Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Theoretical Biology 554 (2022) 111278

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Doubling time of infectious diseases

Asami Anzai, Hiroshi Nishiura *

Kyoto University School of Public Health, Yoshida-Konoe, Sakyo-ku, Kyoto 606-8601, Japan

ARTICLE INFO

This manuscript was submitted as part of a theme issue on "Modelling COVID-19 and Preparedness for Future Pandemics".

Keywords:
Exponential growth
Intrinsic growth rate
Mathematical model
Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)
Transmission dynamics

ABSTRACT

The concept of doubling time has been increasingly used since the onset of the coronavirus disease 2019 (COVID-19) pandemic, but its characteristics are not well understood, especially as applied to infectious disease epidemiology. The present study aims to be a practical guide to monitoring the doubling time of infectious diseases. Via simulation exercise, we clarify the epidemiological characteristics of doubling time, allowing possible interpretations. We show that the commonly believed relationship between the doubling time and intrinsic growth rate in population ecology does not strictly apply to infectious diseases, and derive the correct relationship between the two. We examined the impact of varying (i) the growth rate, (ii) the starting point of counting cumulative number of cases, and (iii) the length of observation on statistical estimation of doubling time. It was difficult to recover values of growth rate from doubling time, especially when the growth rate was small. Starting time period is critical when the statistical estimation of doubling time occurs during the course of an epidemic. The length of observation was critical in determining the overall magnitude of doubling time, and when only the latest 1–2 weeks' data were used, the resulting doubling time was very short, regardless of the intrinsic growth rate r. We suggest that doubling time estimates of infectious disease epidemics should at a minimum be accompanied by descriptions of (i) the starting time at which the cumulative count is initiated and (ii) the length of observation.

1. Introduction

Doubling time, the time it takes for a number of individuals to double, is classically used in the field of population ecology. Nowadays, doubling time is also used in the field of infectious disease epidemiology as a measurement of the spread of disease, representing the time required for the cumulative number of infections to double during the course of an epidemic (University of Cambridge, 2021; Vynnycky and White, 2010). Usually, doubling time is estimated using the time series data of (infected) individuals or from the growth rate of individuals, and the growth rate usually represents the rate at which individuals increase per unit time, i.e., a nearly opposing concept of doubling time (Gotelli, 2008). As a more widely accepted measurement of transmissibility of infectious diseases, the basic reproduction number, which is interpreted as the average number of secondary cases produced by a single primary case in a fully susceptible population, is a well-defined dimensionless quantity. Nevertheless, the basic reproduction number requires us to know the length of the generation time in advance of the estimation, especially to attain a real-time assessment. The importance of estimating an alternative epidemiological metric to assess the transmissibility or the speed of growth of cases in the early stages of the COVID-19 pandemic has been emphasized (Thompson et al., 2020). When the generation time has yet to be fully quantified, the speed of epidemic growth must be measured in real time (Dushoff and Park, 2021; Ridenhour et al., 2014), and doubling time, T_d, is one of the available alternative metrics, and intrinsic growth rate, r, is another possible choice.

In the field of ecology, the relationship between T_d and r has been understood to be simple. Let $C(t)$ be the population size at calendar time t, and $C(0)$ be the initial value of the population. As for the concept of doubling, we have

$$C(t) = C(0)2^{\frac{t}{T_d}}, \tag{1}$$

and for the exponentially growing phase, we have.

$$C(t) = C(0)\exp(rt). \tag{2}$$

The right-hand sides of (1) and (2) are equated, and we then obtain the relationship.

* Corresponding author.
E-mail address: nishiura.hiroshi.5r@kyoto-u.ac.jp (H. Nishiura).

https://doi.org/10.1016/j.jtbi.2022.111278
Received 30 June 2022; Received in revised form 18 August 2022; Accepted 7 September 2022
Available online 13 September 2022

$$T_d = \frac{\ln 2}{r}, \tag{3}$$

which is well known in the field of ecology (Gotelli, 2008). Nevertheless, the Eq. (2) is not strictly the case for the cumulative number of infectious diseases, and the estimation of T_d using (3) is not applicable (see Methods).

The defining relationship between T_d and r has not been well formulated, nor have statistical methods to measure doubling time. Doubling time is estimated using the cumulative number of cases over time, e.g., by using confirmed cases, and can therefore be used to detect rapid increases in the number of cases. In particular, it is attractive that T_d can be measured even when intrinsic characteristics of an infectious disease, such as the generation time or incubation period, remain unknown (Pellis et al., 2021). For instance, during the epidemic of severe acute respiratory syndrome (SARS) from 2002 to 3, time-dependent changes in doubling time value were noted, and possible factors affecting doubling time have been discussed (Galvani et al., 2003). Doubling time has been also used for monitoring the epidemiological dynamics of coronavirus disease-2019 (COVID-19). In China, where COVID-19 was first widespread, the doubling time in the early stages of the epidemic was estimated to be 1.4–3.1 days by province (Muniz-Rodriguez et al., 2020). Following China, the epidemic was seen in Italy (Remuzzi and Remuzzi, 2020; World Health Organization, 2020), and the initial doubling time in Italy was estimated to be 3 days (Riccardo et al., 2020). Doubling time was also employed to measure the spread of the Omicron variant (B.1.1.529) in the United Kingdom during its early phases (UK Health Security Agency, 2021). In addition to monitoring the epidemiological dynamics, some studies have used doubling time to evaluate interventions against the epidemic (Khosrawipour et al., 2020; Liang et al., 2021). In sum, in situations where an epidemic grows rapidly, doubling time has been used to describe the rapidity of increase, and such an exercise has preceded an accurate statistical estimation of the basic (or effective) reproduction number; additionally, no technical discussion took place as to (i) when and how to start counting the cumulative number of cases, (ii) during which epidemic phases the measurement would be deemed useful, or (iii) how to interpret the estimate of doubling time.

Doubling time has been increasingly used since the COVID-19 epidemic began, but its characteristics are not well understood, especially in its application to infectious disease epidemiology. As the measurement relies on the cumulative number of cases, it is vital to understand when and for how long the cumulative count should be taken. Moreover, even provided that a given doubling time is 3 days, we have not firmly understood how to interpret such a value. Considering that doubling time is easy to compute, this measurement will likely continue to be employed as part of epidemiological monitoring. It is vital to understand the epidemiological characteristics of doubling time in advance of such use.

The present study aims to be a practical guide to monitoring the doubling time of infectious diseases. Via simulation exercise, we clarify the epidemiological characteristics of doubling time, allowing possible interpretations.

2 Materials and methods

In the following, we first describe the analytical relationship between T_d and r through a trivial mathematical exercise. Subsequently, we describe the details of a simulation-based investigation.

2.1. Modelling doubling time

Here we describe the analytical relationship. Doubling time T_d is estimated using the cumulative number of infected cases at time t, $C(t)$, i.e.,

$$C(t) = C(0)2^{\frac{t}{T_d}}, \tag{4}$$

where $C(0)$ is the initial value. It should be noted that $C(0)$ cannot be zero for $C(t) > 0$ for $t > 0$, and thus doubling time does not assume that cases are counted from the very beginning of an epidemic.

Next, in the case of infectious diseases, not the cumulative number but the incidence of infection grows exponentially with the intrinsic growth rate r. The number of newly infected cases at time t, $i(t)$, is expressed by

$$i(t) = i(0)exp(rt). \tag{5}$$

where $i(0)$ is the initial value. Taking the integral from time 0 to time t, we obtain the cumulative number of infected individuals, $I(t)$:

$$I(t) = \int_0^t i(s)ds = \frac{i(0)}{r}(exp(rt) - 1). \tag{6}$$

It should be noted that $C(t)$ and $I(t)$ cannot be immediately equated; there must be a clock zero to start counting $C(t)$, the cumulative number of cases, say t_0. At time t_0, we have the relationship:

$$C(0) = i(t_0) = i(0)exp(rt_0). \tag{7}$$

That is, the cumulative counting starts at t_0 and $C(0)$ is the same as the incidence at time t_0, $i(t_0)$. In particular, we assume that $C(0)$ was seen over $t \in [t_0\text{-}1, t_0]$. From the time t_0, the cumulative number of cases at time $t_0+\tau$, following Eq. (4), is

$$C(\tau) = C(0)2^{\frac{\tau}{T_d}} = i(0)exp(rt_0)2^{\frac{\tau}{T_d}}. \tag{8}$$

It should be noted that τ is the time elapsed since the start of counting the cumulative number of cases. As for the cumulative number of cases following Eq. (6), we have

$$I(t_0) = \frac{i(0)}{r}(exp(rt_0) - 1), \tag{9}$$

as the summation from time 0 to t_0. What corresponds to the quantity in the right-hand side of Eq. (8) is the integral of $i(t)$ from time t_0-1 to $t_0+\tau$, thus,

$$C(\tau) = I(t_0 + \tau) - I(t_0 - 1). \tag{10}$$

Notably, $C(0)$ in Eq. (7) was dealt with as the discrete quantity, and therefore the integral from t_0 to $t_0 + \tau$was calculated as the difference between $I(t_0 - 1)$ and $I(t_0 + \tau)$. From Eq. (10), we obtain

$$i(0)exp(rt_0)2^{\frac{\tau}{T_d}} = \frac{i(0)}{r}\left(exp\left(r(t_0+\tau)\right) - 1\right) - \frac{i(0)}{r}\left(exp\left(r(t_0-1)\right) - 1\right), \tag{11}$$

which can be reduced to

$$2^{\frac{\tau}{T_d}} = \frac{exp(r\tau) - exp(-r)}{r}, \tag{12}$$

In the end, we obtain.

