

COVID 19

Estimate Exit of Crisis

The Gompertz Function $y = Ka^{b^x}$ may help you estimate such an exit date of the COVID 19 crisis for a region where certain data with statistics on the cumulative number of infected patients are available. This report exemplifies cases from New Zealand, China, several European countries, the United States, Japan, etc., and the corresponding Gompertzian curve shows the status of novel coronavirus infections with explosive expansion to date. Also expect end date ?

20 May, 2020

Reported by:

Kunio Hirotsu

Syngas Associates FE ("SAFE")

Narashino-shi, Chiba, Japan



SAFE

Syngas Associates FE
Consultant of Syngas Technology

COVID 19 – When will the crisis end?

INTRODUCTION

As of 19 May 2020, the cumulative number of COVID 19 patients have crossed 4.7 million worldwide with counting 317,000 deaths. Can someone accurately predict when this pandemic crisis will end?

A Singapore-based University of Technology and Design (“SUTD”) has predicted the end of the coronavirus pandemic in 131 countries with the help of artificial intelligence (“AI”). The SUTD’s study used data based on the SIR (Susceptible-Infected-Recovered) model to predict the key dates of transition during the coronavirus lifecycle across the world¹⁾.

In this connection, Jianxi Luo²⁾ cited the following researchers who have adopted classic or latest data science and AI techniques to forecast the ending of COVID 19 since the outbreak in January 2020.

- Institute of Health Metrics and Evaluation (IHME) at University of Washington
- MRC Centre for Global Infectious Disease Analysis at the Imperial College London
- University of Texas COVID-19 Modelling Consortium
- MIT IDSS Isolate

Note that, as far as we refer to the cited papers by Jianxi Luo, all of the above researchers do not use classic models anymore but either the SIR model or the SEIR (Susceptible-Exposed-Infected-Removed) model³⁾.

Jianxi Luo pointed out that, even the most noticeable forecasting method from the IHME, it has been found with model issues and large errors between predicted data and later coming actual “future” data. Then he added that researchers are learning and improving the methods and tools to make more accurate predictions.

SAFE applies a specific classic model for the subject purpose of predicting the end of COVID 19. In doing so, we pay close attention to the error between the correlation equation and the actual data collected. We, chemical engineers, are good at finding the most appropriate formulas available today. Finding the theoretical implications of applying them is also a fun task. It is further important to use the simplest formula possible so that everyone can easily understand and work with it.

GOMPERTZ FUNCTION

The Gompertz function is a useful mathematical model for a time series analyses and is named after Benjamin Gompertz (1779-1865)⁴⁾. It is a classical model which has been widely used in many aspects of biology, and frequently used to describe the growth of animals and plants, as well as the number or volume of bacteria and cancer cells⁵⁾. Keiji Matsui, MD, PhD, confirms that cancer growth obeys Gompertzian model⁶⁾.

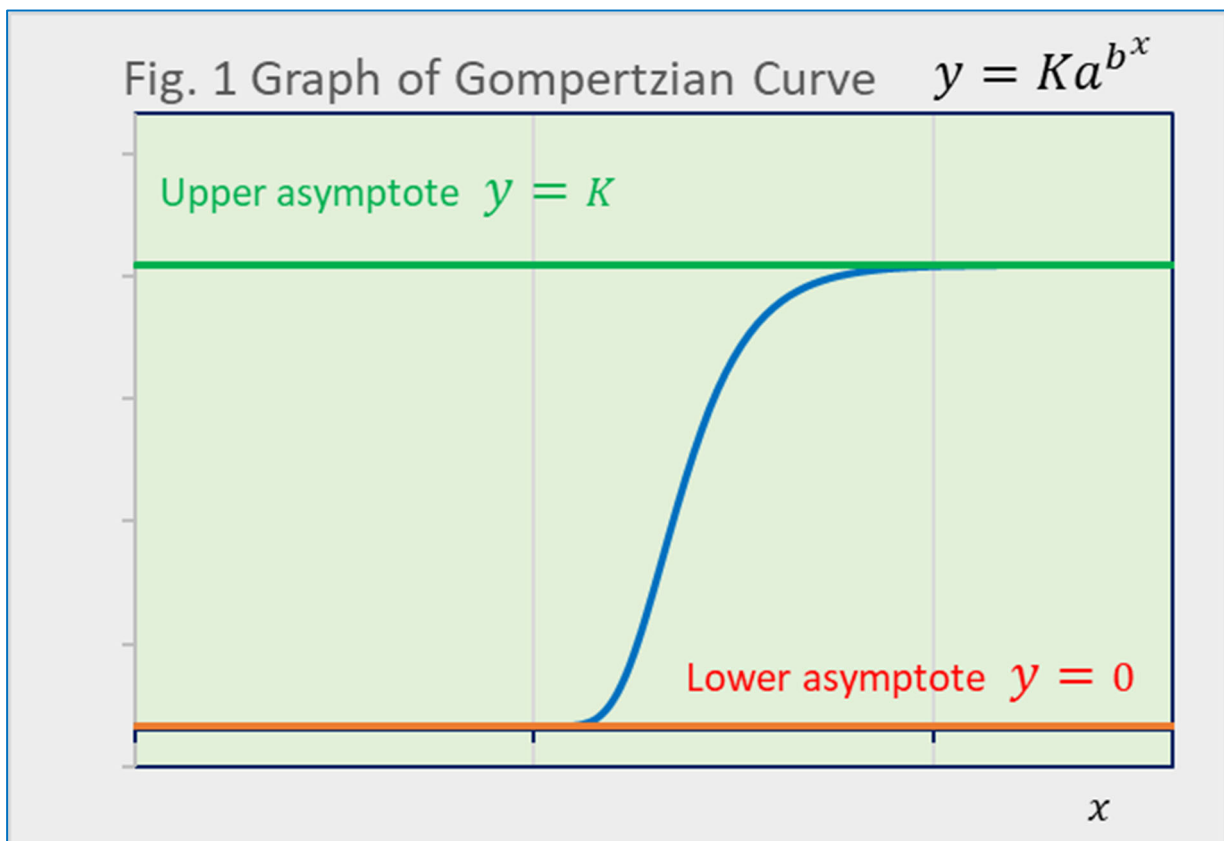
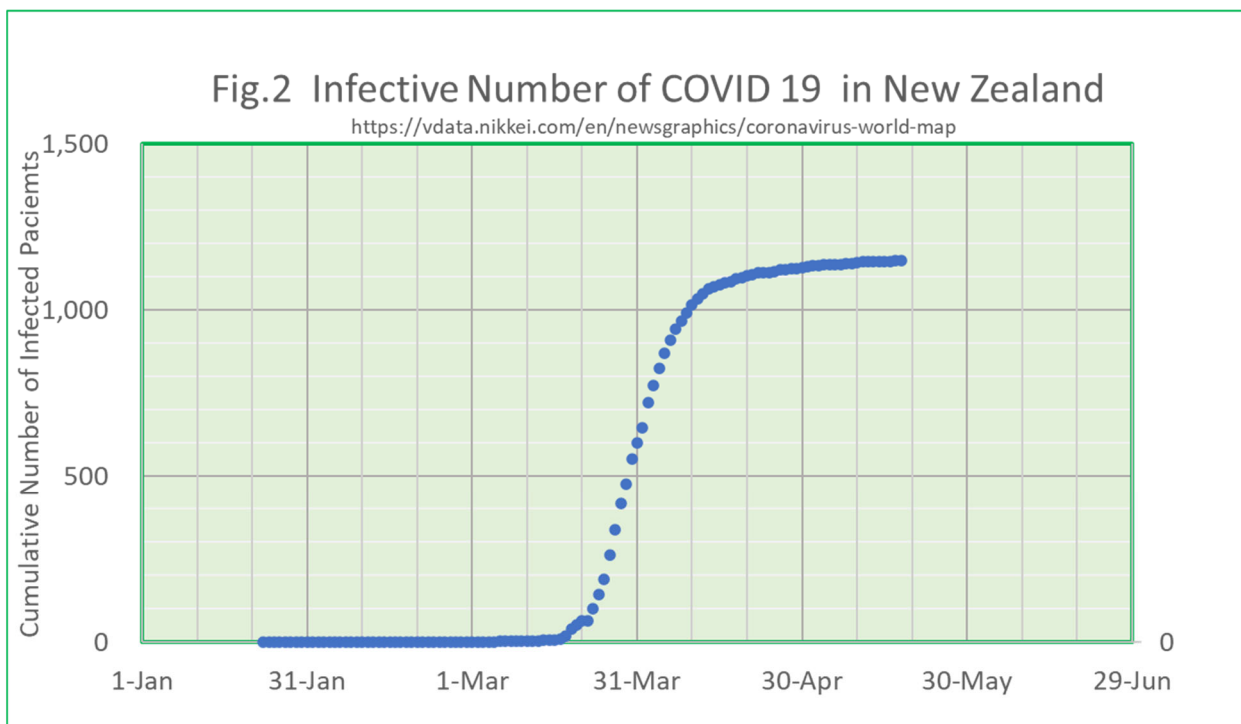


Fig. 1 above expresses a curve corresponding to a standard formula of the Gompertz function, $y = Ka^{bx}$ (subject to $0 < a < 1$ and $0 < b < 1$), where an upper asymptote is K and a lower asymptote is 0 as follows.

$$\lim_{x \rightarrow \infty} Ka^{bx} = Ka^0 = K$$

$$\lim_{x \rightarrow -\infty} Ka^{bx} = 0$$

At first, let us have a look at the records of the cumulative number of infected patients from this January to May for New Zealand in Fig. 2 below. The country's Prime Minister Jacinda Ardern described almost five weeks of the strict level 4 lockdown measures as "the strictest constraints placed on New Zealanders in modern history", and on Tuesday, 27 April 2020, the country eased into a less restrictive lockdown, with 400,000 more New Zealanders heading back to work and 75% of the country's economy operating. There is a series of useful data available from the beginning to the end of the COVID 19 epidemic.



