Mathematical Biosciences 223 (2010) 24–31

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Some model based considerations on observing generation times for communicable diseases

Gianpaolo Scalia Tomba [a,*], Åke Svensson [b], Tommi Asikainen [c], Johan Giesecke [c]

[a] Dept. of Mathematics, University of Rome "Tor Vergata", Italy
[b] Dept. of Mathematics, Stockholm University, Sweden
[c] ECDC, Sweden

ARTICLE INFO

Article history:
Received 3 March 2009
Received in revised form 11 October 2009
Accepted 16 October 2009
Available online 23 October 2009

Keywords:
Epidemic model
Statistics
Generation time
Serial interval

ABSTRACT

The generation time of an infectious disease is usually defined as the time from the moment one person becomes infected until that person infects another person. The concept is similar to "generation gap" in demography, with new infections replacing births in a population. Originally applied to diseases such as measles where at least the first generations are clearly discernible, the concept has recently been extended to other diseases, such as influenza, where time order of infections is usually much less apparent.

By formulating the relevant statistical questions within a simple yet basic mathematical model for infection spread, it is possible to derive theoretical properties of observations in various situations e.g. in "isolation", in households, or during large outbreaks. In each case, it is shown that the sampling distribution of observations depends on a number of factors, usually not considered in the literature and that must be taken into account in order to achieve unbiased inference about the generation time distribution. Some implications of these findings for statistical inference methods in epidemic spread models are discussed.

1. Introduction

In this paper, we discuss statistical and stochastic properties related to generation times and serial intervals as used in models for spread of communicable diseases. Both concepts concern the time between the infection of a primary case and a secondary case. The term generation time refers to the actual time (although usually unobservable) between the moments of infection, while serial interval refers to the time between two similar, well defined, observable events, such as appearance of symptoms.

The notion of generation time has seen a recent increasing use in descriptions and models of epidemic processes. It plays an important role in analyses of the SARS epidemic [1] and in planning for pandemic influenza [2]. A general discussion of generation times and serial intervals which includes a survey of several infections is given by Fine [3]. Various aspects and uses of the concept have recently been discussed [4–6].

In this paper, we will focus on the effects of different observation schemes on the resulting sampling distributions of generation times and serial intervals. Statistical methods both to infer important parameters related to the transmission of infections and to make predictions from the early development of an epidemic are also considered. The simple stochastic SIR model is used to highlight the special features and problems related to the studied observation schemes. We then discuss how conclusions from the simple model will be relevant in more complex, and realistic, situations.

The concept of generation time and much of the mathematical treatment is modelled on demography. Generations in demography are defined by parents and children. Births are taken to correspond to transmission of the infection, a secondary case of infection is seen as a descendant of a primary case. This makes it possible, in theory, to use much of the advanced theory developed in demography [7,8] and, more generally, in the theory of branching processes [9]. However, one important difference that has to be considered is that the dynamics of an epidemic is influenced by the depletion of susceptibles. This phenomenon has no natural counterpart in the demographic theory. Another important difference is that infection events are essentially not observable, as opposed to births, leading to the use, in practice, of serial intervals as proxies for generation times.

Theoretical arguments are illustrated and supported by simulation results. All simulations have been performed using an individual (agent) based program written in C, where individual infection chains and times are identifiable and thus available for analysis.

* Corresponding author. Address: Dept. of Mathematics, University of Rome "Tor Vergata", Via Ricerca Scientifica 1, 00133 Roma, Italy.
E-mail address: scaliato@mat.uniroma2.it (G. Scalia Tomba).

0025-5564/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2009.10.004

2. Some general considerations about generation times

As in most statistical problems, it is important to distinguish between observations and models (statistical, probabilistic, analytical) for observations.

Seen as observations, generation times are, at least conceptually, i.e. disregarding the practical problem of observing them, well defined if we assume that disease is transmitted from person to person with well defined moments of infection. Abstractly, this means that the disease spread can be seen as a graph with nodes, representing individuals, that also contain information at least about time of infection; it must also be assumed that there is a specific infector for each infectee, i.e. the graph is a tree. If the spread is considered within a given population, all cases except the introductory ones will have an infector but some infected individuals will have no secondary cases (infectees).

To any given outbreak, there corresponds a set of generation times, where each time is defined by the pairs infector–infectee and their respective infection times. It should be noted that generation times genuinely depend on two individuals, in the sense that they cannot be measured (are not defined) on a single isolated infected individual. This is different from e.g. infectious periods, which can be measured by observing the timing of secondary cases but also, in theory, by a reliable proxy such as viral shedding, that can be measured in the primary case independently of the presence of other individuals. It should also be noted that generation times are lengths of certain intervals and thus have no "natural" time coordinate of occurence. This means that there is a choice about when, during the course of the outbreak, a given generation time can be considered to have occured.

