

Mathematical Epidemiological Model of the Propagation and Prevention of Infectious Diseases Considering COVID-19 Vaccination

Kazumi Omata (✉ komata@hosp.ncgm.go.jp)

National Center for Global Health and Medicine Research Institute

Hiroaki Mitsuya

National Center for Global Health and Medicine Research Institute

Research Article

Keywords: COVID-19 pandemic, epidemiological studies, social distancing, vaccination, future scenarios.

Posted Date: July 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-709496/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Mathematical epidemiological model of the propagation and prevention of infectious diseases considering COVID-19 vaccination

Kazumi Omata^{1,2} and Hiroaki Mitsuya^{2,3,4}

¹Center for Clinical Sciences, National Center for Global Health and Medicine, 1-21-1 Toyama,
Shinjuku, Tokyo, 162-0052 Japan

²Departments of Hematology, Rheumatology, and Infectious Diseases, Kumamoto University
Hospital, 1-1-1 Honjo, Chuo-ku, Kumamoto, Japan

³Department of Refractory Virus Infections, National Center for Global Health and Medicine
Research Institute, 1-21-1 Toyama, Shinjuku, Tokyo, 162-0052 Japan

⁴Experimental Retrovirology Section, Center for Cancer Research, National Cancer Institute,
National Institutes of Health, Bethesda, MD, USA

Abstract (199/200 words)

The duration of the necessity of current restrictions and preventive countermeasures against the COVID-19 pandemic is of great concern. While several previous epidemiological studies have discussed controlling the course of the epidemic with regard to social distancing, vaccination, care capacities, and future scenarios, we adopt an alternative approach to provide insights into the time scale of epidemic propagation in human populations. We show that the reproduction number modified by preventive countermeasures R_m implies a threshold \tilde{R}_m to reach the herd immunity level at time T_h . While T_h varies moderately for large R_m , it increases sharply around $R_m = \tilde{R}_m$, and T_h is infinite below $R_m = \tilde{R}_m$. The transition region for this increase is minute, demonstrating that prevention of infectious diseases must consider the relatively asymptotic nature of their propagation, which varies unpredictably between steady transmission and explosive outbreaks. These results suggest the continuation of preventive countermeasures to suppress the transmission of COVID-19 for many years; if they were discontinued or reduced such that R_m exceeded \tilde{R}_m , the disease would be transmitted throughout the considered community, and the required herd immunity level would be reached within 250 days if $R_m > 1.6$. However, the implementation of vaccination programs could drastically alter this dynamic.

Main text (4500/4500)

Introduction

At present, the containment of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection that causes coronavirus disease 2019 (COVID-19) is the most urgent issue worldwide [1]. To date, the statistics concerning the global pandemic remain unprecedentedly grim as the largest global public health crisis in over a century continues. As of July 7, 2021, the US had suffered the largest reported damage; the cumulative number of confirmed cases was 33,451,965 (a ratio to the total population of 10.2%). In contrast, the size of the epidemic outbreak in Japan has been relatively small, with 813,976 confirmed cases (0.64%) [1].

Expectations concerning the course and potential end of the current global pandemic remains unclear, and the question of how long the restrictions on private and economic activities will continue remains as well. Socially, the question arises of whether to give priority to our activities or to the elimination of the pandemic. Four scenarios for the end of local epidemics in specified communities may be considered [2], including (1) the eradication of the epidemics by the successful development of effective vaccines and vaccination programs that completely end the crisis, (2) the attenuation of the infection force of the viruses by interventions including vaccination, leading to remarkable reduction in mortality and morbidity, (3) acceptable successes in the development of vaccines and drugs alongside preventative measures to reduce the severity of the epidemics such that they remain as a relatively small problem, and (4) the acquisition of herd immunity. We may be progressing to scenario (1); however, this is an ideal goal, and it is not yet known whether vaccines will be well-distributed globally or whether the developed vaccines will retain their effectiveness when facing virus mutations. Regarding mutations, the realization of scenario (2) also involves uncertainties. The realization of scenario (4) implies that many individuals must acquire immunity, and society must be ready to accept many fatalities unless sufficient vaccination programs are implemented. These three scenarios are so uncertain that it may be necessary to proceed with scenario (3). In this case, the most immediate concern regarding the COVID-19 pandemic is the duration of the necessity of continued preventive measures such as social distancing, remote work, shutdown or curtailment of opening hours of public facilities, restrictions on social events, and restrictions on travel.

Regarding this question, the present work investigated the times required to achieve herd immunity with tools for mathematical epidemiology and obtained intriguing results. The herd immunity level h_c is defined as a level of population immunity such that the spread of the disease will decline and cease even after all preventive measures have been relaxed [3]; stated differently, h_c provides a threshold on which an epidemic starts to decline under the hypothetical condition of no preventive measures being applied.

According to mathematical epidemiology, the herd immunity level is expressed as $h_c = 1 - 1/R_0$, where R_0 denotes the basic reproduction number [3,4,5,6]. This number is an important variable for

the epidemiology of infectious diseases [7,8,9], defined as the average number of secondary infections produced when one infected individual is introduced into a host population where all members are susceptible. As an epidemic spreads, the number of susceptible individuals S decreases, and the number of infectious individuals I increases. When one infectious individual passes pathogens on to R_0 individuals, and $R_0 - n$ individuals among them are already immune, the effective number of infected individuals is n . The fraction of immunized individuals is $1 - n/R_0$. If n is more (less) than 1, an epidemic grows (declines). Hence, an epidemic threshold can be defined at $n = 1$. This provides the herd immunity level as $h_c = 1 - 1/R_0$.

Suppose that the ratio of n to R_0 equals the ratio of the total number of susceptible individuals S to the total population N : $S/N = n/R_0$. Then, $S/N = 1/R_0$ on the threshold because $n = 1$. According to the conventional compartmental model (see Method), $dI/dS = -1 + \gamma N/\beta S$, where β and γ denote contact and recovery rates, respectively. This equation indicates that I takes on a maximum value at $S = \gamma N/\beta$ such that I increases when S is greater than $\gamma N/\beta$ and decreases when S is less than this value. Therefore, we obtain $R_0 = \beta/\gamma$, which is consistent with the calculation of the herd immunity level and leads to $R_0 S = N$ on the threshold. If all members of the host population are susceptible, $S = N$, and thus $R_0 = 1$ on the threshold; this is in accordance with the definition of the basic reproduction number R_0 .

Apart from the basic reproduction number R_0 , the present work considers two other reproduction numbers: the effective reproduction number defined as $R_e = R_0 S(t)$, and the modified reproduction number R_m . The former considers the decrease in the number of susceptible individuals S , which occurs as an epidemic develops. When an epidemic grows perceptibly and enters into community transmission in the host population, the effective reproduction number is given by analysis of case reports. The latter includes a modification in the contact rate β representing the impact of non-pharmaceutical interventions and preventive measures, such as social distancing and isolation. Therefore, we can write $R_m = \beta_m/\gamma$ using the modified contact rate β_m .

