Evaluating a novel reproduction number estimation method: A comparative analysis

KATSURO ANAZAWA

Department of Natural Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan

Email: anazawa@k.u-tokyo.ac.jp, anazawa@g.ecc.u-tokyo.ac.jp

This paper presents practical methodologies and demonstrations for determining reproduction number, offering valuable insights to researchers and public health officials. Multiple approaches for simplified estimation techniques are proposed for the reproduction number of infectious diseases, and their effectiveness is compared. Methods assuming either exponential or delta distributions for the generation time of infectious diseases offer convenience by enabling the calculation of the reproduction number based solely on the mean of the generation time and number of new infection cases. However, the former tends to underestimate the reproduction number when the variance of the generation time distribution is small, while the latter tends to overestimate it when the variance is large. Conversely, the method assuming a normal distribution may underestimate the reproduction number depending on the growth rate. However, the estimation method assuming a gamma distribution provides reliable values in all scenarios. These estimation formulas should be applied judiciously, considering the characteristics of the generation time distribution of infectious diseases.

Keywords: Reproduction number; generation time; gamma distribution; Euler-Lotka equation; SARS-CoV-2.

1. Introduction

The pandemic of the novel coronavirus infection (COVID-19) has once again underscored the importance of infectious disease control (Stawicki *et al.*, 2020). The rapid spread of such pathogens has profound implications for individuals, society, and the global economy. Unprecedented concern about infectious diseases has led to the widespread adoption of mathematical pathology terminology, including population immunity, incubation period, reproduction number, and doubling rate (Clement *et al.*, 2021; Riccaboni & Verginer, 2022). Among these, reproduction numbers such as the basic reproduction number R_0 and effective reproduction number R_t are frequently used to elucidate the dynamics of infectious disease transmission and contraction. It is defined as the number of secondary infections caused by one infected person at a given point in time, or the mean number of new infections throughout the infection period, and if this value is less than 1, the infection is under control, while if it exceeds 1, it indicates the continuation of the epidemic of infection (Diekmann *et al.*, 1990; Roberts & Heesterbeek, 2003; Bacaër & Guernaoui, 2006; Inaba, 2012). The concept of reproduction numbers, notable for its simplicity and clarity, serves as a valuable indicator for assessing the risk of an epidemic or the interventions required to curb infection spread (Anderson & May, 1991; Roberts & Heesterbeek, 2003; Heffernan *et al.*, 2005). However, a singular calculation method does not exist for the exact determination of the reproduction number (R). Consequently, numerous estimation methods have been proposed, resulting in different values calculated by various institutions and researchers (Annunziato & Asikainen, 2020; Gostic *et al.*, 2020; Arvanitis *et al.*, 2021; Khailaie *et al.*, 2021; Bsat *et al.*, 2022). Moreover, the exact calculation of the reproduction number necessitates advanced knowledge of mathematics and programming, thus rendering it a “black box” for many researchers. In many cases, existing software tools such as those developed by Wallinga and Teunis (2004) or Cori *et al.* (2013) are employed for these calculations (Thompson *et al.*, 2019; Nash *et al.*, 2022; Bhatia *et al.*, 2023). Conversely, to meet societal demands, simplified methods have been proposed by reputable institutions such as the Robert Koch Institute (RKI) in Germany (Koch-Institut, 2020) or the Joint Research Centre (JRC) (An der Heiden & Hamouda, 2020), enabling manual computation of reproduction numbers from newly reported infection cases. Nevertheless, few studies compare the practicality of these simplified approaches (Annunziato & Asikainen, 2020), and they appear to lack a solid theoretical foundation. Thus, systematic organization of easily calculable methods for estimating reproduction numbers is believed to be beneficial for future societal strategies against infectious diseases. Therefore, this paper proposes multiple novel estimation equations for reproduction numbers, offering a comparative analysis of their attributes alongside existing methods and demonstrating their practical applications. The novel methodologies and equations for reproduction number estimation may address the aforementioned challenges.

2. Relationship between the generation time distribution *g(t)* and the reproduction number R

Let $f(a)$ represent the probability that an individual infected with a certain infectious disease infects others during a given time, a. Furthermore, let $I(t)$ denote the number of new infections in a population at time t. The number of new infections caused by new infections at time $(t-a)$ (the number of infected individuals is $I(t-a)$) after a time a has elapsed since infection (i.e., at time t) is $I(t-a)f(a)$. The sum of this quantity across all time intervals a subsequent to infection yields the total number of infections $I(t)$ at that time (Feller, 1941).

$$I(t) = \int_0^\infty I(t-a) f(a) da \qquad (2.1)$$

The number of secondary infections generated by a single infected individual (primary case) within a given time can be expressed as the basic reproduction number, *R*, using the following equation:

$$R = \int_0^\infty f(t) dt \qquad (2.2)$$

The probability density function, *g(t)*, representing the distribution of the time interval from primary to secondary infection can be expressed by normalizing *f(t)* as shown in the following equation:

$$g(t) = \frac{f(t)}{\int_0^\infty f(a) da} \qquad (2.3)$$

This *g(t)* is referred to as the generation time. From equations (2.1), (2.2), and (2.3), the following relationship is derived:

$$R = \frac{I(t)}{\int_0^\infty I(t-a) g(a) da} \qquad (2.4)$$

The number of infections exhibits exponential growth or decline (Kermack & McKendrick, 1927). Denoting the growth rate as λ, the number of new infections, *I(t)*, can be expressed as follows (Lipsitch *et al.*, 2003):

$$\frac{dI(t)}{dt} = \lambda I(t) \qquad (2.5)$$

Solving this equation yields the following expression:

$$I(t) = I(t-a) \exp(\lambda a) \qquad (2.6)$$

From equations (2.1) and (2.6), equation (2.7) is obtained:

$$I(t) = \int_0^\infty I(t) \exp(-\lambda a) f(a) da \qquad (2.7)$$

Dividing both sides of equation (2.7) by *I(t)* yields the Euler-Lotka equation (Sharpe & Lotka, 1911; Lotka, 1913):

$$1 = \int_0^\infty \exp(-\lambda a) f(a) da \qquad (2.8)$$

Dividing the Euler-Lotka equation (2.8) by the definition equation (2.2) of the reproduction number *R* and using equation (2.3) to transform *f(a)* into *g(a)*, *R* can be expressed as follows:

$$\frac{1}{R} = \int_0^\infty \exp(-\lambda a) g(a) da \qquad (2.9)$$

Therefore, given the generation time distribution *g(t)*, the relationship between the growth rate λ and the

reproduction number R can be determined from equation (2.9) (Wallinga & Lipsitch, 2007). In the context of infectious disease transmission, the generation time distributions commonly used encompass the exponential distribution (equivalent to the SIR model), normal distribution, delta distribution, Weibull distribution, gamma distribution, and log-normal distribution. Among these, the exponential distribution, normal distribution, and gamma distribution allow the description of the reproduction number R solely based on the expected value (mean) and variance of the generation time distribution (Chen *et al.*, 2022; Lippiello *et al.*, 2022). In this study, an examination is conducted on estimation formulas for the reproduction number derived from these distributions.

3. Relationship between Generation Time Distribution *g(t)* and Reproduction Number

3.1. *Exponential Distribution*

The simplest description of the Kermack–McKendrick theory, encompassing the SIR model, classifies the study population into susceptible, infected, and recovered individuals. The spread of the infection via contact between infected and susceptible individuals can be described by the following differential equations (Kermack & McKendrick, 1932; Anderson & May, 1991; Kermack & McKendrick, 1927):

$$\frac{dS(t)}{dt} = -\beta S(t) I(t) \qquad (3.1)$$

$$\frac{dI(t)}{dt} = \beta S(t) I(t) - \gamma I(t) \qquad (3.2)$$

$$\frac{d\,\mathrm{Re}(t)}{dt} = \gamma I(t) \qquad (3.3)$$

Here, $S(t)$ represents the size of the susceptible population at time t, $I(t)$ represents the size of the infected population, and $Re(t)$ represents the recovered or removed population. β is the transmission coefficient representing the infection rate through contacts between susceptible and infected individuals. γ is the rate constant at which infected individuals recover or are isolated.

