

The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends

Hiroshi Nishiura and Gerardo Chowell

Abstract Although the basic reproduction number, R_0 , is useful for understanding the transmissibility of a disease and designing various intervention strategies, the classic threshold quantity theoretically assumes that the epidemic first occurs in a fully susceptible population, and hence, R_0 is essentially a mathematically defined quantity. In many instances, it is of practical importance to evaluate time-dependent variations in the transmission potential of infectious diseases. Explanation of the time course of an epidemic can be partly achieved by estimating the effective reproduction number, $R(t)$, defined as the actual average number of secondary cases per primary case at calendar time t (for $t > 0$). $R(t)$ shows time-dependent variation due to the decline in susceptible individuals (intrinsic factors) and the implementation of control measures (extrinsic factors). If $R(t) < 1$, it suggests that the epidemic is in decline and may be regarded as being *under control* at time t (vice versa, if $R(t) > 1$). This chapter describes the primer of mathematics and statistics of $R(t)$ and discusses other similar markers of transmissibility as a function of time.

1 Introduction

The basic reproduction number, R_0 (pronounced as *R nought*), is a key quantity used to estimate transmissibility of infectious diseases. Theoretically, R_0 is defined as the average number of secondary cases generated by a single primary case during its entire period of infectiousness in a fully susceptible population [14]. The reproduction number, R , is directly related to the type and intensity of interventions necessary to control an epidemic since the objective of public health efforts is to achieve $R < 1$ as soon as possible. One of the best known utilities of R_0 is in determining the critical coverage of immunization required to eradicate a disease in a randomly mixing population. When an effective vaccine is available against the disease in question, it is of interest to estimate the critical proportion of the

H. Nishiura (✉)

Theoretical Epidemiology, University of Utrecht, Yalelaan 7, 3584 CL, Utrecht, The Netherlands
e-mail: h.nishiura@uu.nl

population that needs to be vaccinated (*i.e.* vaccination coverage) in order to attain $R < 1$ [3, 4, 33]. Considering the so-called *control relation*, $1 - \frac{1}{R_0}$, the protection conferred to the population by achieving a critical vaccination coverage, *herd immunity*, yields the threshold condition for the eradication of a disease [18, 28]. As it is extensively discussed elsewhere [13, 14], the mathematical definition of R_0 is given by using the next generation matrix where R_0 is in the simplest case calculated as the dominant eigenvalue (see Chapter 1). In addition to the threshold phenomena, R_0 has been classically used to suggest the *severity* of an epidemic, because the proportion of those experiencing infection at the end of an epidemic (*i.e.* *final size*) depends only on R_0 [23]. The basic statistical methods to estimate R_0 from observed epidemiological datasets have been reviewed by Klaus Dietz elsewhere [15].

Although R_0 may be useful for understanding the transmissibility of a disease and designing various intervention strategies, the classic threshold quantity theoretically assumes that the epidemic first occurs in a *fully* susceptible population, and hence, R_0 is essentially a *mathematically defined* quantity. In addition to R_0 , it is of practical importance to evaluate time-dependent variations in the transmission potential. Explanation of the time course of an epidemic can be partly achieved by estimating the effective reproduction number, $R(t)$, defined as the actual average number of secondary cases per primary case at calendar time t (for $t > 0$) [6–10, 22, 29, 30, 35]. $R(t)$ shows time-dependent variation due to the decline in susceptible individuals (intrinsic factors) and the implementation of control measures (extrinsic factors). If $R(t) < 1$, it suggests that the epidemic is in decline and may be regarded as being *under control* at time t (vice versa, if $R(t) > 1$). Even when effective interventions against a specific disease are limited, it is plausible that the contact frequency leading to infection varies as a function of time owing to the recognition of epidemics and/or dissemination of the relevant information through mass media. In this chapter, we show how $R(t)$ is mathematically defined and how it can be estimated from the observed epidemiological datasets. In addition, other similar time-dependent threshold quantities, which have been proposed in a few practical settings, are discussed.

2 Renewal Equation Offers the Conceptual Understanding of $R(t)$

2.1 Infection-Age Structured Model

To understand the theoretical concept of $R(t)$, we first consider an *infection-age* structured epidemic model. Hereafter, *infection-age* stands for the time elapsed since infection. Whereas the simple modified version (or widely known form) of the Kermack-McKendrick model is governed by ODEs (e.g. SIR and SEIR models), the very initial model employed the infection-age structured assumption from in 1927 [24]. Nevertheless, the mathematical importance of the original model was recognized only after the 1970s [12, 27]. We denote the numbers of susceptible and

recovered individuals by $S(t)$ and $U(t)$ (Note: to avoid any confusions between the effective reproduction number and the recovered individuals, we denote the recovered individuals by $U(t)$ hereafter). Further, let $i(t, \tau)$ be the density of infectious individuals at calendar time t and infection-age τ . The infection-age structured SIR model is given by

$$\begin{aligned} \frac{dS(t)}{dt} &= -\lambda(t)S(t) \\ \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau}\right) i(t, \tau) &= -\gamma(\tau)i(t, \tau) \\ i(t, 0) &= \lambda(t)S(t) \\ \frac{dU(t)}{dt} &= \int_0^\infty \gamma(\tau)i(t, \tau) d\tau \end{aligned} \tag{1}$$

where $\lambda(t)$ is referred to as the *force of infection* (foi) at calendar time t (i.e. foi is defined as the rate at which susceptible individuals get infected) which is given by:

