people in one week, is also found to show universal behavior. The mechanism for the Gompertz function to appear is discussed from the time dependence of the produced pion numbers in nucleus-nucleus collisions, which is also found to be described by the Gompertz function.

1. Introduction

The COVID-19 pandemic is the worst disease spread in this century. As of May 20, 2020, over 4 million people have tested positive in the world, and the number of infected people $N(t)$ at the time t is still increasing rapidly. In order to control the spread of infection, it is desired to understand the diffusion mechanism of COVID-19.

Recently, a double exponential function called the Gompertz function is found to catch the features of $N(t)$ [1–20]. The Gompertz function appears when the infection probability per infected people exponentially decreases as a function of time. With the Gompertz function, the daily increase of infected people ($\simeq dN(t)/dt$) shows asymmetric time-profile rather than the symmetric one found in the prediction of the Susceptible-Infected (SI) model [21], one of the standard models of the spread of infection. The Gompertz function was proposed by B. Gompertz in 1825 to discuss the life contingencies [22]. It is interesting to find that the Gompertz function also appears as the number of tumors [23] and the number of detected bugs in a software [24], as well as particle multiplicities at high energies [25].

The exponential decrease of the infection probability is also important to deduce when the restrictions can be relaxed. For example, Nakano and Ikeda [26] found that the total number of infected people of COVID-19 is well characterized by the newly proposed indicator K ,

[†]Report number: YITP-20-82

which represents the increasing rate of infected people in one week. The indicator K takes a value between zero and unity, is not affected by the weekly schedule of the test, and is found to decrease almost linearly as a function of time in the region $0.25 < K < 0.9$ provided that there is only a single outbreak affecting the infection. In order to understand the linearly decreasing behavior of K , Nakano and Ikeda assumed that the infection probability decreases exponentially as a function of time. Since the indicator K is expected to be useful to predict the date when the restrictions can be relaxed as $K(t) \simeq 0.05$, it would be valuable to analyze its solution, the Gompertz function, in more detail.

In this article, we analyze the number of infected people by using the Gompertz function. We examine that $N(t)$ in one outbreak is well described by the Gompertz function in several countries. Then with the time shift and the scale transformation of time and $N(t)$, the data are found to show universal behavior; they are on one-curve described by the basic Gompertz function, $\exp(-e^{-x})$. The newly proposed indicator K [26] also shows universality. We further discuss that the number of produced pions in nucleus-nucleus collisions is described by the Gompertz function. This similarity may be helpful to understand the mechanism of COVID-19 spread.

This article is organized as follows. In Sec. 2, we give a brief review of the Gompertz function and its relevance to the disease spread. In Sec. 3, we show the comparison of the number of infected people and the Gompertz function fitting results. We demonstrate that the numbers of infected people in many countries show universal scaling behavior. In Sec. 4, we show that the number of pions in nuclear collisions is well described by the Gompertz function, and we deduce the mechanism to produce the time-dependence described by the Gompertz function. In Sec. 5, we summarize our work.

2. Gompertz function, indicator K and scaling variables

When the infection probability $k(t)$ is given as a function of time, the evolution equation for the number of infected people $N(t)$ and its solution are given as

$$\frac{dN(t)}{dt} = k(t)N(t), \quad N(t) = N(t_0) \exp \left[\int_{t_0}^t dt' k(t') \right]. \quad (1)$$

Adopting an exponentially decreasing function as $k(t)$, the solution is found to be

$$k(t) = k(t_0)e^{-\gamma(t-t_0)}, \quad N(t) = N(t_0) \exp \left[\frac{k(t_0)}{\gamma} \left(1 - e^{-\gamma(t-t_0)} \right) \right]. \quad (2)$$

By choosing the reference time t_0 so that $k(t_0) = \gamma$, the solution is given as

$$N(t) = N_0 \exp(e^{-\gamma(t-t_0)}), \quad (3)$$

where $N_0 = N(t_0) \exp(k(t_0)/\gamma) = eN(t_0)$ is the asymptotic value of $N(t)$, $N_0 = \lim_{t \rightarrow \infty} N(t)$. Here, N_0 can be interpreted as “terminal velocity”, since the evolution equation (1) has an analogy to the equation of motion of particle feeling a viscous resistance, where its coefficient is not constant but exponentially decreasing.

The double exponential function appearing in Eq. (3) is called the Gompertz function [22],

$$f_G(x) = \exp(-e^{-x}). \quad (4)$$

By using the Gompertz function, $N(t)$ is given as

$$\frac{N(t)}{N_0} = f_G(x), \quad x = \gamma(t - t_0), \quad (5)$$

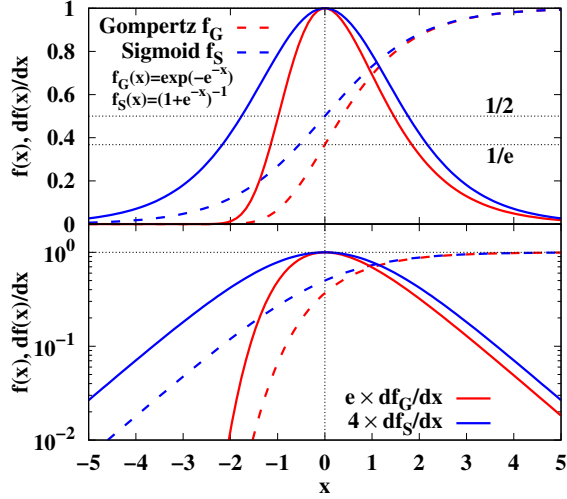


Fig. 1 Comparison of Gompertz $f_G(x)$ and sigmoid $f_S(x)$ functions and their derivatives, $df_G(x)/dx$ and $df_S(x)/dx$, in the linear (top) and logarithmic (bottom) scales.

where x is the scaling variable for $N(t)$.

One of the characteristic features of the Gompertz function is the asymmetry of its derivative in the early ($x < 0$) and late ($x > 0$) stages. In Fig. 1, we show the Gompertz function $f_G(x)$ and its derivative,

$$f'_G(x) = \frac{df_G(x)}{dx} = e^{-x} \exp(-e^{-x}), \quad (6)$$

as functions of x . The derivative takes the maximum at $x = 0$, and the asymmetry in the negative and positive x region is clearly seen.

This asymmetry should be compared with the solution of the SI model [21], in which the susceptible people ($S(t)$) are infected by the infectious people ($I(t)$) at the rate proportional to the product $S(t)I(t)$,

$$\frac{dS}{dt} = -kSI, \quad \frac{dI}{dt} = kSI, \quad (7)$$

where k is a constant. The solution is found to be

$$I(t) = \frac{N_0}{1 + \exp[-\gamma(t - t_0)]} = N_0 f_S(\gamma(t - t_0)), \quad (8)$$

where $N_0 = S + I$ is a constant, $\gamma = N_0k$, and t_0 is the time when half of the people are infected, $I(t_0) = N_0/2$. The characteristic function $f_S(x)$ is called the sigmoid function, which is also referred to as the logistic function,

$$f_S(x) = \frac{1}{1 + e^{-x}} = \frac{1}{2} [1 + \tanh(x/2)], \quad f'_S(x) = \frac{df_S(x)}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{4 \cosh^2(x/2)}. \quad (9)$$

Then the daily number of infected people is a symmetric function of $t - t_0$,

$$\frac{dI(t)}{dt} = N_0 \gamma f'_S(\gamma(t - t_0)) = \frac{N_0 \gamma}{4 \cosh^2[\gamma(t - t_0)/2]}, \quad (10)$$

where $f'_S(x) = df_S(x)/dx$. In Fig. 1, we also show $f_S(x)$ and $df_S(x)/dx$ by blue curves. As already noticed [1–20] and is discussed later, the asymmetry is clearly found in $dN(t)/dt$

of COVID-19, so $N(t)$ is better understood by the Gompertz function than by the sigmoid function.

Once the solution is given, one can obtain the K value [26], increasing rate of the infected people in a week, as

$$\begin{aligned} K(t) &= \frac{N(t) - N(t-w)}{N(t)} = 1 - \exp \left[-(e^{\gamma w} - 1)e^{\gamma(t-t_0)} \right] \\ &= 1 - f_G(x_K), \end{aligned} \quad (11)$$

where x_K is the scaling variable for $K(t)$,

$$x_K = \gamma(t - t_0 - \Delta t), \quad (12)$$

$$\Delta t = w + \frac{1}{\gamma} \log(1 - e^{-\gamma w}), \quad (13)$$

with $w = 7$ days.

3. Comparison with data

3.1. Adopted dataset

Let us now examine the universal behavior given by the Gompertz function in real data. We use the data given in Ref. [27] as of May 21, which contain the $N(t)$ data from Dec. 31, 2019 ($t = 0$) till May 20 ($t = 141$). Throughout this article, we measure the time t in the unit of day. In order to avoid the discontinuity coming from the definition change, we have removed the spikes in the daily increase ($dN(t)/dt$) data in Japan (April 12, $t = 103$) and China (February 13, $t = 44$), and instead the daily numbers in previous days are increased by multiplying a common factor, which is determined to keep the total number of infected people in the days after the spike. The number of infected people ($N(t)$) is obtained as the integral of thus-smoothen $dN(t)/dt$. In addition, seven-day averages (± 3 days) are considered in the analysis of $dN(t)/dt$ in order to remove the fluctuations in a week. Thus-smoothen $dN(t)/dt$ data are available in $3 \leq t \leq 138$.

We have chosen the countries with a large number of infected people (the USA, Russia, Brazil, and the UK as of May 20, 2020), the first three Asian countries where the COVID-19 spread explosively (China, South Korea, and Japan), the first two countries of spread in Europe (Italy and Spain), a country with a unique policy (Sweden), and a country with somewhat different $dN(t)/dt$ profile (Indonesia).

3.2. Number of infected people and its daily increase

In the left panel of Fig. 2, we show the daily increase of the infected people $dN(t)/dt$ given in Ref. [27] with the smoothing mentioned above. The legends stand for the abbreviation of the country name (internet country domain code, see Table 1). The $dN(t)/dt$ data show there is one big peak in each country, and the shape of the peak is asymmetric; fast rise and slow decay. In many of the countries, there are several other peaks, which are smaller than the dominant one but visible at least in the log-scale plot. The fitting results using the Gompertz function are shown by dotted lines, and are found to explain the dominant peak region of data well.

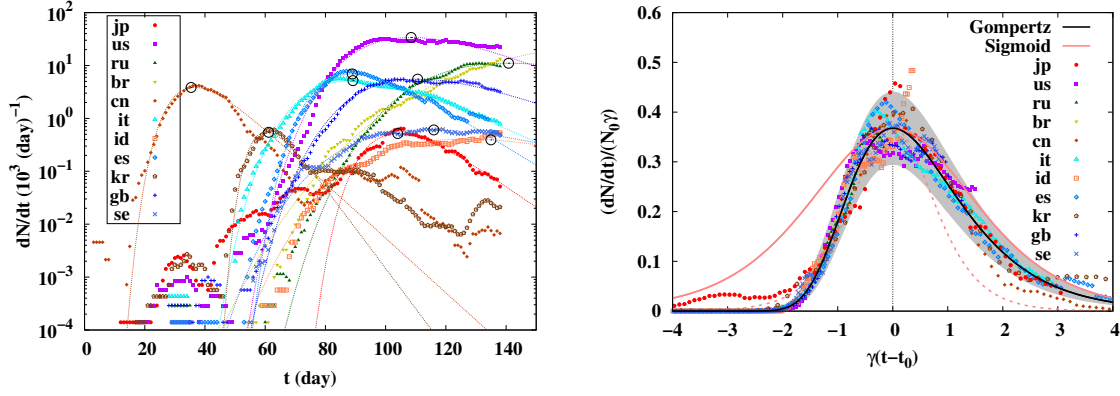


Fig. 2 Derivative of the number of infected people $dN(t)/dt$ (left, symbols) and the scaling behavior (right, symbols). In the left panel, dotted lines show the fitting results in the derivative of the Gompertz function, $dN(t)/dt \simeq N_0 \gamma f'_G(\gamma(t-t_0))$, and open circles show the peak points in the fitting function, $(t_0, N_0 \gamma f'_G(t_0))$. In the right right panel, solid black curve shows the derivative of the Gompertz function $f'_G(x)$, and the grey band shows its region with 5 % uncertainty in γ and 20 % in N_0 . The orange solid (dashed) curve shows the derivative of the sigmoid function normalized to reproduce the peak height, $4f'_S(x)/e$ ($4f'_S(2x)/e$).

Table 1 Fitting results with the fitting ranges of t . Parameters in the line with (*) are adopted to draw figures.

Country	γ [%/day]	t_0 [day]	N_0 [10^3]	ΔN	fitting range
Japan	9.1 ± 0.2	104.0 ± 0.3	15.4 ± 0.3	1554 ± 22	85– 150 (*)
	8.7 ± 0.3	103.8 ± 0.4	15.6 ± 0.4	0.5 ± 1.3	0– 150
USA	7.2 ± 0.1	100.3 ± 0.3	1202 ± 23	-800 ± 240	75– 100
	5.2 ± 0.1	108.5 ± 0.4	1780 ± 28	0.1 ± 2.2	0– 150 (*)
Russia	3.7 ± 0.1	140.8 ± 0.9	819 ± 28	0.1 ± 0.3	0– 150 (*)
Brazil	1.3 ± 0.1	245 ± 15	15100 ± 5600	-806 ± 45	74– 150
	1.9 ± 0.1	193.2 ± 4.3	3890 ± 510	-0.7 ± 0.6	0– 150 (*)
China	11.9 ± 0.2	36.4 ± 0.2	88.5 ± 1.5	52.6 ± 9.5	0– 50 (*)
	14.1 ± 0.2	35.4 ± 0.2	78.9 ± 1.7	73 ± 19	0– 150
Italy	6.0 ± 0.1	89.1 ± 0.2	234.6 ± 2.0	0.0 ± 0.4	0– 150 (*)
Indonesia	2.7 ± 0.2	135.0 ± 3.0	42.4 ± 3.5	-173 ± 12	75– 150
	3.4 ± 0.1	126.7 ± 1.2	32.8 ± 1.2	-0.2 ± 0.2	0– 150 (*)
Spain	8.1 ± 0.1	88.9 ± 0.2	231.5 ± 2.6	0.1 ± 0.4	0– 150 (*)
S.Korea	17.5 ± 0.4	60.4 ± 0.1	8.5 ± 0.2	0.2 ± 0.3	0– 75 (*)
	14.1 ± 0.5	61.1 ± 0.3	9.1 ± 0.4	0.2 ± 0.7	0– 150
UK	4.8 ± 0.1	110.6 ± 0.3	310.1 ± 3.5	0.1 ± 0.4	0– 150 (*)
Sweden	3.5 ± 0.1	116.0 ± 0.8	46.3 ± 1.1	-36.1 ± 5.7	60– 150
	3.7 ± 0.1	114.9 ± 0.6	44.7 ± 0.8	-0.0 ± 0.2	0– 150 (*)