$$T_d = \frac{\tau}{log_2 \frac{(exp(r\tau) - exp(-r))}{r}}, \tag{13}$$

which we suggest replacing the Eq. (3) in the case of infectious disease epidemiology. It should be noted that the Eq. (13) is not influenced by the initial value at time t_0, and therefore the doubling time is still independent of the empirically observed data; however, the Eq. (13) certainly contains τ, the time elapsed since starting to count cumulative number of cases.

2.2. Computational exercise

While the abovementioned Eq. (13) provides the analytical relationship between T_d and r, the finding is restricted to exponential growth phase, and moreover, we have not secured a stable interpretation of T_d.

What if the starting time to count cumulative number of cases is varied? What if the length of observation τ is very short? We tackle these points via numerical simulations.

There have been several different methods for estimating the doubling time. The first is to use Eq. (3). As we discuss with Eq. (13), Eq. (3) may not be directly applicable to infectious disease. The second method is to use two data points: supposing that the cumulative number of cases at time t_1 and time t_2 were $C(t_1)$ and $C(t_2)$, respectively, the doubling time is calculated as

$$T_d = (t_2 - t_1)\frac{\ln 2}{\ln \frac{C(t_2)}{C(t_1)}}. \quad (14)$$

The Eq. (14) is preferred when the calculation must be kept simple (e.g., when using a spreadsheet program). Nevertheless, this method forces us to select two specific time points arbitrarily, and such choice leads to biased estimation when the number of cases remains small and stochasticity cannot be ignored. The third method is to fit Eq. (4), i.e., the equation of the cumulative number of cases, to the empirical data. This method uses an additional number of data points and can thereby avoid potential inflation of T_d. Here we estimated doubling time T_d from simulated epidemic data using Eq. (4), assuming that the variations in cases are captured by Gaussian distribution and employing a maximum likelihood method. In the present study, doubling time is defined as the time that the cumulative number of cases from the estimated start time doubles. The doubling time would always take a positive value, because the cumulative number of cases increases even when the growth rate is negative and the incidence is decreasing.

2.2.1. Varying interpretations during exponential growth

First, we explored how T_d varies by varying (i) the growth rate, (ii) the starting point, and (iii) the time period used in the calculation on the doubling time (i.e., the length of observation τ) during the exponential growth phase. Consider an epidemic in which the number of new infected cases $i(t)$ at time t is described by $i(t) = i(0)e^{rt}$. The initial value is not influential and is therefore set at $i(0) = 1$. Then, we varied r from $-0.1, -0.01, 0, 0.01$ to 0.1 per day, and the length of observation τ was also varied from 0 to 50 days. We examined how the estimate of T_d differed when doubling time was calculated for each given growth rate for different periods of time, and when doubling time was calculated for the latest 7, 14, and 21 days. In the case of an infectious disease epidemic with a single peak, all datasets from day zero would be used to calculate the doubling time. However, in the case of COVID-19 with multiple epidemic waves, it has been practically the case that only the datasets of the latest few weeks were subject for analysis. The intrinsic growth rate was calculated from the doubling time using Eq. (13), and we examined how well the estimated doubling time reflected the growth of cases.

2.2.2. Doubling time during the course of an epidemic

Second, we conducted simulations using the susceptible-infectious-recovered/removed (SIR) model written by ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t),$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t),$$

$$\frac{dR(t)}{dt} = \gamma I(t), \quad (15)$$

where β is the transmission coefficient and γ is the rate of recovery. Using the numerical solution of the incidence from this model, we examined how the estimates of T_d varied depending on how the starting point was taken during the course of epidemic (e.g., during the increasing phase, near the peak, or during the decreasing phase) and what period of the epidemic we used as the length of observation τ. When the SIR model was employed, we performed simulations with $r = 0.1$ and the mean generation time $T_g = 5$ days (assumed to be identical to the mean infectious period), thus in the SIR model written by three ordinary differential equations, $R_0 = 1 + rT_g = 2$. Similar to the simulation with the exponential growth, T_d was estimated by varying (i) the starting point and (ii) the time period used in the calculation of the doubling time.

3. Results

Fig. 1 shows the effect of exponential growth rate on the statistical estimate of doubling time during the exponential growth phase. The simulation was conducted for a total of 50 days, and the doubling time was measured in four different ways, i.e., using data from time 0 to time 50 (entire period), using the data for the latest 7 days (Day 44–50), the latest 14 days (Day 37–50), and the latest 21 days (Day 30–50). In the all-50-days case, T_d was estimated to range from 4.4 to 12.5 days depending on the growth rate. Similarly, when the latest 7-day data were used, T_d was estimated to range from 1.7 to 2.1 days; when the latest 14-day or 21-day data were used, T_d was estimated to range from 2.5 to 3.5 days and 3.1 to 5.5 days, respectively. That is, as the length of observation τ was shortened, the resulting doubling time was estimated to be shorter. Even when r was negative (i.e., in the decreasing phase), values of T_d close to other growth rates were obtained. Especially when the latest 7-day or 14-day data only were used, the resulting doubling time values were close to each other and tended to be very short (i.e., on the order of a few days).

Table 1 summarizes the estimate of doubling time during the exponential growth phase from a different angle from Fig. 1. In Table 1, the simulation was conducted for the total of 60 days, and thus, using the data for the latest 7 days, 14 days and 21 days represent Days 54–60, Days 47–60 and Days 40–60, respectively. Qualitatively, similar patterns to Fig. 1 were observed. In addition to Fig. 1, neither the estimate and the uncertainty bounds (i.e., 95 % confidence intervals) were very sensitive to the time at which estimation was conducted. That is, the results of estimation using 0–20 days, 0–40 days and 0–60 days were not very variable. When the length of observation was short (e.g., 7 days), the impact of variations in r was again minimal.

Fig. 2 shows estimates of T_d during the course of an epidemic. As was examined for the exponential growth model, we varied starting time to count the cumulative number of cases and the length of observation

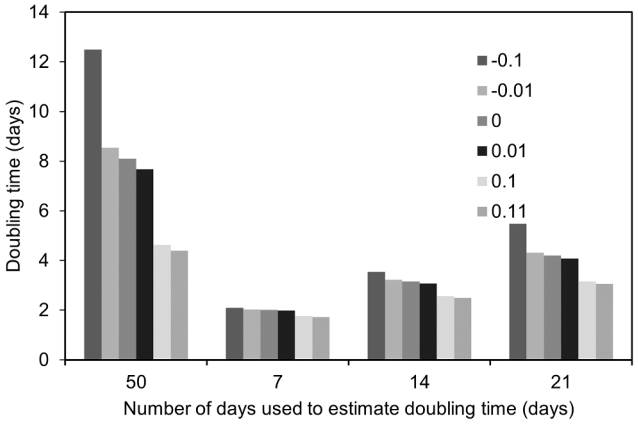


Fig. 1. Impact of varying growth rates and length of observation on doubling time. Estimates of doubling time by different exponential growth rates. The horizontal axis represents the length of observation used for estimation of doubling time. The simulation was conducted for 50 days, and the doubling time was measured in four different ways, i.e., using data from time 0 to time 50 (all period), using the data for the latest 7 days (Day 44–50), the latest 14 days (Day 37–50), and the latest 21 days (Day 30–50).

Table 1
Estimates of doubling time during exponential growth phase.

Period used to estimate doubling time	All observed days used (60 days)	Number of recent days used 7 days	14 days	21 days
	Growth rate = −0.01			
0 – 20		2.04 (1.85, 2.33)	3.23 (3.02, 3.52)	4.31 (4.08, 4.63)
0 – 40	7.18 (6.89, 7.54)	2.04 (1.85, 2.32)	3.22 (3.02, 3.52)	4.32 (4.09, 4.63)
0 – 60	9.88 (9.54, 10.30)	2.03 (1.85, 2.32)	3.24 (3.03, 3.53)	4.31 (4.08, 4.63)
	Growth rate = 0.01			
0 – 20		1.98 (1.81, 2.25)	3.09 (2.90, 3.35)	4.08 (3.88, 4.34)
0 – 40	6.55 (6.34, 6.82)	1.98 (1.81, 2.25)	3.09 (2.91, 3.35)	4.08 (3.88, 4.34)
0 – 60	8.74 (8.51, 9.01)	1.99 (1.81, 2.25)	3.10 (2.91, 3.35)	4.08 (3.88, 4.34)
	Growth rate = 0.1			
0 – 20		1.76 (1.63, 1.94)	2.55 (2.44, 2.69)	3.15 (3.06, 3.26)
0 – 40	4.27 (4.22, 4.32)	1.76 (1.63, 1.94)	2.55 (2.45, 2.69)	3.15 (3.06, 3.26)
0 – 60	4.91 (4.88, 4.93)	1.76 (1.63, 1.94)	2.55 (2.45, 2.69)	3.15 (3.06, 3.26)

Numbers in parenthesis represent 95% confidence intervals as computed via the profile likelihood method. During the maximum likelihood estimation, daily cases were assumed to follow Gaussian distribution.

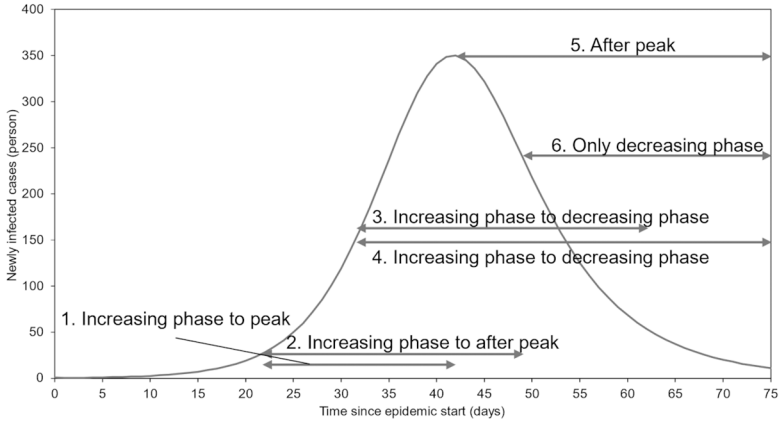
during the course of epidemic as described by the SIR model. Doubling time during the increasing phase (from the beginning of the epidemic to the peak, i.e., Day 22 to 42) yielded the shortest estimate, and the value calculated after the peak (Day 49–75) yielded the longest T_d. As the length of observation was shortened from 21 days to 7 days, T_d became shorter.

Table 2 shows the estimated doubling time during the course of an epidemic using the SIR model. Using the nonlinear epidemic model, Table 2 shows the impact of involving datasets during sub-exponential and decreasing phases on the estimate of T_d, which can practically take place. The longer the time period used for estimation of T_d, the longer the doubling time estimate T_d would be. This is theoretically understandable because the highest growth rate is attained during the initial exponential growth phase, and the growth rate monotonically decreases subsequently. Using shorter lengths of observation, shorter estimates of T_d were obtained.

Table 3 shows the results of calculating the intrinsic (exponential) growth rate from the estimated doubling time using Eq. (13), exploring whether the growth rate can be recovered. The uncertainty bound of r informs the uncertainty of T_d. This was examined only during the exponential growth phase (for the total of 75 days). Notably, the growth rate was not successfully recovered when the length of observation was short or when the growth rate was small. Especially in cases of negative growth rate it became too difficult to recover the growth rate from doubling time. When $r > 0$, a longer time period of observation was required to obtain a value close to the original exponential growth rate. Fig. 3 compares Eqs. (3) and (13), attempting to recover the estimate of the intrinsic growth rate (set at 0.01 per day) from doubling time. The estimated intrinsic growth rate using Eq. (13) was closer to the original value compared with those recovered from the widely used estimator (3).

4 Discussion

The present study characterized the doubling time of infectious diseases, frequently used as an epidemiological measurement to quantify the speed of epidemic growth. Through a short analytical exercise, we have shown that the commonly believed relationship in population ecology (i.e., Eq. (3)) is not strictly the case for infectious diseases, and instead Eq. (13) should be used to describe the relationship between T_d and r. In addition, we examined the impact of varying (i) the growth rate, (ii) the starting point of counting the cumulative number of cases,



Period used to estimate doubling time	All observed days used	Number of recent days used 7 days	14 days	21 days	
22 – 42	-	1.84 (1.69, 2.08)	2.43 (2.32, 2.58)	2.71 (2.64, 2.81)	1. Increasing phase to peak
22 – 47	3.13 (3.04, 3.25)	2.06 (1.86, 2.38)	2.78 (2.62, 3.02)	3.05 (2.93, 3.20)	2. Increasing phase to after peak
32 – 62	5.03 (4.77, 5.40)	2.38 (2.11, 2.82)	3.95 (3.61, 4.49)	5.00 (4.63, 5.55)	3. Increasing phase to decreasing phase
32 – 75	7.02 (6.64, 7.54)	2.38 (2.11, 2.82)	4.06 (3.70, 4.62)	5.76 (5.29, 6.46)	4. Increasing phase to decreasing phase
42 – 75	7.82 (7.28, 8.60)	-	-	-	5. After peak
49 – 75	7.07 (6.52, 7.87)	-	-	-	6. Only decreasing phase

Fig. 2. Impact of varying period of observation and length of observation on doubling time during the course of an epidemic. The doubling time T_d was estimated during the course of an epidemic, simulated by the SIR model, varying the period of observation and the length of observation. Doubling times were estimated in six different time periods, including those at the start of the exponential phase, near the epidemic peak, and after the peak of the epidemic. The peak incidence was observed on Day 42.