$$\lambda(t) = \int_0^\infty \beta(\tau)i(t, \tau) d\tau \tag{2}$$

and $\beta(\tau)$ and $\gamma(\tau)$ are the rates of secondary transmissions per single infectious case and recovery at infection-age τ , respectively. It should be noted that the above model has not taken into account the background host demography (i.e. birth and death). In a closed population, the total population size N is thus given by

$$N = S(t) + \int_0^\infty i(t, \tau) d\tau + U(t) \tag{3}$$

which is independent of calendar time t . The system (1) can be reasonably integrated along the characteristic line

$$i(t, \tau) = \Gamma(\tau)j(t - \tau) \tag{4}$$

for $t - \tau > 0$ (and $\frac{\Gamma(\tau)}{\Gamma(\tau-t)}j_0(\tau - t)$ for $\tau - t > 0$) where

$$j(t) = i(t, 0) \tag{5}$$

and

$$\Gamma(\tau) = \exp\left(-\int_0^\tau \gamma(\sigma) d\sigma\right) \tag{6}$$

and $j_0(\tau)$ informs the infection-age distribution of initially infected individuals at the beginning of an epidemic. Accordingly, the number of new infections at calendar time t , $j(t)$, is referred to as the *incidence of infection*. It is not difficult to derive

$$S(t) = S(0) - \int_0^t j(\sigma) d\sigma \quad (7)$$

from (1). Thus, the subequation of $i(t, 0)$ in system (1) is rewritten as

$$j(t) = \lambda(t) \left[S(0) - \int_0^t j(\sigma) d\sigma \right] \quad (8)$$

Taking into account the initial condition in (4), Equation (8) is rewritten as

$$j(t) = \left[S(0) - \int_0^t j(\sigma) d\sigma \right] \left[G(t) + \int_0^t \psi(\tau) j(t - \tau) d\tau \right] \quad (9)$$

where

$$\psi(\tau) = \beta(\tau)\Gamma(\tau) \quad (10)$$

$$G(t) = \int_0^\infty \beta(\sigma + t) \frac{\Gamma(\sigma + t)}{\Gamma(\sigma)} j_0(\sigma) d\sigma \quad (11)$$

Considering the initial invasion phase (i.e. exponential growth phase of an epidemic), we get a linearized equation

$$j(t) = S(0)G(t) + S(0) \int_0^t \psi(\tau) j(t - \tau) d\tau \quad (12)$$

The Equation (12) represents Lotka's integral equation, where the basic reproduction number, R_0 , is given by

$$R_0 = S(0) \int_0^\infty \psi(\tau) d\tau \quad (13)$$

Thus, the epidemic will grow if $R_0 > 1$ and decline to extinction if $R_0 < 1$. Assuming that the infection-age distribution is stable, we get a simplified renewal equation

$$j(t) = \int_0^\infty A(\tau) j(t - \tau) d\tau \quad (14)$$

where $A(\tau)$ is the product of $\psi(\tau)$ and $S(0)$, indicating the rate of secondary transmissions caused by a single primary case at calendar time 0 and infection-age τ . Assuming that we observe an exponential growth of incidence during the initial

phase (i.e. $j(t) = k \exp(rt)$ where k and r are, respectively, a constant ($k > 0$) and the intrinsic growth rate), the following relationship is obtained:

$$j(t) = j(t - \tau) \exp(r\tau) \quad (15)$$

Replacing $j(t - \tau)$ in the right hand side of (14) by (15), we get

$$j(t) = \int_0^\infty A(\tau) j(t) \exp(-r\tau) d\tau \quad (16)$$

Removing $j(t)$ from both sides of (16), we get the Euler-Lotka characteristic equation:

$$1 = \int_0^\infty e^{-r\tau} A(\tau), d\tau \quad (17)$$

Further, we consider a probability density of the *generation time* (i.e. the time from infection of a primary case to the infection of a secondary case by the primary case [34]), denoted by $w(\tau)$:

$$w(\tau) := \frac{A(\tau)}{\int_0^\infty A(x)dx} = \frac{A(\tau)}{R_0}. \quad (18)$$

Using (18), the Equation (17) is replaced by

$$\frac{1}{R_0} = \int_0^\infty \exp(-r\tau) w(\tau), d\tau \quad (19)$$

The Equations (15), (16), (17), (18), (19) are what Wallinga and Lipsitch have discussed, revisiting the classical theory of Lotka [16, 36], which reasonably suggests the relationship between the generation-time distribution and R_0 . Accordingly, the estimator of R_0 using the intrinsic growth rate is given by:

$$\hat{R}_0 = \frac{1}{M(-r)}, \quad (20)$$

where $M(-r)$ is the moment generating function of the generation-time distribution $w(\tau)$, given the intrinsic growth rate r [36]. Equation (20) significantly improved the issue of estimating R_0 using the intrinsic growth rate alone, because (20) permits validating estimates of R_0 by various different distributional assumptions for $w(\tau)$. The importance of realistic assumptions for the distributions of latent and infectious periods has been emphasized in recent studies [25, 26, 32, 37, 39] and indeed, this point is addressed by (20) to gain robust estimate of R_0 . It should be noted that the convolution of latent and infectious periods yields $w(\tau)$. Since the assumed lengths of generation time most likely yielded different estimates of R_0 , for example, for

Spanish influenza by different studies [30], Equation (20) highlights a critical need to clarify the generation time distribution using observed data.