The temporal change in the number of infected individuals $I(t)$ until recovery or isolation at time t is described by the following equation, which is derived from equation (3.2):

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

Here, assuming the number of infected individuals at time $t = 0$ is $I(0)$, $I(t)$ described by the equation (3.2), is expressed by the following equation (3.4), indicating that the number of infected individuals decreases exponentially:

$$\mathrm{I(t)} = \mathrm{I(0)}\exp(-\gamma t) \qquad (3.4)$$

At this juncture, I(t), the quantity of individuals infected by others, is considered to be proportional to the

probability of infection *f(t)*. Therefeore, by normalizing *I(t)* using equation (2.3), the probability density function *g(t)* for the generation time can be expressed as follows:

$$g(t) = \frac{I(t)}{\int_0^\infty I(a)da} \qquad (3.5)$$

Substituting equation (3.4) into equation (3.5), the following equation is obtained:

$$\mathrm{g(t)} = \gamma \exp(-\gamma\, t) \qquad (3.6)$$

By substituting equation (3.6) into equation (2.9), *R* can be expressed as follows:

$$\frac{1}{R} = \gamma \int_0^\infty \exp\{(-\lambda - \gamma)a\}da$$

Thus, the relationship in equation (3.7) is obtained:

$$R = \frac{\lambda}{\gamma} + 1 \qquad (3.7)$$

Furthermore, using equation (3.6), $1/\gamma$ is expressed as the mean value (expected value) of the generation time, *Tg*, as follows:

$$\mathrm{T}g = \int_0^\infty t g(t)dt$$

$$= \int_0^\infty t\, \gamma \exp(-\gamma\, \mathrm{t})dt$$

$$= \frac{1}{\gamma} \qquad (3.8)$$

From equations (3.7) and (3.8), the reproduction number *R* can be described by the following simple equation:

$$R = \lambda\, Tg + 1 \qquad (3.9)$$

In order for $R > 0$, the condition for this equation to hold is $\lambda Tg > -1$. In this case, the variance σ^2 is described by the following equation:

$$\sigma^2 = \int_0^\infty (t - Tg)^2\, g(t)dt$$

By substituting equations (3.6) and (3.8) into this equation, the following relationship is obtained:

$$\sigma^2 = Tg^2 \qquad (3.10)$$

Specifically, when *g(t)* is expressed by an exponential distribution, the standard deviation σ is equal to the mean value *Tg*.

3.2. *Normal and Delta Distributions*

If the generation time distribution *g(t)* follows a normal distribution with a mean value of *Tg* and a standard deviation of σ, *g(t)* can be described by the following equation:

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(t-Tg)^2}{2\sigma^2}\right) \qquad (3.11)$$

By substituting equation (3.11) into equation (2.9) and calculating, the reproduction number *R* is expressed by the following equation:

$$R = \exp(\lambda\ Tg - (1/2)\lambda^2\sigma^2) \qquad (3.12)$$

In the case where the variance $\acute{o}^2$ of the generation time distribution *g(t)* approaches *0*, i.e., assuming a delta distribution, equation (3.12) can be further simplified as follows:

$$R = \exp(\lambda\ Tg) \qquad (3.13)$$

This equation assumes a variance $\sigma^2 = 0$ for *g(t)*. Therefore, when the variance σ^2 of *g(t)* is large, it leads to an overestimation of the value of the reproduction number *R*.

3.3. *Gamma Distribution*

According to a review study on SARS-CoV-2, the generation time is often assumed to follow a gamma distribution (Hart *et al.*, 2022). Out of 90 references, 33 studies used the gamma distribution, 30 studies used the normal distribution, 8 studies used the Weibull distribution, and 11 studies used the log-normal distribution (Jusot, 2022).

When the probability density function of the generation time distribution *g(t)* follows a gamma distribution with a shape parameter *m* and a scale parameter *ç*, *g(t)* is expressed by the following equation:

$$g(t) = \frac{1}{\Gamma(m)\eta}\left(\frac{t}{\eta}\right)^{m-1}\exp\left(-\frac{t}{\eta}\right) \qquad (3.14)$$

Here, $m > 0$ and $\eta > 0$. The expected value of the probability distribution represented by equation (3.14), that is, the mean value of the population Tg and the variance σ^2, are $Tg = m\eta$ and $\sigma^2 = m\eta^2$. Therefore, the shape parameter m and the scale parameter η are expressed by the following equations:

$m = Tg^2 / \sigma^2$ (3.15)
$\eta = \sigma^2 / Tg$ (3.16)

Furthermore, by substituting equation (3.14) into equation (2.9), the following equation is obtained (Yamauchi, 2020):

$$R = (1 + \lambda \eta)^m \qquad (3.17)$$

By substituting equations (3.15) and (3.16) into equation (3.17), the estimation formula for the reproduction number R is obtained:

$$R = (1 + \frac{\sigma^2}{Tg} \lambda)^{\frac{Tg^2}{\sigma^2}} \qquad (3.18)$$

Since $R > 0$, the condition for this equation to hold is as follows:

$$\frac{\sigma^2}{Tg} \lambda > -1 \qquad (3.19)$$

These estimation formulas for the reproduction number R are summarized in Table 1.

3.4. *Comparison of Reproduction Number (R) Estimation Methods*

Fig. 1 shows the $g(t)$ distributions for different values of the standard deviation σ ranging from 0 to 5, with the expected value Tg of the generation time distribution $g(t)$ set to 5. Since generation time is often estimated by serial interval (Lehtinen *et al.*, 2021), those values were established based on the reported serial interval distributions of SARS-CoV-2, Tg = 4.8, σ = 2.3 (Nishiura *et al.*, 2020), and Tg = 5.1, σ = 5.3(Ali *et al.*, 2020). When $\sigma = 0$, all distributions except the exponential distribution with a constant σ value ($\sigma = Tg$) converge to the delta distribution. Moreover, when $\sigma = 2$, all distributions except the exponential distribution and the delta distribution exhibit similar shapes. When $\sigma = Tg\ (= 5)$, the gamma distribution and the Weibull distribution overlap with the exponential distribution, while the normal distribution forms an asymmetric low peaked mountain shape. When σ is extremely small, the computed value of R assuming the delta distribution approaches the true value, and when σ is large and close to the mean value Tg, the computed value assuming the exponential distribution is expected to approach the true value.

For $Tg \geq ó \geq 0$, the order of sizes is Delta distribution ≥ Gamma distribution ≥ Exponential distribution, which can be expressed as:

$$R : Delta[\exp(\lambda\ Tg)] \geq Gamma[(1+\frac{\sigma^2}{Tg}\lambda)^{\frac{Tg^2}{\sigma^2}}] \geq Exponential[1+\lambda\ Tg] \qquad (3.20)$$

When $lim(\sigma \rightarrow 0)$, Delta and Gamma formula become equal. Furthermore, when $Tg = \sigma$, the Gamma and Exponential formula become equal.

Delta distribution $exp(\lambda Tg)$ exhibits the largest value among the estimation methods based on these distributions. Additionally, as the variance σ^2 approaches 0, both the normal distribution and the gamma distribution approach the delta distribution.

Furthermore, when $\lambda = 0$, all the above estimation formulas yield $R = 1$. Therefore, as R approaches 1, the estimated values of R from these methods will approach each other.

4. Estimation Formula for Growth Rate $\lambda(t)$

Subsequently, the calculation method for determining the growth rate λ that changes over time is examined. The growth rate of the number of new infections $I(t)$ at a certain time t is denoted as $\lambda(t)$. By designating $\lambda(t)$ at the median of the observation interval s and rearranging equation (2.5), $\lambda(t)$ can be determined through the following equation:

$$\lambda(t-\frac{s-1}{2}) = \frac{1}{s}\log_e\left(\frac{I(t)}{I(t-s)}\right) \qquad (4.1)$$

The value of $\lambda(t)$ is defined at the median of the observation interval, and since a single day is quantified as the minimum time unit, the time given to λ is delayed by $(s-1)/2$ days. $\lambda(t)$ can be obtained from the actual measured values of the number of infected people using this equation. Furthermore, the number of new infections measured varies by the day of the week, especially between weekdays and holidays, with observation bias (Fig. 2). To mitigate such observation bias, some form of smoothing is necessary (Eales *et al.*, 2022). One method of smoothing involves the use of a 7-day moving average to stabilize the variations associated with different days of the week, yielding the following equation (Bonifazi *et al.*, 2021):

$$\lambda(t-\frac{s-1}{2}-3) = \frac{1}{s}\log_e\left(\frac{\sum_{i=0}^{6} I(t\ -i)}{\sum_{i=0}^{6} I(t-s-i)}\right) \qquad (4.2)$$

When taking a 7-day moving average, there is a 3-day delay in the time of the λ value given by calculation.

5. Simple formula for estimating the reproduction number $R(t)$ from the number of new infections $I(t)$

Subsequently, a formula for estimating the reproduction number from the number of new infections is examined.

The time function of the reproduction number at a certain time is denoted as *R(t)*, and the estimation formula for the reproduction number is derived for each of the exponential distribution, normal distribution, delta distribution, and gamma distribution when assuming the generation time distribution *g(t)* in the previous $\lambda(t)$ formula.

5.1. *Estimation formula derived from exponential distribution*

If *g(t)* is assumed to be an exponential distribution (equivalent to SIR model) with a mean value of *Tg*, the reproduction number is expressed as $R(t) = 1 + \lambda(t)Tg$ according to equation (3.9).

When equation (4.1) is substituted into equation (3.9), equation (5.1) is obtained:

$$R(t-\frac{s-1}{2}) = \frac{Tg}{s}\log_e\left(\frac{I(t)}{I(t-s)}\right)+1 \qquad (5.1)$$

Here, *I* is the number of new infections, *t* is any observation day, and *s* is the observation interval (days). This formula represents a general form of the formula proposed by the Joint Research Centre (JRC) (Annunziato & Asikainen, 2020). JRC sets *Tg* and *s* to 7 days. The reason why they set *s* to 7 days is presumably to eliminate fluctuations due to the day of the week. Employing a 7-day moving average for *I(t)* to suppress variations originating from weekdays yields the following equation:

$$R(t-\frac{s-1}{2}-3) = \frac{Tg}{s}\log_e\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)+1 \qquad (5.2)$$

In this case, given the utilization of a 7-day average, an approximate 3-day lag in the obtained value can be expected. The method of calculating *R(t)* using this formula will be referred to as Exponential method.