The fitting to $dN(t)/dt$ data is carried out by using the derivative of the Gompertz function,

$$\frac{dN(t)}{dt} = N_0 \gamma f'_G(\gamma(t - t_0)). \quad (14)$$

In order to concentrate on the dominant peak, we first limit the time region of the fit to $t_{\min} \leq t \leq t_{\max}$, which covers it. Next, the fitting time region is extended to the whole range, $0 \leq t \leq 140$. In the fitting procedure, we have assumed a Poisson distribution for the daily number of newly infected people, then the uncertainty in $dN(t)/dt$ is assumed to be $\sqrt{dN(t)/dt + \varepsilon}$, where $\varepsilon = 0.1$ is introduced to avoid zero uncertainty in the case of zero daily number. We summarize the obtained parameters (N_0, γ, t_0) in Table 1, and parameters in the line with (*) are adopted to draw the figures. When the χ^2 value is smaller in the whole range analysis and the obtained parameters in the two cases are similar, the single outbreak assumption is supported and we show only the results in the whole range analysis. In other cases, we in principle adopt the results giving the smaller reduced χ^2 . The exception is the USA, where the reduced χ^2 is larger but we adopt the whole range analysis results. This is closely related to the multiple outbreaks, and will be discussed in Appendix A.

In the right panel of Fig. 2, we show the normalized daily numbers, $(dN(t)/dt)/(N_0\gamma)$, as functions of the scaling variable, $x = \gamma(t - t_0)$. Most of the data points are around the derivative of the Gompertz function and inside the gray band, which shows the region with 5% uncertainty in γ and 20% uncertainty in N_0 .

We also show the sigmoid function by the orange curves in the right panel of Fig. 2. It is clear that a single sigmoid function cannot describe the behavior of the $dN(t)/dt$ data. If we try to fit $dN(t)/dt$ data by the sigmoid function, we need to adopt larger γ in the negative x region as shown by the orange dashed curve in Fig. 2, $f'_S(2x)$, while $f'_S(x)$ approximately agrees with $f'_G(x)$ in shape in the positive x region.

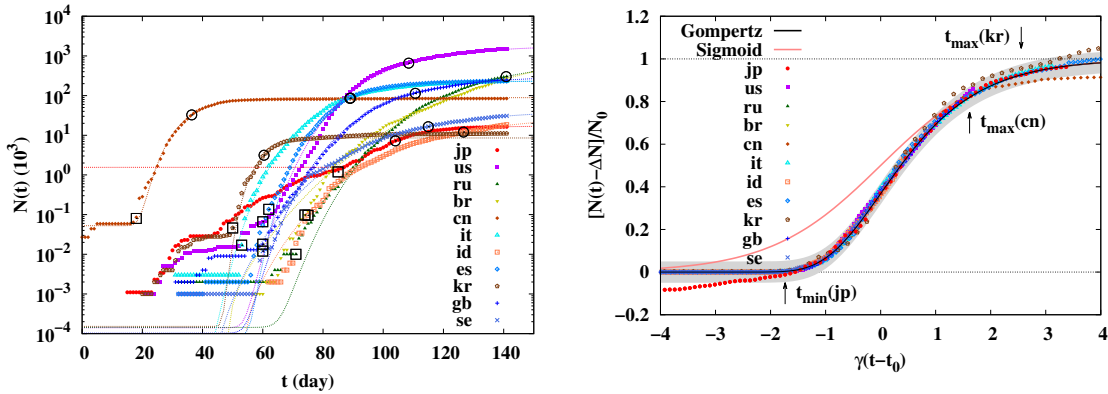


Fig. 3 The number of infected people $N(t)$ data (left) and the scaling behavior (right). In the left panel, dotted lines show the Gompertz function fit, $N_0 f_G(\gamma(t - t_0))$, open circles show the reflection points, $(t_0, N_0/e)$, and open squares show the offset points, $(t_{\text{offset}}, N_{\text{offset}})$. In the right panel, the black solid curve shows the Gompertz function $f_G(x)$ and the grey band shows the region $f_G(x) \pm 0.05$. The orange solid curve shows the sigmoid function $f_S(x)$.

The left panel of Fig. 3 shows the number of the infected people $N(t)$, obtained as the integral of $dN(t)/dt$ data after removing the spike. We also show the Gompertz function results with the parameters (N_0, γ, t_0) determined from the $dN(t)/dt$ data, and an integration constant ΔN ,

$$N(t) = N_0 f_G(\gamma(t - t_0)) + \Delta N, \quad (15)$$

where ΔN is obtained by fitting to the $N(t)$ data. In most of the countries, the Gompertz function with ΔN explains the data in the large $N(t)$ region and deviations are found only in the region with small $N(t)$. In Japan, the fitted time range is limited to be $t \geq 85$ and earlier time data are not fitted to. Thus deviations at $t < 80$ are visible in the log scale, while the value is less than 10% of the total number of infected people at $t = 140$.

We show the normalized $N(t)$ as functions of the scaling variables in the right panel of Fig. 3. We subtract ΔN from $N(t)$. It is interesting to find that most of the world data are on the Gompertz function $f_G(x)$. In Japan, China and South Korea, the fitted time range is limited and deviation from the Gompertz function results are found in the earlier times (Japan) and in later times (China and South Korea). It should be noted that the agreement at $x < 0$ owes largely to the large denominator compared with the number of infected people in the early stage. Compared with the exponentially grown number of infected people, the number of infected people in the early stage is much smaller and the ratio is seen to be very small. Nevertheless, it is impressive to find the agreement of the observed number of infected people after scaling and the Gompertz function.

3.3. K value

We now proceed to discuss the K value. We choose the offset day when explosive spread started, as shown by open squares in Fig. 3. The offset day and offset number of infected people, $(t_{\text{offset}}, N_{\text{offset}})$, are summarized in Table 2. With these offset parameters, the K value is obtained as

$$K(t) = \frac{N(t) - N(t - 7)}{N(t) - N_{\text{offset}}}. \quad (16)$$

Since N_{offset} is generally much smaller than the number of infected people after explosive spread, the K value is not sensitive to the choice of the offset parameters as long as we discuss the long-time behavior.

In the left panel of Fig. 4, we show the K factor as a function of time. Data of $K(t)$ are explained by the prediction from the scaling function, $1 - f_G(x_K)$, while the fluctuations around the predictions are large compared with the number of infected people, $N(t)$. In the right panel of Fig. 4, we show K values as functions of the scaling variable $x_K = \gamma(t - t_0 - \Delta t)$. Except for several countries, the scaling behavior in K is observed.

In Ref. [26], $K(t)$ is found to show the linear dependence on time in the large $K(t)$ region, $0.25 < K(t) < 0.9$. Actually, the Gompertz function $f_G(x)$ shows the linear dependence on time in the small $|x_K|$ region. When the scaling variable is small, the first-order Taylor expansion would work,

$$K(t) = 1 - \exp(-e^{-x_K}) = \left(1 - \frac{1}{e}\right) - \frac{x_K}{e} + \mathcal{O}(x_K^2). \quad (17)$$

The precision of this first order approximation is around 1% (10%) for $|x_K| \leq 0.5$ ($|x_K| \leq 1$), where $1 - f_G(x_K)$ amounts to be 0.45 – 0.81 (0.28 – 0.93). When K is smaller ($K < 0.25$),

Table 2 Offset parameters ($t_{\text{offset}}, N_{\text{offset}}$) used to evaluate the K value.

Country	t_{offset} [day]	N_{offset}
Japan	85	1193
USA	60	66
Russia	71	10
Brazil	74	98
China	18	80
Italy	53	17
Indonesia	75	96
Spain	62	136
S.Korea	50	46
UK	60	18
Sweden	60	12

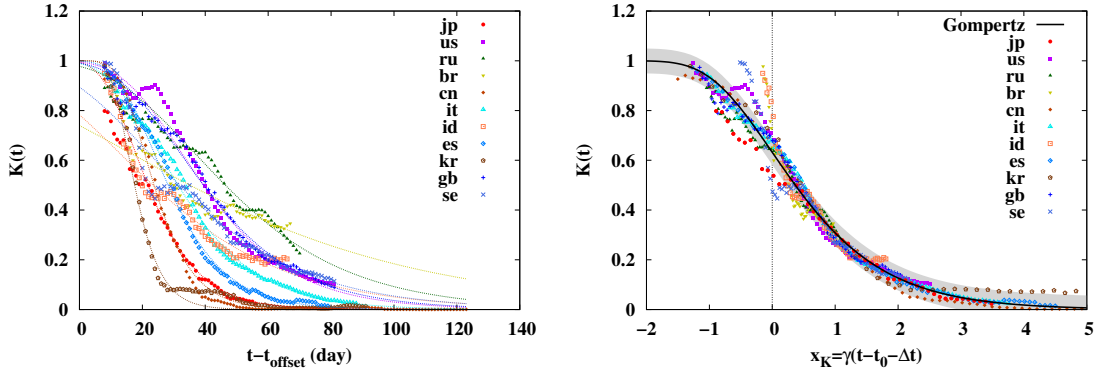


Fig. 4 K value as a function of t (left) and as a function of the scaling variable x_K . In the left panel, dotted lines show the K value from the Gompertz function fit. In the right panel, the black solid curve shows $1 - f_G(x_K)$ and the grey band shows the region $1 - f_G(x_K) \pm 0.05$.

e^{-x_K} should be also small and the Taylor expansion with respect to e^{-x_K} may work. Then K damps exponentially,

$$K(t) = 1 - \exp(-e^{-x_K}) = 1 - (1 - e^{-x_K}) + \mathcal{O}((e^{-x_K})^2) \simeq e^{-x_K}. \quad (18)$$

With the sigmoid function, while scaling behavior is observed in $N(t)$ and $dN(t)/dt$, K does not scale as a function of a single scaling variable. In Fig. 5, we compare the functions in the K value derived from the Gompertz and sigmoid functions for $N(t)$ as functions of $x = \gamma(t - t_0)$. With the Gompertz function, the shift of the scaling variable is enough for the K value to be described by the basic Gompertz function. In contrast, the K value from the sigmoid function reads

$$K(t) = \frac{f_S(\gamma(t - t_0)) - f_S(\gamma(t - t_0 - w))}{f_S(\gamma(t - t_0))} = \frac{1 - e^{-\gamma w}}{e^{x - \gamma w} + 1}. \quad (19)$$

In addition to the shift in the scaling variable, the amplitude also depends on $\gamma w = 7\gamma$.

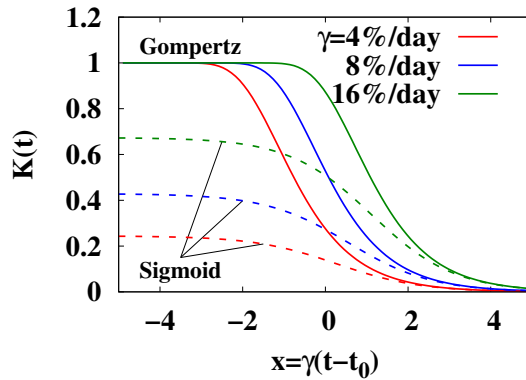


Fig. 5 Comparison of K values from the Gompertz and sigmoid functions in terms of $x = \gamma(t - t_0)$.

3.4. Days of expected relaxing COVID-19 restrictions from the K value

When the $K(t)$ value goes down to be around 0.05 [26], it would be possible to relax COVID-19 restrictions such as lifting the lockdown of the city or relaxing the state of emergency. Let us call those days as t_{relax} . The days of relaxing the restrictions in several countries roughly correspond to the time at $K(t) = 0.05$. With this condition, the number of infected people in the current outbreak will increase by around 5% and the infection in the current outbreak will converge. In terms of the scaling variable, this corresponds to the solution of $1 - f_G(x_K) = 0.05$ and is found to be $x_K = x_R \simeq 2.97$. Thus the expected day of relaxing the restrictions can be evaluated to be around,

$$t_{\text{relax}} = \frac{x_R}{\gamma} + t_0 + \Delta t(\gamma), \quad (20)$$

provided that there will be no further outbreaks. In Table 3, we summarize the expected day of relaxing COVID-19 restrictions t_{relax} from the Gompertz function analyses in comparison with the relaxed day t_{relaxed} in several countries.

In Fig. 6, we show the expected days of relaxing the restrictions in comparison with some of the relaxed days. The lockdown in the Wuhan city was lifted on May 10, 2020 ($t = 131$) in China, the lockdown was relaxed on May 2, 2020 ($t = 123$) in Spain, May 4, 2020 ($t = 125$) in Italy, and May 11, 2020 ($t = 132$) in the UK. Restrictions were relaxed in part on May 6, 2020 ($t = 127$) in South Korea, May 12, 2020 ($t = 133$) in Russia, and May 20, 2020 ($t = 141$) in the USA. In Japan, the state of emergency was declared on April 7, 2020 ($t = 98$) and canceled on May 25 ($t = 146$). The relaxed days in Italy, Spain, Japan, the UK, and the USA are close to those expected from the Nakano-Ikeda model analyses. In China and Korea, the relaxed days were significantly later than the expectations from the model. One can guess that the governments tried to be on the safe side in these first two countries of COVID-19 spread.

The expected day of relaxing COVID-19 restrictions strongly depends on the value of the damping rate of the infection probability, γ , which people and governments should try

Table 3 Expected day of relaxing COVID-19 restrictions t_{relax} and the relaxed day t_{relaxed} .

Country	t_{relax} [day]	t_{relaxed} [day]
Japan	$135.4^{+1.5}_{-0.7}$	146
USA	$149.9^{+2.0}_{-0.7}$	141
Russia	$188.3^{+3.7}_{-1.2}$	133
Brazil	$246.9^{+15.2}_{-4.2}$	–
China	$63.6^{+0.8}_{-0.4}$	131
Italy	$127.8^{+0.9}_{-0.4}$	125
Indonesia	$175.5^{+5.2}_{-1.5}$	–
Spain	$122.2^{+0.7}_{-0.3}$	123
S.Korea	$82.4^{+0.7}_{-0.4}$	127
UK	$153.3^{+1.5}_{-0.5}$	132
Sweden	$162.3^{+2.7}_{-0.8}$	–

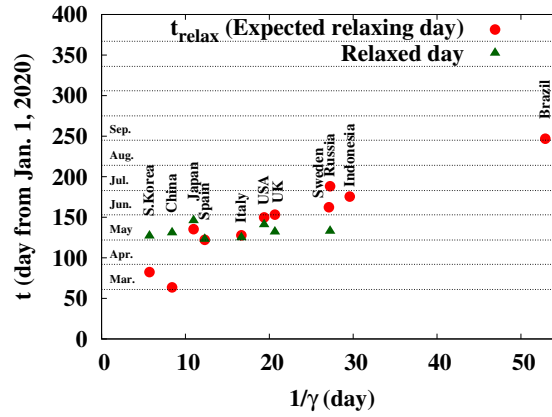


Fig. 6 Expected days of relaxing COVID-19 restrictions and the relaxed days.

to enhance. In South Korea and China, the damping rate is larger than 0.1. Then the infection probability decreases by a factor of $1/e$ within 10 days. In these countries, the test-containment processes have been performed strongly. In many of the countries under consideration (Japan, USA, UK, Italy, and Spain), the damping rates take the value between 4 – 10%/day. In these countries, many of the restaurants and shops are closed and people are requested to stay home for one month or more. Sweden may be an interesting example. The Swedish government does not require restaurants and shops to be closed and does not ask people to stay home. The government asks people to be responsible for their behavior and social distancing is encouraged. The damping rate in Sweden, $\gamma = 3.7\%$ /day, may be regarded as a value representing the intrinsic nature of COVID-19.

It would be valuable to comment on the use of the effective reproduction number, R_e , which is defined as the ratio of the newly infected people in one week to that in the previous week and has the merit that it is not necessary to define the offset. For a constant infection probability per infected people, $k(t) = k_0 = \text{const.}$ in Eq. (1), R_e is found to be

$R_e = \exp(k_0 w)$ and we can guess k_0 from R_e . In contrast, when the number of infected people is given by the Gompertz function, $N(t) = N_0 f_G(x)$, the effective reproduction number $R_e(t)$ is a function of two variables, x_K and γw ,

$$R_e(t) \equiv \frac{N(t) - N(t-w)}{N(t-w) - N(t-2w)} = \frac{1 - f_G(x_K)}{f_G(x_K)[1 - f_G(x_K - \gamma w)]}, \quad (21)$$

while $K(t)$ is a function of a single scaling variable x_K . Thus the universality observed in $K(t)$ is lost in $R_e(t)$, and it would be less easy to give a prediction of t_{relax} from $R_e(t)$ than from $K(t)$. In Fig. 7, we show $R_e(t)$ from the Gompertz function at $\gamma = 4, 8$ and 16% /day. The value of $R_e(t)$ at $x_K = x_R$ depends on γ , the value of x_K at $R_e(t) = 1$ is different from x_R , $x_K = 0.49$ (1.42) for $\gamma = 16\%$ (4%)/day, and apparent large values of $R_e(t)$ appear in the early stage, $x_K < 0$ or $t < t_0 + \Delta t(\gamma)$.

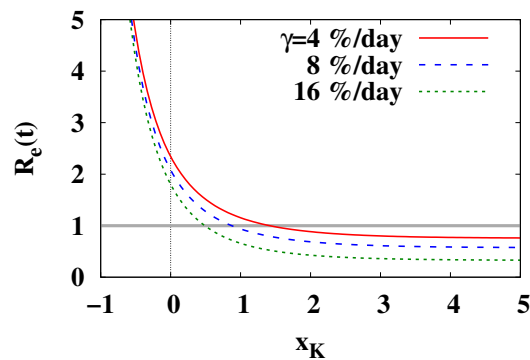


Fig. 7 The effective reproduction number R_e as a function of the scaling variable x_K .

4. Mechanism of appearance of the Gompertz function

Fast and slow rises in the early and late stages are found in many physical processes such as the particle production in nuclear collisions. In Fig. 8, we show the number of Δ particles ($\Delta^{++}, \Delta^+, \Delta^0$ and Δ^-) and π particles (π^+, π^0 and π^-) produced in central Au+Au collisions at the incident energy of 1 GeV/nucleon [28]. Histograms show the calculated results by using the hadronic transport model JAM [29]. The main production mechanism of π particles at this incident energy is the Δ production and its decay. In the early stage ($t < 15$ fm/c), nucleons are excited to resonances such as the Δ particles in the nucleon-nucleon collisions, $NN \rightarrow N\Delta$. Produced Δ s collide with other nucleons and Δ s, some of them produce additional Δ particles, $\Delta N \rightarrow \Delta\Delta$, and some of them are deexcited to nucleons in the Δ absorption processes such as $\Delta N \rightarrow NN$. The Δ particles decay and produce π particles, $\Delta \rightarrow N\pi$. Produced π particles may collide with other nucleons, Δ s and π s, and occasionally produce additional Δ , $\pi N \rightarrow \pi\Delta$. In the later stage, the system expands, particle density decreases, and interaction rate goes down. When the density becomes low enough, all Δ particles decay to $N\pi$ with the lifetime $\tau_\Delta = \hbar/\Gamma_\Delta \simeq 2$ fm/c and π particles go out from the reaction region and are detected.

The number of π and Δ particles in the above nucleus-nucleus collision is found to be well fitted by the Gompertz function $f_G(x)$ and its derivative $f'_G(x)$. Curves in Fig. 8 show the

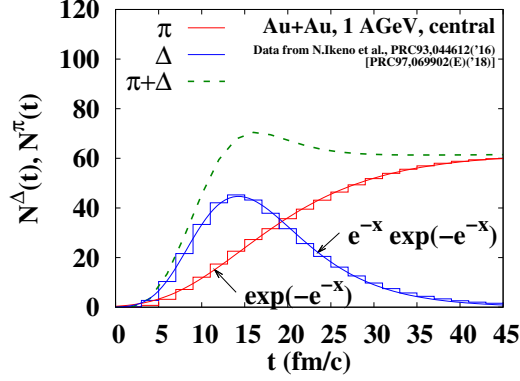


Fig. 8 The number of π and Δ particles as functions of time t in central Au+Au collisions at the incident energy of 1 GeV per nucleon. Histograms show the π (red) and Δ (blue) numbers taken from [28], and solid curves show the fitting results using the Gompertz function. The green dashed curves shows the sum of π and Δ numbers.

results of fit by the Gompertz function and its derivative,

$$N^\pi(t) = N_\pi f_G(\gamma_\pi(t - t_\pi)) , \quad N^\Delta(t) = N_\Delta \gamma_\Delta f'_G(\gamma_\Delta(t - t_\Delta)) , \quad (22)$$

where the parameters are obtained as $(N_\pi, \gamma_\pi, t_\pi) = (62.0, 0.113 \text{ (fm/c)}^{-1}, 14.6 \text{ fm/c})$ and $(N_\Delta, \gamma_\Delta, t_\Delta) = (760, 0.160 \text{ (fm/c)}^{-1}, 14.3 \text{ fm/c})$. Compared with the lifetime of Δ , the number of Δ during nucleus-nucleus collisions has a longer tail. This may be because of the relativistic effects, resonance mass dependence of the width, and sequential decay and production of Δ such as $\Delta \rightarrow \pi N$ followed by $\pi N \rightarrow \Delta$.

It should be noted that one Δ particle mostly decays into $N\pi$ and additionally produced number of π is less than unity in the present pion production in nucleus-nucleus collisions. A rough estimate of the upper bound of the basic reproduction number R_0 from Δ to π may be obtained as follows. The lower bound of the number of produced Δ particles is the peak number of Δ , which is $N^\Delta = 45.3$ at $t = 14 \text{ fm/c}$ in the calculated data and is expected to be $N^\Delta(t_\Delta) = N_\Delta \gamma_\Delta / e \simeq 44.7$ at $t = t_\Delta$ from the Gompertz function. Then the upper bound of the additionally produced π number is given as $N^\pi(t = \infty) - N^\Delta(t_\Delta) \simeq 17$. Consequently, the upper bound of R_0 is given as $R_0 = [N^\pi(t = \infty) - N^\Delta(t_\Delta)] / N^\Delta(t_\Delta) \simeq 0.38$, which is less than unity. As a result, the number of pions in the final state is already determined in the early stage. The green dashed curve in Fig. 8 shows the " π -like" particle number, $N^{\pi\Delta} = N^\pi + N^\Delta$, which shows a peak at $t \simeq 16 \text{ fm/c}$, gradually decreases by the Δ absorption processes, and converges to $N^\pi(t = \infty)$.

Based on the success in describing $N^\pi(t)$ by using the Gompertz function, we may make a conjecture of the correspondence of the COVID-19 spread and the π production in nuclear collisions. Let us assume that π particles correspond to the infected people who tested positive and Δ particles correspond to the exposed (infected) people who have not tested positive yet. The people are exposed in the initial dense stage (Δ production), occasionally infect other susceptible people (additional Δ production) or are recovered (Δ absorption), develop symptoms and test positive ($\Delta \rightarrow N\pi$). Number of infected people including those have not tested positive, $N^\pi + N^\Delta$, grows rapidly in the early dense stage but does not

change much in the later stage. Hence, except for the early dense stage, the basic reproduction number would be less than unity.

5. Summary

We have analyzed the number of infected people of COVID-19, one of the coronaviruses, as a function of time, $N(t)$, by using the double exponential function referred to as the Gompertz function, $f_G(x) = \exp(-e^{-x})$. The Gompertz function appears when the infection probability is an exponentially decreasing function in time. One of the characteristic features of the Gompertz function is the asymmetry of its derivative, $f'_G(x) = df_G(x)/dx$, fast rise and slow decay.

This feature is found in the daily cases of infected people, $dN(t)/dt$. We have assumed that the number of infected people from one outbreak is given as $N(t) \simeq N_0 \exp[-e^{-\gamma(t-t_0)}]$, where N_0 , γ and t_0 are the total number of infected people, damping rate of the infection probability, and the time where $N(t)$ becomes N_0/e . These parameters are obtained by the χ^2 fitting to the $dN(t)/dt$ data. Then we have found that $N(t)$ and $dN(t)/dt$ show universal scaling, $N(t)/N_0 = f_G(x)$ and $(dN(t)/dt)/(N_0\gamma) = f'_G(x)$, where $x = \gamma(t - t_0)$ is the scaling variable. The K value, the increasing rate of infected people in one week, is also found to show the scaling behavior, $K(t) = 1 - f_G(x_K)$, where $x_K = \gamma(t - t_0 - \Delta t(\gamma))$ is the scaling variable for $K(t)$ with $\Delta t(\gamma)$ being a given function of γ .

We have also found that the time dependence of the produced pion number in nucleus-nucleus collisions is described by the Gompertz function. Since both of the COVID-19 spread and the pion production are transport phenomena, the mechanism of the former may be similar to the latter. If this is the case, there is a possibility that the basic reproduction number is high only in the initial stage of the outbreak.

Throughout this article, we have imposed the single-outbreak assumption. Since this assumption may be too restrictive, we show the results of multiple-outbreak analyses in Appendix A. The multiple-outbreak analyses also show that the COVID-19 spread in one outbreak is well described by the Gompertz function.

Acknowledgments

The authors thank T. Kunihiro and T. T. Takahashi for constructive comments and suggestions. They also thank N. Ikeno, A. Ono and Y. Nara for useful discussions. This work is supported in part by the Grants-in-Aid for Scientific Research from JSPS (Nos. 19H01898 and 19H05151).

A. Multiple-outbreak model analysis

In the analyses in the main text, we have assumed that there is only one dominant outbreak. Let us consider here the multiple-outbreak cases, where we assume the number of infected people is described by the sum of several Gompertz functions,

$$N(t) = \sum_{i=1}^n N_i f_G(\gamma_i(t - t_i)) + \Delta N . \quad (\text{A1})$$

As in the single-outbreak model discussed in the main text, we analyze the daily number of newly infected people, $dN(t)/dt$,

$$\frac{dN(t)}{dt} = \sum_{i=1}^n N_i \gamma_i f'_G(\gamma_i(t - t_i)) , \quad (\text{A2})$$

where $f'_G(x) = df_G(x)/dx$.

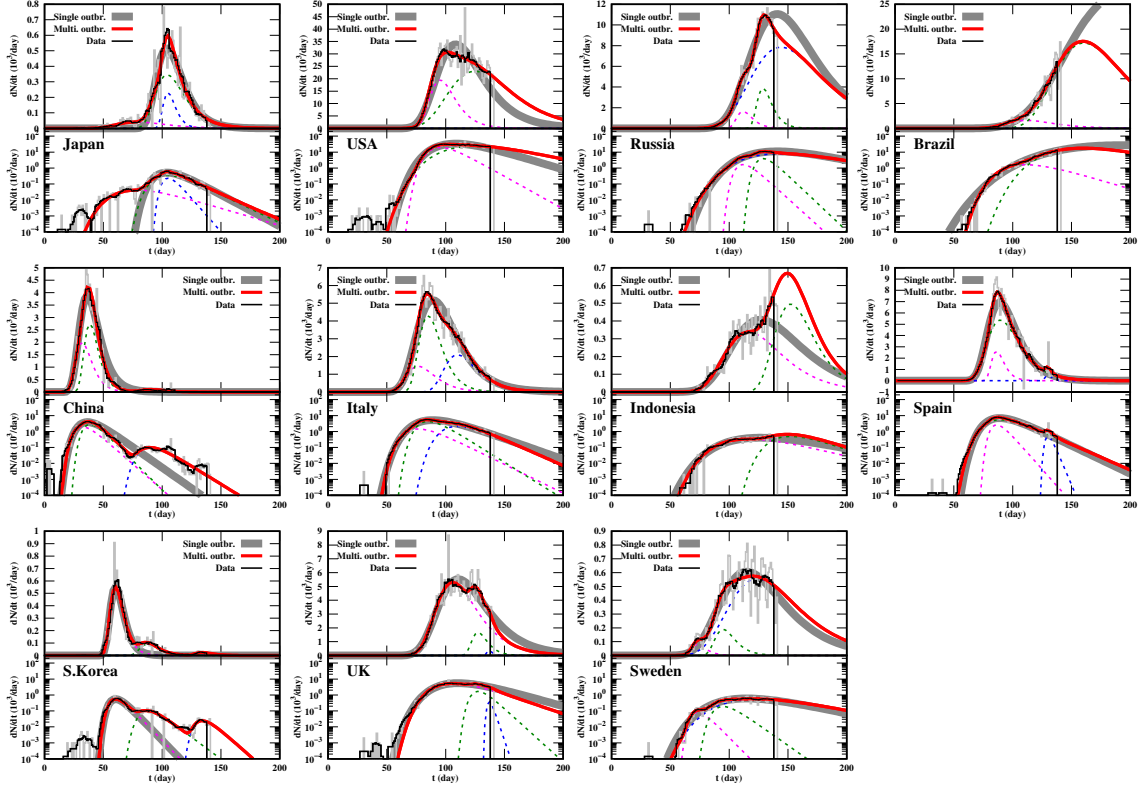


Fig. A1 Derivative of the number of infected people dN/dt with single-outbreak (grey) and multiple-outbreak (red) model analyses in the linear (top) and logarithmic (bottom) scales. Magenta, green and blue dotted curves show the contributions from the first, second and third outbreaks, respectively.