The daily records of cumulative number plotted in Fig. 2 keeps quite a similarity to Fig. 1 in the character of formula ("S-curve"). The

cumulative value does not decrease with the passage of time. This property leads to considerably small error as average value, and the more cumulative number, or the more time, the smaller error. It becomes effective statistical data for such a case as pandemic or epidemic where there is large fluctuation in the daily data. The corresponding curves

- are sandwiched between two asymptotes,
- draw a monotonically increasing curve with the inflection point near the midpoint,
- when it starts to increase, rapidly increases up to the inflection point, and starts to decrease rapidly after the inflection point (the differential takes the maximum value at the inflection point), and
- asymmetric with respect to the inflection point.

By replacing the variables with $x = t - c$ and $y = \mathbf{CNP}_{1M}(t)$ our Gompertz function is expressed as follows with t counting the number of days since a certain designated day.

$$\mathbf{CNP}_{1M}(t) = \mathbf{Ka}^{b^{t-c}}$$

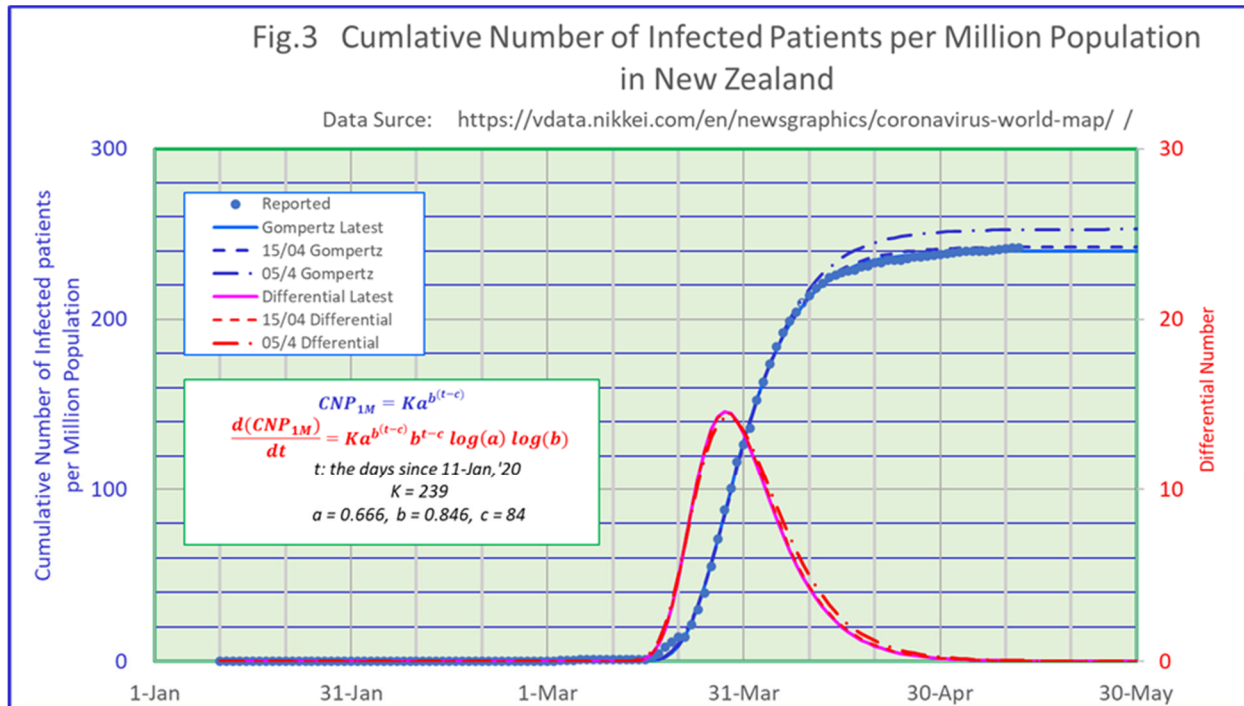
The differential of the above Gompertz function is also derived hereto as follows.

$$\begin{aligned} d\mathbf{CNP}_{1M}(t)/dt &= \mathbf{Ka}^{b^{t-c}} b^{t-c} (\ln a) (\ln b) \\ &= \mathbf{CNP}_{1M}(t) b^{t-c} (\ln a) (\ln b) \end{aligned}$$

The function $\mathbf{CNP}_{1M}(t)$ is standardized as expression of Cumulative Number of Infected Patients per million population (“ \mathbf{CNP}_{1M} ”) for future comparisons between nations or cities or other regional units in the world having differences of population.

As to New Zealand, Fig. 3 is provided above with the Gompertzian fitting curve of $\mathbf{CNP}_{1M}(t)$, for which the coefficients of equation, \mathbf{K} , \mathbf{a} , \mathbf{b} and \mathbf{c} , are all determined by the least square method, together with its differential curve, $d\mathbf{CNP}_{1M}(t)/dt$. In fact, the Gompertzian curve seems to be excellent in fitting the reported data. The maximum value of the

differential curve can be evaluated not only as the position of the inflection point is accurately recognized but also as a numerical value indicating the infectivity (infective strength). It also helps predict the end of the epidemic by analyzing the decay after the inflection point.

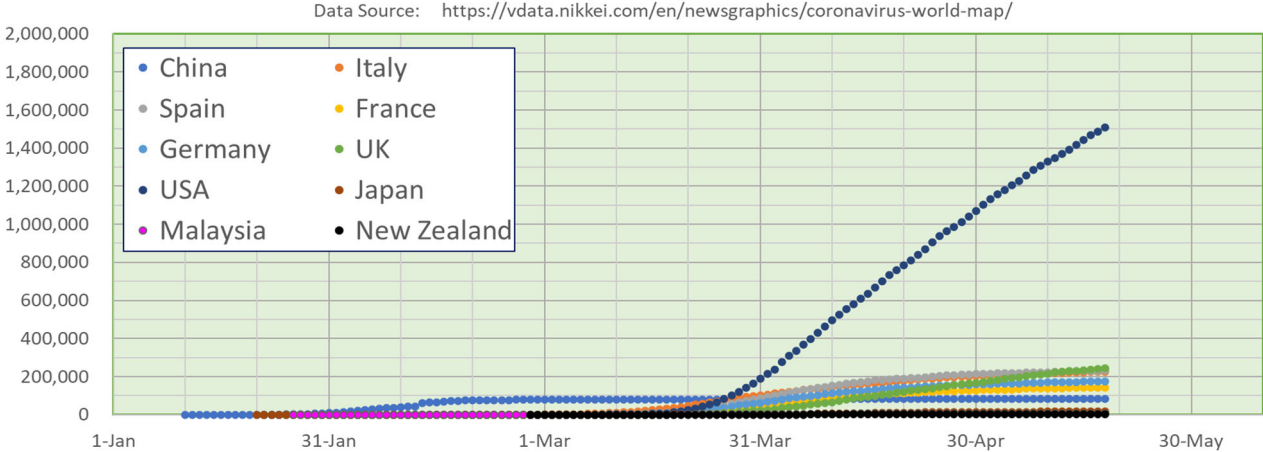


The Gompertzian curve up to 5 April, at just a few days of time after the inflection point, shows some decreasing change. The another data up to 15 April shows that the coefficient K decreases compared to that of up to 5 April, and the data up to 18 May (“latest” at the time writing this report) shows almost no decrease in K . This fact can be used as an end date criterion. The result of the New Zealand is summarized in a table in Example-1 together with others. Suppose a set of data up to the date beyond the inflection point be available, expect that the end date can be predicted with acceptable accuracy?

DETAILS OF GOMPERTZIAN CURVE FITTINGS

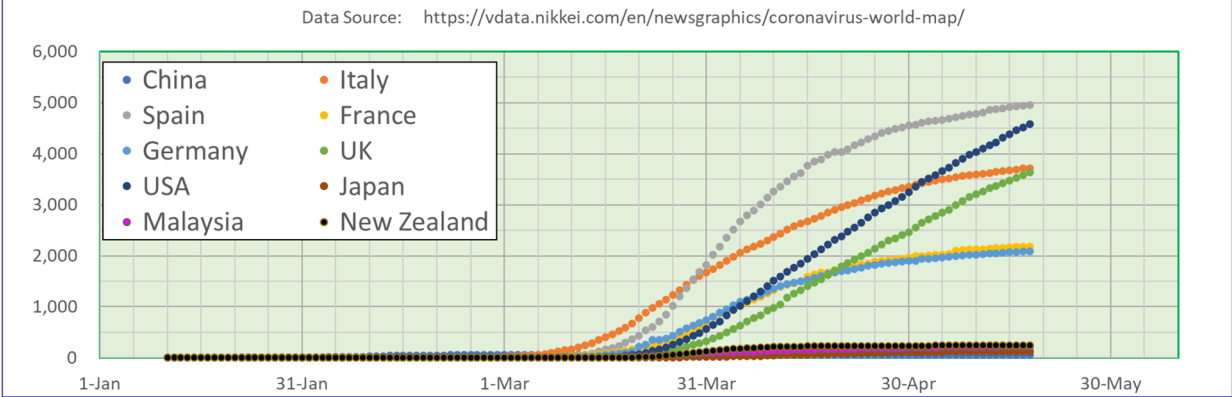
Fig. 4 shows the actual value of the cumulative number of infected patients reported on each country.

Fig. 4 COVID 19- Cumulative Number of Infected Patients



The numbers in Fig.5 are obtained by dividing it by the country's population and multiplying by one (1) million. The actual number of infected patients in the United States in Fig.4 is significantly higher than in other countries, but the United States has the second highest after Spain, for the case of CNP_{1M} number per million population in Fig.5 having crossed Italy at the beginning of May. In the standardized diagram of Fig. 5, you can see that, unlike the imagination expected from Fig. 4, the rate of explosive infections in the United States is more rapid that of European countries except for the United Kingdom only the country with similar behavior to the United States.

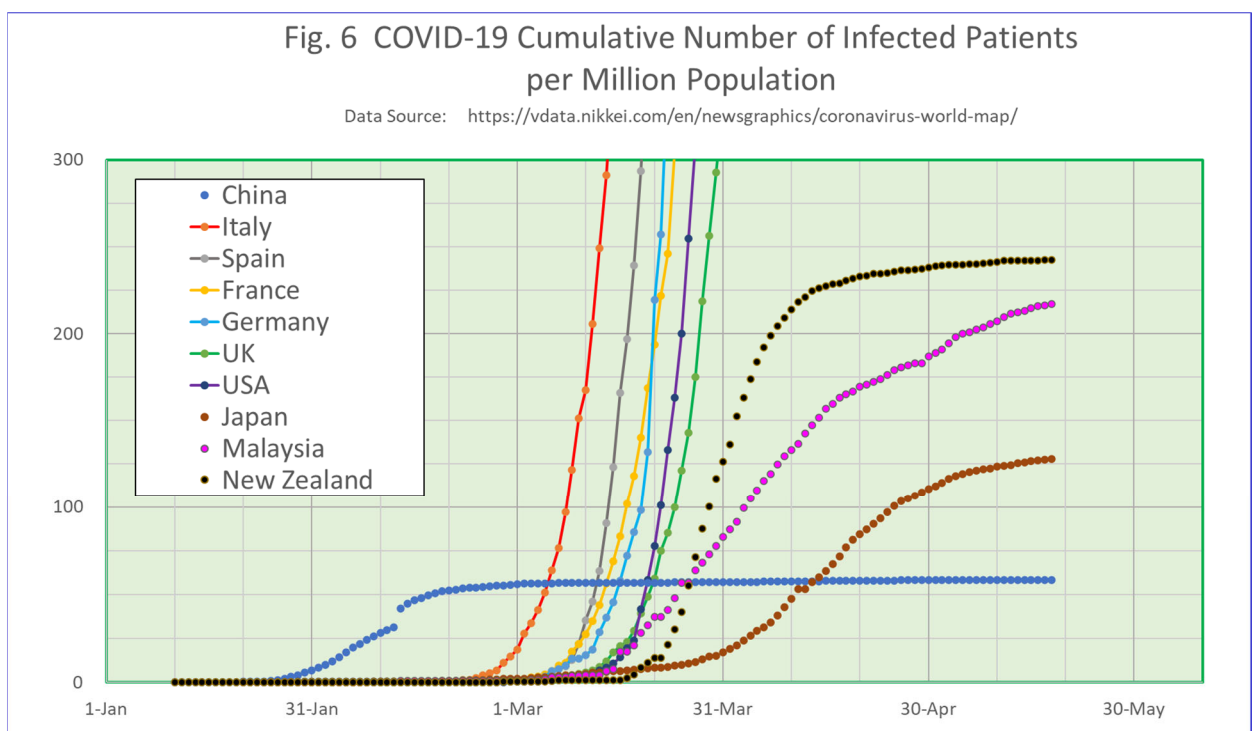
Fig. 5 COVID-19 Cumulative Number of Infected Patients per Milliom of Population



It is understood that the cited United States and all European countries began an outbreak of explosive infection with a similar slope very shortly

after a certain delay period. Among them, according to Fig. 5, the data of latest 10-20 days of France and Germany seem to, or probably Italy and also Spain, show decaying to the ending date while the United Kingdom and the United States seem not.

Fig. 6 is an enlarged graph of the vertical axis in Fig. 5 with the scale 0~300 and also depicts unique situation of Asia Pacific region, i.e. China, Japan, Malaysia, and New Zealand.



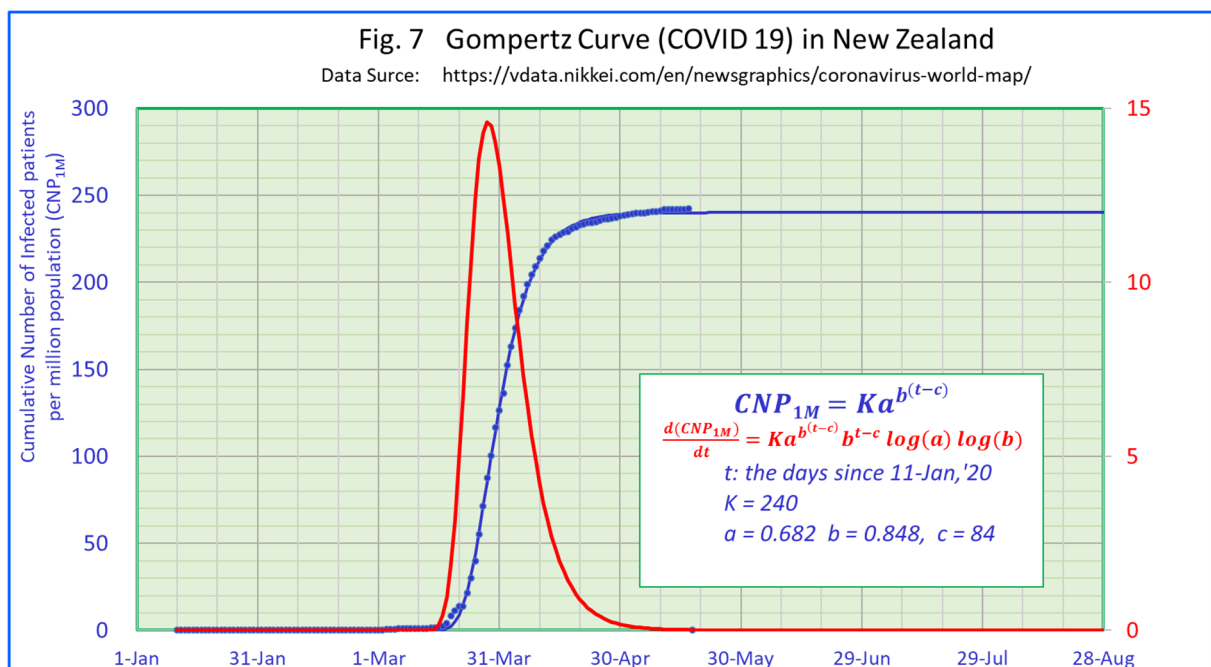
China and New Zealand seem to have already on a promenade to exit the pandemic crisis. China ordered lockdown at early stage quite effectively. New Zealand as mentioned above. Malaysia and Japan will be expected soon to the same road.

The figures in the Examples 1 & 2 below are the results of determining the coefficients of Gompertzian curve for each country. The same scale pairs are used in each example for ease of comparison.

Example 1: China and Asia Pacific

In this region, an explosive COVID 19 infection first occurred in China in January 2020, in Malaysia in the first half of March and in Japan in March consecutively. The same thing happened in New Zealand in late March.

Fig. 7 is the same one as Fig. 3 except for scales of all the three (3) axes for the graphs updated as of 18 May. No other combination of the four (4) coefficients could represent the Gompertz equation fittings of the New Zealand data. As already mentioned, New Zealand was relaxed on 27 April with a loosely closed blockade. The corresponding value of a Gompertzian differential curve on 27 April is calculated as "0.31 patients per million people per day". In addition, the number "1.5 ($\approx 0.31 \times 4.7$) patients per day" is simply derived referring to the population 4.7 million of New Zealand, as an "end of epidemic" judgement specific in this country. The numerical validity itself must be discussed; whether "suitable for such a judgement or not? Who did judge it? How theoretically?", etc.



Anyway, this evidence in the New Zealand epidemic of COVID 19 leads to a guideline how to define the ending date of the epidemic. In this report SAFE defines the end with the minimum value of t_e that satisfies

the following formula with a condition $t_e > c$. To avoid being too optimistic under environment surrounding COVID 19, instead of the above-mentioned value 0.31, we set the criterion value at 0.2 (35% lower) as follows to determine the ending date of epidemic on a differential curve.

$$dCNP_{1M}(t_e)/dt = Ka^{b^{te-c}} b^{te-c} (\ln a)(\ln b) \leq 0.2$$

The ending date for New Zealand, as of the records up to 18 May 2020, is 30 April according to the above criterion. Practically, it is easy to calculate $dCNP_{1M}(t)/dt$ vs. t in an Excel table with the above equation and use “goal seek method” simply to find a variable t_e for the goal, $dCNP_{1M}(t_e)/dt = 0.2$, executable on such a Table-1 as referred to below. There is currently no rationale for “whether or not” which is accurate, but the criterion is that if the number of patients has not increased by more than 1 per 5 million people per day, an epidemic might be considered as an end. Our simulation and the suggested criterion may be able to almost support New Zealand’s decision.

Table-1		$Df Eq(t_e)=$	0.2	Round up	Start-up=	11/01/2020	by Goal Seek with Excel
		$t_e=$	110.2	111	End=	30/04/2020	
With data up to	K	a	b	c	Date# for t_e	End Date	$Df Eq(t_e)$
05/04/20	252.8	0.602	0.857	83	113	02/05/2020	0.193
15/04/20	242.4	0.663	0.849	84	111	30/04/2020	0.189
18/05/20	240.0	0.682	0.848	84	111	30/04/2020	0.174

In Table-1, three (3) samples are summarized for a result simulating the ending date in New Zealand equivalent to the above factor 0.2, using the cumulative data up to 05/04/2020, 15/04/2020 and 18/05/2020 respectively.

5 April is the date that the Gompertzian curve quickly began to converge. Since that date the estimate has become stable soon with accuracy of plus minus a few days (this difference could not be viewed on the figure but only be obtained by assigning a numerical value to the Gompertz

function) and with the cumulated data up to 18 May already exceeded the convergence value on the 30 April data, being able to represent the epidemic case in New Zealand very well.

Fig. 8 and Table-2 introduces the epidemic status in Malaysia similarly. Malaysia has a Gompertz function with a gentler differential curve than New Zealand. It means that speed of infection is slower and rather will have later ending date than New Zealand. Actually the Gompertzian curve indicates the ending date to be 6 June. According to a range of bell shape in the differential curve of the Gompertz function, the infective period continues more than 3 months in Malaysia, and twice longer than in New Zealand. The height of bell shape seems to represent the infective strength which is 1/3 of New Zealand.

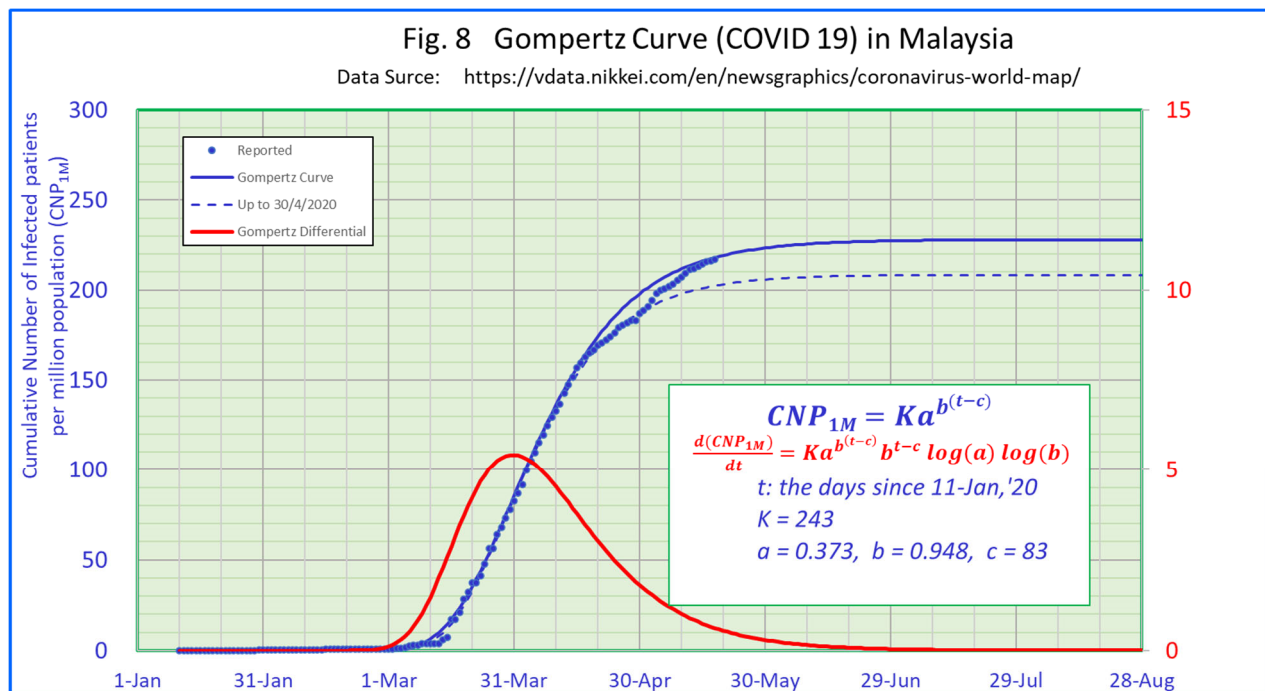
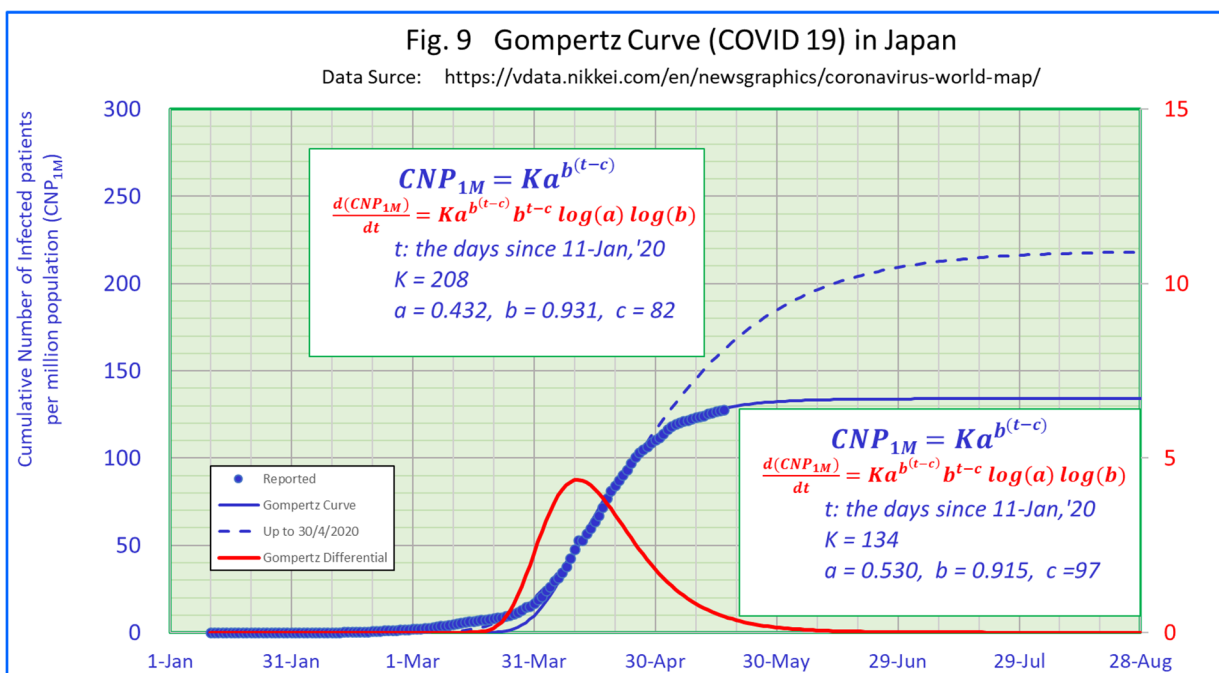


Table-2		$DfEq(t_e)=$	0.2	Round up	Start-up=	11/01/2020	by Goal Seek with Excel
		$t_e=$	147.0	148	End=	06/06/2020	
With data up to	K	a	b	c	Date# for t_e	End Date	$DfEq(t_e)$
30/04/20	208.4	0.532	0.931	86	140	29/05/2020	0.195
09/05/20	218.1	0.445	0.933	83	143	01/06/2020	0.191
18/05/20	243.0	0.373	0.948	83	148	06/06/2020	0.189

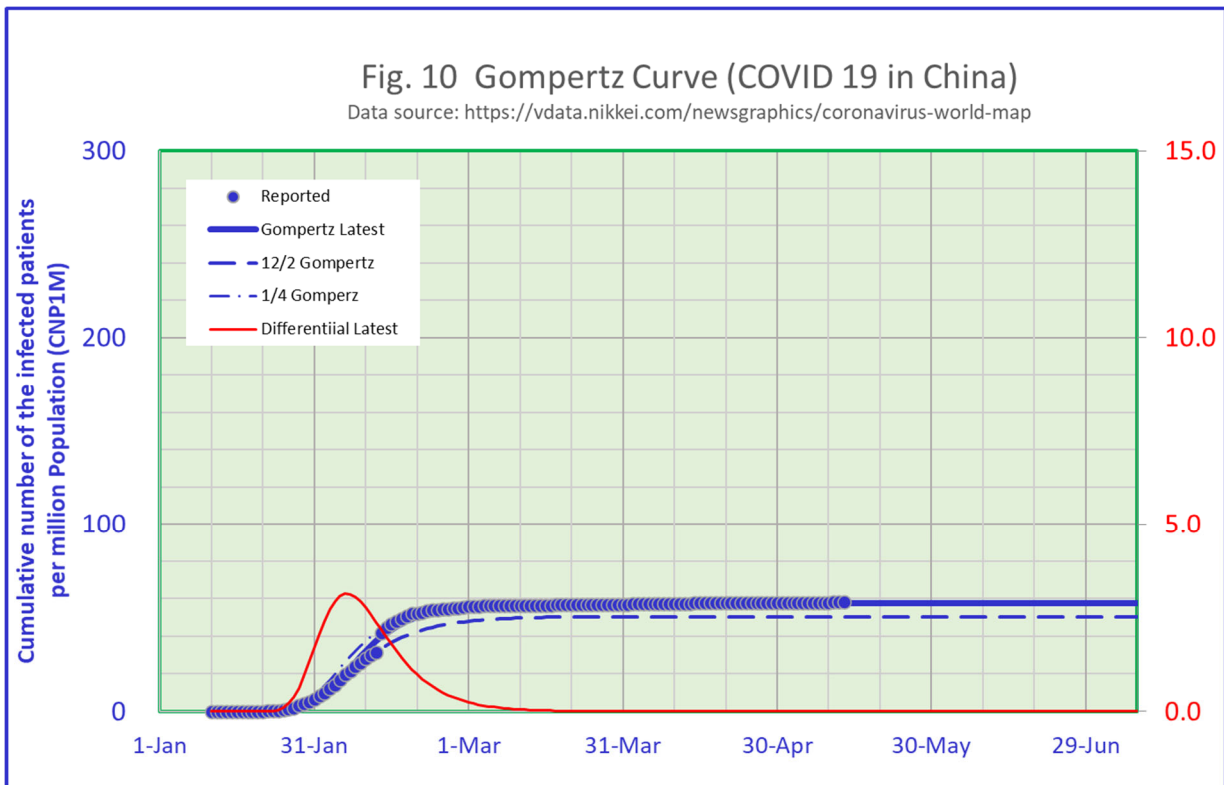
However, there is a part of concern in the range of the latest one (1) month in the Gompertzian curve. Around 20 April, the curves appeared to have begun to converge. The dashed line in Fig. 8 is the Gompertzian curve corresponding to the data set by 30 April, and we predict that the epidemic will end on 29 May, as shown in Table 2. But the curve has changed several times. According to the data up to 9 May and the data up to 18 May, the curve is back to what it was before 20 April. Is the epidemic recurring, or is it wrong counting the number of infections?

Fig. 9 refers to Japan where the Gompertzian curve shows essentially same characteristics as those of Malaysia, i.e. slow speed, lower infectivity and longer period of infection. According to a similar simulation with a Gompertz differential function, the epidemic end will be 16 June 2020.



Note that, if a simulation is executed with a set of data up to 30 April, the Japan epidemic end becomes 16 July as a result following the dashed curve in the figure. Regarding the “COVID 19 Emergency Declaration of Infectious Diseases” issued by the Japanese government on 7 April, the

target area was expanded to all prefectures on 16 April 2020. It may be interpreted that the majority of the Japanese people have made efforts to refrain from going out, etc., and the Gompertzian curve began to converge from 30 April, after so-called "two weeks" time, considering the incubation period of infection.



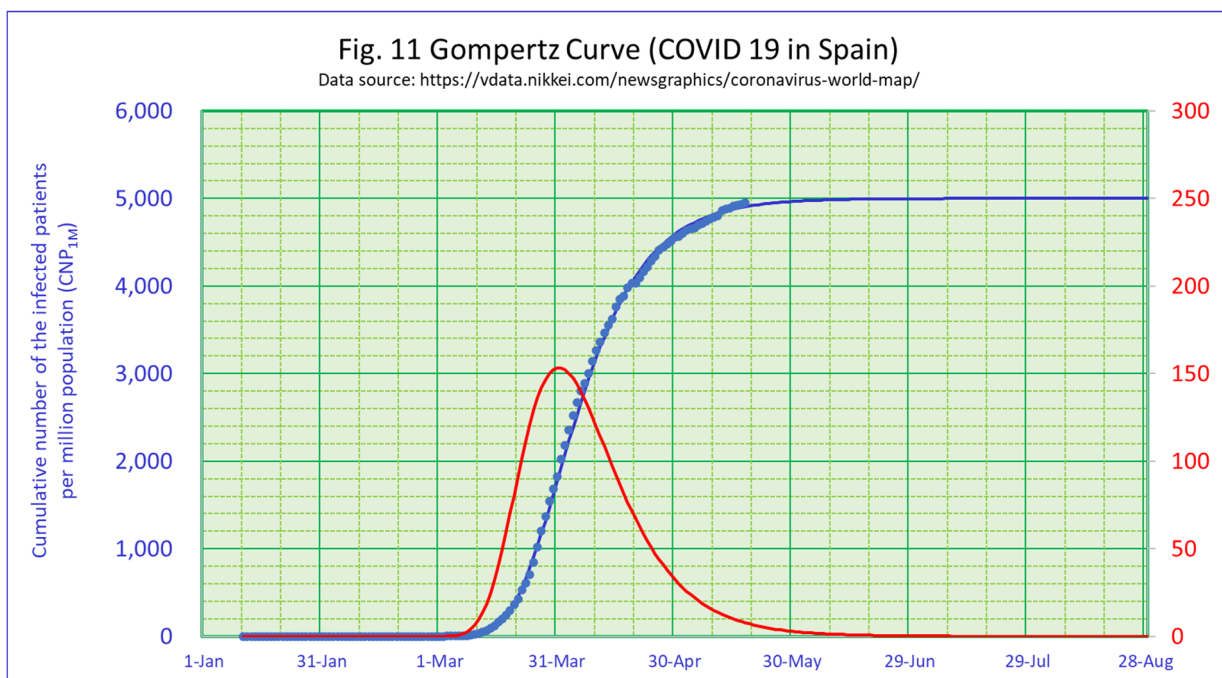
As shown in Fig. 10, all data that was not collected before 13 February was added once in the next day in China. Therefore, we ignored all records of cumulative data on and prior to 12 February and plotted a Gompertzian curve that directly connects the data on the start date of the record and the data on 13 February. For reference only, the cumulative data up to 12 February are also shown, and the corresponding curve is added with dashed line from there.

China appears to be faster infection than New Zealand but less infectivity and shorter infective duration, as expressed in terms of CNP_{1M} number per million population.

Example 2: Europe and North America

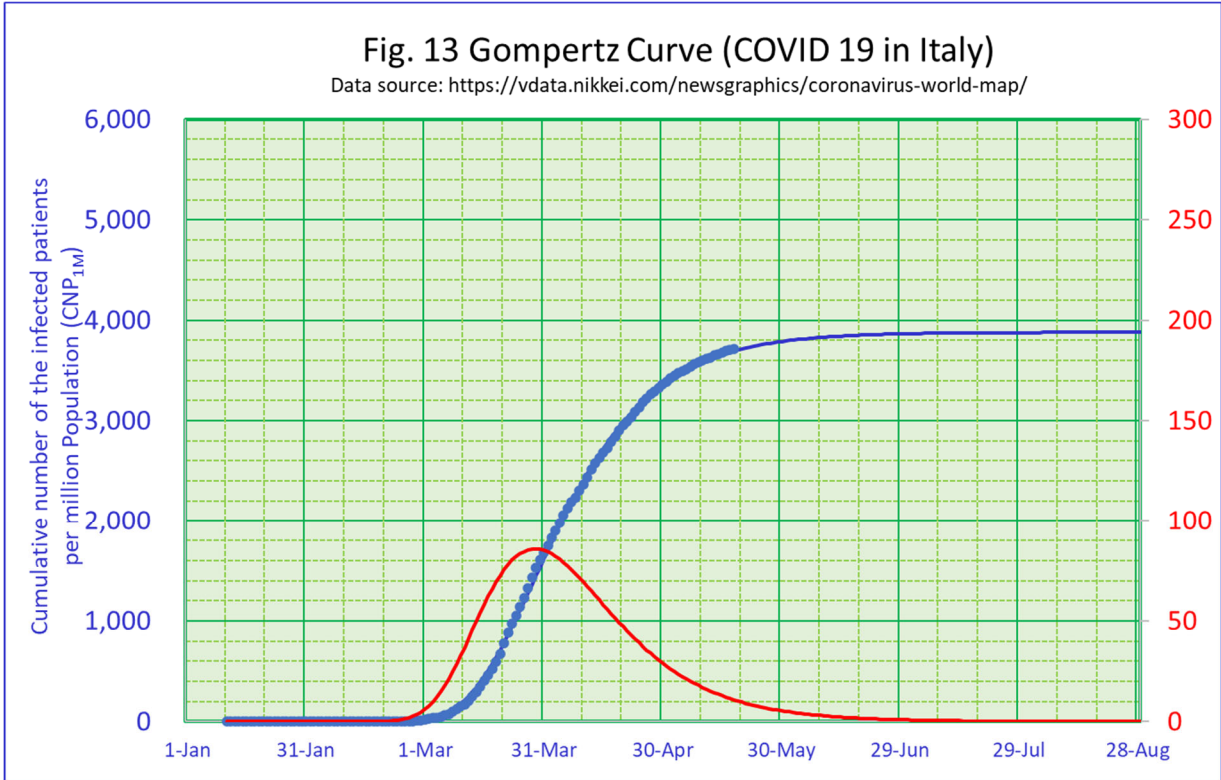
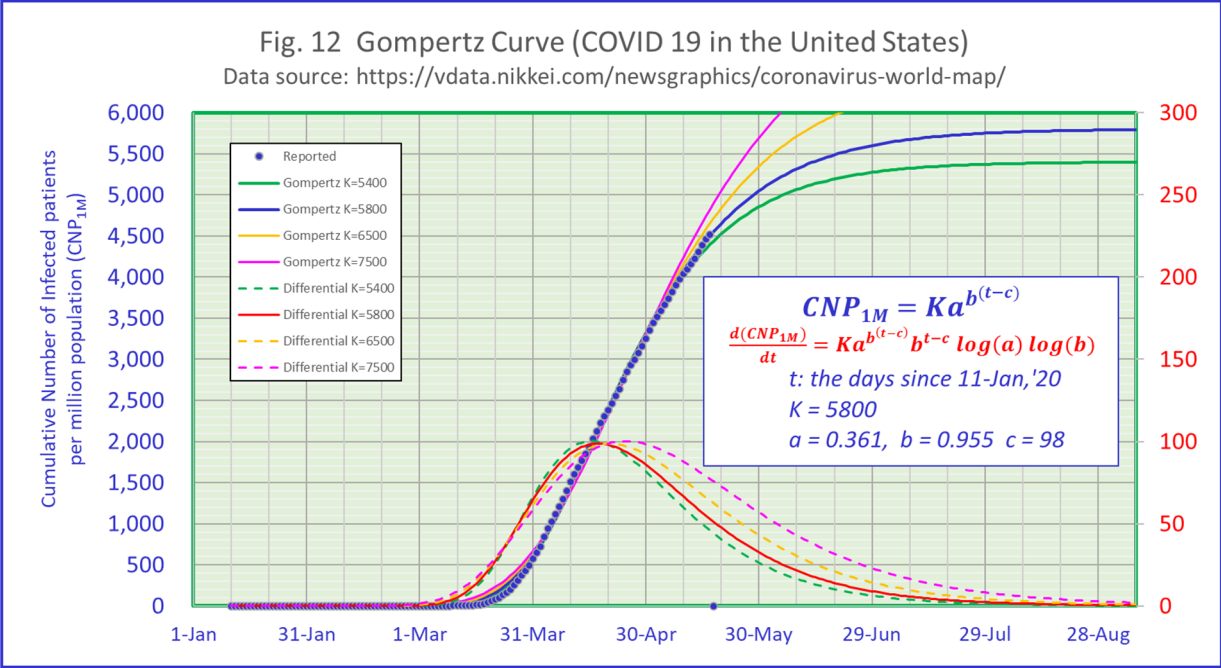
In Europe and the United States, the infective characteristics of COVID 19 are somewhat different. The explosive infection began in early March in Spain almost at the same time as Italy, France, Germany. The explosive infection continued to happen a bit later in the United States in addition to the United Kingdom.

