medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Novel indicator of change in COVID-19 spread status

Takashi Nakano[1*†] & Yoichi Ikeda[2‡§]

[1]*Research Center for Nuclear Physics, Osaka University, 10-1 Mihogaoka, Ibaraki, 567-0047, Osaka, Japan.*

[2]*Department of Physics, Faculty of Science, Kyushu University, Motooka 744, Nishi-ku, Fukuoka-city, 819-0395, Fukuoka, Japan*

As the spread of COVID-19 worldwide progresses, it is important not only to protect human lives, but also to minimize social losses due to economic paralysis by detecting signs of spread at an early stage and predicting future trends accurately. This report introduces a new indicator (K) of the magnitude of the spread of COVID-19 based on daily number of infected people from publicly available data. Transitions of the K can predict effects of measures such as the blockage of cities and social distancing, signs of new spread, and possible regional dependence in the formation of population immunity.

Keywords: COVID-19

Global spread of COVID-19 has resulted in significant human and economic losses worldwide. In order to prevent the spread of infection, it is necessary to restrict social activities by policies such as the blockage of cities and the prohibition of assembly. For the effective implemen-

*Corresponding author.
†E-mail: nakano@rcnp.osaka-u.ac.jp
‡Corresponding author.
§E-mail: ikeda.yoichi@phys.kyushu-u.ac.jp

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

tation of these policies, it is necessary to ascertain the status of spread and to estimate the trend of spread accurately. However, it is often difficult to grasp the severity of spread and estimate the trend by using model calculations because the implementation criteria of PCR testing vary from country to country and a high level of expertise is required to adjust model parameters according to the circumstances of each country and confirm their validity. The threat of COVID-19 has spread over countries that do not have the high-level computing resources, and the development of means to ascertain the status of spread accurately, without relying on specific models, has become an urgent issue.

In this study, we introduce a new indicator called the K value defined by $K(d) = 1 - N(d-7)/N(d)$, where d is the number of days from the reference date, $N(d)$ and $N(d-7)$ are the total number of infected people on days d and $(d-7)$, respectively. Since $N(d)$ is greater than $N(d-7)$ during the period from the initiation of spread to convergence, K takes a value between 0 and 1.

Without loss of generality, the daily evolution of $N(d)$ can be expressed with a time dependent exponential factor $a(d)$ as $N(d+1) = \exp(a(d))\, N(d)$. Our assumption is that $a(d)$ can be expressed by a geometric series with a constant dumping factor k, namely, $a(d+1) = ka(d)$. The simulation study under this assumption found that K can be approximated by a linear function of d in a wide range ($0.25 < K < 0.9$) and the input value of k can be reproduced by $k = 1 + 2.88K'$, where K' is a slope of a straight line obtained by the fit. The validity of the assumption must be be checked by analyzing the existing data[1–3] to see if the K values lie on a straight line in the range

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of $0.25 < K < 0.9$. If the assumption is confirmed, the analysis gives a rate of of convergence of infection. By updating the K value in real time using daily input data, we can identify the current status of spread, estimate the future status of spread, and detect signs of new spread at an early stage.

The analysis using K was first applied to China, which was less susceptible to the impact of other countries, because the spread began ahead of other countries. The K value was calculated using data from January 26 onward, but the data before February 12 were uniformly multiplied by 1.27 in order to correct a sharp increase in the number of infected people artificially caused by the change of certification criteria for SARS-CoV-2 infections in Hubei Province, China on February 13. As shown in Figure 1, the K values is closely approximated by a straight line with a slope (K') of $-0.040/d$.

The linearity of K is also prominent in the United States. After a high K level period indicating successive infectious explosions from mid to late March, the K have continued to decline with a uniform rate of $K' = -0.024/d$. However, it starts to deviate from the straight line around April 19, indicating the slowdown of the convergence speed or new spread.

In order to understand the change of K, the analysis of the data for the United States is performed using the SI epidemic model[4] (see Methods). It is found that we need to take into account at least the four independent sources of the infection to reproduce the data. We find the peak positions of the maximum momentum of the infection with respect to each source as March 24, April 1, 9 and 18 (Fig. 2 **a**). The first peak position clearly corresponds to the date of the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

change in the slope of the K. This shows that the K starts to decrease when the first peak-out takes place. The other sources contribute to keep the K linear as superposition.