In Fig. A1, we show dN/dt in Eq. (A2) in comparison with the daily number of newly infected people. In the multiple-outbreak analysis, we use data in the whole range, $0 \leq t \leq 140$. Two or three outbreaks are considered, and we try to describe the region with large dN/dt by adding outbreaks. Obtained parameters are summarized in Table A1.

In many countries under consideration, dN/dt is decreasing on May 21, 2020, and the parameters are well determined. Then we adopt the three-outbreak model ($n = 3$). In Japan, China, and South Korea, multiple-outbreak structure of dN/dt is clearly seen in the logarithmic plot and can be fitted by using Eq. (A2). In Russia, Italy, Spain, the UK and Sweden, additional outbreaks improve the reduced χ^2 by filling the peaks which are not covered by the single outbreak. In Brazil and Indonesia, where the numbers of infected people are still rapidly increasing, we need at least one outbreak term with $t_i > t_{\text{now}}$. In those cases, parameters generally have large uncertainties, so the third outbreak, if included, has extremely

Table A1 Parameters in multiple-outbreak model analyses.

Country	γ_i [%/day]	t_i [day]	N_i [10^3]	ΔN
Japan	5.0 ± 0.5	78.5 ± 3.0	2.2 ± 0.3	0.3 ± 0.3
	8.2 ± 0.3	104.3 ± 0.2	11.4 ± 0.6	
	19.4 ± 1.8	105.0 ± 0.3	3.2 ± 0.6	
USA	9.4 ± 0.2	95.4 ± 0.4	560 ± 33	0.2 ± 0.5
	3.7 ± 0.1	124.8 ± 1.2	1682 ± 21	
Russia	15.4 ± 0.5	111.8 ± 0.2	28.9 ± 1.8	0.1 ± 0.1
	16.1 ± 0.6	128.8 ± 0.2	65.8 ± 5.1	
	3.3 ± 0.0	144.1 ± 0.9	647 ± 31	
Brazil	5.0 ± 0.4	111.3 ± 3.5	86.6 ± 22.4	0.0 ± 0.1
	3.4 ± 0.4	160.6 ± 4.9	1380 ± 330	
China	14.8 ± 2.1	32.0 ± 1.9	36.5 ± 14.6	52.2 ± 6.6
	17.0 ± 1.3	38.5 ± 0.3	43.8 ± 14.6	
	10.4 ± 0.6	90.0 ± 0.8	2.5 ± 0.2	
Italy	8.3 ± 0.5	76.6 ± 1.8	47.3 ± 11.1	0.0 ± 0.1
	10.6 ± 0.4	85.0 ± 0.3	110 ± 12	
	7.4 ± 0.1	110.0 ± 0.6	76.7 ± 3.7	
Indonesia	4.1 ± 0.1	116.5 ± 1.0	22.3 ± 1.0	-0.1 ± 0.1
	6.0 ± 2.4	152 ± 12	23 ± 23	
Spain	19.1 ± 0.9	86.3 ± 0.2	36.9 ± 3.2	0.1 ± 0.2
	7.5 ± 0.1	89.5 ± 0.1	195.4 ± 3.2	
	39.8 ± 3.1	129.9 ± 0.2	3.6 ± 0.3	
S.Korea	17.5 ± 0.3	60.4 ± 0.1	8.5 ± 0.2	0.2 ± 0.2
	12.6 ± 0.6	87.8 ± 0.5	2.0 ± 0.1	
	15.0 ± 3.7	134.1 ± 1.9	0.4 ± 0.1	
UK	5.7 ± 0.1	106.1 ± 0.3	251.5 ± 3.5	0.1 ± 0.3
	15.5 ± 1.3	127.6 ± 0.4	29.4 ± 3.0	
	50 ± 38	137.0 ± 1.4	2.0 ± 1.9	
Sweden	17.2 ± 1.4	73.2 ± 0.5	1.3 ± 0.2	0.0 ± 0.1
	12.0 ± 1.1	94.9 ± 0.7	4.3 ± 0.8	
	3.4 ± 0.1	124.1 ± 1.7	44.6 ± 0.9	

large uncertainties larger than 100 %. Thus we use the two-outbreak model ($n = 2$) in these countries.

In the USA, there are many centers of outbreaks. In Fig. A2, we show the $dN(t)/dt$ data in New York, Massachusetts and California states [30]. It is possible to fit the data in each state by using the Gompertz function, but the results show significantly different values in (γ, t_0) . This would be the reason of the slow decrease of $dN(t)/dt$ in the USA. As a result, a single-outbreak treatment is not appropriate. By comparison, the $dN(t)/dt$ data are reasonably explained by two outbreaks ($n = 2$). We have used the data in Ref. [27] updated on June

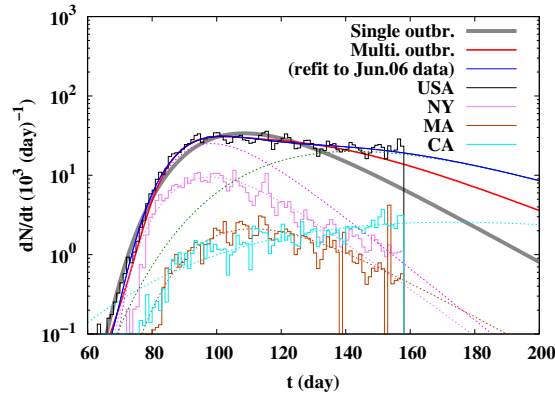


Fig. A2 The daily increase of the number of infected people $dN(t)/dt$ in the USA with single-outbreak (grey) and multiple-outbreak (red and blue) model analyses. The red and blue curves show the fitting results to the data till May 20 and June 6, respectively. Magenta, brown and cyan histograms (curves) show the data (fitting results in the single-outbreak model) in the New York, Massachusetts, and California states, respectively.

7, 2020. The daily number of newly infected people does not decrease and it seems that it takes more time for the settle down.

After obtaining (N_i, γ_i, t_i) by fitting dN/dt data, the constant part (ΔN) is obtained by fitting $N(t)$. Thus obtained multiple-outbreak functions in Eq. (A1) are compared with the data in Fig. A3. In the region with $N(t) > 100$, the multiple-outbreak functions are found to explain the data well. This supports the idea that the number of infected people in one outbreak would be described by the Gompertz function.