5.2. *Estimation formula derived from normal distribution*

If *g(t)*, the generation time distribution, is assumed to be a normal distribution with a mean value of *Tg* and a variance of $\acute{o}^2$, $R = exp(\lambda Tg - (1/2)\lambda^2\acute{o}^2)$ according to equation (3.12). When equation (4.1) is substituted into equation (3.12), the following equation is obtained:

$$R(t-\frac{s-1}{2}) = \left(\frac{I(t)}{I(t-s)}\right)^{\left[\frac{Tg}{s}-\frac{\sigma^2}{2s^2}\log_e\left(\frac{I(t)}{I(t-s)}\right)\right]} \qquad (5.3)$$

To eliminate fluctuations originating from the day of the week and to smooth out data, the adoption of a 7-day moving average results in the following equation:

$$R(t-\frac{s-1}{2}-3) = \left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)^{\left[\frac{Tg}{s}-\frac{\sigma^2}{2s^2}\log_e\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)\right]}$$

(5.4)

The method of calculating *R(t)* using this formula will be abbreviated as Norm method.

5.3. *Estimation formula derived from delta distribution*

If *g(t)* is assumed to be a normal distribution with a mean value of *Tg* and a variance of *0*, which essentially corresponds to a delta distribution, then *R(t) = exp(λTg)* according to equation (3.13). When equation (4.1) is substituted into equation (3.13), the following relationship is obtained:

$$R(t-\frac{s-1}{2})=\left(\frac{I(t)}{I(t-s)}\right)^{\frac{Tg}{s}} \qquad (5.5)$$

Employing a 7-day moving average in equation (5.5) yields the following equation.

$$R(t-\frac{s-1}{2}-3)=\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i))}\right)^{\frac{Tg}{s}} \qquad (5.6)$$

Many simple estimation formulas for *R(t)* that have been published previously can be explained by this formula. The method proposed by the Robert Koch Institut (RKI) in Germany fixes the generation time at 4 days (An der Heiden & Hamouda, 2020) and employs a 7-day summation to eliminate fluctuations due to the day of the week and enhance data smoothness (Bonifazi *et al.*, 2021; Koch-Institut, 2020). This is equivalent to substituting 4 days for *Tg* in equation (5.6). Similarly, the National Institute of Infectious Diseases (NIID) proposed the formula *R(t)* = (recent 7-day new positive reports) / (7-day new positive reports 5 days before), assuming a mean generation time (*Tg*) of 5 days and a reporting interval of *Tg* = 5 days (Ko *et al.*, 2021). This aligns with substituting *Tg* = 5 days for *s* in equation (5.6). Toyo Keizai Shinpo's Corona Dashboard provides daily *R(t)* estimates on its HP using *R(t)* = (recent 7-day new positive cases / previous 7-day new positive cases)^ (mean generation time / reporting interval) (Toyo_Keizai_Online, 2023). This formula is essentially substituting 7 for *s* in equation (5.6). The method of calculating *R(t)* using this original equation (5.6) will be referred to as the Delta method.

5.4. *Estimation formula derived from gamma distribution*

If *g(t)* is assumed to be a gamma distribution with a mean value of *Tg* and a variance of $ó^2$, equation (5.7) is obtained by substituting equation (4.1) into equation (3.18):

$$R(\mathrm{t}-\frac{s-1}{2})=\left\{\frac{\sigma^2}{s\times Tg}\log_e\left(\frac{I(t)}{I(t-s)}\right)+1\right\}^{\frac{Tg^2}{\sigma^2}}$$

(5.7)

Here, I represents the number of new infections and s is the observation interval (days). When using a 7-day moving average to eliminate fluctuations due to the day of the week and smooth out data for this formula, the following equation is derived:

$$R(t-\frac{s-1}{2}-3)=\left\{\frac{\sigma^2}{s\times Tg}\log_e\left(\frac{\sum_{i=0}^{6}I(t-i)}{\sum_{i=0}^{6}I(t-s-i)}\right)+1\right\}^{\frac{Tg^2}{\sigma^2}} \tag{5.8}$$

The method of calculating *R(t)* using this formula will be referred to as the Gamma method. These *R(t)* estimation formulas are summarized in Table 1.

5.5. *Formula without assuming distribution*

Methods have been reported to determine *R(t)* using specific observation days for generation time without assuming generation time distribution *g(t)* (Wallinga & Teunis, 2004; Cori *et al.*, 2013). Among these, the method outlined by Cori et al. (2013) can be expressed by the following formula, which discretely solves equation (2.4) and converts the denominator into a sum:

$$R(t-\frac{Tg-1}{2})=\frac{I(t)}{\sum_{j=0}w(j)\times I(t-j)} \tag{5.9}$$

Here j is a discrete time (e.g., days), and w(j) is a probability distribution, equivalent to g(t) expressed in (2.3) in continuous functions. The delay component *Tg/2* is derived from the denominator and can be deduced from the following relationship:

$$Tg=\sum_{j=0}j\times w(j) \tag{5.10}$$

This formula requires specific observation data to obtain generation time distribution; however, it is expected to yield more accurate *R(t)* values compared to cases where only the mean value of generation time is available. When smoothed with a 7-day moving average, the following formula is obtained:

$$R(t-\frac{\mathrm{Tg}-1}{2}-3)=\frac{\sum_{i=0}^{6}I(t-i)}{\sum_{j=0}\left\{w(j)\sum_{i=0}^{6}I(t-i-j)\right\}} \tag{5.11}$$

The method of calculating *R(t)* using this formula will be referred to as the Cori et al. method.

6. Discussion

6.1. *Synthetic data for new infections I(t)*

Synthetic data for new infections were used to compare and examine the $\lambda(t)$ and *R(t)* estimation methods. The synthetic data were generated assuming the conventional strain of SARS-CoV-2, using the probability density function of generation time *g(t)* and the reproduction number *R(t)*. While the generation time distribution of infectious diseases is generally approximated using the Weibull distribution, log-normal distribution, or gamma distribution, many studies approximating SARS-CoV-2 have employed the gamma distribution (Jusot, 2022; Knight & Mishra, 2020). Therefore, in this study, synthetic data were created using the gamma distribution as the generation time distribution *g(t)*.

For baseline information on the generation time distribution *g(t)* of the conventional SARS-CoV-2 strain, Nishiura *et al.* (2020) reported a mean value of 4.8 days (95% credibility interval (CrI): 3.8, 6.1) and a standard deviation of 2.3 days (95% CrI: 1.6, 3.5), based on a survey of 28 pairs of infected-infected individuals. Ali *et al.* (2020) reported Tg = 5.1 days [95% CrI: 4.7, 5.5] and σ = 5.3 days [95% CrI: 5.0, 5.6] based on a survey of 677 pairs. Although the mean value Tg in both cases is approximately 5 days and similar, the standard deviation exhibits a difference of more than twice, with Ali's σ almost equal to the mean value Tg. Based on these findings, synthetic data were generated by setting the mean value (Tg) of generation time *g(t)* to 5 days and changing the standard deviation (σ) from *0* to 5.

Synthetic data based on the gamma distribution were generated using the inverse function of equation (5.7), which is shown below:

$$\mathrm{I(t)} = \mathrm{I(0)} \exp\left(\frac{\mathrm{s} \times Tg \times \{R(t - \frac{s-1}{2})^{\frac{\sigma^2}{Tg^2}} - 1\}}{\sigma^2} \right) \qquad (6.1)$$

Here, Tg represents the mean value of generation time distribution *g(t)*, and σ^2 is the variance. Moreover, for the observation interval *s*, the shortest time of 1 day was used, and the initial number of infected individuals *I(0)* was set to one. Additionally, *R(t)* for the initial 30 days was set to *R(t)* = 3 based on reports on the basic reproduction number of the new coronavirus (Iyaniwura *et al.*, 2022; Alimohamadi *et al.*, 2020; D'Arienzo & Coniglio, 2020). Assuming a public health intervention for the subsequent 30 days, *R(t)* was set to one and for the subsequent 30 days it was set to 0.5 to generate the total number of new infections *I(t)* for a total of 90 days. Notably, when $\sigma = 0$, equation (5.1) cannot be applied. As σ approaches zero, the gamma distribution overlaps with the delta distribution, so when $\sigma = 0$, *I(t)* was generated using the following formula based on the delta distribution:

$$\mathrm{I(t)} = \mathrm{I(0)} \times R(t - \frac{s-1}{2})^{\frac{s}{Tg}} \qquad (6.2)$$

Moreover, when $\sigma = 5$, the gamma distribution becomes equal to the exponential distribution since $\sigma = Tg$ (mean value). The formula for generating synthetic data based on the exponential distribution is as follows:

$$\mathrm{I(t)} = \mathrm{I(0)} \times \exp\left(\frac{s}{Tg} (R(t - \frac{s-1}{2}) - 1) \right) \qquad (6.3)$$

6.2. *Examination of day-of-week smoothing and observation interval*

The relationship between $\lambda(t)$ and observation interval s, which are elements that make up the $R(t)$ estimation formula, was examined in terms of time function.