2.2 Deriving the Estimator of the Effective Reproduction Number

To further derive an estimator of $R(t)$, we consider the non-linear phase of an epidemic. Derivation of R_0 given by (20) assumes an exponential growth which is applicable only during the very initial phase of an epidemic (or, when the transmission is stationary over time), and thus, it is of practical importance to widen the applicability of the above-described renewal equations in order to appropriately interpret the time-course of an epidemic. We explicitly account for the depletion of susceptible individuals, as we deal with an estimation issue with time-inhomogeneous assumptions. Adopting the *mass action* principle of Kermack and McKendrick, we get:

$$\begin{aligned} j(t) &= S(t) \int_0^\infty \psi(\tau) j(t - \tau) d\tau \\ &= \int_0^\infty A(t, \tau) j(t - \tau) d\tau \end{aligned} \quad (21)$$

where $A(t, \tau)$ is interpreted as the reproductive power at calendar time t and infection-age τ at which an infected individual generates secondary cases. We refer to the Equation (21) as a non-autonomous renewal equation, where the number of new infections at calendar time t is proportional to the number of infectious individuals (as assumed in the renewal equation in the initial phase).

Using Equation (21), the effective reproduction number, $R(t)$ (i.e. the *instantaneous* reproduction number at calendar time t) is defined as:

$$R(t) = \int_0^\infty A(t, \tau) d\tau \quad (22)$$

where $A(t, \tau)$ is, in practical terms, decomposed as

$$A(t, \tau) = S(t)\beta(\tau)\Gamma(\tau) \quad (23)$$

Following (23), we can immediately see that $R(t)$ with an autonomous assumption (i.e. where contact and recovery rates do not vary with time) is given by:

$$R(t) = \frac{S(t)}{S(0)} R_0 \quad (24)$$

which is shown in [14]. In practical terms, Equation (24) reflects the temporal decline in the epidemic due to depletion of susceptible individuals. This corresponds to the classic assumption of the Kermack and McKendrick model.

However, as we discussed in the beginning of this chapter, we postulate that human contact behaviors (and other extrinsic factors) modifies the dynamics as a function of epidemic time, assuming that the decline in incidence does reflect not only depletion of susceptibles but also various extrinsic dynamics (e.g. isolation and contact tracing). Thus, instead of the assumption in (21), we assume time-inhomogeneous $\psi(t, \tau)$; *i.e.*

$$\begin{aligned} j(t) &= S(t) \int_0^\infty \psi(t, \tau) j(t - \tau) d\tau \\ &= \int_0^\infty A(t, \tau) j(t - \tau) d\tau \end{aligned} \quad (25)$$

to describe $A(t, \tau)$.

Even so, it is convenient to assume separation of variables for $A(t, \tau)$ to derive simple estimator of $R(t)$ (implicitly assuming that the relative infectiousness to infection-age is independent of calendar time) [20]. Under this assumption, $A(t, \tau)$ is rewritten as the product of two functions $\phi_1(t)$ and $\phi_2(\tau)$:

$$A(t, \tau) = \phi_1(t)\phi_2(\tau) \quad (26)$$

Arbitrarily assuming a normalized density for $\phi_2(\tau)$, *i.e.*,

$$\int_0^\infty \phi_2(\tau) d\tau \equiv 1 \quad (27)$$

then, it is easy to find that

$$R(t) = \int_0^\infty A(t, \tau) d\tau = \phi_1(t) \quad (28)$$

suggesting that the function $\phi_1(t)$ is equivalent to the (instantaneous) effective reproduction number $R(t)$. Another function $\phi_2(\tau)$ represents the density of infection events as a function of infection-age τ . Accordingly, we can immediately see that $\phi_2(\tau)$ is exactly the same as $w(\tau)$, the generation-time distribution. That is, the above arguments suggest that $A(t, \tau)$ (*i.e.* the rate at which an infectious individual at calendar time t and infection-age τ produces secondary cases) is decomposed as:

$$A(t, \tau) = R(t)w(\tau) \quad (29)$$

Inserting (29) into (25) yields an estimator of $R(t)$ [20]:

$$\hat{R}(t) = \frac{j(t)}{\int_0^\infty j(t - \tau)w(\tau) d\tau} \quad (30)$$

Another type of the effective reproduction number as a function of time considers the number of secondary cases per single primary case as a function of calendar time when the primary case experienced infection. Due to this reason, the reproduction number is referred to as the *cohort* reproduction number, $R_c(t)$, defined as

$$R_c(t) = \int_0^{\infty} A(t + \tau, \tau) d\tau \quad (31)$$

If the separable assumption (28) is the case, Equation (31) is rewritten as

$$R_c(t) = \int_0^{\infty} R(t + \tau)w(\tau) d\tau \quad (32)$$

which is interpreted as a smoothed function of the instantaneous reproduction number [20, 21]. The above Equation (32) is exactly what was proposed in applications to SARS [35] and foot and mouth disease [17]. Preceding these definitions in infectious disease epidemiology [20], both $R(t)$ and $R_c(t)$ have been explicitly defined as the period and cohort total fertility rates, respectively, in mathematical demography [1]. The difference between $R(t)$ and $R_c(t)$ is highlighted when a specific event at calendar time t occurs (e.g. a public health intervention starts at calendar time t). Then, $R(t)$ abruptly varies (e.g. declines) with calendar time t , but $R_c(t)$ smoothly varies, because $R_c(t)$ smooth out the timing (i.e. infection-age) of secondary transmissions among a cohort who experienced infection at calendar time t .