Fig. 11 shows the situation of Spain where the cumulative number of infective patients per million population is worst in the world, as of 18 May 2020. It is observed in Spain that the infectivity is more than three (3) times larger than Malaysia or Japan while the infective period will be continuing same as these countries about three months. The infectious period would end in the beginning of July 2020 in Spain.



In general, the large infectivity is own characteristics of COVID 19 in Europe and the United States. However, compared to Spain, it is still difficult for the United States to predict the end of a pandemic, as Figure 12 shows. The Gompertzian curve has not yet begun to change towards

the end date. Here, for the data of the United States, we added four (4) Gompertzian curves with different values for K , the asymptotic value, namely $K=5400, 5800, 6500$ and 7500 , respectively. The corresponding coefficients a, b and c are adjusted so that the reported values can be approximated sufficiently. It seems us probable to follow a curve for $K = 5800$.



The infectivity is equivalent to the peak value of the differential curve. For the United States it is 2/3 of Spain, and the infective period is one (1) month longer than Spain. If we dare to predict, the end of pandemic in the United States would be after 20 September 2020.

Fig. 13 shows the Gompertzian curve for Italy. Same as Spain the curve has just begun to converge toward the end, but the convergence is milder than Spain due to less infectivity. It will end on 25 July.

Fig. 14 deals with the United Kingdom. Equally same as the United States, a Gompertzian curve has not yet begun to change towards the end date. COVID 19 will end on 13 September, according to forecasts with the same accuracy as in the United States.

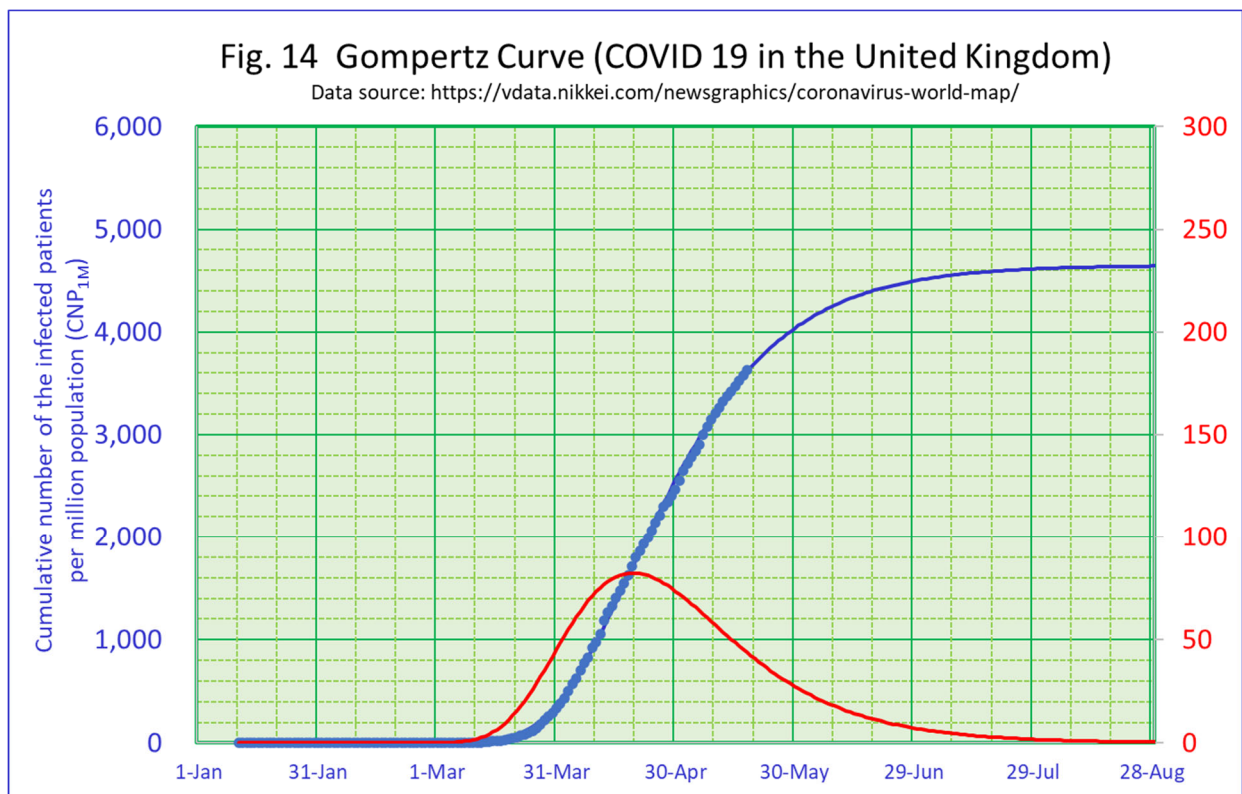
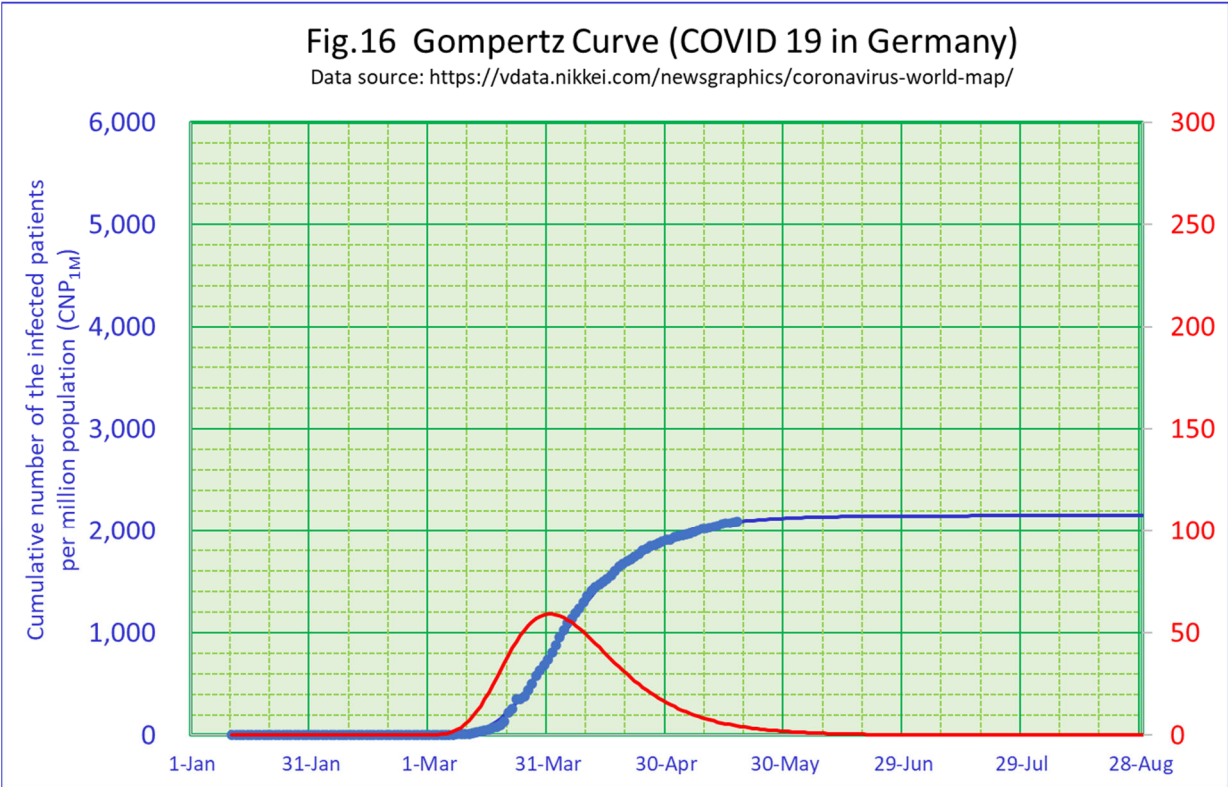
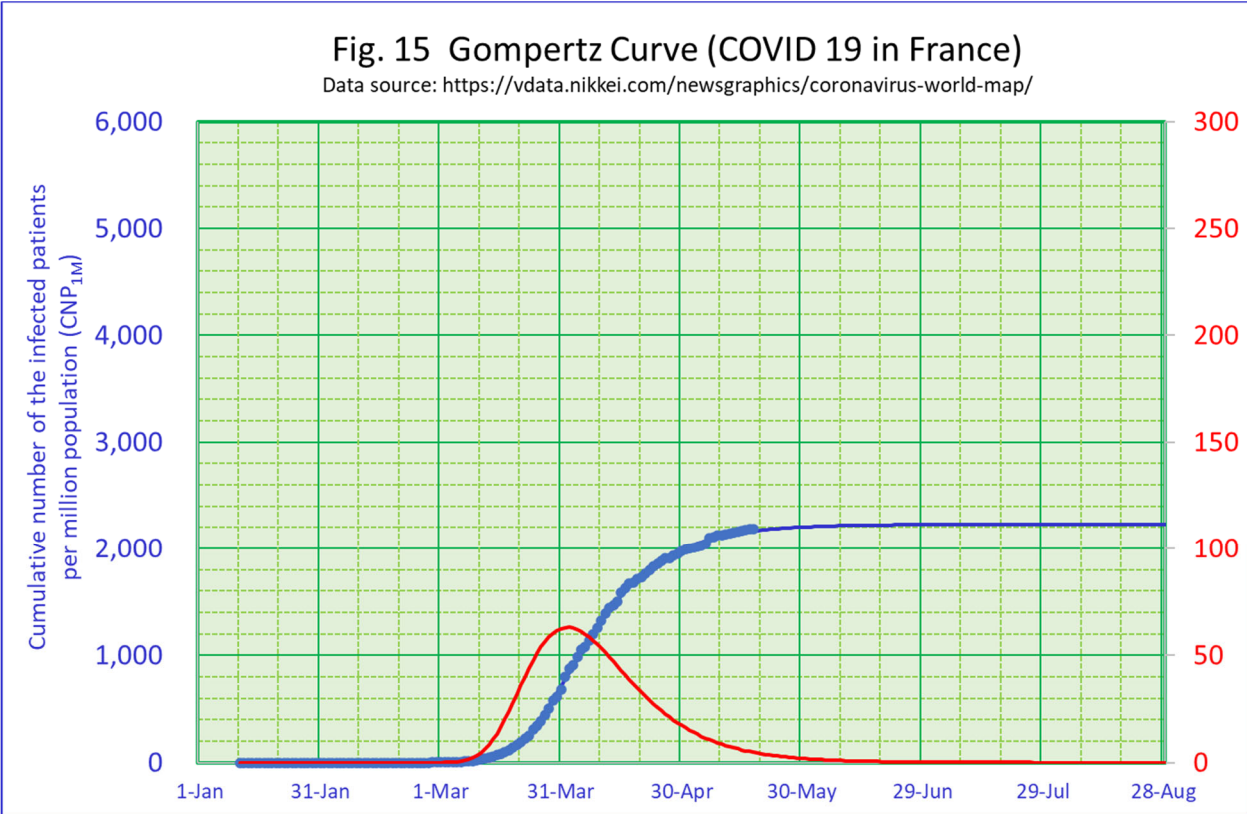