In Italy, where COVID-19 started to spread first in Europe, K' was $-0.014/d$ from March 1 to 23 indicating slow convergence, but the convergence rate of was improved after March 24 resulting in $K' = -0.026/d$. This improvement is most likely due to the containment policy implemented in early March, including the blockade of cities. In France, where the spread started after about 10 days from Italy, K' is $-0.019/d$ from the early stages. The difference from Italy in K' is probably due to the fact that the 10-days delay made it possible to implement aggressive containment measures in the early stages.

Germany and Sweden are examples of countries' policies showing changes in the slope of their K values. On March 16, Germany began measures to prevent the spread of infection, including closure of most retail stores except grocery stores and pharmacies, restriction of operation hours of restaurants, and prohibition of assembly at religious facilities. This resulted in a steep K slope ($K' = -0.025/d$). Sweden, on the other hand, sought to acquire population immunity and took relatively mild measures resulting in a moderate gradient ($K' = -0.018/d$) lasting more than a month. UK took took similar measures in the early stage of the spread resulting in infection explosion with $K' = -0.007/d$. It was tamed down ($K' = -0.020/d$) after they introduced the strict policy. On the other hand, in Russia, the K value stays above 0.6 with a slope of $K' = -0.007/d$, and there is no indication that the spread will converge yet.

In Asian countries close to China, after the first wave originated in China, and the subsequent

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

spread in synchronized with the worldwide spread can be observed as the upward change in K'. In order to accurately estimate the change in the status of the subsequent spread from the K value, it is necessary to set a reference date for counting the total number of infected people at the rise of the second wave. We obtained $K' = -0.029/d$ for Japan by setting the reference date on March 25. The slope is milder than those of Taiwan ($K' = -0.052/d$) and South Korea ($K' = -0.082/d$ and $K' = -0.038/d$). However, the Japanese slope is steeper than those of European countries with more strict social restrictions than Japan, indicating faster convergence. On the other hand, many Asian countries give similar K' values to the Japanese one. For example, the K' value of Thailand is $K' = -0.036/d$. The relatively high absolute K' values of Asian countries from the beginning prior to strict social restrictions suggest that the population immunity may be formed more quickly in these countries than European countries[5,6]. Note that a typical K' value of $-0.035/d$ of Asian countries corresponds to a dumping factor k of 0.90, while a typical K' value of $-0.020/d$ of European countries corresponds to a k value of 0.94. Incidentally, the difference between the two K' values in South Korea may be attributed to the fact that most of the infection routes were clear in the former, whereas the route was often unknown in the latter.

To reproduce Japan data[3] with the SI epidemic model, which has two peaks in the K from March to April, it is necessary to introduce four sources of the spread. The corresponding peak positions are obtained as February 22, March 10, April 3 and 14. In the actual data, the K starts to decrease on March 12 in the first peak and on April 11 in the second one. The peak positions in the model agree with these date (Fig. 2 **b**). Moreover, the third peak position in the model coincides with the date when the K saturate at 0.5.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Comparing with the model analysis of the United States, it is worth mentioning that the parameter, which controls how infections spread, is smaller in Japan than that in the United States. Nevertheless, the steep gradient of the K is observed in Japan. This observation is understood by the deference of the number of the sources of the infection contributing to the K in the both countries. In Japan, the first two and last two sources are well separated in time, so that only the last two affect the decrease of the K as superposition. While, in the United States, all four sources contribute to it. The results show the K plays a crucial role to understand how the infection spreads.

In conclusion, we have demonstrated that the value of K and its slope of K' are crucial for understanding the spread status of COVID-19 and predicting future trends. Analyses with the K and K' values will help us to implement appropriate measures in a timely manner. Since the K' is related with a dumping factor of the exponential constant, a systematic study of K' may reveal the underlying reasons for regional differences in infection rate and mortality of COVID-19. Also, as evidenced by the comparison with the SI model calculations, the linearity of the K value is not trivial but is most likely to be caused by several consecutive infectious explosions. Focusing on the change in the value of K will help to improve and refine epidemiological models of infectious diseases with the same tendency as COVID-19.

References

1. Coronavirus Source Data. https://ourworldindata.org/coronavirus-source-data

   see also, https://web.sapmed.ac.jp/canmol/coronavirus/index.html

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

3. NIPPON TELEVISION NETWORK CORPORATION, COVID-19 special site.

https://www.news24.jp/archives/corona_map/index2.html

4. Hethcote, H. W. Qualitative analysis of communicable disease models. *Math. Biosci.*, **28**, 335-356 (1976).

5. Sala G. & Miyakawa T. Association of BCG vaccination policy with prevalence and mortality of COVID-19. *MedRxiv* April 6, 2020. https://doi.org/10.1101/2020.03.30.20048165 (2020).