Given that changes in the number of susceptible individuals is expected to be negligible during a short time period, the compartmental model provides for the exponential growth of an epidemic; $I(t) =$

$I(0)e^{rt}$, where $r = \beta S/N - \gamma$ represents the growth rate. In the early stage of an epidemic, $S \cong N$, and $r = (R_0 - 1)\gamma$. In contrast, in the later stages $r = (R_0 S - 1)\gamma = (R_e - 1)\gamma$. If the contact rate is modified, R_0 is replaced with R_m in the growth rate.

While it may be difficult to incorporate the wide variety of complex, interacting factors that affect actual epidemics, the techniques of mathematical epidemiology provide us with a simplified perspective from which to analyse the propagation of infectious diseases in various populations [3,4,5,6,13,14,15,16,17,18]. To this end, it is important to observe the essential features of epidemics. The capability of mathematical epidemiology to produce clear and precise projection based on various conjectures is among its key benefits [13-15]. A simple perspective has been provided to model the spread of infectious diseases, e.g. age-dependence of vaccination effects on rubella epidemics [3,19], biennial epidemics of measles [6,20,21], or influence of heterogeneity in sexual activity on HIV epidemics [3]. The present work presents the results of an investigation of mathematical models for epidemics focusing on the three reproduction numbers R_0 , R_m , and R_e , and especially discusses the time scale for disease propagation, as well as the results of a comparison between the actual data on the COVID-19 epidemics in Tokyo [10,11] and NYC [12] and the computational results derived mathematical models.

Since the beginning of 2020, numerous studies have been published on COVID-19 and SARS-CoV-2 in fields ranging from microscopic to macroscopic, including molecular biology, medical and clinical studies, epidemiology, and political decision making, with each field applying its characteristic methodologies. As examples of reports with similar focus and methodology to this study, research on COVID-19 modelling has considered countries or regions such as France [22], Italy [23], the Basque country [24], Spain [25], Portugal [26], Hong Kong [27], China [28,29,30], and the UK [31]. These studies discussed recurrence and rebound of the epidemics, effects of protection measures and lockdown, widespread testing, contact tracing, care capacities, or scenarios of pandemic management. In this context, as an alternative approach to model the progression of the COVID-19 pandemic, the present work has evaluated the time required to reach the herd immunity level in a given community, using the epidemiological compartment model [3,4,5,6]. This approach enabled us to discuss the time

scale of epidemic propagation through populations and the effects of vaccination. We provide insights into factors supporting the spread of infectious diseases and illustrate the implications of vaccination mathematically. The implementation of vaccination not only increases the number of immunized individuals but also accelerates the achievement of the herd immunity level. This latter effect may be rather important for the elimination of epidemics in general, and specifically for the current global pandemic.

Results

Figure 1 shows the results of the numerical calculations of the epidemiological compartmental model (see Method). The curves represent solution trajectories, which are traces of $(S(t), E(t) + I(t))$ derived by solving the associated equations, where S , E , and I denote the numbers of susceptible, exposed (not infectious), and infectious individuals at time t , respectively; $(S(t), E(t) + I(t))$ and the plane consisting of S and $E + I$ are referred to as the representative point and $EI-S$ plane, respectively. Namely, Fig.1 shows $E + I$ as a function of S with t implicitly included. The curves began at $(S(t), E(t) + I(t)) = (1,0)$ on the $EI-S$ phase plane and curved along the arrows depicted in Fig.1 over time. This represents a small number of infectious individuals introduced into a population where everyone is susceptible, causing an epidemic to begin. Therefore, $E + I$ increased and S decreased with time. After the epidemic spread for some time, $E + I$ reached the peak of the trajectory. This peak represents the herd immunity level, and the number of infectious individuals began to decrease at this peak. Subsequently, $E + I$ decreased, and the trajectory descended to the $E + I = 0$ line, indicating the elimination of the epidemic.

The $EI-S$ curve in Fig. 1(a) was calculated with $R_0 = \beta_0/\gamma = 2.5$ (a value in the range plausible for COVID-19 [24,30,32]). When model parameters representing the implementation of preventive measures and changes in the contact rate of individuals were modified, the representative point traced a different trajectory. Figure 1(b) shows trajectories calculated with the modified reproduction number $R_m = \beta_m/\gamma$ due to the reduction in the contact rate from β_0 to β_m . When R_m was set to 2.0, the curve could intersect the herd immunity level (green solid curve). When $R_m = 1.6$, the curve could still pass

across the herd immunity level (green broken curve); however, its final value point was very near the herd immunity level. When $R_m = 1.5$ or less, the curve could not reach the herd immunity level (red curves).

Next, suppose that preventive countermeasures are cancelled in the middle of an epidemic, and the value of the contact rate is restored from β_m to β_0 . Then, the curves that cannot reach the herd immunity level under the preventive countermeasures come to reach that level. Figure 1(c) depicts this situation. For example, the blue curve almost converges to the $E + I = 0$ line around $S = 0.65$, i.e. the epidemic was almost eliminated, and the relaxation of preventative measures caused a further increase followed by an eventual decrease after the herd immunity level was reached. When the timing of countermeasures being rescinded was earlier, the trajectories were closer to the inherent case.

Figure 1(d) represents the effects of vaccination. With increasing vaccination rate, the location of the trajectory was lowered, and the epidemic decreased in size. The vaccination rate v is defined in the SEIR compartment model as the rate of decrease in the number of susceptible individuals, which equals the rate of increase in the number of individuals considered to possess immunity (see Methods). However, the peak position of the trajectories remained at $S = 1/R_0$; this implies that the herd immunity level is unchanged by the implementation of vaccination.

With the derivation of the trajectories, we have calculated the time T_h over which a trajectory reached the herd immunity level. Figure 2(a) shows T_h as a function of R_m (these calculations do not consider the suspension of preventive countermeasures). We found that T_h increased gradually with decreasing R_m ; T_h was less than 250 days when R_m was greater than 1.6 under the condition of $E(0)/N \geq 10^{-7}$ (see Discussion). However, slightly above $R_m = 1.5$, this increase became sharp, and T_h diverged for $R_m < 1.5$. Figure 2(b) shows the effects of vaccination on T_h . First, it is found that the implementation of vaccination tends to transform the curves into a sigmoid form, distorting and flattening them in the region of $R_m < 1.5$. This effect removes the singularity of T_h at $R_m = \tilde{R}_m$. Second, the height of the curves, especially in the flattened part, becomes smaller with increasing vaccination rate. When v was larger than 0.02, the curve was not even sigmoid, but rather almost linear.