Discretizing (30) and (32) to apply them to the daily incidence data (i.e. using j_i incident cases infected between time t_i and time t_{i+1} and discretized generation time distribution w_i),

$$\hat{R}(t_i) = \frac{j_i}{\sum_{j=0}^n j_{i-j} w_j} \quad (33)$$

can be used as the estimator of $R(t)$, and

$$\hat{R}_c(t_i) = \frac{\sum_{m=0}^n j_{i+m} w_m}{\sum_{k=0}^n j_{i+m-k} w_k} \quad (34)$$

as the estimator of $R_c(t)$. However, it should be noted that the study in SARS implicitly assumed that onset data $c(t)$ at calendar time t reflects the above discussed infection event $j(t)$ [35]. That is, supposing that we observed c_i onset cases reported between t_i and t_{i+1} , $R_c(t)$ was calculated as

$$\hat{R}_c(t_i) = \frac{c_i}{\sum_{j=0}^n c_{i-j} s_j} \quad (35)$$

where s_j is the discretized *serial interval* which is defined as the time from onset of a primary case to onset of the secondary cases [19]. The method permits reasonable transformation of an epidemic curve (i.e. temporal distribution of case onset) to the estimates of time-inhomogeneous cohort reproduction number $R_c(t)$. Employing the relative likelihood of case k infected by case l using the density function of serial interval $s(t)$; i.e.,

$$P^{(k,l)} = \frac{s(t_k - t_l|\theta)}{\sum_{m \neq k} s(t_k - t_m|\theta)} \quad (36)$$

the expected value and variance of $R_c(t_i)$ are given by the following

$$\begin{aligned} E(R_c(t_i)) &= \frac{1}{n_t^2} \sum_{l:t_l=t}^{n-q} \sum_{k=1}^{n-q} P^{(k,l)} \\ \text{Var}(R_c(t_i)) &= \frac{1}{n_t^2} \sum_{k=1}^{n-q} \left(\sum_{l:t_l=t} P^{(k,l)}(1 - P^{(k,l)}) - \sum_{l,m:t_l=t_m=t} P^{(k,l)}P^{(k,m)} \right) \end{aligned} \quad (37)$$

where n_t is the total number of reported case onsets at calendar time t [11].

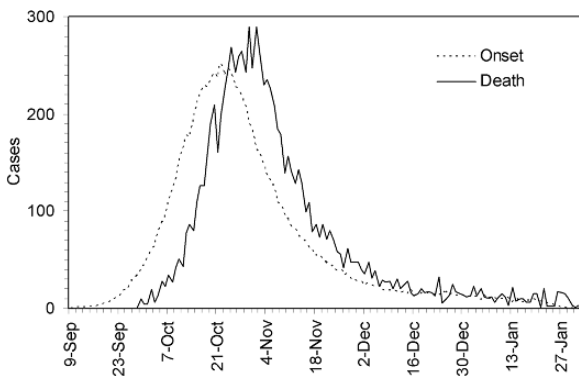
Using the above described methods (or similar concepts with similar assumptions), we can transform epidemic curves into the effective reproduction number and assess the impact of control measures on an epidemic. However, whereas the Equations (33) and (35) are similar in theory, we need to explicitly account for the difference between onset and infection event. In fact, when there are many asymptomatic infections and asymptomatic secondary transmissions, serial interval is not equivalent to the generation time, and thus, directly adopting the above methods would be inappropriate.

3 Applying Theory to the Data

3.1 A Simple Example

Here we consider a simplified example of pandemic influenza from 1918 to 1919 in Prussia, Germany [30]. Medical officers in Prussia recorded the daily number of influenza deaths from 29 September 1918 to 1 February 1919 (Fig. 1) [31]; a total of 8911 deaths were reported. Throughout the pandemic period in Germany, the largest number of deaths was seen in this fall wave. Prussia represents the northern part of present Germany and at the time of the pandemic it was part of the Weimer Republic as a free state following World War I. The death data were collected from 28 different local districts surrounding the town of Arnsberg, which, at the time of the epidemic, had a population of approximately 2.5 million individuals (the mortality rate in this period being 0.36%). Although case fatality for the entire observation

Fig. 1 Epidemic curve of pandemic influenza in Prussia, Germany, from 1918 to 1919. Reported daily number of influenza deaths (*solid line*) and the back-calculated temporal distribution of onset cases (*dashed line*). Daily counts of onset cases were obtained using the time delay distribution from onset to death (Fig. 2). Data source: Ref [31]



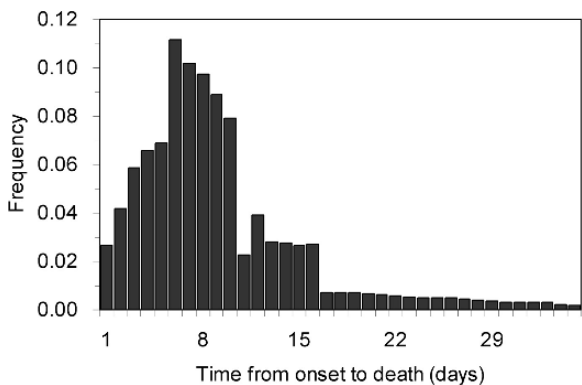
area was not documented, the numbers of cases and deaths during part of the fall wave were recorded for 25 districts. Among a total of 61,824 cases, 1609 deaths were observed, yielding a case fatality estimate of 2.60% (95% CI: 2.48, 2.73). For simplicity, the inflow and outflow of individuals migrating between Prussia and other areas were ignored in the following analysis.

The daily incidence (i.e. daily case onset) was back-calculated using the daily number of influenza deaths (Fig. 1) and the time delay distribution from onset to death (Fig. 2). Given $f(\tau)$, the frequency of death τ days after onset, the relationship between the reported daily number of deaths, $D(t)$, and daily incidence, $C(t)$, at calendar time t is given by:

$$D(t) = p \int_0^t C(t - \tau) f(\tau) d\tau \tag{38}$$

where p is the case fatality ratio, which is independent of time. Although the case fatality, p , was not taken into account in Fig. 1, the following model reasonably cancels out the effect of p assuming that the conditional probability of death given infection is independent of time.

Fig. 2 Distribution of the time delay from onset to death during the influenza epidemic in Prussia, Germany, from 1918 to 1919. Time from disease onset (i.e. fever) to death is given for 6233 influenza deaths. A simple 5-day moving average was applied to the original data. Data source: Ref [31]



The effective reproduction number can be estimated using estimators (33) and (34), but, unfortunately, detailed information on the distribution of the generation time, $w(\tau)$, has yet to be clarified for pandemic influenza, and historical records often offer only the approximate mean length. Thus, the analyses conducted here simplify the model using various mean lengths of the generation time assumed in previous studies. Supposing that we observed C_i cases in generation i , the expected number of cases in generation $i + 1$, $E(C_{i+1})$ occurring a mean generation time after onset of C_i is given by:

$$E(C_{i+1}) = C_i R_i \tag{39}$$

where R_i is the effective (cohort) reproduction number in generation i . That is, cases in each generation, $C_1, C_2, C_3, \dots, C_n$ are given by $C_0 R_0, C_1 R_1, C_2 R_2, \dots, C_{n-1} R_{n-1}$ and also by $C_0 R_0, C_0 R_0 R_1, C_0 R_0 R_1 R_2, \dots, C_0 \prod_{k=0}^{n-1} R_k$, respectively. By incorporating variations in the number of secondary transmissions generated by each case into the same generation (referred to as the offspring distribution), the model can be formalized using a discrete-time branching process [5]. The Poisson process is conventionally assumed to model the offspring distribution, representing stochasticity (i.e. randomness) in the transmission process. This assumption indicates that the conditional distribution of the number of cases in generation $i + 1$ given C_i is given by:

$$C_{i+1} | C_i \sim \text{Poisson}[C_i R_i] \tag{40}$$

For observation of cases from generation 0 to N , the likelihood of estimating R_i is given by:

$$L = \text{constant} \times \prod_{j=0}^{N-1} (C_j R_j)^{C_{j+1}} \exp(-C_j R_j) \tag{41}$$

Since the Poisson distribution represents a one parameter power series distribution, the expected values and uncertainty bounds of R_i can be obtained for each generation. The 95% CI were derived from the profile likelihood. Since the length of the generation time in previous studies ranged from 0.9 to 6 days, three different fixed-length generation times (i.e. 1, 3 and 5 days) are assumed for Equation (41) with respect to the observed data. Although application of the delta function for the generation time suffers some overlapping of cases in successive generations, this exercise ignored this and, rather, focused on the time variation in transmissibility using this simple assumption. That is, assuming that the generation-time distribution of length τ , $w(\tau)$, is given by the following delta function with the mean length 1, 3 or 5 days,

$$w(\tau) = \infty, \quad \text{for } \tau = 1, 3 \text{ or } 5 \tag{42}$$

and $w(\tau) = 0$ otherwise, and for each assumption of the mean length, the daily number of cases was grouped by the determined generation time length. Whereas the choice of generation time therefore affects estimates of R_t , it does not affect the ability to predict the temporal distribution of cases. It should be noted that this simple model assumes a homogeneous pattern of spread.

Figure 3 shows time variations in the estimated effective reproduction numbers obtained assuming three different generation times (i.e. 1, 3 and 5 days) compared with the corresponding epidemic curve. Epidemic date 0 represents 9 September 1918 when the back-calculated onset of cases initially yielded a value the nearest integer of which was 1. Since the precision of the estimate is influenced by the observed number of cases, wide 95% confidence intervals were observed for estimates using a short generation time. However, these time variations in $R(t)$

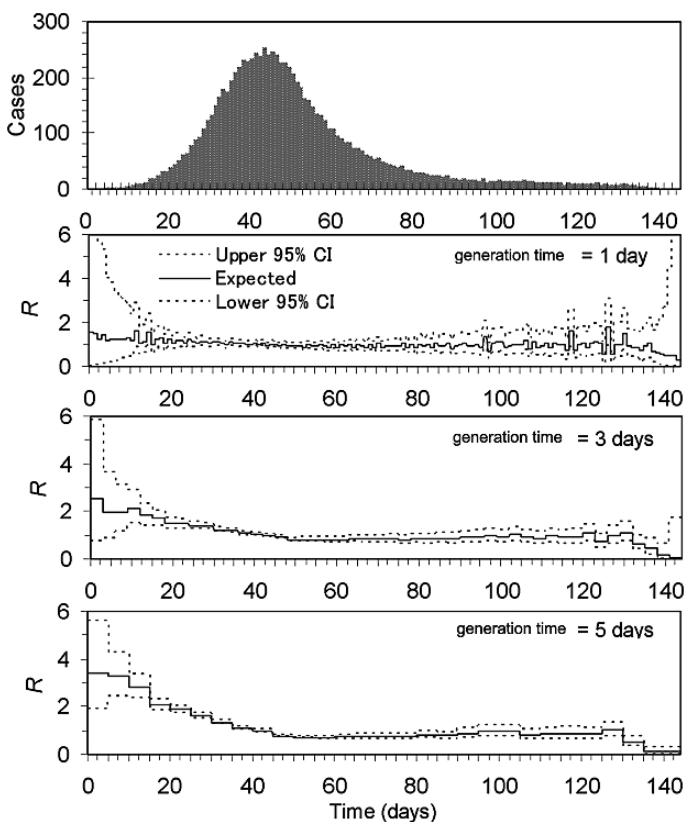


Fig. 3 Epidemic curve and the corresponding effective reproduction numbers (R) with variable generation times. Time variation in the effective reproduction number (the number of secondary infections generated per case by generation) assuming three different generation times is shown. The generation time was assumed to be 1 (second from the top), 3 (lower middle) and 5 days (bottom). Days are counted from September 9, 1918, onwards

exhibited similar qualitative patterns: (i) although the $R(t)$ was highest at the beginning of the epidemic, the estimates fell below 1 when the epidemic curve came close to the peak (i.e. Days 45–50). For example, the estimated $R(t)$ at Day 50 was 0.92 (95% CI: 0.79, 1.06), 0.82 (0.75, 0.89) and 0.72 (0.67, 0.78), respectively, for a generation time of 1, 3 and 5 days. This period corresponds to the time when public health measures were instituted, e.g. obligatory case reporting, encouragement of mask wearing, and closing of public buildings such as churches and theaters [31]. (ii) Thereafter, $R(t)$ stayed slightly below unity, reflecting a slow decline in the number of onset cases. (iii) Shortly before the end of the epidemic (i.e. Days 90–120), $R(t)$ increased again above 1. (iv) Finally, the expected values of $R(t)$ fell below 1 very close to the end of the epidemic. In this stage, estimates assuming a short generation time exhibited wide uncertainty bounds, reflecting stochasticity due to the small number of cases.

Figure 4 compares the expected values of $R(t)$ assuming each of the generation times employed. Although the possibility of individual heterogeneity (e.g. potential superspreaders in the early stage) cannot be excluded, $R(t)$ at calendar time $t = 0$ is theoretically equivalent to R_0 . Assuming generation times of 1, 3 and 5 days, R_0 was estimated to be 1.58 (95% CI: 0.03, 10.32), 2.52 (0.75, 5.85) and 3.41 (1.91, 5.57), respectively. It is remarkable, therefore, to see that $R(t)$ largely depends on the assumed length of the generation time. That is, the longer the generation time, the higher the $R(t)$. It should also be noted that the relationship between $R(t)$ and the generation time is reversed when the epidemic is under control (i.e. when $R(t) < 1$ in the later stage of the epidemic). The finding is analytically interpretable from Equation (20) which suggests that the absolute number of the reproduction number

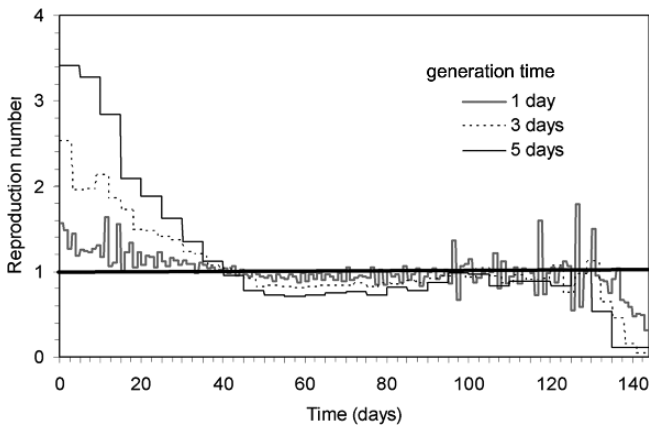


Fig. 4 Comparison of the effective reproduction number assuming different generation times. Expected values of the effective reproduction number with a generation time of 1 (grey), 3 (dashed black) and 5 days (solid black). The horizontal solid line represents the threshold value, $R = 1$, below which the epidemic will decline to extinction. Days are counted from September 9, 1918, onwards

is informed by the growth rate of an epidemic as well as the shape and scale of the generation time distribution [32, 36].

3.2 What to do with the Coarsely Reported Data?

Although we usually seek for precisely reported data (e.g. daily counts of cases) to estimate the reproduction number as a function of time, it is impractical in many instances to report observations every day (or to be more precise). If the datasets are reported in a very coarse interval, we have to consider alternative simple algorithms to deal with interval censoring. There are two approaches.

The first is the geometric approximation. As above, we consider that the expected number of cases in generations $0, 1, 2, \dots, i$ follows a simple geometric series, but with a constant reproduction number, R_k , in a single reporting interval k :

$$a, aR_k, aR_k^2, \dots, aR_k^i \quad (43)$$

where a denotes the number of index cases in the first generation of reporting interval k . As a special case, suppose that the reporting interval, Δt , is exactly a multiple of the mean generation time (i.e. $\Delta t = ng$ where g and n are the mean generation time and an integer, respectively). In that case, the numbers of cases in k -th and $(k + 1)$ -th reports, J_k and J_{k+1} , are

$$J_k = a + aR_k + aR_k^2 + \dots + aR_k^{n-1} \quad (44)$$

$$= a \sum_{i=0}^{n-1} R_k^i \quad (45)$$

and

$$J_{k+1} = aR_k^n + aR_k^n R_{k+1} + aR_k^n R_{k+1}^2 + \dots + aR_k^n R_{k+1}^{n-1} \quad (46)$$

$$= aR_k^n \sum_{i=0}^{n-1} R_{k+1}^i \quad (47)$$

where R_k and R_{k+1} are the effective reproduction numbers in reporting intervals k and $k + 1$, respectively. Thus, given an observation of J_k cases in interval k , the expected number of cases in the next interval $k + 1$, $E(J_{k+1} | J_k)$, is given by

$$E(J_{k+1} | J_k) = \frac{(1 - R_k)(1 - R_{k+1}^n)R_k^n J_k}{(1 - R_{k+1})(1 - R_k^n)} \quad (48)$$

It should be noted that n is the number of generations included in each reporting interval.