6. Iwasaki A. & Grubaugh N. D. Why does Japan have so few cases of COVID19? *EMBO Mol Med* (2020) 0. https://doi.org/10.15252/emmm.202012481 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Methods

We assumed that a convergent series of the total number of infected people ($N(d)$) is given by

$$N(d+1) = e^{a(d)} N(d) \tag{1}$$

and

$$a(d+1) = ka(d), \tag{2}$$

where k is a constant ($k < 1$). Then K values for $d > 7$ were calculated by $K(d) = 1 - N(d-7)/N(d)$ for the cases with $k = 0.890$, 0.895, 0.900, 0.905, 0.910, 0.915, 0.920, 0.925, 0.930, 0.935, 0.940 and 0.950 (Extended Data Fig. 1). The fitting region of $0.25 < K < 0.9$. with a linear function were determined so that the maximum deviation of data points from the straight line is less than 11% of the data value. The relation between the slope of the straight line K' and k were examined to obtain a linear relation of $k = 1 + 2.88K'$ (Extended Data Fig. 1). Note that the relation does not depend on $a(0)$ as far as the first K value is larger than 0.9. We set $a(0) = 0.5$ for the all cases. For application of real-data analyses, once k is known from K', the daily evolution of $N(d)$ can be recursively calculated from the equations (1) and (2).

In order to understand the behavior of the K, we analyze the public data of COVID-19 in the United States and Japan employing the SI epidemic model [4], which consists of the infectives(I) and susceptibles($S = N - I$) with N being the final number of infectives. In the both countries, several changes in the slope of the K are found, it is natural to introduce several sources of COVID-19: The K is a monotonically decreasing function if a single source is taken into account in the SI

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

model. The infectives with respect to each source i is described as

$$\frac{dI_i(t)}{dt} = \frac{a}{N_i} I_i(t) \left(N_i - I_i(t)\right). \tag{3}$$

The model parameters a and N_i control the spreading speed of COVID-19 and the final number of infectives caused by the source i, respectively. The analytic solution for I_i is found as $I_i(t) = N_i/2\left[1 + \tanh\left(a/2\left(t - t_i^{\mathrm{c}}\right)\right)\right]$ with t_i^{c} denoting the peak time of the infection spreading. The total number of infected people in a country at time t are then obtained as $I(t) = \sum_i^{n_{\mathrm{src}}} I_i(t)$ and finally reach $N = \sum_i^{n_{\mathrm{src}}} N_i$. The optimal solution of the parameter set $\boldsymbol{P} = (a, \{N_i\}, \{t_i^{\mathrm{c}}\})$ together with the number of sources n_{src} is found by minimizing the weighted mean-square deviation $L(\boldsymbol{P}) = \sum_{d \in \boldsymbol{D}} \left(I(d) - N(d)\right)^2 / N(d)$ in the fit range $\boldsymbol{D}$. We choose $\boldsymbol{D}$ from February 23 to April 20 (58 days) for Japan and March 15 to April 20 (37 days) for the United States, respectively. We find the optimal number of sources for both countries as $n_{\mathrm{src}} = 4$, and the resulting parameters are given by $(a, N_1, N_2, N_3, N_4, t_1^{\mathrm{c}}, t_2^{\mathrm{c}}, t_3^{\mathrm{c}}, t_4^{\mathrm{c}})$ = (0.38, 92k, 232k, 258k, 294k, 03/24, 04/01, 04/09, 04/18) for the United States and (0.24, 0.2k, 0.7k, 3.1k, 8.7k, 02/22, 03/10, 04/03, 04/14) for Japan, respectively. Equipped with the parameters above, the number of daily new cases dI_i/dt and the K value are presented in Fig. 2.