To investigate the actual situation of the COVID-19 epidemics, we estimated the effective

reproduction numbers using public data on Tokyo [10,11] and NYC [12]. Figure 4 displays the reported numbers of newly confirmed cases (blue bars), the numbers of individuals who initially contracted the infection adjusted by the back-calculation method (orange dots and line), and R_e numbers (black dots and line) in Tokyo (Fig.4a) and NYC (Fig.4b) as of January 31, 2021. In Tokyo, the adjusted epidemic curve (orange dots and line) included a few peaks reflecting the growth of the epidemic, with a large one occurring in March and April 2020, and another in January 2021. In contrast, the epidemic curve in NYC showed two large peaks in April 2020 and January 2021. The cumulative numbers of confirmed cases were approximately 99,800 and 532,500 in Tokyo and NYC, respectively, as of January 31, 2021.

These epidemic curves provide the epidemic growth rates, which allowed us to calculate the R_e numbers as a function of time (black dots and line). The values at representative times are marked with arrows. In Tokyo, there were peaks on February 24 and March 16; the R_e values were 5.13 (95% CI, 3.86-6.4) and 4.19 (95% CI, 3.63-4.75), respectively (Fig 3a). The values were 1.98 (95% CI, 1.95-2.02) and 0.82 (95% CI, 0.73-0.90) on March 25 (on which the Tokyo Metropolitan government put Tokyo under a stay-at-home rule) and April 7 (on which the state of emergency was declared by the Japanese government), respectively. After May, the R_e values rose again, forming distinct peaks (e.g. May 18 [$R_e = 2.56$ (95% CI, 2.19-2.93)] and June 4 [$R_e = 2.38$ (95% CI, 1.84-2.91)]); on May 25, the day on which the state-of-emergency declaration was lifted, the value was 1.37 (95% CI, 1.19-1.54). After June, the R_e number fluctuated until a relatively larger peak arose on December 28 [$R_e = 2.32$ (95% CI, 2.15-2.50)], which gave rise to a large increase in the number of infections in January 2021. In NYC (Fig 3b), at the beginning of March 2020, the reproduction number was as high as 11.5 (95% CI, 10.5-13.3). However, the R_e number exhibited no large peak since then and subsequently remained at approximately 1.0. On March 7 and 12, on which a state of emergency was declared by NYC and New York State, respectively, the R_e numbers were 5.60 (95% CI, 5.01-6.17), and 1.68 (95% CI, 1.63-1.73), respectively. When the number of new infections per day was the largest on April 1, the R_e number was 0.74 (95% CI, 0.66-0.81). After May, the R_e number fluctuated and a large increase in

the number of infections occurred in December; however, this increase proceeded slowly, and therefore, very large R_e numbers did not arise.

The frequency distribution of the R_e numbers estimated for the period until January 2021 are shown for Tokyo (Fig. 4(c)) and NYC (Fig.4(d)). This estimate did not consider the period after February 2021 to compare with the mathematical model excluding the mixing effects of reduction in the contact rates and vaccination programs. The mean value was 1.07 and 1.00, and the standard deviation was 0.47 and 0.25 for Tokyo and for NYC, respectively, for the period from April 2020 to January 2021.

Discussion

These results suggest that if the basic reproduction number was modified by preventive countermeasures against epidemics and the modified reproduction number R_m was reduced while not exceeding \tilde{R}_m , it would be necessary to continue the countermeasures to suppress the propagation of the disease. The general emergency condition created by the COVID-19 pandemic would thus be protracted for many years. In contrast, if preventive countermeasures were relaxed and R_m exceeded \tilde{R}_m , the epidemics would propagate rapidly, and the herd immunity level would be reached in a short period of time (less than 250 days in the present calculations for COVID-19). These results indicate the classically explosive nature of infectious disease propagation, in which either low levels of steady community transmission or asymptotic acceleration of the rate of transmission tend to prevail. Because the transition region for the sharp increase around $R_m = \tilde{R}_m$ is very small, it is difficult to realize intermediate propagation speeds. It is almost impossible to realize epidemic situations preferable to us, in which a moderate degree of deceleration of the propagation of the disease is combined with a similar progressive reduction in the degree of preventive countermeasures. Clearly, continuously holding R_m slightly above \tilde{R}_m may be considered impractically difficult.

Because the proportion of the cumulative total number of individuals infected with COVID-19 was less than 1% and 2% as of January 2021 in Tokyo and NYC, respectively [10,11], the effective reproduction number R_e can be approximated by R_m . Except in the early stages of the epidemics, R_e fluctuated between 0.0 and 3.0 in Tokyo and between 1.0 and 2.0 in NYC (Fig. 3), i.e. R_e has

continued to increase and decrease across the value of 1.5, which is the threshold for the divergence of T_h . This fluctuation reflects the dilemma that periods of rapid epidemic propagation and of associated social restrictions have repeatedly alternated since the onset of the pandemic. The distributions of R_e (Fig.4) also illustrate the challenge of determining whether to prioritize the prevention of disease or the continuance of economic activities, demonstrating the circumstances under which both cities have found the COVID-19 pandemic unmanageable. It may be considered likely that many countries have been experiencing similar situations. The abovementioned scenario (3) would oblige us to continue the general living conditions of the current emergency for many years, and it may not be possible to eliminate the pandemic within a few years. This may be a matter of concern in countries where vaccination programs are not progressing.

However, it is probable that the fluctuation of R_e around 1.5 in Fig.3 was not intended but rather a fortuity because individuals generally do not know that $R_m \cong 1.5$ is a threshold for T_h . This suggests that the threshold for T_h may be of benefit as a reference value for evaluating the future course of the COVID-19 pandemic. Regarding this point, Fig. 3 illustrates that the propagation speed at the early stage of an epidemic is small, not depending on R_m . Thus, unless we do not know the existence of a threshold for T_h , we cannot find any singularity with asymptomatic increase in the epidemic growth at the early stage, which indicates the threshold. Figure 3 also shows that the spreading velocity remained small until half of T_h , and then a notable increase occurred. Therefore, the trajectories in Fig.1(b) are traced by the representative point in the latter half of T_h .

The mathematical structure of the EI - S plane is described as follows. The compartmental model (method) shows that the EI - S plane includes a fixed point $(1/R_0, 0)$, which plays a role to attract the representative point $(S(t), E(t) + I(t))$ and impose it to rotate around the fixed point (such a fixed point is referred to as a spiral sink). As the representative point moves, $E(t) + I(t)$ reaches the maximum trajectory, and eventually, the representative point descends to and converges with the $E + I = 0$ line. The maximum point $(1/R_0, (E + I)_{max})$ is located just above the fixed point $(1/R_0, 0)$.