6.2.1 *Examination of time delay using synthetic data*. $\Lambda(t)$ encompasses the observation interval s and forms a key part of the R(t) estimation formula. Therefore the impact of varying the observation interval s on λ(t) was explored helping evaluate bias originating from the day-of-week effect. The simplest method to reduce the day-of-week effect is to use the corresponding day of the week from one week ago. Consequently, calculations were conducted with $s = 7$ in equation (4.1). Additionally, λ(t) was calculated using equation (4.2) with actual new infection data using 7-day moving average and setting observation interval s from 1 day to 7 days at intervals of 1 day. The reference value for λ was calculated using equation (4.1) with $s = 1$.

The $\lambda(t)$ was calculated with and without considering the time delay within the parentheses in equations (4.1) and (4.2). When model data was generated under the conditions of $Tg = 5$, $\sigma = 2$, and varying $R(t)$ values from 3 to 1.5 and then to 0.5, $\lambda(t)$ changed from 0.24 to 0.08 and then to -0.13 (Fig. 3).

As shown in Fig. 3 (a), the $\lambda(t)$ value lags as the observation interval s increases in cases without considering the time delay. By contrast, when delay is considered, almost all $\lambda(t)$ values are consistent and intersect at the centre of the transition with the reference λ line as shown in Fig. 3(b). These results suggest that considering delay is a reasonable approach.

When the $R(t)$ value was changed from 3 to 1.5, $\lambda(t)$ changed from 0.26 to 0.09, taking 7 days in the case without using a 7-day moving average with $s = 7$ and when using a 7-day moving average with $s = 1$. Moreover, it took 12 days to take a 7-day moving average with $s = 7$. Additionally, the time taken to reach half the value, i.e., 0.17, was 4 and 6 days, respectively. This demonstrates that the time delay indicated by [$-(s-1)/2$] or [$-(s-1)/2-3$] in equations (4.1) or (4.2) aligns with the number of days required to reach the halfway point. Similar results were obtained for other synthetic data.

6.2.2 *Examination of smoothing of λ(t) using actual data*. Changes in $\lambda(t)$ due to differences in observation interval s were examined using actual data. The actual data used were the daily reported number of new infections $I(t)$ in Japan provided by the Ministry of Health, Labour and Welfare (Ministry of Health. Labour and Welfare, 2023). When examining the transition of new infections $I(t)$, the ratio of new infections $I(t)$ on observation day t to new infections $I(t-7)$ 7 days before t, both without and with a 7-day moving average with $s = 1$ for $\lambda(t)$, revealed strong

periodic fluctuations each week (Fig. 4). While it is possible to calculate $\lambda(t)$ manually by substituting $I(t)$ data every 7 days into equation (4.1), it should be noted that considerable day-of-week fluctuations are introduced. Almost all day-of-week fluctuations can be eliminated when s is approximately ≥ 5 (Fig. 4). However, as s increases to achieve greater smoothing, so does the progression of the delay $(s-1)/2$. In this study, a 7-day moving average was applied to the number of infected individuals $I(t)$, and calculations were performed with an observation interval $s = 7$ days.

6.3. *Comparison of R(t) estimation methods*

6.3.1 *Comparison of R(t) estimation methods using synthetic data*. Synthetic data were used to compare the estimated value of $R(t)$ obtained through the Cori et al., Exponential, Delta, Norm, and Gamma methods. A 7-day moving average of the new infected number $I(t)$ was used for the calculation. For all methods except the Cori et al. method, the observation interval s was set to 7 days. For the Cori et al. method, the coefficients $w(i)$ were calculated for i = 1-20 using the synthetic data. For all $R(t)$ estimation methods, the time delay was approximately 12 days, as discussed in the section on λ(t).

When the reproduction number $R(t)$ was set to 3 and synthetic data with a generation time distribution $g(t)$ of $Tg = 5$ days and $\sigma = 0$ days (i.e., delta distribution) were used to generate the number of new infections $I(t)$, all estimation methods for $R(t)$ except the Exponential method, accurately yielded a value of 3.0 as shown in Fig. 5(a). Conversely, the Exponential method produced a lower value of 2.1. When $R(t) = 1.5$ or 0.5, the Exponential method yielded relatively low values of 1.41 and 0.31, respectively, whereas the other methods produced accurate values.

For the estimated $R(t)$ values obtained from synthetic data with $Tg = 5$ days $\sigma = 2$ days, when $R(t) = 3$, the Cori et al., Norm, and Gamma methods yielded an accurate value of 3.0, while Delta method slightly overestimated with a value of 3.3, and the Exponential method considerably underestimated with a value of 2.2 as shown in Fig. 5(b). When $R(t) = 1.5$, the Cori et al. method, Norm method, and Gamma method accurately estimated a value of 1.5. The Delta method also yielded an almost accurate value of 1.52, while the Exponential method slightly underestimated with a value of 1.42. When $R(t) = 0.5$, the Cori et al, Norm, and Gamma methods provided accurate values of 0.50 each, and the Delta method yielded an almost accurate value of 0.52. By contrast, the Exponential method significantly underestimated with a value of 0.34.

Conversely, for the synthetic data with $Tg = 5$ and $\sigma = 5$, which had a standard deviation equal to the mean value (i.e., data following an exponential distribution), when $R(t) = 3$, the Delta method yielded a markedly higher value of 7.4 (Fig. 5(c)). Similarly, the Cori et al. method produced a relatively high value of 4.1. By contrast, the Norm method yielded an erroneous value of 1.0. In comparison, the Exponential and Gamma methods produced accurate values of 3.0. When $R(t) = 1.5$, the Cori et al. and Delta methods generated high values of 1.76 and 1.65, respectively. The Gamma and Exponential methods remained accurate with values of 1.5, and the Norm method yielded an almost accurate value of 1.45. When $R(t) = 0.5$, the Cori et al. and Delta methods generated higher values of 0.62 and 0.61, respectively, while the Gamma and Exponential methods yielded accurate values of 0.5, and the Norm method remained nearly accurate with a value of 0.54. In the case of $\sigma = Tg$, representing an exponential distribution, $\sum(1\text{-}20)w(i) = 0.90$ in the Cori et al. method, indicating a discrepancy of 0.04, because about 10% of the data is not reflected in the calculation. These results encapsulate the characteristics of each probability density function. Notably, when the standard deviation σ of the generation time distribution takes on large values close to the mean value, the reproduction number $R(t)$ approximates the value from the Exponential method. Conversely, when the

standard deviation σ is small and close to zero, it converges towards the value of the Delta method.
The Gamma method, which incorporates the properties of these distributions, consistently produces appropriate *R(t)* values in both cases. The Norm method exhibited an extremely low value of 1.0 when $R(t) = 3$. With a large σ relative to *Tg*, the Norm method introduced a negative term and a decrease that is absent in other methods. As shown in Fig. 1(c), when σ becomes large in a normal distribution, it is assumed to have a negative value in the time domain. In this study, *g(t)* is discussed as a generation time distribution constrained within the range of $t > 0$. However, if *g(t)* is defined based on a serial interval distribution that considers $t < 0$ instead of a generation time distribution, correlation with the Cori et al. method could potentially be explained by a normal distribution.

6.3.2 Comparison of R(t) estimation methods using actual data. A comparison was conducted among each *R(t)* estimation method using actual data on new infections in Japan. After the SARS-CoV-2 outbreak, various strains have emerged, and the generation time distribution *g(t)* differs for each strain. Therefore, data concerning the number of new infections before the mutations were used for comparing each estimation method. The period considered was from 1 March 2020 to 28 February 2021 (105th Novel Coronavirus Infection Control Advisory Board, 2022). This period encompasses the so-called first wave to the third wave, showcasing the peak of infections in Japan stemming from the conventional strain of SARS-CoV-2. The actual data on new infections were derived from openly accessible information on daily COVID-19 infections reported in Japan, published by the Ministry of Health, Labour and Welfare of Japan (Ministry of Health. Labour and Welfare, 2023). Mean generation time (*Tg*) and standard deviation (σ) values were set at $Tg = 4.8$, $\sigma = 2.3$ (Nishiura *et al.*, 2020), and $Tg = 5.1$, $\sigma = 5.3$ (Ali *et al.*, 2020). The generation time distribution data for *w(i)* in the Cori et al. method were also obtained from these literature sources. When *R(t)* was calculated with a mean generation time of $Tg = 4.8$ and a standard deviation of $\sigma = 2.3$, the Exponential method yielded smaller values of *R(t)* than the other methods in the initial stage of the first wave from 1 March 2020 to 31 May 2020 (Fig. 6(a)). Similarly, when *R(t)* estimation was performed with $Tg = 5.1$ and $\sigma = 5.3$ during the same period, the Delta method yielded significantly larger *R(t)* values than the other methods (Fig. 6(b)).
The differences in *R(t)* obtained by each method were quantitatively analysed through correlation coefficients (Table 2). All methods exhibited high correlations of more than 0.94 with each other. In calculations with $Tg = 4.8$ and $\sigma = 2.3$, the Exponential and Delta method showed slightly lower correlations of 0.98 with the Cori et al. method compared with the other methods. Similarly, in calculations with $Tg = 5.1$ and $\sigma = 5.3$, the Delta method exhibited a slightly lower correlation of 0.94 with the Cori et al. method compared with the other methods.
A paired one-way analysis of variance (ANOVA) was applied to the mean *R(t)* values with $Tg = 4.8$ and $\sigma = 2.3$. Statistically significant difference were observed in the mean *R(t)* values ($P < 0.01$), prompting multiple comparison tests using the Bonferroni, Sidak, and Holm methods. These tests consistently yielded the same results. The mean *R(t)* values obtained by each estimation method, along with the standard deviation and standard error, are presented in Table 3. Significantly different results were observed between the Cori et al. method (set as level 1) and the other estimation methods (set as level 2) using Bonferroni's method ($P < 0.01$), indicated by ** symbols. Specifically, the Exponential method yielded a significantly lower mean *R(t)* value (mean:1.057, $P < 0.01$), while the Delta method produced a significantly higher value (mean:1.097, $P < 0.01$).
Similarly, when employing paired one-way ANOVA with $Tg = 5.1$ and $\sigma = 5.3$, significant differences were also observed in the mean *R(t)* values ($P < 0.01$). Multiple comparison tests using Bonferroni's, Sidak's, and Holm's