References

- [1] Michael Levitt, online lecture, <http://med.stanford.edu/levitt.html> (2020).
- [2] P. Castorina, A. Iorio, and D. Lanteri (2020), arXiv:2003.00507 [physics.soc-ph].
- [3] Lin Jia, Kewen Li, Yu Jiang, Xin Guo, and Ting zhao (2020), arXiv:2003.05447 [q-bio.PE].
- [4] Giuseppe Dattoli, Emanuele Di Palma, Silvia Licciardi, and Elio Sabia (2020), arXiv:2003.08684 [q-bio.PE].
- [5] D. Lanteri, D. Carco', and P. Castorina (2020), arXiv:2003.12457 [q-bio.PE].
- [6] José Miguel Ponciano, Juan Adolfo Ponciano, Juan Pablo Gómez, Robert D. Holt, and Jason K. Blackburn (2020), arXiv:2005.11201 [q-bio.PE].
- [7] Jiri Mazurek and Zuzana Nenickova (2020), <https://doi.org/10.13140/RG.2.2.19841.81761>.
- [8] Wuyue Yang, Dongyan Zhang, Liangrong Peng, Changjing Zhuge, and Liu Hong, medRxiv (2020), <https://doi.org/10.1101/2020.03.12.20034595>.
- [9] Ali Ahmadi, Yasin Fadaei, Majid Shirani, and Fereydoon Rahmani, medRxiv (2020), <https://doi.org/10.1101/2020.03.17.20037671>.
- [10] Yuri Tani Utsunomiya, Adam Taiti Harth Utsunomiya, Rafaela Beatriz Pintor Torrecilha, Silvana Cassia Paulan, Marco Milanese, and Jose Fernando Garcia, medRxiv (2020), <https://doi.org/10.1101/2020.03.30.20047688>.
- [11] Daren J Austin, medRxiv (2020), <https://doi.org/10.1101/2020.04.09.20059402>.
- [12] Jean Roch Donsimoni, René Glawion, Bodo Plachter, Constantin Weiser, and Klaus Wälde, medRxiv (2020), <https://doi.org/10.1101/2020.04.10.20060301>.
- [13] Thomas Klabunde and Clemens Giegerich, medRxiv (2020), <https://doi.org/10.1101/2020.04.14.20064790>.
- [14] Renato Rodrigues Silva, Wisley Donizetti Velasco, Wanderson da Silva Marques, and Carlos Augusto Goncalves Tibirica, medRxiv (2020), <https://doi.org/10.1101/2020.04.19.20071852>.
- [15] Carlos Maximiliano Dutra, medRxiv (2020), <https://doi.org/10.1101/2020.04.22.20074898>.
- [16] A M C H Attanayake, sanjeewa Perera, and Saroj Jayasinghe, medRxiv (2020), <https://doi.org/10.1101/2020.04.22.20074898>.

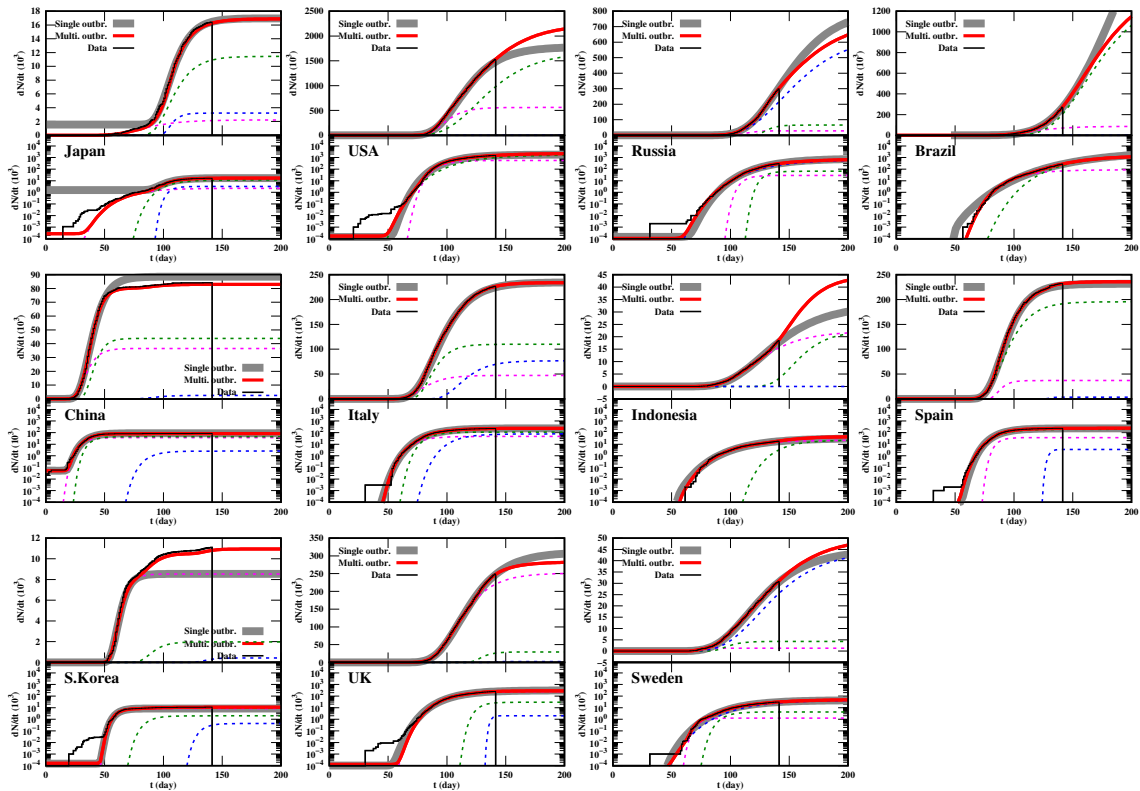


Fig. A3 The number of infected people $N(t)$ with single-outbreak and multiple-outbreak model analyses in the linear (top) and logarithmic (bottom) scales. Magenta, green and blue dotted curves show the contributions from the first, second and third outbreaks, respectively.

1101/2020.05.04.20091132.

- [17] Marti Catala, Sergio Alonso, Enrique Alvarez Lacalle, Daniel Lopez, Pere-Joan Cardona, and Clara Prats, medRxiv (2020), <https://doi.org/10.1101/2020.05.13.20101329>.
- [18] Rhodri P Hughes and Dyfrig A Hughes, medRxiv (2020), <https://doi.org/10.1101/2020.05.15.20102764>.
- [19] Christophe Z Z Guilmoto, medRxiv (2020), <https://doi.org/10.1101/2020.05.17.20097410>.
- [20] Nicola Bartolomeo, Paolo Trerotoli, and Gabriella Serio, medRxiv (2020), <https://doi.org/10.1101/2020.05.20.20108241>.
- [21] D Dagon, C Zou, W Lee, Proc. 13th Annual Network and Distributed System Security Symposium (NDSS'06), Feb., pp 235-249, 2006.
- [22] B. Gompertz, Philosophical Transactions of the Royal Society of London, **115**, 513 (1825).
- [23] A. K. Laird, British Journal of Cancer, **13**, 490 (1964).
- [24] K. Ohishi, H. Okamura, and T. Dohi, Journal of Systems and Software, **82**, 535 (2009).
- [25] R. Chawla and M. Kaur, Adv. High Energy Phys., **2018**, 5129341 (2018).
- [26] T. Nakano and Y. Ikeda, medRxiv (2020), <https://doi.org/10.1101/2020.04.25.20080200>.
- [27] Coronavirus Source Data, <https://ourworldindata.org/coronavirus-source-data> (2020).
- [28] N. Ikeno, A. Ono, Y. Nara, and A. Ohnishi, Phys. Rev. C, **93**, 044612 (2016) [Erratum: Phys. Rev. C, **97**, 069902(E) (2018)].
- [29] Y. Nara, N. Otuka, A. Ohnishi, K. Niita, and S. Chiba, Phys. Rev. C, **61**, 024901 (2000).
- [30] The COVID Tracking Project, <https://covidtracking.com/> (2020).