If the contact rate is modified, R_0 is replaced with R_m , shifting the peak position and landing point of the representative point (Fig.1(b)). This shift of the I - S trajectory leads to the sharp increase

in T_h (Fig.2) with the following mathematical explanation. Two factors affect this sharp increase. First, when the landing point is shifted to the region of S larger than the herd immunity level ($S = 1/R_0$), the representative point cannot reach the herd immunity level and T_h becomes infinitely large. Second, the compartmental model shows that the rate of change of the number of infectious individuals I is proportional to that number itself (see Methods). This is a typical rate equation. Therefore, the rate of change of I was small when I is small, implying that the motion of the representative point decreases as it approached the $E + I = 0$ line. If R_m was lowered due to the reduction in the contact rate, the landing point of the representative point shifted to the vicinity of the herd immunity level. Then, a remarkable decrease in the rate of change of the representative point occurs in the vicinity of the herd immunity level, leading to the sharp increase in T_h .

In Fig. 2(c), we have considered cases in which the intrinsic contact rate β_0 is restored in the middle of an epidemic, the course of which was previously modified with preventive countermeasures. If the epidemic had not yet reached the herd immunity level, the representative point moved again to a position near the original fixed point $(1/R_0, 0)$, which is located on the left-hand side of the modified fixed point $(1/R_m, 0)$, and a renewed increase in the number of infectious individuals occurred. This accounts for the recurrence of the epidemic. As seen in the trajectories of the representative point, the maximum values of I on the herd immunity level were less than the value for the inherent trajectory indicating a reduction in the size of the epidemic, indicating that preventive countermeasures did have lasting effects despite being ceased in the middle of the epidemic. If the contact rate was restored after an epidemic reached the herd immunity level, such a recurrence did not arise. This agrees with the definition of the herd immunity level.

The trajectories displayed in Fig.1(d) have the same value of S at the peaks irrespective of the vaccination rates, demonstrating that the implementation of vaccination did not change the herd immunity level [3]. This is because vaccination does not reduce the contact rate (hence, the basic reproduction number) but rather the number of susceptible individuals. It is important to understand that a reduction in the contact rate β is different from a reduction in the number of susceptible individuals S ; however, these two factors decrease the effective reproduction number and the epidemic

size, as indicated by the effective reproduction $R_e = R_0 S = \beta_0 S / \gamma$. The compartmental model including the effect of vaccination provides a rate equation for S with $dS/dt = -\beta SI - vS$, where v denotes a vaccination rate. It is noted that, in this equation, the vaccination rate is not included in the first term of the right-hand side but rather in the second term. The essential point concerning the effects of vaccination is given below. We found in Fig. 3(b) that the vaccination removes the singularity in T_h at $R_m = \tilde{R}_m$. Therefore, a reduction in R_m did not prevent the representative point from reaching the herd immunity level; specifically, preventive countermeasures did not prevent the epidemic status from reaching the herd immunity level. Here, it is noteworthy that this achievement of the herd immunity level proceeds without considerable epidemic spreading and expansion, demonstrating the special role of vaccination. This stands in drastic contradistinction to the case without vaccination. The mathematical interpretation is that the second term in the equation yielding S plays a role to update the state of infectious individuals I to the state indicating recovered or immunized individuals R (see Method). Apart from the numerical calculations shown in the present study, the removal of the singularity by this vaccination term may be proved by calculus for the equations in the compartmental model.

Recent studies showed that heterogeneity of populations reduces the herd immunity level [32]. By introducing age and activity heterogeneities into the population models for SARS-CoV-2, herd immunity was achieved in the models at a population-wide infection rate of $\sim 40\%$, which is substantially less than the classical herd immunity level of 60% obtained through homogeneous immunization. Although this result is intriguing, such quantitative details of the estimate of the herd immunity level may not affect the qualitative feature obtained in the present work because the sharp increase in the time required to reach the immunization level occurs in the interrelationship and superposition of the fixed point and the landing point of the representative point, as described above; accordingly, the detailed location of the herd immunity level does not affect the nature of the sharp increase in T_h . It is interesting, however, to examine effects of heterogeneities by using more complicated mathematical models.

The experimental results suggest that the classical herd immunity level of 60% would be obtained

within a finite period of time if homogeneous immunization was implemented: ~500 days with the rate $\nu = 0.001$, or ~100 days with $\nu = 0.02$ (Fig.2b).

To compare the model calculations with the actual course of the COVID-19 pandemic, this study has estimated the effective reproduction numbers using reported cases for Tokyo [10,11] and NYC [12]. However, several points must be noted. First, conditions such as age, gender, address, or ethnic groups were not recorded in the datasets used. Therefore, if further quantitatively precise comparisons are required, cohort studies should be conducted under well-controlled conditions. Second, we can expect further improvements for the input parameters incorporated in the calculations. The parameters used to represent the COVID-19 pandemic in the present work were based on estimates given by a study on the outbreak in Wuhan [33] because it was the earliest epidemic, and thus the estimated parameters reflect the intrinsic state of the COVID-19 epidemics, with the caveat that these values might be altered in practice by the characteristics of variant strains.

Today, many viral diseases other than COVID-19 are under control; however, their impacts on the economy and human life are smaller than those of COVID-19. For example, influenza characterized by frequent viral mutations is controlled by annual vaccination and drugs, and social distancing is conducted in accordance with the situation; HIV/AIDS is managed by efforts toward increasing access to testing and initiating pharmaceutical therapy at early stages of the disease, and such campaigns are an important tool; however, they themselves are not control measures.

This study has considered the time required to reach the herd immunity level as an index for evaluating the time scale of epidemic propagation and has provided several insights to illustrate the processes of such propagation. Further studies may yield a clearer understanding regarding this problem by using different indices.

In addition, the results derived herein are applicable to infectious diseases, which may emerge or re-emerge in the future. Although each infectious disease possesses a distinctive parameter set for the compartmental model that determines the herd immunity level, the characteristic that T_h increases sharply in a narrow range of R_m arises irrespective of the parameter set. Therefore, precise values of T_h , as a function of R_m , can be calculated only by adjusting the parameter set to the target infectious

disease and solving the compartmental model.

The results of the present work suggest that R_m , which is connected with T_h , may be appropriately used as a reference for the ascertainment of epidemic status and the timing of related decision-making. When vaccination campaigns are conducted, it is important to estimate both R_m and the vaccination rate and to inspect the status of the epidemic carefully. In addition, the results demonstrated that the implementation of vaccination not only increases the number of immunized individuals but accelerates the achievement of the herd immunity level. This latter effect may be of considerable importance for the elimination of epidemic diseases including the COVID-19 pandemic.