methods consistently indicated significantly higher values for the Delta method (mean:1.147, $P < 0.01$) compared with the other methods. Conversely, the Norm and Gamma methods did not show significant differences from the Cori et al. method.

When organizing these methods for estimating *R(t)* by assuming some distribution for the generation time distribution *g(t)*, the Gamma method showed the most versatile and reliable value. The Norm method yielded negative values with high variance in generation time *g(t)*. The Delta method tended to overestimate *R(t)* when the standard deviation (σ) of the generation time distribution was large, while the Exponential method underestimated *R(t)* when standard deviation σ was small. These *R(t)* estimation methods have the advantage of enabling relatively easy calculation of the effective reproduction number *R(t*), even when daily information on new infection counts is not available and the data are limited to intervals of one week or more.

7. Conclusion

This study introduces novel methodologies and presents clear and comprehensible equations for estimating the reproduction number (*R(t)*) of infectious diseases. The formulas for *R(t)* estimation were compared for each assumed generation time distribution, *g(t)*. When a reliable histogram of the generation time distribution is available, and continuous data on daily new infections are obtainable, the Cori et al. formula, which does not presuppose a generation time distribution, proves useful for determining *R(t)*. However, situations may arise where these conditions are not met, necessitating the utilization of estimation methods that rely on certain assumed generation time distributions as examined in this study. The assumptions made regarding the generation time distribution *g(t)* —whether as delta distribution or an exponential distribution— offer the advantage of allowing the straightforward estimation of *R(t)* using only the mean generation time *Tg* and the number of new infections *I(t)*. However, *R(t)* calculated using the delta distribution-based estimation formula tends to result in higher values, whereas those computed using the exponential distribution-based formula tend to yield lower values. When the standard deviation is small compared to the mean generation time, the Delta method yields more accurate values, but when the standard deviation is large and close to the mean value, the Exponential method yields more accurate values. By contrast, when using the *R(t)* estimation formula based on the gamma distribution, a reliable *R(t)* can be obtained in either case. The normal distribution-based estimation formula may produce negative values when the standard deviation of the generation time is large, but it may be effective when using a serial interval including negative values instead of the generation time. Regardless of the method employed, the pivotal factor is the generation time distribution *g(t)*. If an accurate generation time distribution for a pathogen or variant is not available, it is imperative to recognize that the *R(t)* value estimated, irrespective of the selected method, could exhibit considerable error.

Conflict of interest

The author has no conflict of interests to declare.

Acknowledgements

English language editing was provided by Editage (www.editage.com).

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Supplementary material

Supplementary material related to this article is available on https://doi.org/10.5281/zenodo.10077233.

Figure Captions

Figure 1. Generation time distributions $g(t)$ with a mean (Tg) of 5 days and standard deviation (σ) values of 0, 2, and 5 days.

Figure 2. Trend in the number of newly confirmed cases of COVID-19 during the initial wave of infections in Japan (Ministry of Health. Labour and Welfare, 2023).

Figure 3. Estimation of the parameter $\lambda(t)$ based on synthetic data, assuming gamma distribution with parameters $Tg = 5$ and $\sigma = 2$.

Figure 4. Relationship between the observation interval (s) and the growth rate $\lambda(t)$.

Figure 5. Estimation of the reproduction number $R(t)$ using various methods, based on synthetic data.

Figure 6. Estimation of the reproduction number $R(t)$ based on real data of newly confirmed cases of COVID-19 in Japan.

REFERENCES

105TH NOVEL CORONAVIRUS INFECTION CONTROL ADVISORY BOARD, M. O. H., LABOUR AND WELFARE (2022) Document 3-8-②, Comparison of 3rd, 5th, 6th, and 7th waves (summary). *Shingata coronavirus kansensho taisaku advisory board.* vol. **2023**. Tokyo, Japan, Ministry of Health, Labour and Welfare.

ALI, S. T., WANG, L., LAU, E. H., XU, X.-K., DU, Z., WU, Y., LEUNG, G. M. & COWLING, B. J. (2020) Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science*, **369**(6507), 1106-1109.

ALIMOHAMADI, Y., TAGHDIR, M. & SEPANDI, M. (2020) Estimate of the basic reproduction number for COVID-19: a systematic review and meta-analysis. *J. Prev. Med. Public. Health*, **53**(3), 151.

AN DER HEIDEN, M. & HAMOUDA, O. (2020) Schätzung der aktuellen Entwicklung der SARS-CoV-2- Epidemie in Deutschland – Nowcasting. *Epidemiologisches Bulletin*, **2020**(17), 10-15. https://doi.org/http://dx.doi.org/10.25646/6692.2 (in German).

ANDERSON, R. M. & MAY, R. M. (1991) *Infectious diseases of humans: dynamics and control*. Oxford university press, pp. 757.

ANNUNZIATO, A. & ASIKAINEN, T. (2020) *Effective reproduction number estimation from data series*, vol. JRC121343. Ispra (VA), Italy: Joint Research Centre, European Commission.

ARVANITIS, A., FURXHI, I., TASIOULIS, T. & KARATZAS, K. (2021) Prediction of the effective reproduction number of COVID-19 in Greece. A machine learning approach using Google mobility data. *J. Decis. Anal. Intell. Comput.*, **1**(1), 1-21.

BACA R, N. & GUERNAOUI, S. (2006) The epidemic threshold of vector-borne diseases with seasonality: the case of cutaneous leishmaniasis in Chichaoua, Morocco. *J. Math. Biol.*, **53**(3), 421-436.

BHATIA, S., WARDLE, J., NASH, R. K., NOUVELLET, P. & CORI, A. (2023) Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study. *Epidemics*, **44**, 100692.

BONIFAZI, G., LISTA, L., MENASCE, D., MEZZETTO, M., PEDRINI, D., SPIGHI, R. & ZOCCOLI, A. (2021) A simplified estimate of the effective reproduction number Rt using its relation with the doubling time and application to Italian COVID-19 data. *Eur. Phys. J. Plus*, **136**(4), 1-14.

BSAT, R., CHEMAITELLY, H., COYLE, P., TANG, P., HASAN, M. R., AL KANAANI, Z., AL KUWARI, E., BUTT, A. A., JEREMIJENKO, A. & KALEECKAL, A. H. (2022) Characterizing the effective reproduction number during the COVID-19 pandemic: Insights from Qatar's experience. *J. Glob. Health*, **12**.

CHEN, D., LAU, Y.-C., XU, X.-K., WANG, L., DU, Z., TSANG, T. K., WU, P., LAU, E. H., WALLINGA, J. & COWLING, B. J. (2022) Inferring time-varying generation time, serial interval, and incubation period distributions for COVID-19. *Nat. Commun.*, **13**(1), 7727.

CLEMENT, J. C., PONNUSAMY, V., SRIHARIPRIYA, K. & NANDAKUMAR, R. (2021) A survey on mathematical, machine learning and deep learning models for COVID-19 transmission and diagnosis. *IEEE Rev. Biomed. Eng.*, **15**, 325-340.

CORI, A., FERGUSON, N. M., FRASER, C. & CAUCHEMEZ, S. (2013) A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am. J. Epidemiol.*, **178**(9), 1505-1512.

https://doi.org/10.1093/aje/kwt133.

D'ARIENZO, M. & CONIGLIO, A. (2020) Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy. *Biosaf. Health.*, **2**(2), 57-59.

DIEKMANN, O., HEESTERBEEK, J. A. P. & METZ, J. A. (1990) On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, **28**(4), 365-382.