In simple disease spread models, it is usually assumed that, say, infectious periods are random but that periods belonging to different individuals are independent and identically distributed. This is an assumption of the model, not a result. This makes it possible to talk about "the distribution of the infectious period" in an unequivocal way. Generation times are "results", with properties to be derived from other assumptions in the model. We will show that generation times are not, in general, identically distributed during an outbreak and that the statistical distribution of observations critically depends on how the observations have been collected. Thus, statements about the probability distribution of generation times need to be qualified in a meaningful way, detailing which generation times are intended and, as we will see later, how they have been observed.

Some of the results about the effects of different observation schemes are intimately linked to the statistical concept of "size biased sampling" [10] or, equivalently, to the probabilistic "waiting time paradox" in stochastic process theory [11].

3. Generation times and size biased sampling

There are two different definitions of generation time in demography that are relevant to the infectious disease situation [5,7,8]. Below, we summarize the main notions and relations. If we denote the probability of survival at age a by $l(a)$ (= probability that the infectious period is longer than a) and the childbearing intensity at age a by $m(a)$ (= intensity of infectious contact at infection age a), we first define the intrinsic growth rate of the population, also called Malthusian parameter and denoted by r, as the unique value of $r \geqslant 0$ satisfying

$$1 = \int_0^\infty e^{-ra} l(a)m(a)da.$$

We also note that

$$\int_0^\infty l(a)m(a)da = R_0,$$

the expected total offspring of an individual (= basic reproduction number R_0 in a totally susceptible population, in the disease spread setting). We then define the "cohort generation time distribution" as

$$g(a) = \frac{l(a)m(a)}{R_0}$$

and the "average cohort generation time" as the expected value in this distribution, i.e.

$$T_c = \int_0^\infty ag(a)da$$

usually explained as "the average age of mothers calculated over all births in a cohort". This cohort generation time distribution seems to be the de facto accepted concept in the theory of disease spread models, also because of the useful relation between Malthusian parameter, generation time distribution and R_0 given by (see [4] for a review of related results)

$$\frac{1}{R_0} = \int e^{-ra} g(a)da.$$

There is also the concept of "average age at childbearing in the stable age distribution of an exponentially increasing population", defined as

$$T_A = \int_0^\infty ae^{-ra} l(a)m(a)da$$

representing the "average age of a mother chosen from the stable age distribution in an exponentially growing population (with intrinsic growth parameter = r)". It is also well known that T_A and T_C do not in general agree and that, in fact, $T_A \leqslant T_C$.

With the infection spread interpretation in mind, it is quite easy to see that T_A might be applicable while the disease spread experiences an exponential growth at constant rate, which, as well known from theory, only happens during the initial phase of an epidemic in a large population.

It is usually not realized that there is a statistical implication of both these definitions, namely that the generation time is to be observed starting from (in infectious disease terminology) an infectee, finding the related infector and finally measuring the interval between their respective moments of infection. This is different from following an infected individual and observing the time points of his secondary cases, if any. Infectee based observations and averages allow infectors with long infectious periods and thus many secondary cases to appear several times in the calculations (one for each infectee), while, in an infector based perspective, every infector only appears once in the global average, if an average generation time has been calculated among his secondary cases. This is the sampling effect known as "size biased sampling", namely that a quantity does not appear with the "natural" frequency, but with a frequency modified by weighting the natural frequencies with the size of the values to be sampled, thus putting more emphasis on larger values.

Let us start this investigation by studying generation times from the point of view of a newly infected individual and what this individual should expect, in terms of secondary cases, during his infectious period. We will refer to this calculation as the "single infector" model.

Because of its simplicity, we will initially consider the simple Markov SIR model, i.e. we assume that there is no latent period, that the infectious period is exponentially distributed with expected length $1/\mu$ and that infectious contacts occur according to a Poisson process with constant intensity β during the infectious

period. These contacts lead to secondary cases if the contacted individual is susceptible, which has a given probability (depending on the outcome of the outbreak up to the considered time point) depending on time. We will first assume that this probability remains constant during the infectious period of the considered individual. The following well known results will then hold:

(a) given the infectious period T, contacts are created according to a Poisson process on $[0,T]$, thus S = number of contacts will be Po(βT);
(b) given $S = k$, the conditional distribution of T is Gamma distributed with parameters $k+1$ and (intensity) $\mu+\beta$;
(c) $P(S=k) = \mu\beta^k/(\mu+\beta)^{k+1}$ (i.e. a geometric distribution);
(d) $E(T|S>0) = 1/(\mu+\beta) + 1/\mu$;
(e) given T and $S = k > 0$, the k time points of events are independent and identically distributed with uniform distribution on $[0,T]$, independently of the level of β;
(f) let G be a randomly chosen point among these time points. Then $E(G|S>0) = E(T|S>0)/2$.