Fig.15 and Fig. 16 are respectively for France and Germany. The both curves are exactly same as if they are a twin. According to the Gompertzian curve for the both countries, the end date is 29 June in France and 30 June Germany.



In Table-3 the coefficients of the Gompertzian curve for each country exemplified in Example 1 and Example-2 are summarized below with each estimated date of ending infection. The maximum value of each

differential in the Gompertz curve is entered in the rightmost column. Assuming this to be the numerical value of the infectivity as mentioned before, it is considered that the infection was strong in the order of Spain, the United States, Italy, and the United Kingdom as a relative evaluation.

Table-3		$DfEq(t_e)$	0.2	Round up	Start-up=	11/01/2020	Differential Data	
Update: 19/5/2020		$t_e=$			End=			
With data up to	K	a	b	c	Date# for t_e	End Date	$DfEq(t_e)$	$DfEq(t_M)$
China	58	0.457	0.862	29	53	03/03/2020	0.187	3.2
Italy	3879	0.482	0.942	84	197	25/07/2020	0.191	85.8
Spain	5000	0.505	0.920	86	174	02/07/2020	0.185	153.3
France	2228	0.540	0.926	89	171	29/06/2020	0.193	63.0
Germany	2150	0.460	0.928	85	172	30/06/2020	0.187	59.1
UK	4650	0.258	0.953	94	247	13/09/2020	0.194	82.2
USA	5800	0.361	0.955	98	254	20/09/2020	0.200	98.8
Japan	134	0.530	0.915	97	158	16/06/2020	0.195	4.4
Malaysia	243	0.373	0.948	83	148	06/06/2020	0.189	5.4
New Zealand	240	0.682	0.848	84	111	30/04/2020	0.174	14.6

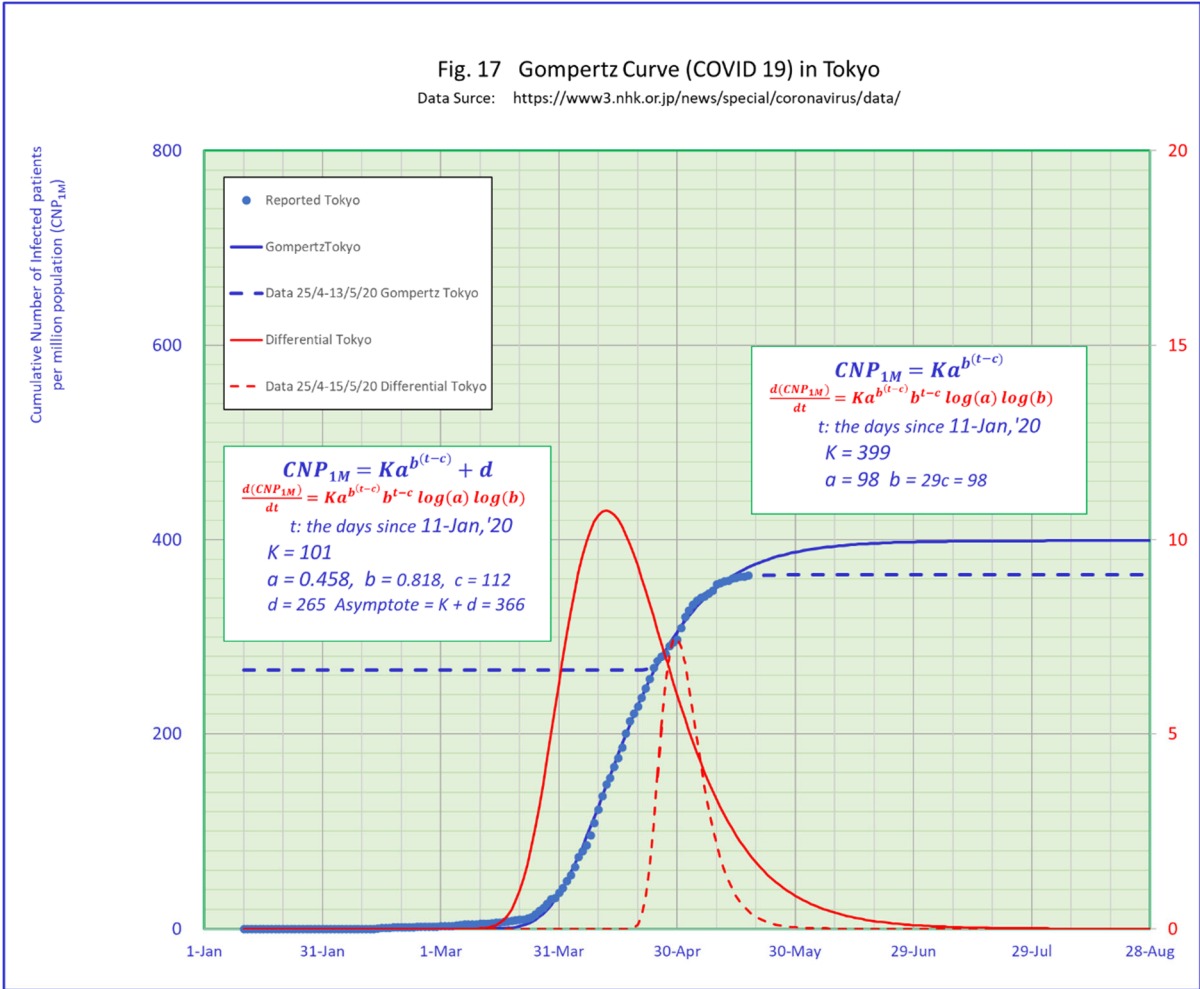
Example 3: Tokyo

Fig. 17 shows the epidemic of Tokyo. The data in the latest 10 days seems to indicate a convergence to the ending date. The final bend, which marks the beginning of convergence, appears to be a sharp curvature. Unfortunately, we could not find a Gompertzian curve that could fit the data up to that point and at the same time cover this bend. Therefore, the dashed line is just a Gompertzian curve that represents only the last part from 25 April to 18 May.

Fig. 18 is an enlarged view of the horizontal axis for this part, and Table-4 examines what was happening substantially on epidemic in Tokyo. The solid line is the result of the entire series of data through 18 May, with the end of the epidemic in Tokyo ending on 20 June, while the ending date following the dashed curve is 23 May. It seems that the effectiveness of the "COVID 19 Emergency Declaration of Infectious

Diseases" being expanded all over Japan on 16 April 2020, is also confirmed here. Many of us have had the longest "Golden Week" vacation in history, but this year we were only at home from Saturday, 25 April to Sunday, 10 May. The number of people using the Shinkansen dropped dramatically at Tokyo Station during the same period.

However, the question is whether it can be determined that the Gompertzian curve changed actual expression from a solid line to a dashed line between 25 April and 9 May. Since a small change on the graph will improve the forecast for the end of the epidemic by a scale of one month, the judgement must be scientifically reliable. We think we would have to prove the validity of such a way of analysis in the next step.



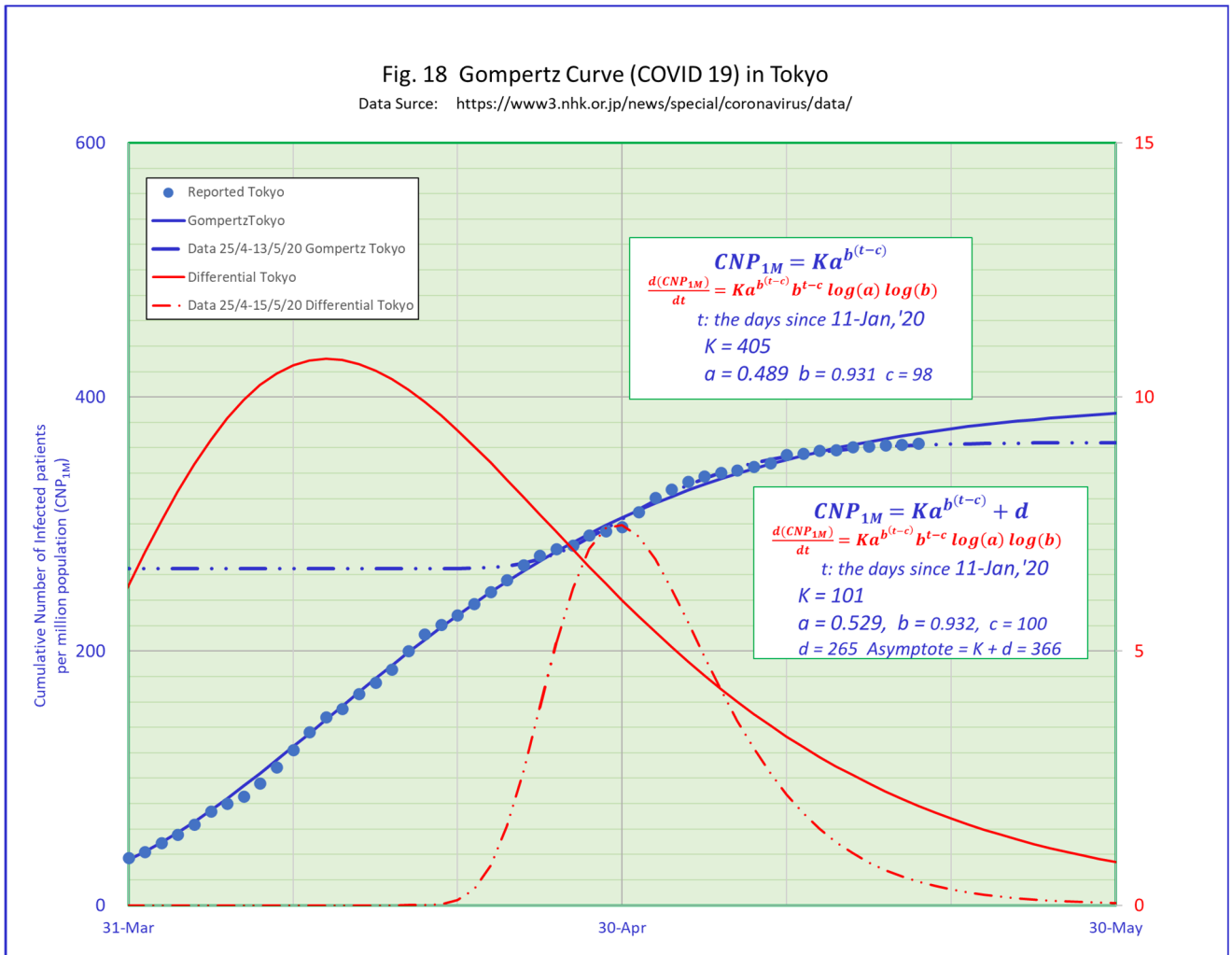


Table-4 Update: 19/5/2020		DfEq(t_e)	0.2	Round up	Start-up=	11/01/2020	Differential Data	
		t_e =			End=		DfEq(t_e)	DfEq(t_M)
With data up to	K	a	b	c	Date# for t_e	End Date		
China	58	0.457	0.862	29	53	03/03/2020	0.187	3.2
Italy	3879	0.482	0.942	84	197	25/07/2020	0.191	85.8
Spain	5000	0.505	0.920	86	174	02/07/2020	0.185	153.3
France	2228	0.540	0.926	89	171	29/06/2020	0.193	63.0
Germany	2150	0.460	0.928	85	172	30/06/2020	0.187	59.1
UK	4650	0.258	0.953	94	247	13/09/2020	0.194	82.2
USA	5800	0.361	0.955	98	254	20/09/2020	0.200	98.8
Japan	134	0.530	0.915	97	158	16/06/2020	0.195	4.4
Malaysia	243	0.373	0.948	83	148	06/06/2020	0.189	5.4
New Zealand	240	0.682	0.848	84	111	30/04/2020	0.174	14.6
Japan	134	0.530	0.915	97	158	16/06/2020	0.195	4.4
Tokyo	399	0.498	0.929	98	162	20/06/2020	0.186	10.8
Tokyo (Partial)	* 101	0.458	0.818	112	134	23/05/2020	0.188	7.4

Note*) For partial data fitting, K is not equivalent to the Upper asymptote which is K(101) + Lower Asymptote(265)

SAMMARY

It is confirmed that the Gompertzian Curve shall correlate and represent the cumulative number of the COVID 19 patients somehow accurately. By using the Gompertzian curve as a fitted curve, it was possible to accurately reproduce the cumulative number of COVID 19 patients from the beginning to the end of the infection in major countries of the world. As a result, it was found that there is a possibility that the end date of infection could be specified by using at least the approximate curve after passing through the inflection point.

However, it has not been clarified yet what kind of environment in which this novel coronavirus is placed determines the coefficients a and b of the Gompertz curve. In the future, it will be a challenge to pursue further research in this area and to prevent the explosive spread of infection and to predict the end of infection without waiting for the day when the inflection point has passed.

References:

- 1) <https://www.wionews.com/world/when-will-coronavirus-end-in-us-uk-france-india-study-predicts-dates-month-295324>
- 2) Jianxi Luo, "Predictive Monitoring of COVID-19", Working draft, updated on May 2, 2020, Data-Driven Innovation Lab, Singapore University of Technology and Design
- 3) James Holland Jones, "Notes On R0", Department of Anthropological Sciences Stanford University, May 1, 2007
<https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf>
- 4) https://en.wikipedia.org/wiki/Gompertz_function as of 03/05/2020
- 5) Kathleen M. C. Tjørve and Even Tjørve, "The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family", Kathleen M. C. Tjørve, Even Tjørve, 5 June 2017
<https://doi.org/10.1371/journal.pone.0178691>
- 6) Kenji Matsui, "Gompertz Curve", <http://gompertz-matsui.la.coocan.jp/Gompertz/English.html>

- 7) <https://edition.cnn.com/2020/04/28/asia/new-zealand-coronavirus-outbreak-elimination-intl-hnk/index.html>
- 8) United Nation Statistics <https://www.globalnote.jp/post-1555.html>