Table 2
Doubling time by the time period and length of observation during the course of an epidemic based on SIR epidemic model.

Period used to estimate doubling time	All observed days used	Number of recent days used		
		7 days	14 days	21 days
0 – 25	3.04 (3.02, 3.06)	1.55 (1.46, 1.67)	2.09 (2.03, 2.15)	2.40 (2.37, 2.44)
0 – 30	3.13 (3.12, 3.15)	1.58 (1.48, 1.71)	2.11 (2.05, 2.18)	2.43 (2.39, 2.47)
0 – 35	3.23 (3.22, 3.25)	1.64 (1.53, 1.79)	2.18 (2.11, 2.27)	2.49 (2.45, 2.54)
0 – 40	3.37 (3.35, 3.40)	1.77 (1.63, 1.97)	2.33 (2.24, 2.46)	2.63 (2.57, 2.70)
0 – 45	3.57 (3.54, 3.61)	1.97 (1.79, 2.26)	2.62 (2.48, 2.82)	2.89 (2.80, 3.02)
0 – 50	3.82 (3.78, 3.88)	2.19 (1.96, 2.56)	3.06 (2.85, 3.36)	3.34 (3.18, 3.55)
0 – 55	4.12 (4.05, 4.19)	2.32 (2.06, 2.74)	3.54 (3.25, 3.97)	3.98 (3.74, 4.31)
0 – 60	4.43 (4.36, 4.53)	2.37 (2.11, 2.81)	3.87 (3.54, 4.39)	4.72 (4.39, 5.22)
0 – 65		2.38 (2.11, 2.83)	4.02 (3.67, 4.58)	5.34 (4.92, 5.96)
0 – 70		2.38 (2.11, 2.82)	4.06 (3.70, 4.63)	5.66 (5.20, 6.34)

Numbers in parenthesis represent 95% confidence intervals as computed via the profile likelihood method. During the maximum likelihood estimation, daily cases were assumed to follow Gaussian distribution.

Table 3
Recovery of exponential growth rate by the time period and length of observation during the course of an epidemic using Eq. (13).

Period used to estimate doubling time	Exponential growth rate (/day)					
	−0.1	−0.01	0	0.01	0.1	0.15
Most recent 7 days (69–75)	0.000991	0.0922	0.102	0.113	0.193	0.243
Most recent 14 days (62–75)	−0.0356	0.0429	0.052	0.0612	0.141	0.186
Most recent 21 days (55–75)	−0.0475	0.0278	0.035	0.0438	0.124	0.17
0–25	−0.052	0.0209	0.0291	0.0372	0.117	0.164
0–30	−0.0548	0.0161	0.0244	0.0331	0.113	0.161
0–35	−0.0568	0.0131	0.0214	0.0297	0.11	0.158
0–40	−0.0582	0.0106	0.0189	0.0273	0.108	0.156
0–45	−0.059	0.00885	0.0169	0.0253	0.107	0.155
0–50	−0.0598	0.00729	0.0155	0.0238	0.106	0.154
0–55	−0.0603	0.00606	0.0141	0.0225	0.105	0.154
0–60	−0.0607	0.00495	0.0131	0.0213	0.104	0.154
0–65	−0.061	0.00408	0.0121	0.0204	0.104	0.153
0–70	−0.0612	0.00329	0.0113	0.0197	0.103	0.153
0–75	−0.0614	0.00257	0.0107	0.0189	0.103	0.153

and (iii) the length of observation τ on the statistical estimate of T_d. The growth rate did not easily recover from T_d, especially when r was small (e.g., negative). The starting time period is critical when the statistical estimation of T_d is undertaken during the course of an epidemic. The length of observation τ was critical in determining the overall magnitude of T_d, and when only the data from the latest 1–2 weeks were used, the resulting T_d appeared to be very short, regardless of the intrinsic growth rate r. Without accounting for these findings, our simulations indicated

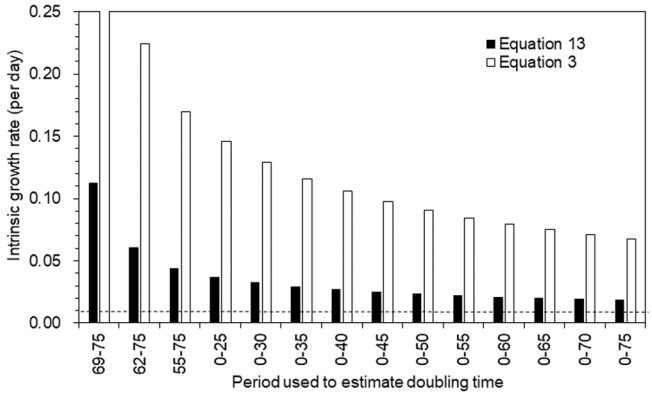


Fig. 3. Recovery of intrinsic growth rate from doubling time. The intrinsic growth rate r was calculated using two types of methods (Eqs. (3) and (13) in the main text, represented by unfilled and filled bars) using doubling time. The growth rate was assumed to be 0.01 per day (horizontal dashed line).

that it was fairly difficult to objectively interpret empirically estimated values of T_d in the epidemiology of infectious diseases. Compared with T_d, the intrinsic growth rate may be regarded as a less biased metric to describe the increase and decrease in the incidence.

To our knowledge, the present study is the first to have derived the relationship between T_d and r in epidemic data and to have clarified that T_d estimated from an identical r value may yield a completely different value if the empirical settings of observation are varied. These findings stem from the fact that doubling time is not calculated in a single unique way. If these issues persist, it is very difficult for an epidemiologist to judge whether a T_d estimate of 2–3 days is an alarming signal of epidemic growth.

To resolve this issue, we propose some friendly guidance for estimating T_d. Three tips from our exercise can contribute to it: First, whenever doubling time is presented, the starting time at which the cumulative count is initiated must be described (and the interpretation should take particular care on this point). It must also be remembered that the use of data from the latest 1–2 weeks alone tends to yield very short T_d estimates, as shown in Figs. 1 and 2. Second, the optimal length of time to assess T_d should be discussed in relation to the intrinsic transmission dynamics (or the natural history) of an infectious disease. Table 1 indicates that T_d could be extended, if the entire epidemic curve is used and the length of observation is extended. Third, it is very difficult to translate the doubling time value to the intrinsic growth rate, especially when the growth rate is small. Of course, the large intrinsic growth rate still has a potential to be recovered from the doubling time, as shown in Table 3, and therefore the estimation of r from T_d would be still sound when the actual growth of cases is fairly fast. In summary, we suggest that any T_d estimates of an infectious disease epidemic should at a minimum be accompanied by descriptions of the time at which the cumulative count is initiated and the length of observation.

Whereas we have shown that calculating doubling time can result in considerably different values by varying empirical estimation settings, T_d is very easy to calculate for monitoring infectious disease epidemics, especially when the natural history of the disease, including the generation time, is yet to be known. However, because of this easy-to-calculate feature, it is necessary to clearly specify the settings when using doubling time. Perhaps these points raised can be better understood if we imagine measuring the growth of cases by exponential growth rate (rather than T_d): we would have to specify (i) for what time we have used the data and (ii) how long the exponential growth rate was assumed to continue. The same applies to T_d. Of course, it is still useful to continuously monitor the same geographic area with a rapid increase in cases with a T_d of 2–3 days (as with COVID-19) to measure the increase using reasonable calculations. However, once the increase in the number of cases slows down, the use of T_d becomes complicated: T_d estimates

based on long-term data may not adequately represent the epidemiological situation, and therefore only the latest data may be used, but the use of the latest 1–2 weeks' data alone tends to result in smaller T_d estimates than those based on longer observation.

There are number of technical limitations that should be discussed. First, our study rests on a simulation study with limited parameter space, and it should be noted that the extent of bias depends on intrinsic transmission dynamics. Second, we have not been able to explicitly account for heterogeneity in measuring the doubling time. Alternative metrics other than a single growth rate or a single doubling time would be merited when the incidence is structured (e.g., by age group). Third, doubling time depends on cases, and when enormous growth is observed, the empirical data are influenced by ascertainment bias; e.g., once an epidemic is recognized, ascertainment of cases abruptly improves and the growth rate may be estimated as very large (i.e., doubling time is estimated as very short) because of this improved ascertainment over time. Without additional data, ascertainment bias cannot be addressed.

5 Conclusions

The present study characterized the doubling time of infectious diseases, frequently used as an epidemiological measurement to quantify the speed of epidemic growth. Cautions must be exercised when handling empirical data and providing estimate of a representative epidemiological metric. Eq. (13) should be used to describe the relationship between T_d and r for infectious diseases. Doubling time should be interpreted with caution, especially when estimated using data from the most recent few weeks. We suggest that T_d estimates of infectious disease epidemic should be accompanied at a minimum by descriptions of the starting time at which cumulative count is initiated and the length of observation.

CRediT authorship contribution statement

Asami Anzai: Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Hiroshi Nishiura:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

A.A. received funding from the Kyoto University Medical Student Support-Fund and JSPS KAKENHI (22J14304). H.N. received funding from Health and Labour Sciences Research Grants (20CA2024, 20HA2007, 21HB1002, and 21HA2016), the Japan Agency for Medical Research and Development (JP20fk0108140 and JP20fk0108535), the JSPS KAKENHI (21H03198), the Environment Research and Technology Development Fund (JPMEERF20S11804) of the Environmental Restoration and Conservation Agency of Japan, the Japan Science and Technology Agency CREST program (JPMJCR1413), and the SICORP program (JPMJSC20U3 and JPMJSC2105). We thank the local governments, public health centers, and institutes for surveillance, laboratory testing, epidemiological investigations, and data collection. We also thank John Daniel from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Dushoff, J., Park, S.W., 2021. Speed and strength of an epidemic intervention. Proc. R. Soc. B 288. https://doi.org/10.1098/RSPB.2020.1556.

Galvani, A.P., Lei, X., Jewell, N.P., 2003. Severe acute respiratory syndrome: temporal stability and geographic variation in death rates and doubling times. Emerg. Infect. Dis. 9, 991. https://doi.org/10.3201/EID0908.030334.

Gotelli, J.N., 2008. A Primer of Ecology, 4th Edition. Sinauer Associates, Sunderland, Mass.

Khosrawipour, V., Lau, H., Khosrawipour, T., Kocbach, P., Ichii, H., Bania, J., Mikolajczyk, A., 2020. Failure in initial stage containment of global COVID-19 epicenters. J. Med. Virol. 92 (7), 863–867.

Liang, L.L., Kao, C.T., Ho, H.J., Wu, C.Y., 2021. COVID-19 case doubling time associated with non-pharmaceutical interventions and vaccination: A global experience. J. Glob. Health 11, 1–10. https://doi.org/10.7189/JOGH.11.05021.

Muniz-Rodriguez, K., Chowell, G., Cheung, C.-H., Jia, D., Lai, P.-Y., Lee, Y., Liu, M., Ofori, S.K., Roosa, K.M., Simonsen, L., Viboud, C., Fung, I.-H., 2020. Doubling time of the COVID-19 epidemic by province. China Emerg. Infect. Dis. 26 (8), 1912–1914.

Pellis, L., Scarabel, F., Stage, H.B., Overton, C.E., Chappell, L.H.K., Fearon, E., Bennett, E., Lythgoe, K.A., House, T.A., Hall, I., 2021. Challenges in control of COVID-19: short doubling time and long delay to effect of interventions. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 376 (1829) https://doi.org/10.1098/RSTB.2020.0264.

Remuzzi, A., Remuzzi, G., 2020. COVID-19 and Italy: what next? Lancet 395, 1225–1228. https://doi.org/10.1016/S0140-6736(20)30627-9.

Riccardo, F., et al., 2020. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. Eurosurveillance 25, 1–11. https://doi.org/10.2807/1560-7917.ES.2020.25.49.2000790/CITE/PLAINTEXT.

Ridenhour, B., Kowalik, J.M., Shay, D.K., 2014. Unraveling R0: considerations for public health applications. Am. J. Public Health 104, S445–S454. https://doi.org/10.2105/AJPH.2013.301704.

Thompson, R.N., et al., 2020. Key questions for modelling COVID-19 exit strategies. Proc. Biol. Sci. 287 (20201405), 2022. https://doi.org/10.1098/RSPB.2020.1405.

UK Health Security Agency, 2021. Omicron daily overview: 24 December 2021. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1043866/20211224_OS_Daily_Omicron_Overview.pdf (accessed 13 May 2022).

Vynnycky, E., White, R.G., 2010. An Introduction to Infectious Disease Modelling. Oxford University Press, Oxford.

World Health Organization, 2020. Novel Coronavirus (2019-nCoV): situation report. URL https://apps.who.int/iris/handle/10665/330988 (accessed 13 May 2022).

University of Cambridge, 2021. How to work out doubling time. Plus Mag. URL https://plus.maths.org/content/calculating-doubling-time (accessed 13 May 2022).