Considering points (d) and (f), we see that the average generation time in this model is slightly more than half the expected length of the infectious period (the exceedance comes from conditioning the infectious periods to contain at least one secondary case).

In order for contacts to become cases, the contacts will be thinned with the probability that a randomly chosen person at the time the contact is made is susceptible. If this probability can be assumed to be constant in $[0,T]$, the thinning reduces to changing β but otherwise leaving the intensity constant, and thus the conditional distribution (see point (e) above) of infection times remains the same, i.e. uniform on $[0,T]$; if however the probability of finding a susceptible decreases markedly in $[0,T]$, then the conditional distribution will be "tilted towards 0", since generation times will then tend to occur more frequently towards the start of the infectious period than towards the end. The shortening increases with increasing slope, but under the assumption of linear decrease of the thinning probability, amounts to at most 1/3 of the value that is obtained under the assumption of constant thinning. A more formal treatment of the shortening phenomenon is possible [4].

However, if we assume that the intervals above are chosen with weights proportional both to the number of secondary cases that they contain, i.e. based on infectees instead of infectors, and to the original geometric probabilities, then the distribution of T, which now by definition always contains at least one secondary case, becomes Gamma distributed with parameters 2 and μ, and the expected value of a generation time, still half of the expected length of an infectious period (but, in this case, an infectious period selected according to the above rule), becomes exactly $1/\mu$, i.e. the same as the expected value of an infectious period (c.f. definition of average generation time in Anderson and May [12]), and almost twice what is expected when sampling infectors instead of infectees. Again, in practice, the reduction of susceptibles may result in slightly shorter generation times.

4. Sampling and representing generation times during an epidemic outbreak in a large population

Let us now show how the infector and infectee sampling schemes related to the "cohort" definition of generation time as well as the definition related to exponential growth all naturally appear when sampling and/or representing the generation times that occur during a large epidemic outbreak in different ways. In addition, we will also find that the dynamics of disease spread have a modifying effect on generation times.

While remaining in the simple Markov SIR framework, we now distinguish between two different ways of observing generation times, namely the "forward" observation, i.e. following the infective career of an infective and, in case there is at least one secondary case, recording the times of his secondary cases, and the "backward" observation, i.e. starting from the moment of infection of an infective and finding the time of infection of his infector. In both cases, the generation time will be the difference between time of infection of infectee and infector, but we have the choice of the reference time point at which we consider the generation time to have occured or to have been observed. There is also the option, in the case of forward observations, to treat each single generation time related to the same infector as a separate data points or to summarize these single values by their mean and then consider only one data point per infector. The sampling scheme that comes closest to the cohort based generation time distribution definition is the forward sampling scheme where each single generation time related to the same infector is treated as a separate data point. The forward sampling scheme where the mean generation time is calculated within each infector's set of generation times and then used as a single data point per infector corresponds to the single infector model treated above, which yields smaller averages because of the absence of size bias sampling effects. The backward sampling scheme represents contact tracing data where an infector is identified starting from a case.

In the following Figures, the generation times recorded in a simulation of the basic stochastic model with $\mu = 1$ and $\beta = 2$, i.e. $R_0 = 2$, in a population of 10,000 individuals, given a large outbreak, are shown and summarized in different ways. The initial number of infectives is 1, the final size is 7888, close to the expected final fraction, the outbreak lasts 19.6 time units (1 time unit = 1 average infectious period), the "large" part of the epidemic occurs between times 4.5 and 9.2 (times at which the # of susceptibles is 9000 and 3000, respectively), the maximal incidence of infections occurs at time 6.6, the maximal number of active infectives is recorded at time 7.4, the last infection occurs at time 16.5 and the last removal at time 19.6. Out of the 7888 infectees, only 3895 were also infectors.

Forward observation of generation times is subject to the changing numbers of susceptibles during different phases of an epidemic. As discussed above, the observations can either be considered as separate data points for each infection or be averaged for each infector before final representation and averaging. The results are shown in Figs. 1 and 2. In the first graph, data points are individual infections and each generation time is plotted at the moment of infection of the infector. This observation scheme corresponds to the cohort generation time. The average level is approximately equal to 1, as predicted by the analysis of the cohort based generation time, because each infector is plotted with a multiplicity equal to his number of secondary cases since one point for each infectee has been drawn. However, closer examination (Fig. 2) shows that the average generation time is always slightly less than predicted by the simple theory because the proportion of susceptibles is always declining to some extent, the effect being most marked near the peak of the epidemic, when the decrease of susceptibles, i.e. incidence of infection, is at its maximum [4]. This moment is somewhat earlier than the peak of infectives, at least in the standard deterministic model, but also in the simulation represented in Figs. 1–3. This shortening has also been observed by Kenah et al. [6], who however attribute it to the effect of a large number of infectives. There is a subtle difference between the two explanations, since, according to the decrease in susceptibles approach, generation times would be expected to become smaller, even in the absence of large numbers of infectives, if some other event sharply decreased the number of susceptibles, such as a quick vaccination of part of the population.

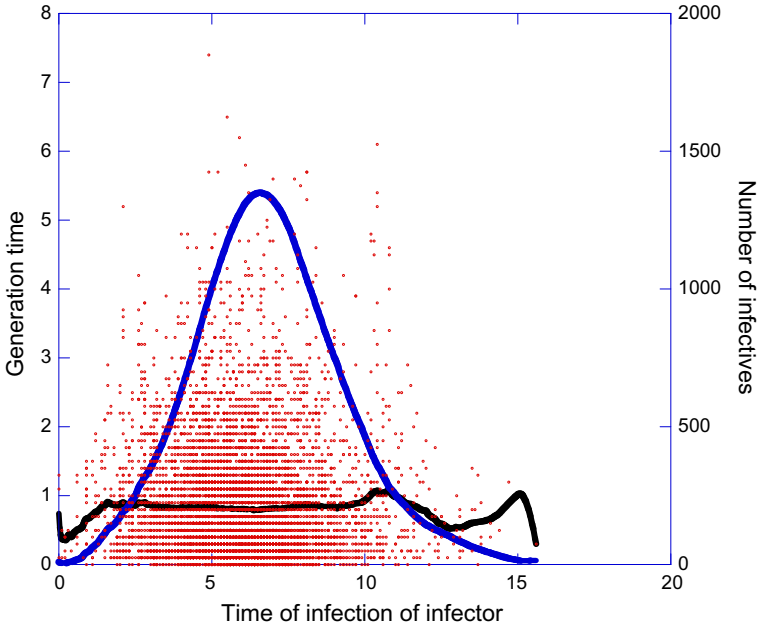

Fig. 1. Generation times (dots) plotted as single points at time of infection of infector. The almost constant solid line is a smoothed version of the local average of the generation times (left y-axis), while the other solid line is a smoothed version of the number of infectives during the outbreak (right y-axis).

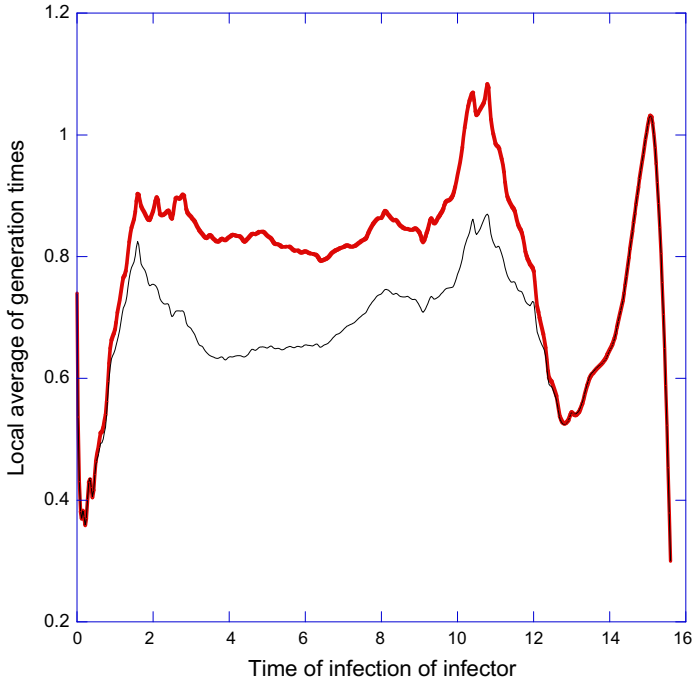

Fig. 2. Blow-up of local average generation times plotted by time of infection of infector. The solid line upper curve corresponds to the average curve shown in Fig. 1, while the other lower curve shows the average generation time when the average is first computed within infector and then plotted as a single point per infector.

In Fig. 2, the curve representing the average generation time per infector is also shown. In this case, the average level is approximately 2/3, as predicted by the "single infector" model. The "theoretical" average level is successively modified by two counteracting effects, a slight increase of the average generation times during the epidemic, due to the decrease of the effective infectious contact rate which leads to longer and longer intervals (dependence on β), when conditioning on at least one secondary case, and the previously discussed shortening effect.

Both curves are furthermore very unstable at both ends, in the intervals [0,3] and [11,20], say, but examination of the epidemic curve in Fig. 1 shows that the numbers of infectives involved in both cases are quite small.

It should also be observed that, in practice, forward observation could be difficult to implement due to a number of factors, including selection of cases for which forward observation is possible, censoring of infectious periods and missing/unobserved secondary cases.

Backward observation of generation times involves a selection effect, namely "finding" the infector of an infectee among available infectives. In the standard Markovian formulation, for instance, each infective has a probability proportional to his infection or

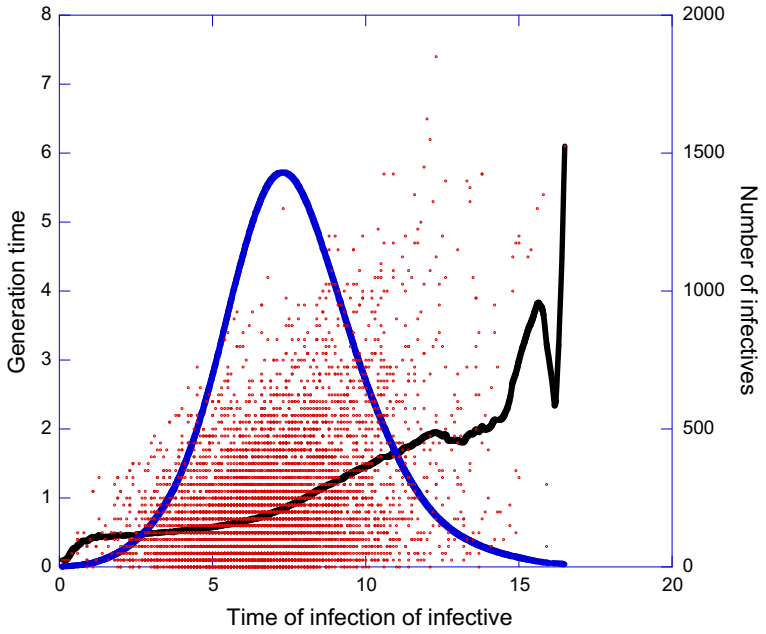

Fig. 3. Generation times (dots) plotted at time of infection of infectee. The increasing solid line is a smoothed version of the local average of the generation times (left y-axis), while the other solid line is a smoothed version of the number of infectives during the outbreak (right y-axis).

contact intensity of being the infector of a given infectee, assuming that the infector is not known. Thus, assuming that all infectives have the same contact intensity, all available infectives are equally probable and their ages of infection will be the possible generation times. However, the distribution of age of infection at a given point of time during the epidemic will depend on both the distribution of the infectious period ("still being active") and preceding dynamics of incidence of infectives. In terms of the deterministic approximation to the epidemic, the functions $s(t)$, $i(t)$ and $r(t)$, $t \geqslant 0$, that solve the ODE system

$$\begin{aligned} &s' = -\beta si,\\ &i' = \beta si - \mu i,\\ &r' = \mu i,\\ &s(0) = 1-\varepsilon, \quad i(0) = \varepsilon, \quad r(0) = 0 \quad (\text{for } \varepsilon > 0 \text{ but very small}), \end{aligned} \tag{1}$$

the density $h_t(a),\ 0 \leqslant a \leqslant t$, of generation times of length a at time t since the start of the epidemic will be proportional to $\beta s(t-a)i(t-a)l(a)$, where $l(a)$ denotes the probability of having an infectious period $\geqslant a$ and the term βsi accounts for the incidence of infection. For instance, in the beginning of the epidemic, according to the standard Markov model, it is well known that the number of infectives increases exponentially with exponent $(\beta-\mu)t$, while the number of susceptibles remains approximately constant and $l(a) = \exp(-\mu a)$. The result is that the distribution of generation times will be approximately exponential with expected value $1/\beta$, in the initial phase of the outbreak. This is only an approximation, however, because of the stochastic variability of incidence and infectious periods. Thus the relevant theory for this kind of representation is, during the early exponential phase of the epidemic, the "mean age at childbearing in the stable age distribution related to exponential growth" and an adaptation of this theory to the changing rate of incidence during the subsequent phases of the epidemic. The result is, as can be seen in Fig. 3, a steadily increasing average generation time, starting from about 1/2 and reaching 2 and more towards the end of the epidemic. This behaviour can be intuitively understood by considering that the rate of increase of infectives is steadily decreasing during the whole outbreak and that the infectives present at a given time will less and less consist of newly infected individuals.

Once again, backward sampling may be complicated by several factors, in practice. Essentially, the observations will be the result of backwards contact tracing, so many possible selection effects will operate on subjects for whom this contact tracing is attempted and successful. It may also be difficult to accurately assess the actual phase of the epidemic in which the observation was made. Finally, it should be remembered that contact tracing is not usually performed, in practice, in a strict backward or forward manner. Thus a possible mixture of the situations discussed above might be needed to model observations.

In conclusion, the presentation and analysis of observations has implications for what will be seen. For instance, since the set of generation times in an outbreak is the same, whether plotted by the time of infection of infector or infectee, Figs. 1 and 3 above show the same generation time values, just plotted in two different ways. However, this equality breaks down if only a subinterval of time is considered. While this last observation may seem paradoxical, it essentially shows that observation schemes "reorder" the set of observables or, in the case not all possible observations are included, "select" from it and great care must be taken to clearly understand what "subset" of observables has been obtained at a given time by a certain observation scheme. There is thus a qualitative difference between the "shortening of generation times" discussed in the "forward observation" case, where the distribution of generation times is modified by the susceptible dynamics, and the "backward observation" scheme that reorders the same observations so that the average becomes progressively larger as the epidemic progresses.

5. More general models

It is useful to consider, at least qualitatively, how the findings above will transfer to epidemic spread models with more features than the simple SIR model.

In many models, it is assumed that there is a latent period between the moment of infection and the beginning of the infectious period (SEIR model). As a rule, assuming a fixed transmission

intensity and distribution of the infectious period, the presence of a latent period does not affect R_0, but decreases the Malthusian parameter, compared to the same situation without latent period. All the effects described above will still hold within the infectious period, but the generation time now consists of the sum of the latent period and a part of the infectious period. In a "forward" looking perspective, it is probable that the "observation modifying effects" discussed above will be relatively smaller since the latent period is unaffected by the sampling and analysis scheme. In fact, in an extreme case, the Reed–Frost model, the latency time is assumed to be constant and the infective period extremely short. In that case, all observed generation times, independently of method of observation, analysis or representation, will be constant. Thus, the relation between the lengths of the latent and infectious periods will be of importance for the size of possible "modifying effects" affecting observations of generation times.

In the "backward" observation scheme, the basic idea presented in the previous section remains valid, assuming that individuals are equally infective during their infectious periods. Given the time of infection of an infectee, the generation time will be equal to the age of infection of a randomly chosen infective at that moment of time and, in the initial phase of an epidemic, there will be relatively more people who have recently become infective than in later phases of the spread. However, the distribution of latency times will play a role in the relevant formulae, since it introduces a delay between time of infection and start of infectivity for infected individuals and since it will play a role in the formula for the Malthusian parameter.

One may also consider allowing the infectious period to have a general probability distribution. As long as the infectivity is constant during the infective period, the "single infector" generation time will on average be half the expected length of an infectious period conditioned on containing at least one secondary case (plus an average latent period, if any). The effect of this conditioning depends in a specific way on the distribution.

The size bias effect of infectee based sampling is however preserved in a simple way, transforming the original density of the infectious period, $f(t)$ say, to $tf(t)/E(T)$ and resulting in an average interval length of $E(T) + \mathrm{Var}(T)/E(T)$. An average generation time will be half of this length.

If, finally, variable infectivity during the infectious period is considered, either as a model component or because the "shortening effect" of quick susceptible reduction is to be explored, it becomes difficult to make general statements. In general, a peaked infectivity will reduce the variability of the infection times of secondary cases, thus mimicking the effect of longer latent period and shorter infectious period. A decreasing infectivity during the infectious period will result in more secondary cases with short generation times and thus decrease the average generation time, compared to constant infectivity.

In [4], a general theoretical framework is formulated within which the combined effects on the behaviour of generation times of different assumptions about latent period, infectious period and varying infectivity during the infectious period can be analysed. In this formulation, it is possible to express important parameters quite easily, but at the same time to show that there is an important difference between what could be called individual quantities and population level quantities. This difference is exemplified, in this paper, by the different results obtained in forward sampling, with the two different kinds of averaging observations.

We have also simulated various combinations of latent and infective period distributions (not reported here) and observed that the phenomena discussed above for the simple SIR model always arise to some extent. It is worth noting that a given value of R_0, for instance, allows for many different values of r, depending on the choices of distributions of latent and infectious periods and also profile of infectivity within the infective period. The same is true for outbreaks with given r, with respect to R_0. Also, as observed above, if the latent period is long compared to the infectious period, sampling effects on generation times become less pronounced. However, the results are quite strongly dependent on the specific components of each model.

6. Generation times and epidemic prediction

In emerging epidemics, it is of fundamental value to be able to draw conclusions about the potential impact of the epidemic spread from the early assessment of the epidemic.

According to most epidemic models, the number of infected individuals up to time t will increase exponentially in the beginning of the epidemic, with rate of increase equal to the Malthusian parameter r.

One of the main reasons for the interest in generation times is that the Euler–Lotka equation establishes a connection between the Malthusian parameter r, observable in the initial phase of an epidemic, and R_0, given the generation time distribution. Approximations to this relationship have also been proposed. For instance, in Ferguson et al. [2], the approximate relationship

$$R_0 \approx 1 + r \times \text{average generation time}$$

is used. Other approximations are discussed by Wallinga and Lipsitch [5]. However, these relationships assume that the distribution or at least the average generation time, corresponding to the one used in the Euler–Lotka equation, is known (or, at least, have been estimated).

There are however potentially relevant situations where this may not be the case. As an example, suppose that some generation times have been observed in the beginning of an epidemic, in the same time period as the exponential increase of infected is observed, using "backwards" contact tracing. Then, if we denote the standard generation time distribution by $g(a)$, the observations will instead follow the modified distribution

$$g^*(a) = R_0 e^{-ra} g(a).$$

For the simple SIR model, where $R_0 = \beta/\mu$, $r = \beta - \mu$ and $g(a) = \mu e^{-\mu a}$ the expected value in $g^*(a)$ turns out to be $1/\beta$. This result would, somewhat unexpectedly, give the possibility to estimate the transmission intensity directly from the initial phase generation time measurements. Furthermore, considering the explicit expression for the Malthusian parameter r, we find that the relationship between R_0 and average observed initial generation time becomes

$$R_0 \approx 1/(1 - r \times \text{average initial generation time}).$$

More interestingly, in general, denoting by $T_1, \ldots, T_k$ the observations of generation times performed in this initial phase of the epidemic, it can be seen that the expression

$$\frac{1}{k}\sum_{i=1}^{k} e^{rT_i}$$

is an unbiased estimator of R_0, since the expected value of e^{rT}, with T having the distribution g^*, is precisely R_0. It can be observed that this result, i.e. the unbiasedness of e^{rT} for R_0, is valid in the SEIR model as well. With the appropriate redefinition of the generation time distribution $g(a)$ and the Malthusian parameter r in the SEIR model (see [5]), it can be seen that $R_0 g(a)$ is equal to $P(L \leqslant a \leqslant L + I)$, with L = latency time and I = infectious period, i.e. the probability that an individual is infectious at infection age a, and that, in consequence, the "initial generation time distribution" will again satisfy

$g^*(a) = R_0 e^{-ra} g(a)$, where the exponential term expresses the increasing initial incidence of infected individuals and R_0 is the proper scaling constant, taken from the Malthusian equation.

7. Generation times in households and endemic situations

A typical source of observations of infection chains might be the household setting. It has been observed [3] that infection probabilities within households may be larger than between members of the general population and, may be, that infectious doses may be larger and therefore that incubation and/or latent periods may be shorter, but, independently of the question whether infection in households is representative of infection in the general population, we may ask what kind of observations of generation times may be made in such a setting. Once again we will find that the rapid changes in susceptibles have a strong effect on observable generation times.

In the household setting, it is difficult to derive closed expressions or relevant approximations for many relevant model quantities. We will therefore illustrate the concept with a simple example. Extending the theory to larger households or households with structure and/or external infection can, in theory, be done quite easily by simulation.

Consider a household with one initial infective and two susceptibles. Assume, as before, that infectious periods have exponential duration with expected value $1/\mu$ and that all individuals in the household have potentially infectious contacts with each other with intensity β. The expected times until transitions and their probabilities now depend on the state of the process. For instance, the first event, whether removal of the primary case or infection of a secondary case, occurs at expected time $1/(2\beta + \mu)$. The time until the next event depends on what happened at the first event, etc.

Even without considering individual identities and just labelling by the order of infection, there are 19 possible orderings of infections and removals, that can be classified in seven main cases. Explicit calculation of case probabilities and expected generation times in the different cases yields an expression for the expected generation time $E(G)$ in this model (with $R = \beta/\mu$):

$$E(G) = \frac{1}{\mu}\frac{5R^3 + 18R^2 + 12R + 2}{2(R+1)^3(2R+1)}.$$

The rational function above assumes values between 37/48 and 0 for $R \geqslant 1$. There is thus no fixed relation between this value and, say, $1/\mu$, which could be considered as the expected generation time in this situation.

However, by considering an explicit statistical model for household data, that takes the changing numbers of susceptible into account, as in e.g. [13], it is possible to achieve correct inference about the infectious period and the contact intensity parameters.

Finally, it is perhaps not likely that the same disease can be observed both in endemic and epidemic conditions, but one can still investigate what would be observed if a disease were in endemic equilibrium in a population. In this case, there is no commonly accepted simple stochastic model for the fluctuations and there is a practical risk that "mini epidemics" may occur cyclically instead of a really stationary behaviour. However, assuming that there is no latent period, and since the numbers of susceptibles and infectives are assumed to stay at constant levels, there will be no "forward" shortening, but still a difference between infector and infectee based statistics, as described by the size biased sampling effect. "Backward" observations will now have a density proportional to just $l(a)$, the survival function corresponding to the distribution of infectious periods, in the deterministic approximation, which yields an average generation time equal to the $(E(IP) + \mathrm{Var}(IP)/E(IP))/2$. Interestingly, this average is the same as the one that will be observed in the infectee based "forward" scheme, but not in the "single infector" model.

8. Inference about generation times

In the above treatment, it has been assumed that infectious periods and generation times are observable. In reality, most observations will probably refer to serial intervals, i.e. the interval between two similar observable events, such as appearance of symptoms, in the progress of disease of infector and infectee. In Svensson [4], it is argued that it is natural to assume that generation times and the corresponding serial intervals have the same expected value, but otherwise different distributions, making it theoretically impossible, without further assumptions, to estimate further characteristics of the generation time distribution. These difficulties do not only involve, say, the variance but also other nonlinear characteristics such as the Malthusian parameter. In addition, since serial intervals are essentially generation times observed with an error with mean zero, all the observational problems afflicting the mean of generation times and other distributional aspects will of course also be relevant for serial intervals.

Inference about various model parameters may be approached either by maximum likelihood methods in a rather fully specified model or by counting process and martingale methods, see e.g. [14]. Either way, the dynamics of the disease and the observation plan must be accounted for, in order to avoid bias in the estimates.

In a fully parametrized model, inference may be carried out using ML or MCMC techniques, but the results will always be dependent on the more or less correct representation of reality by the model components, such as constant or variable infectivity during the infectious period, the precise form of joint distribution for latent and infectious period, etc. In a counting process approach, many of these quantities can be estimated non-parametrically, thus allowing data to express the behaviour of some model components.

As an example, following the general construction of an infectivity function in [4], assume that we may observe the full transmission chain in a household. For each individual in that household, we can construct a counting process $N(t)$ = the number of transmissions of disease by this individual up to time t after infection. This process has intensity $\lambda S(t)k(t)$, with λ and k denoting total infectivity and time evolution of infectivity, respectively, and $S(t)$ denoting the number of susceptibles left in the household just before time t (still time since infection of the index infective). It is then possible to derive, using counting process martingale theory, a crude, but unbiased, estimator of

$$\int_0^t \lambda k(s) I(S(s) > 0) ds$$

by calculating

$$\int_0^t \frac{dN(s)}{S(s-)}$$

for $0 \leqslant t \leqslant T$ = time, if ever, at which $S(T) = 0$ for the first time. It is also possible to estimate the (time dependent) variance of this estimator. By combining such estimates from individuals in different households, assumed independent, and also making the assumption that the parameters (λ, k) are iid for all individuals and denoting $E(\lambda k(s))/E(\lambda)$ by $v(s)$, it is now possible to produce a precise estimate of

$$R_0 \int_0^t v(s) ds,$$

i.e. the cumulated expected individual force of infection (including possible latent period and time varying infectivity).

9. Conclusions and discussion

The concept of generation time of an infectious disease has several advantages: it has an obvious meaning, is observable, at least approximately, in the form of serial interval, and, in theory, determines an important part of the temporal evolution of spread. However, as has been shown in this paper, observations will be influenced by spread dynamics and sampling schemes and the precise relation between generation times and overall temporal evolution is model dependent. In this paper, the problems inherent in the observation and interpretation of generation times have been approached from the point of view of stochastic models for infectious disease spread. The conclusions regarding the analysis of observations of generation times is that a simple presentation of data will almost never give an unbiased picture of the generation time distribution and its main characteristics; more elaborate statistical methods are required. Some parametrical and non-parametrical statistical methods have already been proposed in the literature and some further suggestions are given in this paper.

It should also be observed that the concept of generation time for a disease was first introduced for diseases such as measles, where latency times are much longer than infectious periods (see e.g. [15]). In such cases, at least for some time in the beginning of the disease spread, generations of infected are well separated and therefore generation times are ‘‘easy” to estimate and have a clear meaning . This may not be the case, however, for important diseases such as influenza.

Acknowledgments

Financial support from the ECDC project ‘‘The relevance and interpretation of generation times in epidemic modelling” is thankfully acknowledged. Some of the initial work on the subject has been supported by the DG SANCO financed project MODELREL and the FP6 research project INFTRANS. Many enlightening discussions with research colleagues, in particular Jacco Wallinga, have accompanied the progress of this work. The helpful comments of two anonymous referees are also acknowledged.

References

[1] J. Wallinga, P. Teunis, Different epidemic curves for Severe Acute Respiratory Syndrome reveal similar impacts of control measures, Am. J. Epidemiol. 160 (2004) 509.
[2] N. Ferguson et al., Strategies for containing an emerging influenza pandemic in Southeast Asia, Nature 437 (2005) 209.
[3] P. Fine, The interval between successive cases of an infectious disease, Am. J. Epidemiol. 158 (2003) 1039.
[4] Å. Svensson, A note on generation times in models, Math. Biosci. 208 (2007) 300.
[5] J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers, Proc. R. Soc. B 274 (2007) 599.
[6] E. Kenah et al., