Available online at www.sciencedirect.com
ScienceDirect
Mathematical Biosciences 208 (2007) 300–311
Mathematical Biosciences
www.elsevier.com/locate/mbs

A note on generation times in epidemic models

Åke Svensson *

Mathematical Statistics, Stockholm University and the Swedish Institute for Infectious Disease Control, SE-106 91 Stockholm, Sweden

Received 8 March 2006; received in revised form 25 July 2006; accepted 18 October 2006
Available online 9 November 2006

Abstract

The time between the infection of a primary case and one of its secondary cases is called a generation time. The distribution (and mean) of the generation times is derived for a rather general class of epidemic models. The relation to assumptions on distributions of latency times and infectious times or more generally on random time varying infectiousness, is investigated. Serial times, defined as the times between occurrence of observable events in the progress of an infectious disease (e.g., the onset of clinical symptoms), are also considered.

Keywords: Epidemic models; Generation times; Serial times; Transmission intervals

1. Introduction[1]

The reason for writing this short note is to try to provide some understanding of statistical and stochastic properties related to generation (and serial) times as used in models for spread of communicable diseases.

* Tel.: +468164569.
E-mail address: akes@math.su.se

[1] This report has been financially supported by the DG Sanco project MODELREL and the DG Research project INFTRANS. The author is member of the Stockholm group of epidemic modelling (www.s-gem.se). The author has benefitted from comments by Jacco Wallinga and Tom Britton on an earlier draft of this paper. The anonymus referees are also thanked for valuable comments that has improved the paper.

0025-5564/$ - see front matter
doi:10.1016/j.mbs.2006.10.010

Traditionally a main input to epidemic models is assumptions on the distributions of individual latency times, infectious times and infectiousness. These concepts are all related to the progress of an infection within an infected individual. They describe the natural history of an infection and should, in principle, be observable from studies of single infected persons. We will call these aspects of the infection inter-individual. Medical handbooks, that are focused on clinical aspects of infections and infectious diseases provide some information about these. In addition to inter-individual aspects of infections an epidemic model also has to describe how, why and when infections are transmitted between individuals.

Lately the notion of generation times has been popular and now is an important concept in models of the SARS epidemic and pandemic influenza [1–3]. It has been used, e.g., for estimating the basic reproductive number (R_0) for the 1918 pandemic influenza by relating assumptions on the mean generation time to observations on doubling times of the epidemic curve in an early stage of the epidemic. Generation times reflect the time between the infections of a primary case and its secondary cases. In this paper we use the term generation time for the time between the infection of a primary case and a secondary case. By serial time is meant the time between two similar events (e.g., onset of disease) in the progress of the infection of a primary and secondary case.

A number of other terms, such as generation interval and transmission time or transmission interval are also in use. A general discussion of the concept which includes a survey of several infections is given by [4]. Fine uses the term transmission interval.

Generation times are related to the duration of the latent and infectious periods as well as time-variation in the infectivity that an infected person emits. The concept of a generation time involves a primary case and a secondary case. Thus, it can only be understood in relation to a model for transmission between individuals. One way to give generation time a well-defined, and general meaning is to consider it in the most simple conceivable model. This would be the model where infectious contacts are taken according to a homogeneously mixing assumption, i.e., an infected person is equally likely to contact any other member of the population. The following discussion will assume such a simple model.

We will illustrate that the distribution of generation times will depend on how it is defined. We will compare two different definitions. We will stress that it is important to use an appropriate definition in order not to introduce biases in estimates of mean generation times and other quantities related to its distribution.

The derivation will be based on a random measure that describes how infectiousness of an infected person changes over time (see Section 3). Infected persons may have different potential to spread the infection. In the simple model, with homogeneous mixing, the variability infectivity will only involve differences between the inter-individual progress of an infection. In more complex models it will also involve differences in how individuals spread the infection to other persons (see also Section 8).

The possibility to transmit an infection depends on the proportion of susceptible individuals in the population. This proportion may change during an epidemic, due to the possibility that infected persons may obtained immunity. This will influence observed generation times, in particular in the intense period of an emerging epidemic. For this reason we will restrain most of the analysis to the beginning of an epidemic outbreak or to situations where an infected individual's infectivity is not influenced by changes in immunity in the surrounding population.

However, in Section 6 we will consider how generation times may change throughout an epidemic outbreak.

The following discussion is based on a general model for infectious spread (Section 3). In Section 4 results on distributions and moments for generation times are derived. These results are applied to some special models in Section 5. Some remarks regarding the possibility to observe generation times in studies of real epidemic outbreaks are made in Section 6. Some comments of the more general concept of serial times are given in Section 7.

A large part of the discussion that follows is based on well-known results from the theories of demography [5] and branching processes [6–8]. For formal proofs of technical results we will refer to these texts. We will leave out standard derivations of distribution functions and means.

2. Generation times

Definitions of generation time formulated in the literature, are in most cases essentially verbal. They may look common-sense and straightforward, but it is not obvious how to translate them to mathematical well-defined concepts.

We quote some definitions:

- '*The sum of the average latent and the average infectious period is referred to as the average generation time of the infection* [9]'.
- '*The time between the appearance of similar symptoms (e.g., rash, coughs) in successive generations* [10]'.
- '*The generation interval is the time from symptoms onset in a primary case to symptom onset in a secondary case* [1]'.
- '*The average time taken for the secondary cases to be infected by a primary case* [11]'.

These definitions are all made in the context of particular assumptions of duration of latency, infectious times and transmission. One purpose of this paper is to point out that they should not be read as definitions that can be generally applied in any situation. With the exception of the definition of [9] none of these definitions is stated in mathematical terms.

It should also be observed that two of the quoted definitions [9] and [11] relate to times between successive infections and the other [10] and [1] refer to any identifiable event in the progress of an infection. This is a more general definition and we will here refer to such intervals as serial times.

From a statistical viewpoint there are important differences between generation and serial times. In the following we will assume that the transmission of infections is homogeneous in time (and space). As will be seen generation time distributions can be derived from how infectiousness progresses in a single individuals. However, serial time are mixtures of processes that relates to two infected individuals. The serial time between onset of disease in a primary and secondary case depends on the infectors incubation time and his/her progress of infectiousness as well as on the incubation time of the infected. Some consequences of this difference are discussed in Section 7.

In most models (and most real applications) there are infected persons that causes no secondary cases. It is possible to define generation time conditional on that there is at least one secondary

case. Since generation times connected to the same primary case may be dependent it is important to consider the simultaneous distributions.

It turns out that it is a better to define 'backward' generation times and relate the term to the secondary cases rather than to the primary cases. In that case the generation time for any infected person is the time that has evolved since his infector was infected. This may seem like a trivial rewording, but it is not and it has essential implications.

3. A general formulation of an epidemic model

The first task when formulating an epidemic model is to consider how many persons an infected person may infect and when the infections may occur. In order to do this we will consider two random entities, λ and K. Here λ is a non-negative (random) number that decides the 'total amount of infectivity' spread by an infected person, K is a (random) positive measure, with total mass 1, defined on $[0,\infty]$ which measures how the infectiousness is distributed in time. The assumption that K is a random measure implies that it is not the same for each infected individual. It is chosen (independently for all persons) according to a distribution on all possible measures.

A basic assumption is that for a given individual the number of infectious contacts in the time interval $I=(a,b)$ after infection, is Poisson distributed with mean $\lambda K(I)$ (conditional on λ and K). Let $K(t)=K(0,t)$. For simplicity we assume that there exist a density so that

$$K(I)=\int_I k(s)\,\mathrm{d}s. \tag{3.1}$$

This implies that, given K, the secondary infections occurs according to a time inhomogeneous Poisson process with intensity $\lambda k(t)$. An important result from the theory of Poisson process is that a time transformation of this process with K^{-} yields a Poisson process with constant intensity equal to λ.

An infectious contact results in a secondary case if the contact is taken with a susceptible person. In a simple homogeneous model the contacted persons are chosen randomly in the population. As explained above this part of the modelling will not be the concern of the present discussion, that focuses on the inter-individual properties of the infection.

The basic reproduction number R_0, that usually is defined as the mean number of infections caused by an infected person in a totally susceptible population, is an important parameter in most epidemic models. In this case it is straight-forward to prove that

$$R_0=E(\lambda). \tag{3.2}$$

This follows since K is assumed to have total mass 1.

4. Simultaneous distributions of generation times

Assume that a Poisson process, with constant intensity, has m events in the interval $[0,\lambda]$. It is well-known that conditional on m, the times of the events are distributed as $(\lambda U_1,\ldots,\lambda U_m)$ where

$U_1, \ldots, U_m$ are independent uniformly distributed random variables. (Observe that we have not assumed that the U_i's are ordered).

This result can be translated, by using the time transformation K^{-1}, to find the distribution of the times $(\tau_1, \ldots, \tau_m)$ of infections of the secondary cases. It follows that, conditionally on K, the times $\tau_1, \ldots, \tau_m$ are distributed as

$$(K^{-1}(U_1), \ldots, K^{-1}(U_m)). \tag{4.1}$$

Thus conditional on the number of secondary cases and K

$$Pr(\tau_1 \leqslant s_1, \ldots, \tau_m \leqslant s_m \mid m, K) = \prod_{j=1}^{m} K(s_j). \tag{4.2}$$

Given the actual random measure K the times are independent. If we loosen this restriction the times are no longer independent and

$$Pr(\tau_1, \leqslant s_1, \ldots, \tau_m \leqslant s_m \mid m) = \int \prod_{j=1}^{m} K(s_j)\, \mathrm{d}P(K). \tag{4.3}$$

Thus times till infections generated by the same infector are exchangeable but not independent. However, they will all have the same mean (regardless of m).

$$E(\tau \mid K) = \int_0^\infty a\, \mathrm{d}K(a) = \int_0^\infty ak(a)\, \mathrm{d}a. \tag{4.4}$$

The overall mean, integrating over the distribution of K is

$$E(\tau) = \int \left(\int_0^\infty ak(a)da \right) dP(K). \tag{4.5}$$

We have considered the time between a primary and one of its secondary cases, given that such exists.

In order to understand the statistical properties of observations from an epidemic process we will consider two kinds of observations that will yield different distributions related to K in different ways.

First define a random variable, T_p, where p stands for primary, in the following way: First choose at random a infected person that spreads the infection to at least one other person, then choose one of its secondary cases at random. The time between the infection of the primary and the secondary case is T_p.

We will denote the distribution function of T_p by F_p and the density by f_p. From the discussion above it follows that

$$F_\mathrm{p}(a) = E(K(a)) = \int K(a)\, \mathrm{d}P(K). \tag{4.6}$$

and

$$f_\mathrm{p}(a) = E(k(a)) = \int k(a)\, \mathrm{d}P(K). \tag{4.7}$$

Another, perhaps more natural, possibility is to chose a primary case proportionally to the number of secondary cases, and then consider the distribution of the time till a secondary case. We denote the corresponding random variable by T_s, where s stands for secondary, and its distribution function by F_s. The density function of T_s is $f_s = F_s'$.

The expected number of secondary cases will be proportional to λ. Thus

$$F_s(s) = \frac{E(\lambda K(s))}{E(\lambda)}, \tag{4.8}$$

and the mean

$$E(T_s) = \frac{E(\lambda \int_0^\infty ak(a)da)}{E(\lambda)}. \tag{4.9}$$

The random variable T_s can be also be obtained by choosing a case at random and consider how long its infector has been infected.

Observe that the means of T_p and T_s are based on the same distributions, K. However, T_s is the result of a size-biased sampling procedure, where primary cases with many secondary cases are given a greater weight.

Obviously it is important to understand the difference between these two distributions when doing statistical analyzes based on observations from a real epidemic. It is also necessary to understand how the different distributions are related to other interesting features of an epidemic. It will be pointed out below, that the distribution of T_s is the one that should be used when calculating the Malthusian parameter, that decides, e.g., the initial doubling speed of an emerging epidemic.

5. Some special models

5.1. Infectiousness constant in time

A common assumption in epidemic models is that each individual has a random latent time, X, a random infectious time Y, and a random infectivity γ. The infectivity is assumed to be constant during the infectious time, with a level given by γ. For simplicity we assume, as often is done, that the three random variables involved are independent. Let g denote the density of the random variable X and h the density of Y. We will also use the distribution function of Y that is denoted by H.

We can rewrite these assumptions using the notation in the previous section:

$$\lambda = \gamma Y, \tag{5.1}$$

and

$$k(t) = \begin{cases} 1/Y & \text{if } X < t \leqslant X + Y, \\ 0 & \text{otherwise,} \end{cases} \tag{5.2}$$

or with an alternative expression

$$k(t) = \frac{I(X < t \leqslant X + Y)}{Y}. \tag{5.3}$$

It follows directly that

$$R_0 = E(\gamma)E(Y). \tag{5.4}$$

Simple calculation yields that the mean of T_p is

$$E(T_p) = E(X) + E(Y)/2. \tag{5.5}$$

Inserting (5.3) into (4.6) we find that the density of T_p equals $g*h_p$ (here $*$ stands for convolution) where

$$h_p(t) = \int_t^\infty \frac{h(a)}{a}\,da. \tag{5.6}$$

Inserting (5.3) into (4.8) we find that the density of T_s is $g*h_s$ where

$$h_s(t) = \frac{1 - H(t)}{E(Y)}. \tag{5.7}$$

Thus

$$E(T_s) = E(X) + \frac{E(Y^2)}{2E(Y)}. \tag{5.8}$$

As an example we can consider the possibility that if Y is gamma-distributed with parameters $(\alpha, \alpha/E(Y))$. All these distributions have the same mean. However, the variance equals $\mathrm{Var}(Y) = (E(Y))^2/\alpha$. Simple calculations yield that

$$E(T_s) = E(X) + E(Y)\frac{\alpha + 1}{2\alpha}. \tag{5.9}$$

If Y has an exponential distribution, i.e. $\alpha = 1$, then h_s is the density of an exponentially distributed random variable and

$$E(T_s) = E(X) + E(Y). \tag{5.10}$$

Observe that the mean of T_s is decreasing in α and that $E(T_s) > E(T_p)$ for all α.

5.2. Time-varying infectiousness

For many infections it is known that the infectiousness varies during the infectious period. However, it is often difficult to get a realistic model for this variation. To illustrate the possible effect of such variations we will here consider some simple cases. We will assume that the both the total infectiousness, i.e. λ, and the random measure K depends on a real-valued random variable, γ. That is, each individual has its own value of γ, and spreads the total infectivity λ_γ according to the measure K_γ.

In simple cases it is possible to find relations between T_p and T_s. Such a case occurs when both λ_γ and $K_\gamma(t)$, for all t, are monotone in γ. Assume, e.g., that λ_γ is increasing and $K_\gamma(t)$ is decreasing in γ. Then we can apply the well-known inequality ([12], pg. 43) that states that two increasing functions of the same random variable are positively correlated. From the expressions (4.6) and (4.8), it follows for all t that

$$F_s(t) = \frac{E_\gamma(\lambda_\gamma K_\gamma(t))}{E_\gamma(\lambda_\gamma)} \leqslant E_\gamma(K_\gamma(t)) = F_p(t). \tag{5.11}$$

This implies that the random variable T_s, in this case, is stochastically larger than T_p. As a consequence $E(T_s) \geqslant E(T_p)$.

6. Generation times observed from epidemic outbreaks

Statistical analysis of generation times that are derived from observations from an epidemic outbreak requires some care. We can distinguish two different kind of problems. The first occurs if the epidemic is observed during a fixed time at the start. The second problem is associated with observations from the intense part of the outbreak.

If we restrict observations to the start of epidemic the number of primary cases will grow at an exponential rate. This means that at any time a large proportion of the infectors has not been able to spread all their infectivity. Consequently there is an over representation of early infected secondary cases. [13] discuss how to make a statistical analysis that takes this into account. It is still possible to calculate an approximate distribution of the observed ages of the cases that has occurred. We will here quote, without proof, a result from demography which states that the density of the observed generation times, in a branching process, is $e^{-ra} f_s(a)$ and the mean maternal age is

$$E(\lambda) \int a e^{-ra} f_s(a)\, da, \tag{6.1}$$

where r is the Malthusian parameter that solves the equation

$$\int e^{-ra} f_s(a)\, da = 1/E(\lambda), \tag{6.2}$$

(cf. [6] chapter 8.4).

Next, we consider the possibility that secondary cases may be lost due to immunity caused by previous infections. This will occur if we have a large epidemic outbreak that takes place on a time scale such that we may disregard demographic changes (due e.g., to deaths and births) of the population. We will also assume that infection causes immunity through the remaining progress of the epidemic. In short we will consider a fast spreading epidemic in a closed population where all are susceptible before the introduction of the infection. If this is the case an epidemic will go through several stages. It will start slowly (due to few infectors), then go into an intense stage where most of the infectious spread happens, and finally die out slowly (due to few susceptible). A consequence is that observed generation times will tend to be shorter in the intense stage of an epidemic. Let $S(t)$ be the number of susceptible individuals in the population at time t. During the start up and dying out stages $S(t)$ will change slowly in time, whereas during the intense part S will change fast, i.e. S' will be negative.

Let us, for the moment, restrict attention to cases infected by a person infected at time z. Such a person will spread the infection according to the measure defined by

$$K_z(t) = \frac{\int_0^t S(z+a)k(a)\, da}{\int_0^\infty S(z+a)k(a)\, da}, \tag{6.3}$$

with

$$\lambda_z = \lambda \int_0^\infty S(z+a)k(a)\,\mathrm{d}a. \tag{6.4}$$

From (4.8) it follows that the distribution of the generation time associated with a random chosen of these secondary cases is

$$F_{sz} = \frac{E(\lambda \int_0^t S(z+a)k(a)\,\mathrm{d}a)}{E(\lambda \int_0^\infty S(z+a)k(a)\,\mathrm{d}a)}. \tag{6.5}$$

According to the same inequality as used to prove (5.11) we find that

$$\int_0^\infty S(z+a)I(a \leqslant t)k(a)\,\mathrm{d}a \geqslant K(t) \int_0^\infty S(z+a)k(a)\,\mathrm{d}a, \tag{6.6}$$

since both $S(z+a)$ and $I(a \leqslant t)$ are decreasing in a. This implies that

$$F_{\mathrm{sz}}(a) \geqslant F_{\mathrm{s}}(a), \tag{6.7}$$

for all a. The generation time distribution for infections considered during the entire process are weighted combinations of the conditional generation time distributions with weights proportional to the number of secondary cases infected by an infector infected at time z. The inequalities (6.7) are valid for all z. Thus generation times considered during the progress of the epidemic will be stochastically smaller than the generation times at the start.

Heuristically, it will be clear that the generation times will be smallest during the most intense part of the epidemic where the slope of the curve $S(t)$ is steepest. If S decreases fast enough it may substantially reduce the possibilities for late infections and thus favor short generation times. If this is an effect that has to be considered or not depends on the time scale at which the epidemic spreads and the duration of the infectivity of an infected.

It can be observed that the effect on the generation times will decrease after the peak of the epidemic, when $S(t)$ again changes slowly (but on a lower level).

The above discussion concerns large epidemic outbreaks that occur during a relatively short time span. The time variation in the distribution of the generation time will not take place in an 'endemic' situation where the proportion of susceptible in the population is approximately stable over time.

7. Serial times

As an alternative to generation times it has been suggested to consider times between observable events in primary cases and secondary cases. Fig. 1 illustrates this. In the figure the time of infections as well as onset of symptoms of some kind are indicated. Let U denote the time between infection and the observable event, and let V denote the difference between the generation time and the observable event. Following the notation of the figure

$$T_g = U_0 + V_0. \tag{7.1}$$

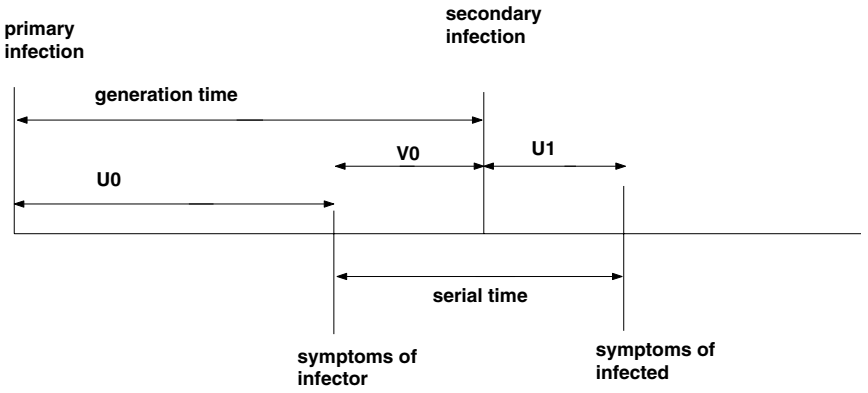

Fig. 1. Generation time and serial time (observe that V_0 might be negative).

However, the serial time is

$$S = V_0 + U_1. \tag{7.2}$$

It is clear that $E(T_g) = E(S)$ if U_0 and U_1 have the same distribution. However, in general, it is possible that T_g and S do not have the same distribution. It is natural to assume that U_0 and V_0 are dependent random variables since they are attached to the same individual and that U_0 and U_1 are independent since they come from two different individuals. If this is the case the distribution of T_g and S will not be the same. This is illustrated already by considering the variances:

$$\mathrm{Var}(T_g) = \mathrm{Var}(S) + 2\,\mathrm{Cov}(U_0, V_0). \tag{7.3}$$

It may also be observed that observations of S may even be negative.

8. Final remarks

As stated in the introduction the progress of an epidemic depends on both inter-individual properties related to infections of single individuals and how the infection is transmitted between members of the population. The discussion, in this paper, is based on a measure K that describes how the infectivity changes in time, and a number λ that measures the total infectivity. They are assumed to be random since it is not realistic that they are the same for all persons. The variability of infectiousness between individuals is, in fact, an important factor for epidemic spread.

To some extent the infectiousness of infected persons can be studied separated from transmission, e.g., by considering virus shedding and individual behavior after infection. Together with a model for transmission this will define the epidemic process. We have here, for simplicity mainly considered homogenously mixing models. For such models the start of an epidemic, is well approximated by a branching process. If the process takes off, i.e. in case the epidemic does not die out early, then it is driven by the measure

$$\mu(t) = E(\lambda K(t)), \tag{8.1}$$

(cf. e.g., [7]).

The speed at which the number of infected grows initially is exponential $\exp(rt)$, where r is the Malthusian parameter that solves the equation:

$$\int_0^\infty \exp(-ra)\,\mathrm{d}\mu(a) = 1, \tag{8.2}$$

(cf. (6.2)).

The inter-individual progress of an infection and the transmission of infections can be studied by observing times between successive infections. We have in this paper suggested a mathematical precise and understandable definition of the concept of generation time. The definition is based on the random variable T_s and its distribution F_s defined in Section 4. It is closely related to the measure μ and other entities that sums up most of the important aspects of how the epidemic will develop initially.

From μ we can derive both the basic reproduction number $R_0 = \mu(\infty)$, and the distribution function of T_s, since

$$F_s(t) = \frac{\mu(t)}{R_0}. \tag{8.3}$$

If we consider epidemic models where the assumptions of homogeneously mixing is not valid the situation is more complicated. A common class of models can be built by dividing the population into subgroups with different transmissibility between groups (e.g., age groups).

In case where the distribution of (λ, K) differs between the groups but when all infected transmit the infection homogeneously over the entire population regardless of group, the discussion above can be straightforwardly applied. The differences between the groups are the accounted for by considering the mixture of the distributions of (λ, K). Of course, it may be interesting to consider different generation times for the different groups.

If the potential for spreading the infection differs between groups, i.e. the probability for transmission depends on who infects whom, the situation is more complicated. In that case we can no longer depend on the regeneration principle, that is the base for the results of elementary branching process theory. Regeneration means that all infected person start similar chains of infections, following stochastically identical rules. Of course, it is still possible to analyze the model, but care has to be taken to when taking spread of infections between groups into account.

References

[1] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, Am. J. Epidemiol. 160 (6) (2004) 509.
[2] C.E. Mills, J. Robins, M. Lipsitch, Transmissibility of 1918 pandemic influenza, Nature 432 (2004) 904.
[3] N. Ferguson, D. Cummings, S. Cauchemez, et al., Strategies for containing an emerging influenza pandemic in southeast asia, Nature 437 (7056) (2005) 209.
[4] P. Fine, The interval between successive cases of an infectious disease, Am. J. Epidemiol. 158 (11) (2003) 1039.
[5] N. Keyfitz, H. Caswell, Applied Mathematical Demography, in: Statistics for Biology and Health, third ed., Springer, Berlin, 2005.
[6] P. Jagers, Branching Processes with Biological Applications, John Wiley and Sons, New York, 1975.
[7] M. Kimmel, D. Axelrod, Branching processes in biology, Interdisciplinary Applied Mathematics, vol. 19, Springer, Berlin, 2002.

[8] P. Haccou, P. Jagers, V. Vatutin, Branching Processes: Variation, Growth, and Extinction of Populations, Volume of Cambridge Studies in Adaptive Dynamics, Cambridge University, Cambridge, 2005.
[9] R. Anderson, R. May, Infectious Diseases of Humans, Oxford University, Oxford, 1991.
[10] J. Giesecke, Modern Infectious Disease Epidemiology, second ed., Arnold, London, 2002.
[11] R. Anderson, C. Fraser, A. Ghani, et al., Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic, Phil. Trans. Roy. Soc. London. B 359 (2004) 1091.
[12] G. Hardy, J. Littlewood, G. Pólya, Inequalities, Cambridge University, Cambridge, 1934.
[13] S. Cauchemez, P.-Y. Boelle, C. Donnelly, et al., Real-time estimates in early detection of SARS, Emerg. Infect. Dis. 12 (1) (2006) 110.