When the geometric approximation is not feasible (e.g. when the reporting interval is not exactly the multiple of the mean generation time), the exponential approximation should replace the geometric approach. Let r_k and r_{k+1} be the constant growth rates of cases in reporting intervals k and $k + 1$, the conditional expectation (48) is replaced by

$$E(J_{k+1} | J_k) = \frac{J_k r_k \exp(r_k \Delta t)}{r_{k+1}} \frac{\exp(r_{k+1} \Delta t) - 1}{\exp(r_k \Delta t) - 1} \quad (49)$$

where Δt is the length of reporting interval. Using the maximum likelihood method, the growth rates r_k are estimated for each reporting interval k . Subsequently, R_k in each reporting interval k is estimated as

$$R_k = \frac{1}{M(-r_k)} \quad (50)$$

which is analogous to Equation (20).

In this way, even when the reporting interval is coarse (e.g. exceeding the mean generation time), we can still get approximate estimates of R_k which is assumed constant during the single reporting interval. Nevertheless, the linear approximation diminishes precision, and thus, it should be remembered that the observation in more precise reporting interval always gives better insights into the time-course of an epidemic.

4 Incidence-to-Prevalence Ratio and the Actual Reproduction Number

As discussed in the last section, it is frequently the case that the generation time for a specific disease has yet to be estimated, and we do not have the relevant data. Previously, another simple method was proposed; namely, the *incidence-to-prevalence ratio* has been employed in interpreting the time course of an epidemic [2, 38]. In particular, the method has been employed to understand the time course of the HIV epidemic. Although it is true that the problem of long generation times for HIV would complicate the interpretation of the simple method (and thus, the instantaneous and cohort reproduction numbers may always provide better information), here we explicitly consider theoretical backgrounds of this simple method.

As we did in the previous sections, we consider the renewal equation:

$$\begin{aligned} j(t) &= S(t) \int_0^\infty \beta(\tau) i(t, \tau) d\tau \\ &= S(t) \int_0^\infty \beta(\tau) \Gamma(\tau) j(t - \tau) d\tau \end{aligned} \quad (51)$$

which employs Kermack and McKendrick type assumption. Following Amundsen et al. [2], here we mathematically define the actual reproduction number, $R_a(t)$. Since the prevalence at calendar time t , $I(t)$, is given by

$$I(t) = \int_0^{\infty} i(t, \tau) d\tau \quad (52)$$

the incidence-to-prevalence ratio, $IPR(t)$, of White et al. [38] at calendar time t is

$$\begin{aligned} IPR(t) &= \frac{j(t)}{I(t)} \\ &= \frac{S(t) \int_0^{\infty} \beta(\tau) i(t, \tau) d\tau}{\int_0^{\infty} i(t, \tau) d\tau} \\ &= S(t) \int_0^{\infty} \beta(\tau) c(t, \tau) d\tau \end{aligned} \quad (53)$$

where

$$c(t, \tau) = \frac{i(t, \tau)}{\int_0^{\infty} i(t, \tau) d\tau} \quad (54)$$

which informs the infection-age distribution (or what we call *age-profile*) of infectious individuals. The actual reproduction number, $R_a(t)$, is defined as

$$R_a(t) = IPR(t)D \quad (55)$$

where

$$D = \int_0^{\infty} \Gamma(\tau) d\tau \quad (56)$$

which informs the average infectious period.

Of course, Equation (55) poses a problem for applying this simple method to HIV epidemiology. If the transmission rate $\beta(\tau)$ was independent of infection-age (and was constant b), $IPR(t)$ would be merely $bS(t)$ and thus

$$R(t) = bS(t)D = R_a(t) \quad (57)$$

Nevertheless, diseases with long generation time usually exhibits strong dependency of infectiousness on infection-age, indicating that the method might not be as useful as the cohort and instantaneous reproduction numbers. Instead, if it is the case that we have both prevalence and incidence in hand for a disease with acute course of illness, $R_a(t)$ still stands as a useful measure of transmissibility as a function of time.

5 Conclusion

In this chapter, we discussed the mathematical and statistical properties of the effective reproduction number as a function of time. We have shown that the renewal theory gives us rich analytical insights into the definition and computation of various time-dependent threshold quantities. The instantaneous and cohort reproduction numbers are explicit measures of the transmissibility, where the former informs the actual number of secondary transmissions at calendar time t , while the latter gives the average number of secondary transmissions among cohort (i.e. infecteds) who were born at calendar time t . These exactly correspond to the period and cohort total fertility rates, respectively, in mathematical demography. The difference between the two is highlighted when a specific event at calendar time t occurs (e.g. a public health intervention starts at calendar time t). Then, $R(t)$ abruptly varies with calendar time t , while $R_c(t)$ smoothly varies. For a disease with long generation time, analysis of both quantities might be called for. We have also provided analytically explicit interpretations of the incidence-to-prevalence ratio and the actual reproduction number. Although it appears that the ratio and the actual reproduction number may not be useful for a disease with long generation time (e.g. HIV/AIDS), these might be extremely useful for a disease with acute course of illness, especially when we have both prevalence and incidence in hand. Applications of the above discussed concepts are seen in other chapters in this volume, and we hope you'll enjoy our statistical approaches to various infectious diseases.

Acknowledgments The work of HN was supported by The Netherlands Organisation for Scientific Research (NWO).

References

1. Alho JM, Spencer BD (2005) *Statistical Demography and Forecasting*. Springer, New York.
2. Amundsen EJ, Stigum H, Rottingen JA, Aalen OO (2004) Definition and estimation of an actual reproduction number describing past infectious disease transmission: application to HIV epidemics among homosexual men in Denmark, Norway and Sweden. *Epidemiology and Infection* **132**:1139–1149.
3. Anderson RM, May RM (1982) Directly transmitted infectious diseases: control by vaccination. *Science* **215**:1053–1060.
4. Anderson RM, May RM (1991) *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.
5. Becker N (1977) Estimation for discrete time branching processes with application to epidemics. *Biometrics* **33**:515–522.
6. Bettencourt LM, Ribeiro RM (2008) Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE* **3**:e2185.
7. Bettencourt LMA, Ribeiro RM, Chowell G, Lant T, Castillo-Chavez C (2007) Towards real time epidemiology: data assimilation, modeling and anomaly detection of health surveillance data streams. in: *Intelligence and security informatics: Biosurveillance*. Proceedings of the 2nd NSF Workshop, Biosurveillance, 2007. Lecture Notes in Computer Science. eds F. Zeng et al. Springer-Verlag, Berlin pp. 79–90.

8. Cauchemez S, Boelle PY, Thomas G, Valleron AJ (2006) Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology* **164**:591–597.
9. Cauchemez S, Boelle PY, Donnelly CA, Ferguson NM, Thomas G, Leung GM, Hedley AJ, Anderson RM, Valleron AJ (2006) Real-time estimates in early detection of SARS. *Emerging Infectious Diseases* **12**:110–113.
10. Chowell G, Nishiura H, Bettencourt LM (2007) Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface* **4**:155–166.
11. Cowling BJ, Ho LM, Leung GM (2007) Effectiveness of control measures during the SARS epidemic in Beijing: a comparison of the Rt curve and the epidemic curve. *Epidemiology and Infection* **136**:562–566.
12. Diekmann O (1977) Limiting behaviour in an epidemic model. *Nonlinear Analysis, Theory, Methods and Applications* **1**:459–470.
13. Diekmann O, Heesterbeek JAP, Metz JAJ (1990) On the definition and the computation of the basic reproductive ratio R_0 in models for infectious diseases. *Journal of Mathematical Biology* **35**:503–522.
14. Diekmann O, Heesterbeek JAP (2000) *Mathematical Epidemiology of Infectious Diseases. Model Building, Analysis and Interpretation*. John Wiley and Sons, New York.
15. Dietz K (1993) The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research* **2**:23–41.
16. Dublin LI, Lotka AJ (1920) On the true rate of natural increase, as exemplified by the population of the United States, 1920. *Journal of American Statistical Association* **150**:305–339.
17. Ferguson NM, Donnelly CA, Anderson RM (2001) Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**:542–548.
18. Fine PE (1993) Herd immunity: history, theory, practice. *Epidemiologic Reviews* **15**:265–302.
19. Fine PE (2003) The interval between successive cases of an infectious disease. *American Journal of Epidemiology* **158**:1039–1047.
20. Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* **2**:e758.
21. Grassly NC, Fraser C (2008) Mathematical models of infectious disease transmission. *Nature Review of Microbiology* **6**:477–487.
22. Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, Wilesmith J, Woolhouse ME (2003) The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proceedings of the Royal Society of London Series B* **270**:121–127.
23. Kendall DG (1956) Deterministic and stochastic epidemics in closed populations. In: Newman P (ed) *Third Berkeley Symposium on Mathematical Statistics and Probability* 4. University of California Press, New York, 1956, pp.149–165.
24. Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics – I. *Proceedings of the Royal Society Series A* **115**:700–721 (reprinted in *Bulletin of Mathematical Biology* **53** (1991) 33–55).
25. Lloyd AL (2001a) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London Series B* **268**:985–993.
26. Lloyd AL (2001b) Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. *Theoretical Population Biology* **60**:59–71.
27. Metz JAJ (1978) The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheoretica* **27**:75–123.
28. Nishiura H, Dietz K, Eichner M (2006) The earliest notes on the reproduction number in relation to herd immunity: theophil Lotz and smallpox vaccination. *Journal of Theoretical Biology* **241**:964–967.

29. Nishiura H, Schwehm M, Kakehashi M, Eichner M (2006) Transmission potential of primary pneumonic plague: time inhomogeneous evaluation based on historical documents of the transmission network. *Journal of Epidemiology and Community Health* **60** (2006) 640–645.
30. Nishiura H (2007) Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918 to 1919. *Theoretical Biology and Medical Modelling* **4**:20.
31. Peiper O (1920) Die Grippe-Epidemie in Preussen im Jahre 1918/19. *Veroeffentlichungen aus dem Gebiete der Medizinalverwaltung* **10**:417–479 (in German).
32. Roberts MG, Heesterbeek JA (2007) Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of Mathematical Biology* **55**:803–816.
33. Smith CE(1964) Factors in the transmission of virus infections from animal to man. *The Scientific Basis of Medicine Annual Reviews* 125–150.
34. Svensson A (2007) A note on generation times in epidemic models. *Mathematical Biosciences* **208**:300–311.
35. Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* **160**:509–516.
36. Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings B* **274**:599–604.
37. Wearing HJ, Rohani P, Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLoS Medicine* **2**:e174.
38. White PJ, Ward H, Garnett GP (2006) Is HIV out of control in the UK? An example of analysing patterns of HIV spreading using incidence-to-prevalence ratios. *AIDS* **20**:1898–1901.
39. Yan P (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *Journal of Theoretical Biology* **251**:238–252.