Acknowledgements

We thank Prof. Y. Kaneda (Graduate School of Medicine, Osaka University), Prof. T. Yoshimori (Graduate School of Frontier Bioscience, Osaka University), and Mr. S. Shimasaki (Embassy of Japan in the United States of America) for helpful discussions.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Authors' contributions

T. Nakano contributed to conceptualization of the K value and Y. Ikeda contributed to the SI model analyses. The both contributed to data analyses and interpretation of the results.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

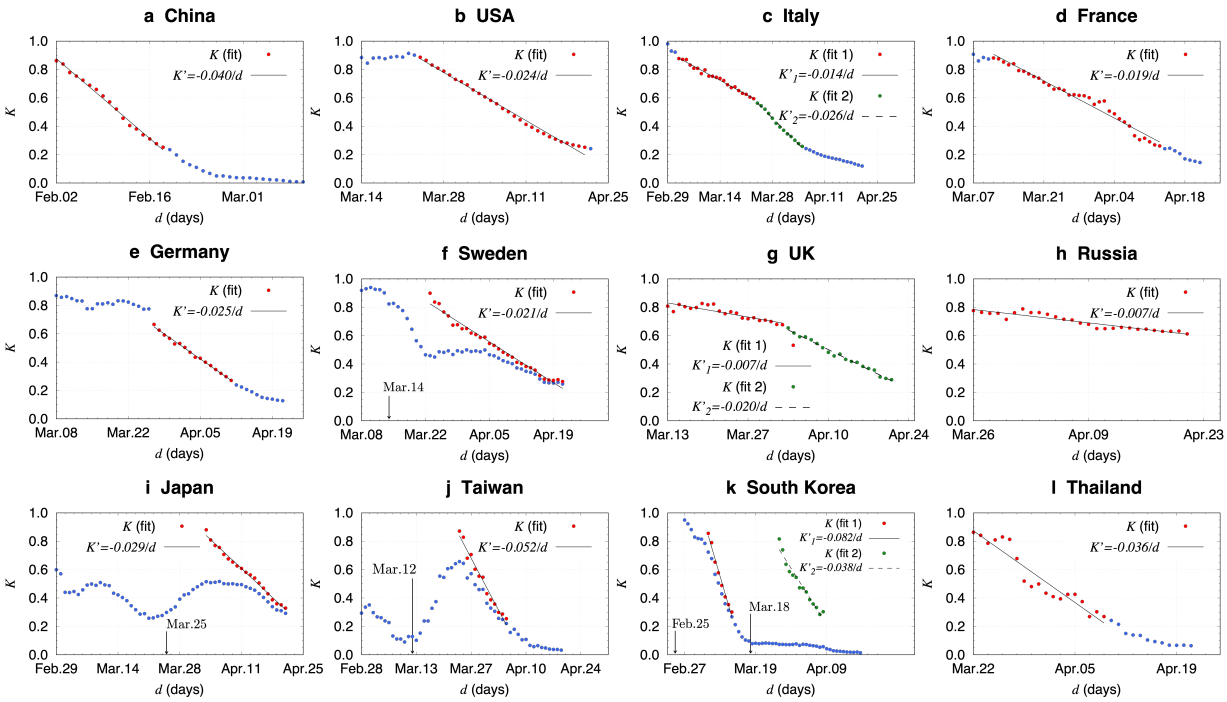


Figure 1: **Transition of the K values from February to April, 2020.** **a**, The K values of China obtained from the daily total number of infected people. The slope K' was obtained by a linear fit in the range of $0.25 < K < 0.9$. The data points used for the fit are indicated by red points. **b**, The K values and the K' value of USA. **c**, The K values and the K' values of Italy. The first and second K' values were obtained by linear fits using red and green points, respectively. **d**, The K values and the K' value of France. **e**, The K values and the K' value of Germany. **f**, The K values and the K' value of Sweden. The total number of infected people were counted from the reference date set on March 14. **g**, The K values and the K' values of UK. **h**, The K values and the K' value of Russia. **i**, The K values and the K' value of Japan. The reference date was set on March 25. **j**, The K values and the K' value of Taiwan. The reference date was set on March 12. **k**, The K values and the K' values of South Korea. The reference dates were set on February 25 and March 18 for the first and second fits, respectively **l**, The K values and the K' value of Thailand.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