Methods

SEIR Compartmental model for infectious diseases

The susceptible-exposed-infectious-recovered (SEIR) compartmental model for infectious diseases [3-6] is described as, $dS/dt = -\beta SI$, $dE/dt = \beta SI - \sigma E$, $dI/dt = \sigma E - \gamma I$, and $dR/dt = \gamma I$, where S , E , I , and R denote the numbers of susceptible, exposed, infectious, and recovered individuals, respectively; β represents rates of contact, and $1/\sigma$ and $1/\gamma$ denote the mean latent and infectious periods, respectively. If the effect of vaccination is included in the SEIR model, the rate equations for S and R are replaced by $dS/dt = -\beta SI - vS$ and $dR/dt = \gamma I + vS$, respectively, where v denotes a vaccination rate [3,6].

The SEIR compartmental model has proven useful for the understanding and prediction of infectious diseases, e.g. this model was able to provide epidemic curves for measles or rubella, which fit to case reports [3,20]. Thereafter, the SEIR model has been applied to other subjects through suitable modifications, e.g. incorporating seasonality into the model gave explanation for chaotic epidemic patterns of measles [6,21,22].

Dataset

The present work used the daily reports of laboratory-confirmed cases. The reports in Tokyo have been published by the Bureau of Social Welfare and Public Health, Tokyo Metropolitan Government [10,11] since January 24, on which the first confirmed case in Tokyo was reported. These reports include the numbers of confirmed cases, fatalities, severe cases, patient discharges, cases by municipality in Tokyo, and continuous tracking of all such data. The daily reports in NYC have been published by the City of New York since February 29 [12]. Those reports include the number of cases by age groups, sex, boroughs, and the number of fatalities.

Back calculation

We used the back-calculation method [34] to estimate the number $i(t_i)$ of infected individuals at time t_i of occurrence of infection: the reported number of individuals who are newly diagnosed as PCR-positive is transformed to the number of individuals at the time when infection occurred. Originally, the back-calculation method was developed to estimate HIV infections in 1980s and is still used.

The number $n(t_i)$ of reported cases is given as $n(t_i) = \sum_j f(t_i - t_j)i(t_j)$, where $f(t_i - t_j)$ denotes the distribution function for time period $t_i - t_j$ from infection at t_j to PCR-testing at t_i . For $f(t_i - t_j)$, we did not have appropriate data and functions; therefore, we supplied this deficiency by substituting the distribution function that was proposed for the serial interval (the time between successive cases in a chain of transmission). The assumption that infections occur at random during an infectious period of an infectious individual leads to the conclusion that the average serial interval is the sum of the average latent period and half the average infectious period. Accordingly, it seems to be relevant to consider that the serial interval is comparative to the period from infection to PCR-testing. For SARS-CoV-2 infections, this distribution is a Gamma distribution with the mean and standard deviations of 7.5 and 3.4 days, respectively [33]. The reported cases were smoothed by 7-day-average, and the above equation was solved by deconvolution.

Effective reproduction numbers

We evaluated effective reproduction numbers from the case reports adjusted by the back-calculation and the SEIR model. The SEIR model provides an effective reproduction number R_e as $R_e = (r + \sigma)(r + \gamma)/\sigma\gamma$ [5]. We set $1/\sigma=5.2$ days and $1/\gamma=4.6$ days, following previous studies for Wuhan

[33].

References

1. World Health Organization. Coronavirus disease (COVID-2019) situation reports. Available from: <https://covid19.who.int/>, 2021
2. Yanagi, Y. *Nishi Nippon Shinbun* (2020) in Japanese. Available from: <https://www.nishinippon.co.jp/item/n/646933>
3. Anderson, R. & May, R. *Infectious diseases of humans: dynamics and control*. (Oxford, UK: Oxford University Press, 1991)
4. Daley, D. J. & Gani, J. *Epidemic Modelling: an Introduction* (Cambridge: Cambridge University Press, 1999)
5. Diekmann, O. & Heesterbeek, J. A. P. *Mathematical Epidemiology of Infectious Diseases* (Chichester: John Wiley & Sons, 2000)
6. Keeling, M. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton: Princeton University Press, 2008)
7. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**(1609), 599–604 (2007). 10.1098/rspb.2006.3754, Pubmed:17476782
8. Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infect. Dis. Modell.* **5**, 129-141 (2020). 10.1016/j.idm.2019.12.009, Pubmed:31956741
9. Thompson, R. Pandemic potential of 2019-nCoV. *Lancet Infect. Dis.* **20**(3), 280 (2020). 10.1016/S1473-3099(20)30068-2, Pubmed:32043978

10. Bureau of Social Welfare and Public Health, Tokyo Metropolitan Government.
Available from: <https://www.fukushihoken.metro.tokyo.lg.jp/hodo/index.html>, 2020
11. Tokyo Metropolitan Government. Updates on COVID-19 in Tokyo, Available from:
<https://stopcovid19.metro.tokyo.lg.jp/en>, 2020
12. City of New York. COVID-19: data archive, Available from:
<https://www1.nyc.gov/site/doh/covid/covid-19-data.page>, 2020
13. May R. M. Uses and abuses of mathematics in biology. *Science* 303, 790-793 (2004)
14. Metcalf, C. J. E., Morris, D. H., Park, S. W. Mathematical models to guide pandemic response. *Science* **369**, 368–369 (2020)
15. Walker et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, 413–422 (2020)
16. Kato, F., Tainaka, K., Sone, S., Morita, S., Iida, H. & Yoshimura, J. Combined effects of prevention and quarantine on a breakout in SIR model. *Scientific Reports* **1**, 10; 10.1038/srep00010 (2011).
17. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **395**(10228), 931-934 (2020). 10.1016/S0140-6736(20)30567-5, Pubmed:32164834
18. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**(6493), 860-868 (2020). 10.1126/science.abb5793, Pubmed:32291278
19. Anderson, R. & May, R. Vaccination against rubella and measles: quantitative investigations of different policies. *J. Hyg. Camb.* **90**, 259-325 (1983).
20. Olsen L. F., Schaffer, W. M. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science* 249(4968), 499-504 (1990).
10.1126/science.2382131

21. Earn, D. J. D., Rohani, P., Bolker, B. M., Grenfell, B. T. A simple model for complex dynamical transitions in epidemics. *Science* **287**(5453), 667-670 (2000).
10.1126/science.287.5453.667
22. Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Rico, M. S., Limosin, F. & Leleu, H. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nat. Med.* **26**, 1801 (2020).
23. Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. & Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **26**, 855-860 (2020).
24. Aguiar, M., Ortuondo, E., Bidaurrezaga, J., Mar, V., & Stollenwerk, N. Modelling COVID 19 in the Basque Country from introduction to control measure response. *Scientific Reports* **10**, 17306; <https://doi.org/10.1038/s41598-020-74386-1> (2020).
25. López, L. & Rodó, X. The end of social confinement and COVID-19 re-emergence risk. *Nat. Hum. Behav.* **4**, 746-755 (2020).
26. Silva, C. J., Cruz, C., Torres, D. F. M., Muñuzuri A. P., Carballosa, A., et al. Optimal control of the COVID-19 pandemic: controlled sanitary deconfinement in Portugal. *Scientific Reports* **11**, 3451; <https://doi.org/10.1038/s41598-021-83075-6> (2021).
27. Cowling, B. J., et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Health* **5**(5), e279-e288 (2020). 10.1016/S2468-2667(20)30090-6, Pubmed:32311320
28. Huang, C., Liu, X., Sun, S., Li, S., Deng, M., He, G., et al. Insights into the transmission of respiratory infectious diseases through empirical human contact networks. *Scientific Reports* **6**,31484; DOI: 10.1038/srep31484 (2016).