EALES, O., AINSLIE, K. E., WALTERS, C. E., WANG, H., ATCHISON, C., ASHBY, D., DONNELLY, C. A., COOKE, G., BARCLAY, W. & WARD, H. (2022) Appropriately smoothing prevalence data to inform estimates of growth rate and reproduction number. *Epidemics*, **40**, 100604.

FELLER, W. (1941) On the integral equation of renewal theory. *Ann. Math. Statist.*, **12**(3), 243-267.

GOSTIC, K. M., MCGOUGH, L., BASKERVILLE, E. B., ABBOTT, S., JOSHI, K., TEDIJANTO, C., KAHN, R., NIEHUS, R., HAY, J. A. & DE SALAZAR, P. M. (2020) Practical considerations for measuring the effective reproductive number, R t. *PLoS Comput. Biol.*, **16**(12), e1008409.

HART, W. S., ABBOTT, S., ENDO, A., HELLEWELL, J., MILLER, E., ANDREWS, N., MAINI, P. K., FUNK, S. & THOMPSON, R. N. (2022) Inference of the SARS-CoV-2 generation time using UK household data. *eLife*, **11**, e70767. https://doi.org/10.7554/eLife.70767.

HEFFERNAN, J. M., SMITH, R. J. & WAHL, L. M. (2005) Perspectives on the basic reproductive ratio. *J. R. Soc. Interface.*, **2**(4), 281-293.

INABA, H. (2012) On a new perspective of the basic reproduction number in heterogeneous environments. *J. Math. Biol.*, **65**, 309-348.

IYANIWURA, S. A., RABIU, M., DAVID, J. F. & KONG, J. D. (2022) The basic reproduction number of COVID-19 across Africa. *PLoS One*, **17**(2), e0264455. https://doi.org/10.1371/journal.pone.0264455.

JUSOT, J.-F. (2022) An update of serial interval estimates for COVID-19: a meta-analysis. *4open*, **5**, 16.

KERMACK, W. O. & MCKENDRICK, A. G. (1927) Contributions to the mathematical theory of epidemics I. *Proceedings of the Royal Society*, **115A**, 700-721.

KERMACK, W. O. & MCKENDRICK, A. G. (1932) Contributions to the mathematical theory of epidemics. II. The problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, **138**(834), 55-83.

KHAILAIE, S., MITRA, T., BANDYOPADHYAY, A., SCHIPS, M., MASCHERONI, P., VANELLA, P., LANGE, B., BINDER, S. C. & MEYER-HERMANN, M. (2021) Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for political measures. *BMC Med.*, **19**(1), 1-16.

KNIGHT, J. & MISHRA, S. (2020) Estimating effective reproduction number using generation time versus serial interval, with application to COVID-19 in the Greater Toronto Area, Canada. *Infect. Dis. Model.*, **5**, 889-896.

KO, Y., ARIMA, Y., SUZUKI, M., SHIMADA, C., FURUSE, Y. & NAKAJIMA, K. (2021) COVID-19 感染報告者数に基づく簡易実効再生産数推定方法 [COVID-19 Simplified method for estimating the effective reproductions number based on the number of reported cases of infection]. *Infect. Agents Surveill. Rep. (IASR)*, **42**(6), 128-129. (in Japanese).

KOCH-INSTITUT, R. (2020) Erläuterung der Schätzung der zeitlich variierenden. Reproduktionszahl R. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/R-Wert-Erlaeuterung.pdf?__b

lob=publicationFile (Accessed 2023/8/21)(in German).

LEHTINEN, S., ASHCROFT, P. & BONHOEFFER, S. (2021) On the relationship between serial interval, infectiousness profile and generation time. *J. R. Soc. Interface.*, **18**(174), 20200756.

LIPPIELLO, E., PETRILLO, G. & DE ARCANGELIS, L. (2022) Estimating the generation interval from the incidence rate, the optimal quarantine duration and the efficiency of fast switching periodic protocols for COVID-19. *Sci. Rep.*, **12**(1), 4623.

LIPSITCH, M., COHEN, T., COOPER, B., ROBINS, J. M., MA, S., JAMES, L., GOPALAKRISHNA, G., CHEW, S. K., TAN, C. C. & SAMORE, M. H. (2003) Transmission dynamics and control of severe acute respiratory syndrome. *Science*, **300**(5627), 1966-1970.

LOTKA, A. J. (1913) A natural population norm. I. *Journal of the Washington Academy of Sciences*, **3**(9), 241-248.

MINISTRY OF HEALTH. LABOUR AND WELFARE (2023) Visualizing the data: information on COVID-19 infections. vol. **2023**.

NASH, R. K., NOUVELLET, P. & CORI, A. (2022) Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. *PLOS Digital Health*, **1**(6), e0000052.

NISHIURA, H., LINTON, N. M. & AKHMETZHANOV, A. R. (2020) Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.*, **93**, 284-286.

RICCABONI, M. & VERGINER, L. (2022) The impact of the COVID-19 pandemic on scientific research in the life sciences. *PLoS One*, **17**(2), e0263001.

ROBERTS, M. & HEESTERBEEK, J. (2003) A new method for estimating the effort required to control an infectious disease. *Proc. R. Soc. Lond. B. Biol. Sci.*, **270**(1522), 1359-1364.

SHARPE, F. R. & LOTKA, A. J. (1911) L. A problem in age-distribution. *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **21**(124), 435-438.

STAWICKI, S. P., JEANMONOD, R., MILLER, A. C., PALADINO, L., GAIESKI, D. F., YAFFEE, A. Q., DE WULF, A., GROVER, J., PAPADIMOS, T. J. & BLOEM, C. (2020) The 2019–2020 novel coronavirus (severe acute respiratory syndrome coronavirus 2) pandemic: A joint american college of academic international medicine-world academic council of emergency medicine multidisciplinary COVID-19 working group consensus paper. *J. Glob. Infect. Dis.*, **12**(2), 47.

THOMPSON, R. N., STOCKWIN, J. E., VAN GAALEN, R. D., POLONSKY, J. A., KAMVAR, Z. N., DEMARSH, P. A., DAHLQWIST, E., LI, S., MIGUEL, E., JOMBART, T., LESSLER, J., CAUCHEMEZ, S. & CORI, A. (2019) Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, **29**, 100356. https://doi.org/https://doi.org/10.1016/j.epidem.2019.100356.

TOYO_KEIZAI_ONLINE (2023) Coronavirus Disease (COVID-19) Situation Report in Japan. vol. **2023**.

WALLINGA, J. & LIPSITCH, M. (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, **274**(1609), 599-604.

WALLINGA, J. & TEUNIS, P. (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.*, **160**(6), 509-516.

YAMAUCHI, S. (2020) 発症間隔分布に基づいた増幅率・再生産数間の関係について(改訂) [Relationship between Amplification Rate and Reproduction Number Based on Serial Interval Distribution (Revised)]. http://sho-yama.c.ooco.jp/covid19/program/r2r.pdf (Accessed 2023/8/21)(in Japanese).

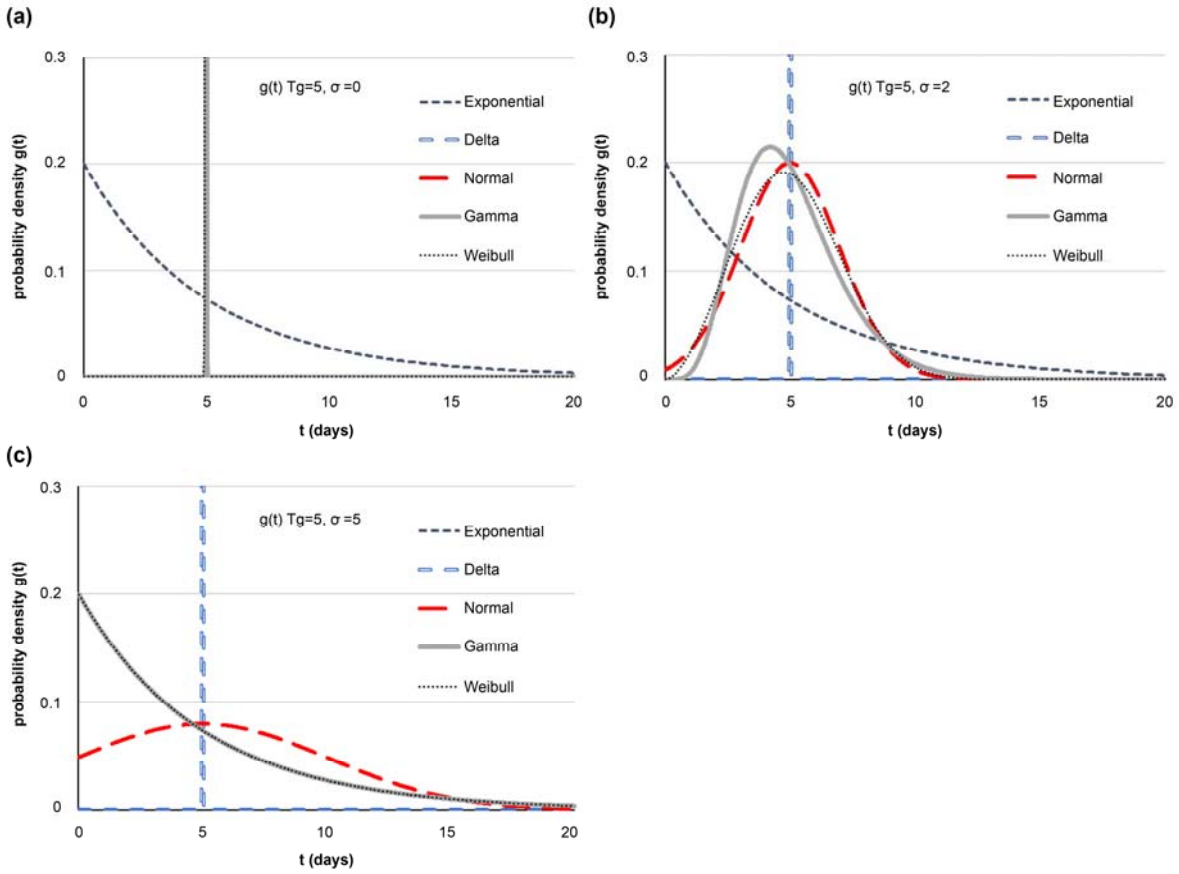

Figure 1. Generation time distributions g(t) with a mean (Tg) of 5 days and standard deviation (σ) values of 0, 2, and 5 days.

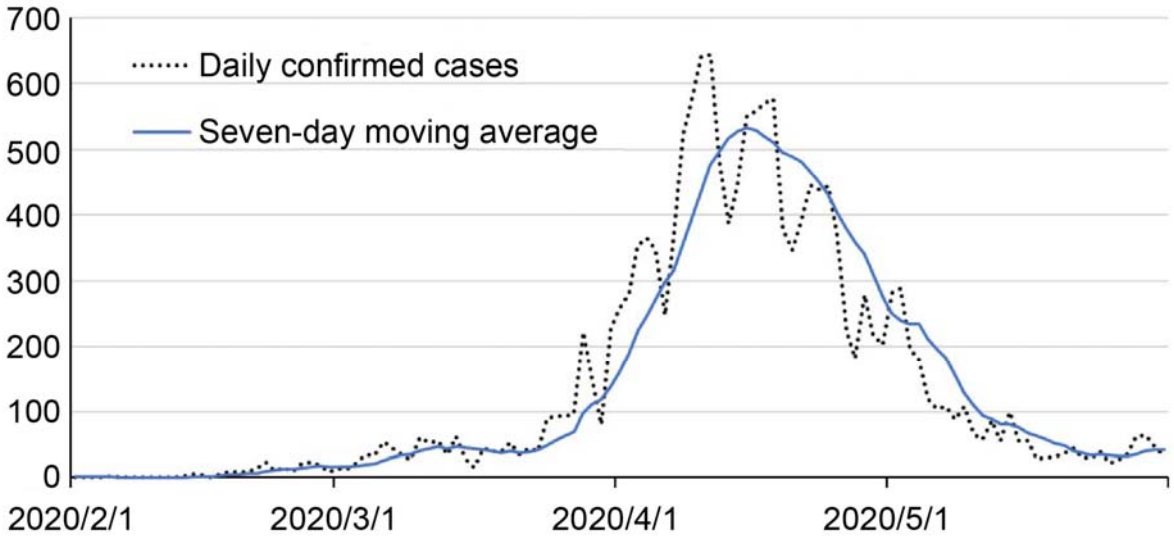

Figure 2. Trend in the number of newly confirmed cases of COVID-19 during the initial wave of infections in Japan (Ministry of Health. Labour and Welfare, 2023).

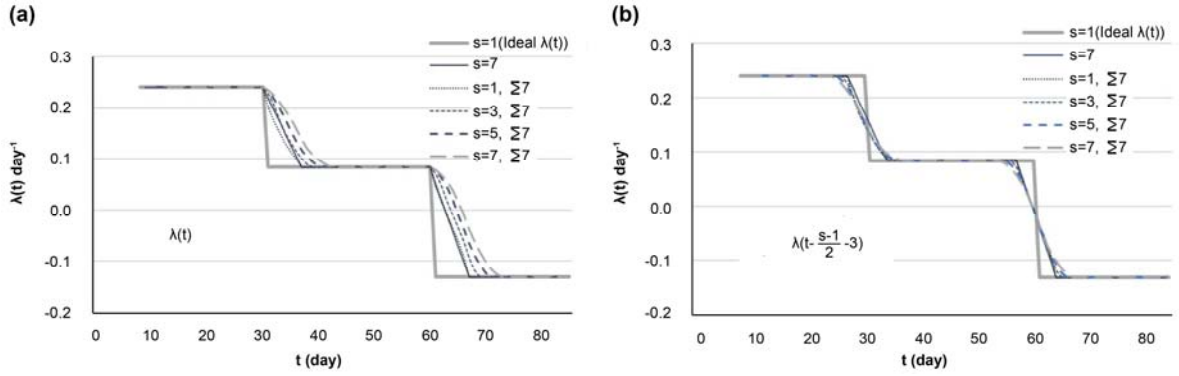

Figure 3. Estimation of the parameter λ(t) based on synthetic data, assuming gamma distribution with parameters Tg = 5 and σ = 2.

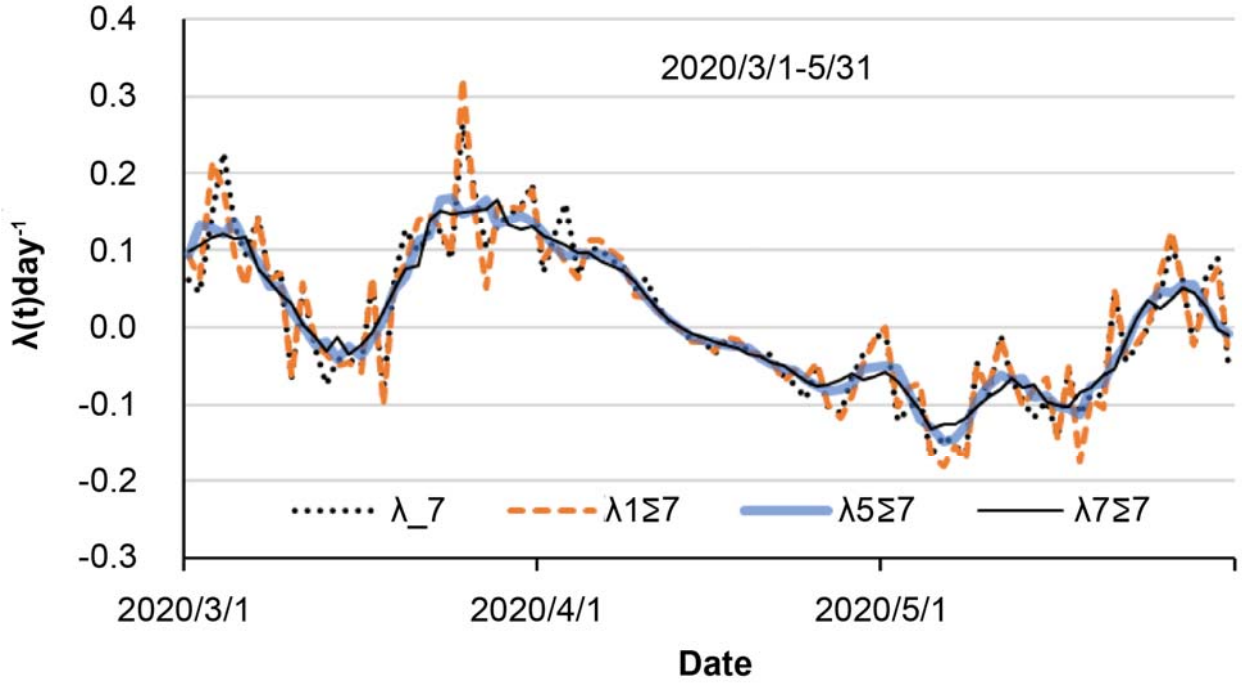

Figure 4. Relationship between the observation interval (s) and the growth rate λ(t).

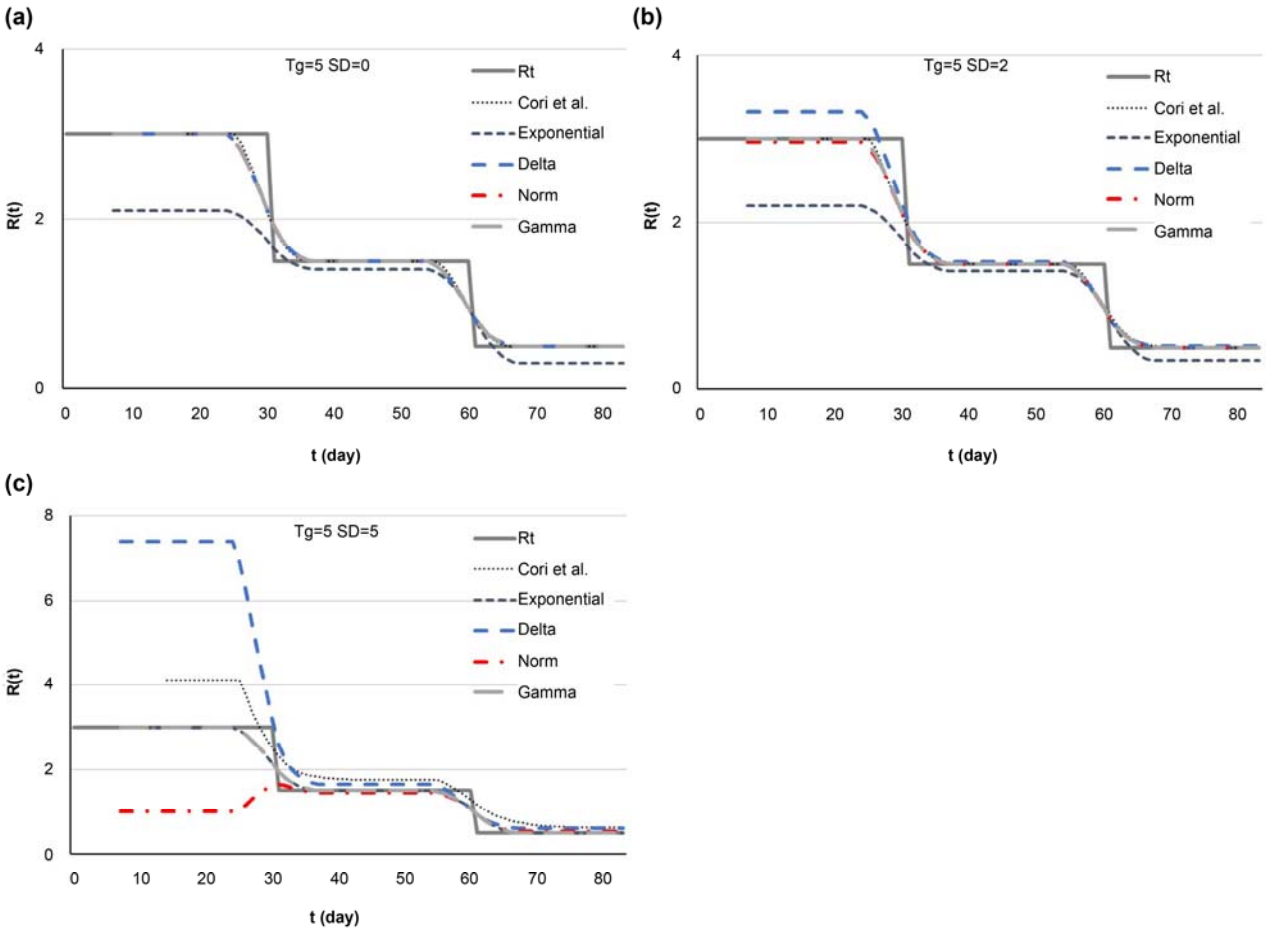

Figure 5. Estimation of the reproduction number R(t) using various methods, based on synthetic data.

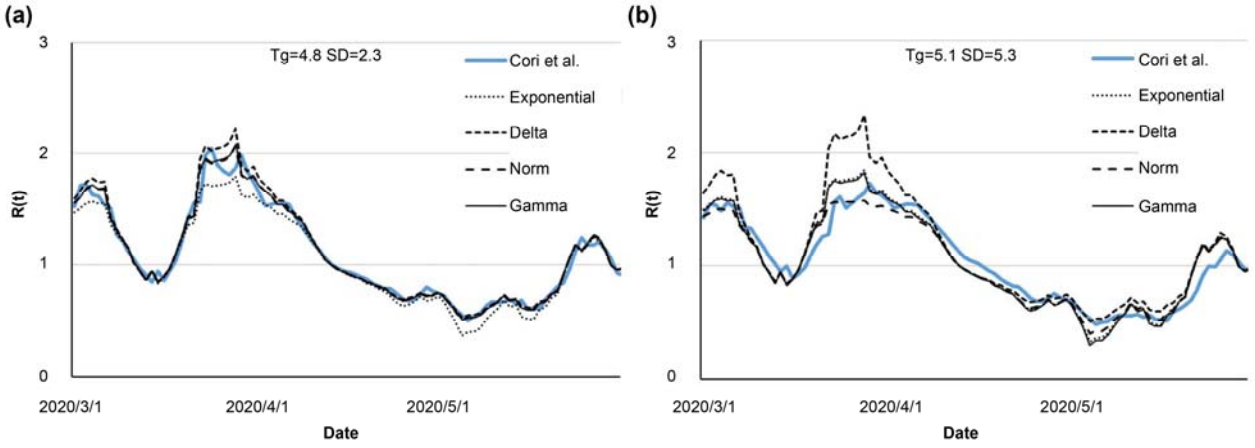

Figure 6. Estimation of the reproduction number R(t) based on real data of newly confirmed cases of COVID-19 in Japan.

Tables

Table 1. Estimation methods of reproduction number *R(t)* and the generation time distributions *g(t)*

Method	*R(t-s/2-3)* seven-day moving average	*R(t-s/2)* by *I(t)*, *Tg* and *σ*	*R* by *λ*, *Tg* and *σ*	*g(t)*
Exponential	$\frac{Tg}{s}\log_e\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)+1$	$\frac{Tg}{s}\log_e\left(\frac{I(t)}{I(t-s)}\right)+1$	$\lambda\ Tg+1$	$\frac{1}{Tg}\exp(-\frac{t}{Tg})$
Delta	$\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i))}\right)^{\frac{Tg}{s}}$	$\left(\frac{I(t)}{I(t-s)}\right)^{\frac{Tg}{s}}$	$\exp(\lambda\ Tg)$	if $t = Tg\ \infty$, else 0
Norm.	$\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)^{\left[\frac{Tg}{s}-\frac{\sigma^2}{2s^2}\log_e\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)\right]}$	$\left(\frac{I(t)}{I(t-s)}\right)^{\left[\frac{Tg}{s}-\frac{\sigma^2}{2s^2}\log_e\left(\frac{I(t)}{I(t-s)}\right)\right]}$	$\exp(\lambda\ Tg-(1/2)\lambda^2\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(t-Tg)^2}{2\sigma^2})$
Gamma	$\left\{\frac{\sigma^2}{s\times Tg}\log_e\left(\frac{\sum_{i=0}^{6} I(t-i)}{\sum_{i=0}^{6} I(t-s-i)}\right)+1\right\}^{\frac{Tg^2}{\sigma^2}}$	$\left\{\frac{\sigma^2}{s\times Tg}\log_e\left(\frac{I(t)}{I(t-s)}\right)+1\right\}^{\frac{Tg^2}{\sigma^2}}$	$(1+\frac{\sigma^2}{Tg}\lambda)^{\frac{Tg^2}{\sigma^2}}$	$\frac{1}{\Gamma(m)\eta}\left(\frac{t}{\eta}\right)^{m-1}\exp\left(-\frac{t}{\eta}\right)$

I(t): the number of new infections in a population at time *t*., *s*: observation interval, *Tg*: mean generation time, *σ*: standard deviation of the generation interval, *Γ()*:gamma function, m: Tg^2/σ^2, *η*: σ^2/Tg

Table 2. Correlation coefficients among estimation methods of reproduction number R(t) based on confirmed COVID-19 case numbers in Japan, from March 1, 2020 to February 28, 2021. n=365

Tg =4.8, σ=2.3	Cori et al.	Exponential	Delta	Norm.	Gamma
Cori et al.	1.00	0.98	0.98	0.99	0.99
Exponential	0.98	1.00	0.98	0.99	0.99
Delta	0.98	0.98	1.00	1.00	1.00
Norm	0.99	0.99	1.00	1.00	1.00
Gamma	0.99	0.99	1.00	1.00	1.00
Tg=5.1, σ=5.3	Cori et al.	Exponential	Delta	Norm.	Gamma
Cori et al.	1.00	0.96	0.94	0.96	0.96
Exponential	0.96	1.00	0.98	1.00	1.00
Delta	0.94	0.98	1.00	0.96	0.98
Norm	0.96	1.00	0.96	1.00	1.00
Gamma	0.96	1.00	0.98	1.00	1.00

Table 3. Summary statistics of estimated reproduction number *R(t)* based on confirmed COVID-19 case numbers in Japan, from March 1, 2020 to February 28, 2021, including mean, standard deviation(SD) and standard error(SE). n=365

Mean and SD of generation time	R(t) estimation method	*R(t)* average	SD	SE
Tg=4.8, *σ*=2.3	Cori et al.	1.08	0.28	0.015
	Exponential	**1.05	0.26	0.014
	Delta	**1.09	0.30	0.016
	Norm	1.08	0.29	0.015
	Gamma	1.08	0.29	0.015
Tg=5.1, *σ*=5.3	Cori et al.	1.05	0.25	0.013
	Exponential	1.06	0.28	0.015
	Delta	**1.10	0.32	0.017
	Norm	1.04	0.25	0.013
	Gamma	1.05	0.28	0.015

** indicates the results of multiple comparison analysis when Cori et al.'s method was specified as Level 1, demonstrating significant differences ($P < 0.01$).