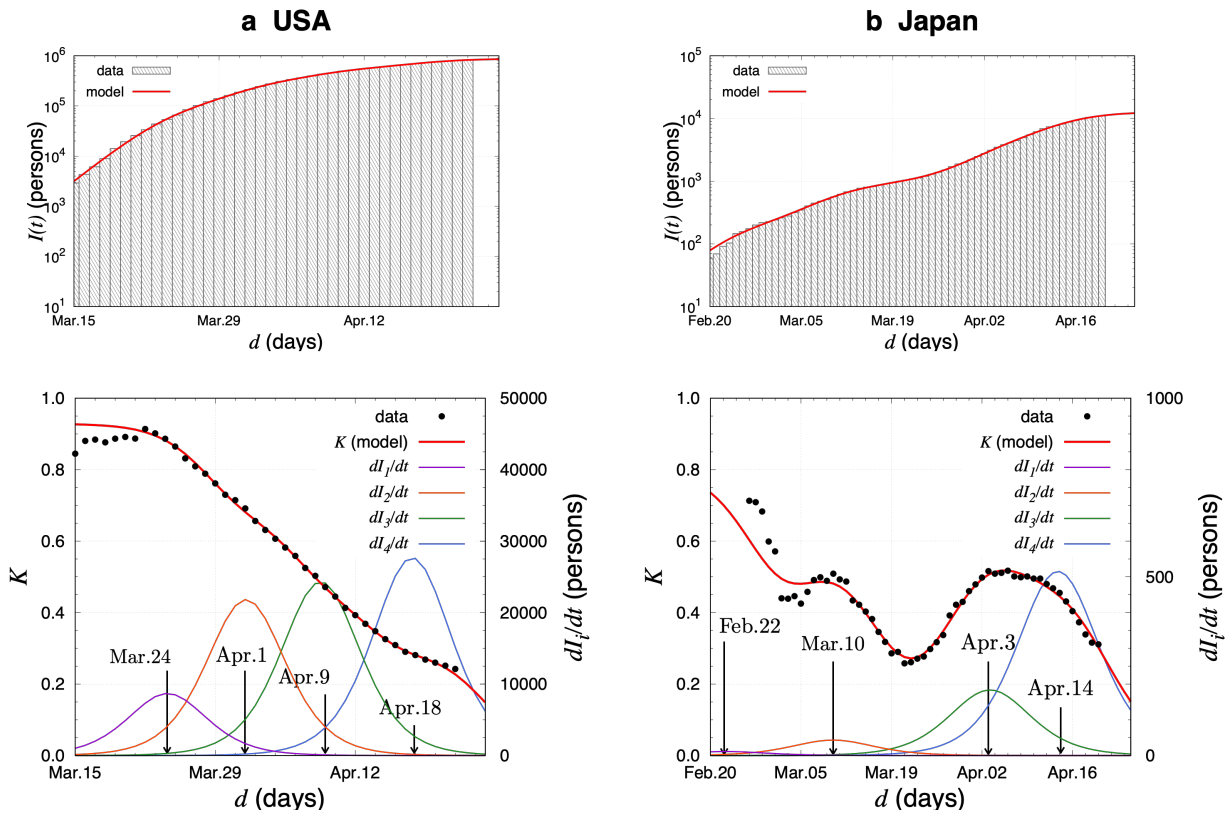
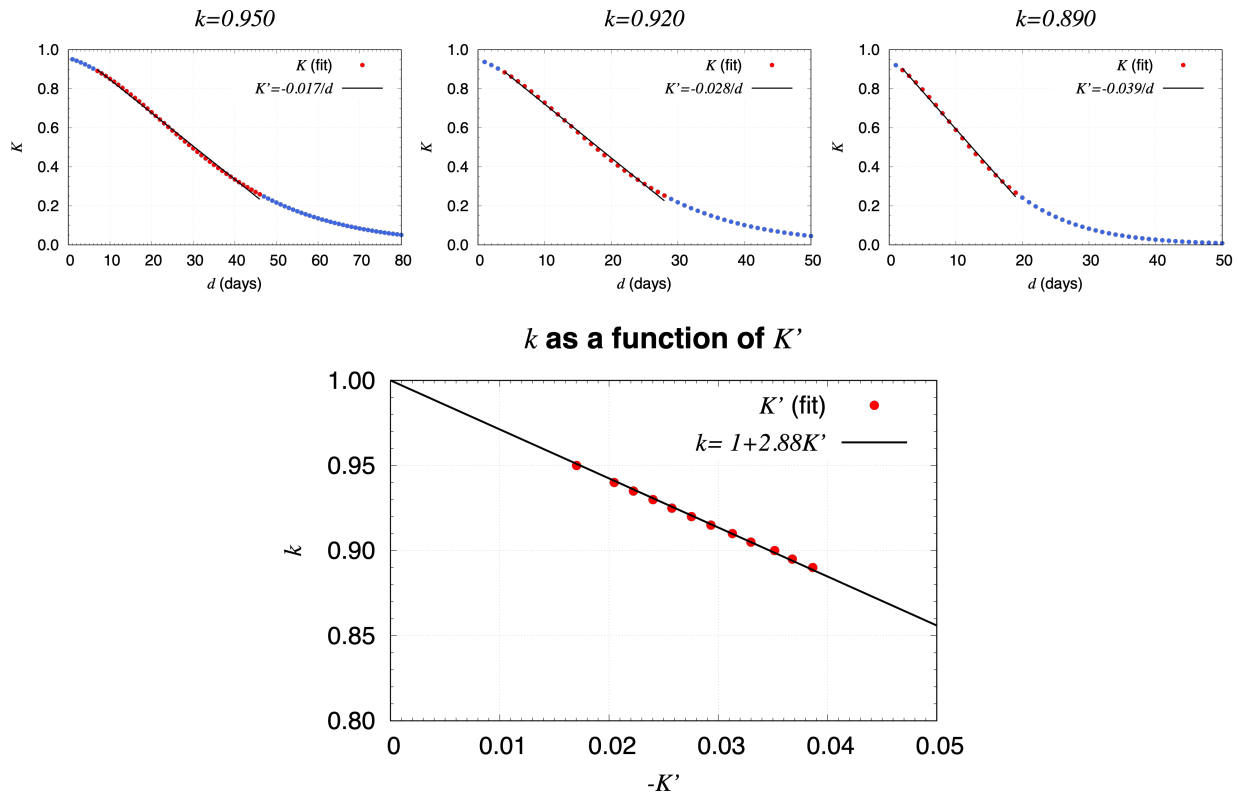


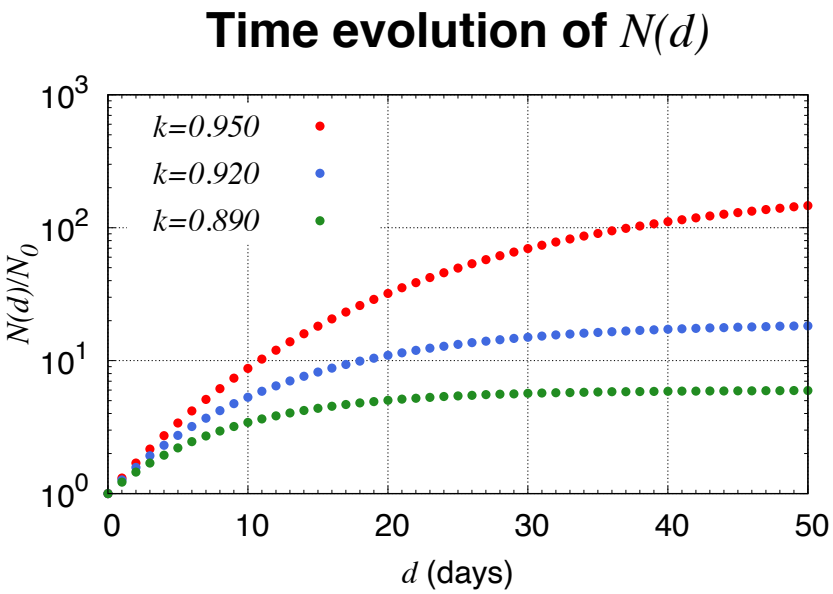
Figure 2: **The model results of total number of infected people and the K value in the United States and Japan.** **a**, The results for the United States. The fit results are given in the top figure. The K value is represented together with the daily new cases dI_i/dt calculated in the SI model in the bottom. The peak positions t_i^c caused by the source i ($i = 1, 2, 3, 4$) are shown by the vertical allows. **b**, The results for Japan. Same as for (a), but for the case of Japan.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 1 | Mock data of the K value and K' dependence of k. The dumping constant k was assumed to be 0.95 (top left), 0.920 (top middle), and 0.890 (top right). The bottom plot shows k as a linear function of K' with a constraint of $k(0) = 1$.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.25.20080200; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 2 | Time evolution of $N(d)$. Time evolution of $N(d)$ for the cases with $k = 0.950$, 0.920, and 0.890 ($K' = -0.017/d$, $-0.028/d$, and $-0.039/d$) normalized to N_0 by setting $K = 0.9$ on the day zero.