29. Hao, X. *et al.* Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**(7821), 420-424 (2020). 10.1038/s41586-020-2554-8, Pubmed:32674112
30. Prem, K., Liu, Y., Russel, T. W., Kucharski, A., Eggo, R. M. & Davies. *et al.* The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Glob. Health* **5**(5), e261-e270 (2020). 10.1016/S2468-2667(20)30073-6, Pubmed:32220655
31. Davies, N. G. *et al.* Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health* **5**(7), e375-e385 (2020). 10.1016/S2468-2667(20)30133-X, Pubmed:32502389
32. Britton, T., Ball, F. & Trapman, P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**(6505), 846-849 (2020). 10.1126/science.abc6810, Pubmed:32576668
33. Li, Q., *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**(13), 1199-1207 (2020). 10.1056/NEJMoa2001316, Pubmed:31995857
34. Brookmeyer, R. & Gail, M. H. Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* **2**(8519), 1320-1322 (1986). 10.1016/s0140-6736(86)91444-3, Pubmed:2878184

Acknowledgements

This work was supported by National Center for Global Health and Medicine, Japan, and the National Institutes of Health, US.

Ethical approval

Ethical approval is not necessary for the present work because the study was based only on the statistical data made available online by the relevant governments.

Author contributions

The authors conceived the study. K.O. performed calculations, data collection, and analyses. K.O. and H.M. interpreted the results and wrote the manuscript. The authors contributed to the draft.

Funding

The present work was supported in part by Health Labour Sciences Research Grant from the Ministry of Health Labour and Welfare (K.O.), from the Intramural Research Program of the Center for Cancer Research, National Cancer Institute, National Institutes of Health (H.M.), and by a grant for Development of Novel Drugs for Treating COVID-19 (H.M., 19A-3001) from the Intramural Research Program of National Center for Global Health and Medicine.

Competing interests

The authors have no conflicts of interest directly relevant to the content of the present study.

Additional information

Correspondence and requests for materials should be addressed to K.O.

Figure legends

Figure 1: Trajectories of the representative point $(S(t), E(t) + I(t))$ in the $EI-S$ plane calculated with the SEIR compartment model, which sets $\sigma = 1/5.2$, $\gamma = 1/2.3$, and $E(0)/N = 10^{-5}$. The values represent the proportions scaled by the total population N . The arrows denote the advancing direction of the representative point. (a) The trajectory with the basic reproduction number $R_0 = 2.5$. The herd immunity level is given by the peak location of the trajectory (broken line). (b) Change in trajectories when R_0 is replaced by R_m : $R_m = 2.0$ (green solid), 1.6 (green broken), 1.5 (red short-broken), 1.4 (red long-broken), and 1.2 (red solid). (c) Trajectories when the reproduction number is restored in the middle of an epidemic from $R_m = 1.2$ to $R_0 = 2.5$ at 300 (red), 400 (green), and 500 (blue) days after the beginning of the epidemic. (d) Effects of vaccination. The vaccination rate $v = 0.002$ (red), 0.003 (green), 0.005 (blue), and 0.01 (orange).

Figure 2: Time to reach the herd immunity level versus the modified reproduction number. Calculations were performed by the SEIR model which sets $\delta = 1/5.2$, $\gamma = 1/2.3$, and $E(0)/N = 10^{-5}$. (a) Calculation with no vaccination. (b) Calculations incorporating vaccination: the vaccination rate $v = 0.001$ (red), 0.003 (green), 0.005 (blue), 0.01 (orange), and 0.02 (violet). The black curve is the same as that in (a).

Figure 3: Velocity of the change in the number of infectious individuals as a function of time. The value of reproduction number is set to $R_0 = 2.5$ (black), $R_m = 2.25$ (red), 2.00 (green), 1.75 (blue), and 1.50 (orange). The curves were calculated until the herd immunity level was reached.

Figure 4: The reported number of PCR confirmed cases (blue), the number of new cases estimated by the back-calculation (orange), and the reproduction numbers (black) in Tokyo (a) and NYC (b). The left-hand and right-hand scales are for the cases and the reproduction numbers, respectively. The frequency distributions of the reproduction numbers are shown in the bar charts for Tokyo (c) and NYC (d). The mean value was 1.07 and 1.00, and the standard deviation was 0.47 and 0.25 for Tokyo and for NYC, respectively, for the period from April 2020 to January 2021.

Figures

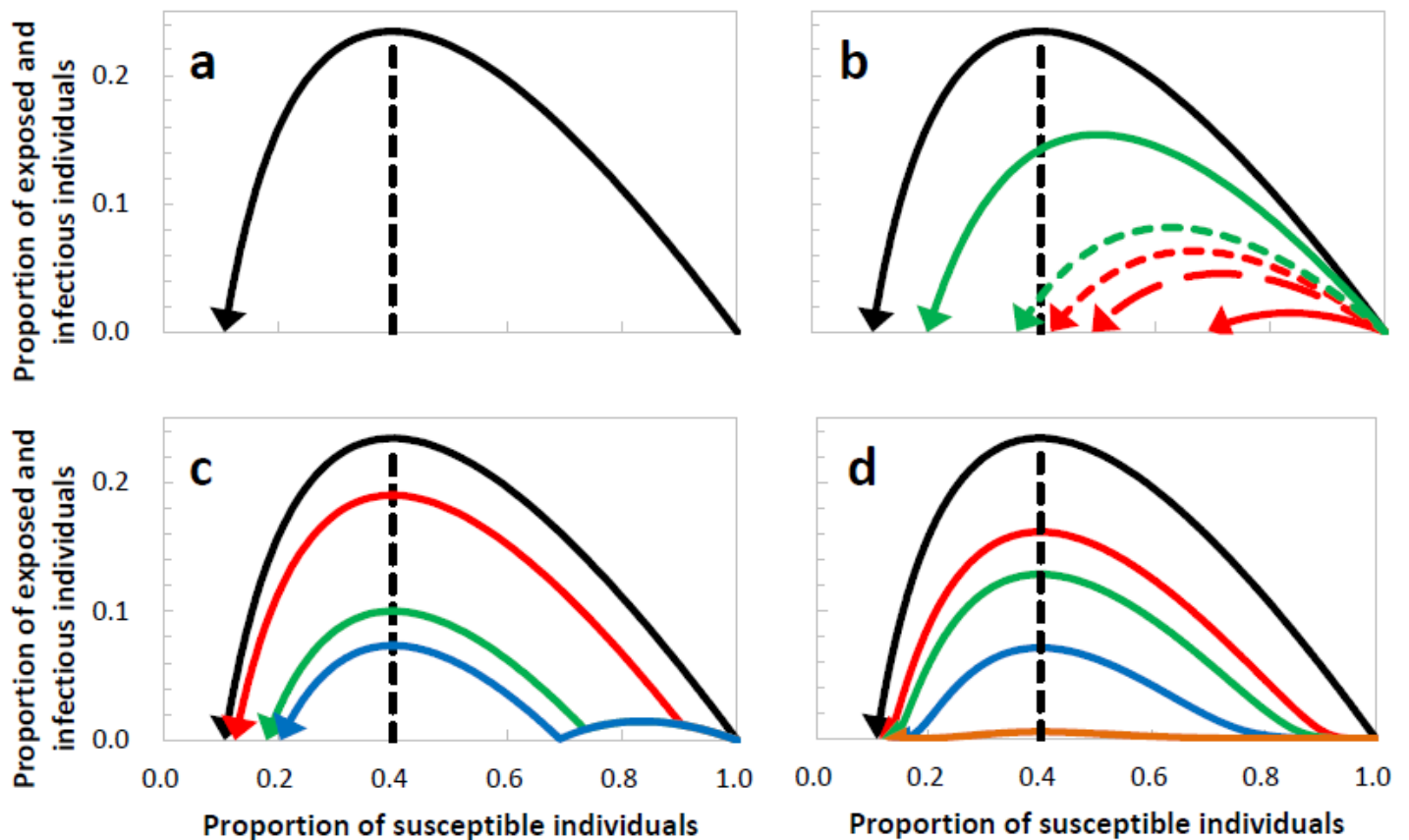


Figure 1

Trajectories of the representative point $(S(t), E(t)+I(t))$ in the EI-S plane calculated with the SEIR compartment model, which sets $\sigma=1/5.2$, $\gamma=1/2.3$, and $E(0)/N=10^{-5}$. The values represent the proportions scaled by the total population N . The arrows denote the advancing direction of the representative point. (a) The trajectory with the basic reproduction number $R_0=2.5$. The herd immunity level is given by the peak location of the trajectory (broken line). (b) Change in trajectories when R_0 is replaced by R_m : $R_m=2.0$ (green solid), 1.6 (green broken), 1.5 (red short-broken), 1.4 (red long-broken), and 1.2 (red solid). (c) Trajectories when the reproduction number is restored in the middle of an epidemic from $R_m=1.2$ to $R_0=2.5$ at 300 (red), 400 (green), and 500 (blue) days after the beginning of the epidemic. (d) Effects of vaccination. The vaccination rate $v= 0.002$ (red), 0.003 (green), 0.005 (blue), and 0.01 (orange).

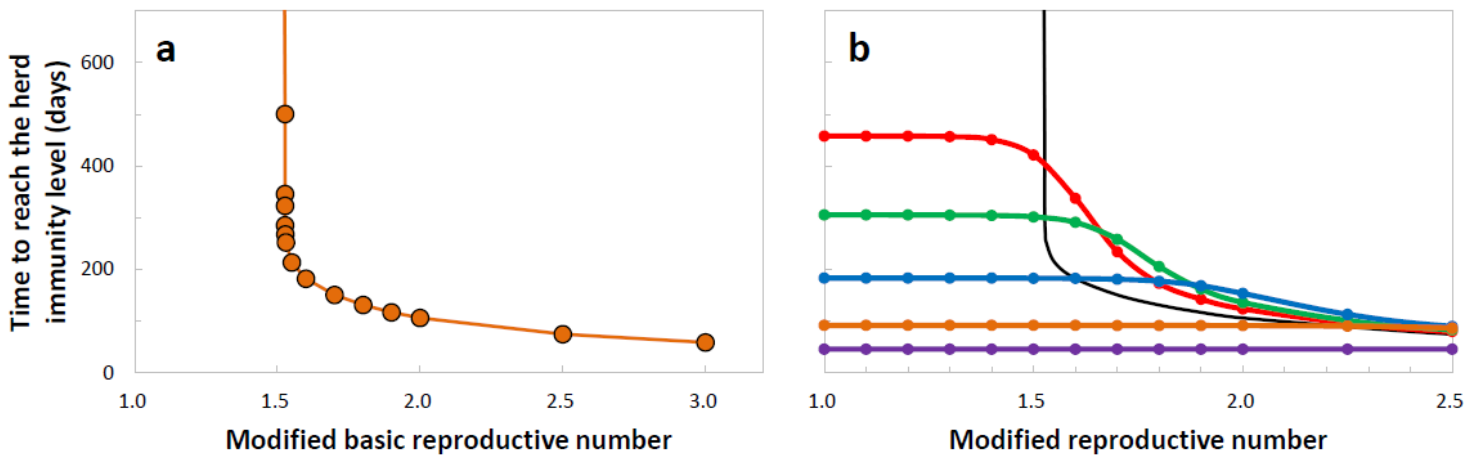


Figure 2

Time to reach the herd immunity level versus the modified reproduction number. Calculations were performed by the SEIR model which sets $\delta=1/5.2$, $\gamma=1/2.3$, and $E(0)/N=10^{(-5)}$. (a) Calculation with no vaccination. (b) Calculations incorporating vaccination: the vaccination rate $v=0.001$ (red), 0.003 (green), 0.005 (blue), 0.01 (orange), and 0.02 (violet). The black curve is the same as that in (a).

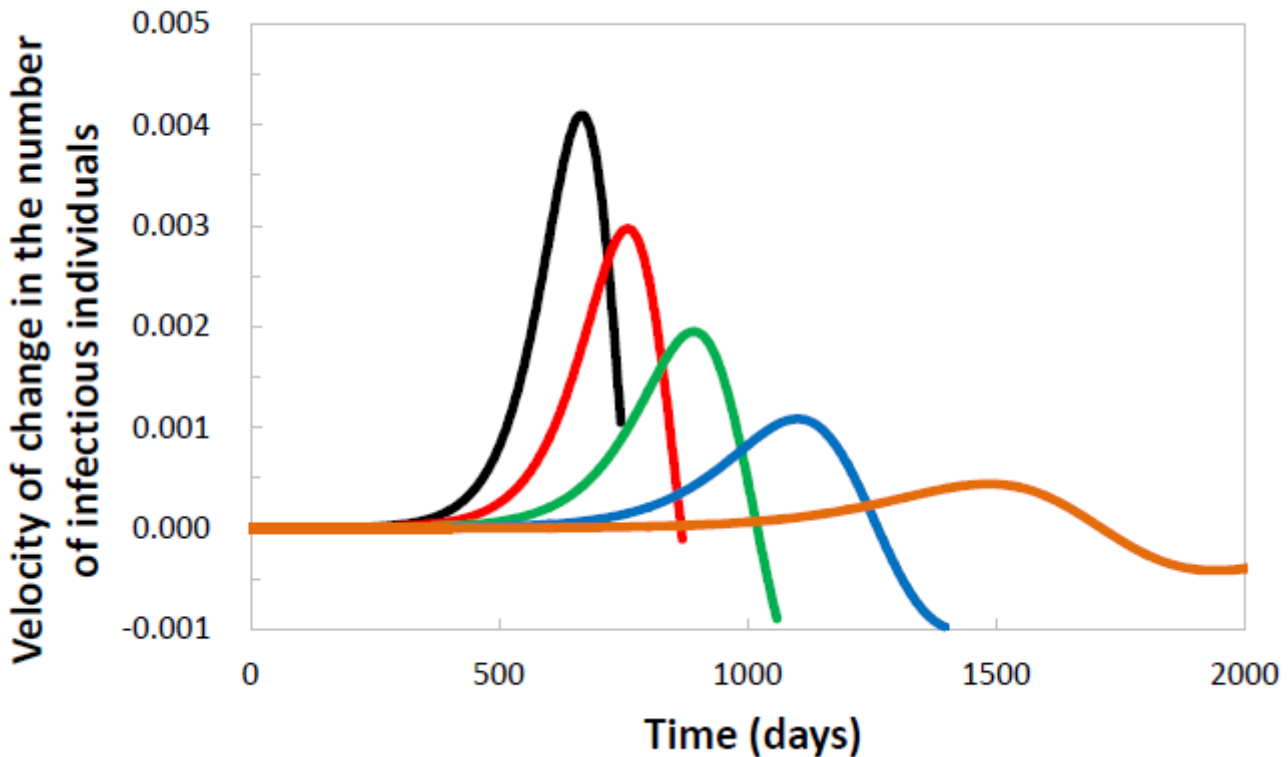


Figure 3

Velocity of the change in the number of infectious individuals as a function of time. The value of reproduction number is set to $R_0=2.5$ (black), $R_m=2.25$ (red), 2.00 (green), 1.75 (blue), and 1.50 (orange).

The curves were calculated until the herd immunity level was reached.

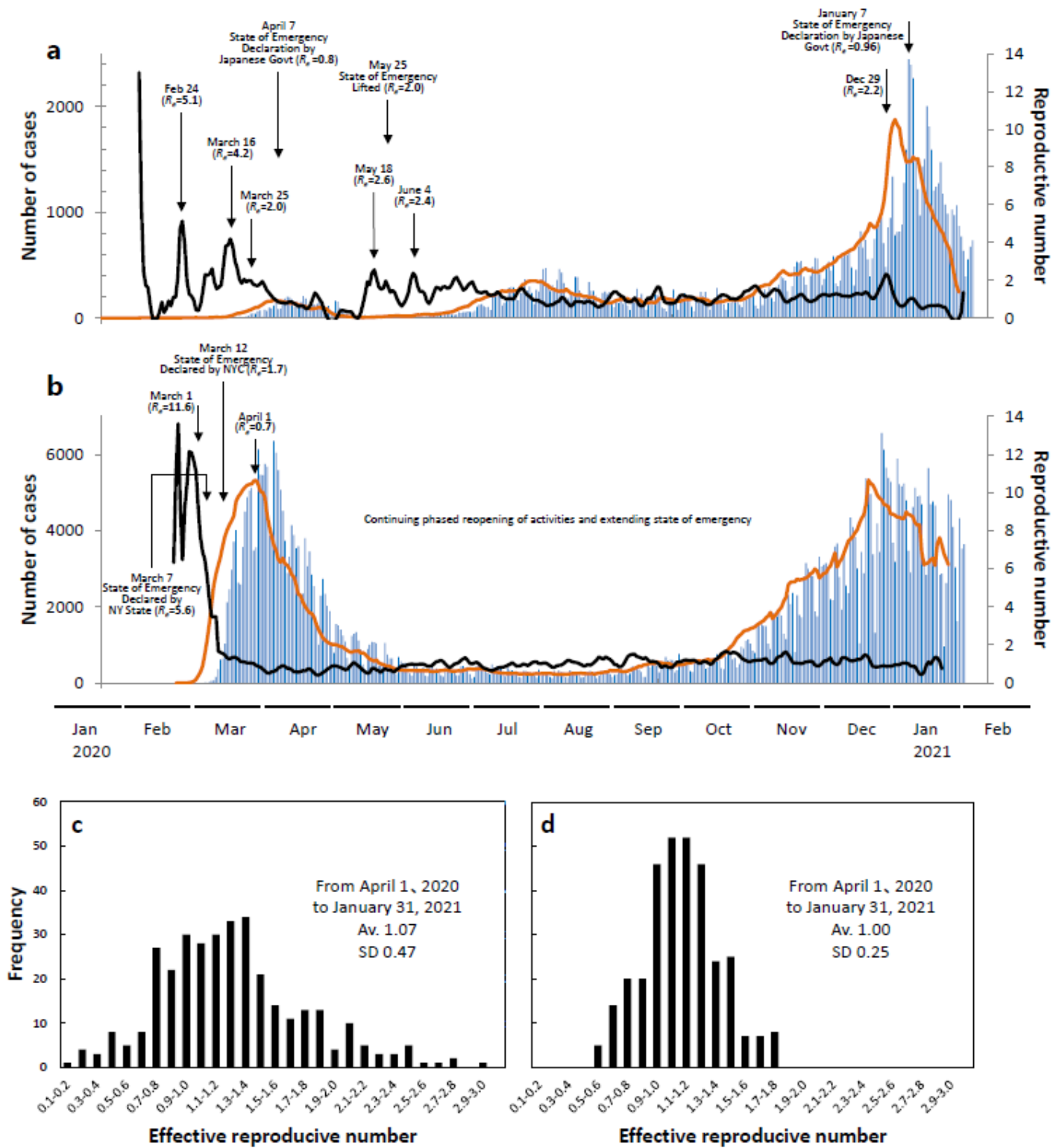


Figure 4

The reported number of PCR confirmed cases (blue), the number of new cases estimated by the back-calculation (orange), and the reproduction numbers (black) in Tokyo (a) and NYC (b). The left-hand and right-hand scales are for the cases and the reproduction numbers, respectively. The frequency distributions of the reproduction numbers are shown in the bar charts for Tokyo (c) and NYC (d). The

mean value was 1.07 and 1.00, and the standard deviation was 0.47 and 0.25 for Tokyo and for NYC, respectively, for the period from April 2020 to January 2021.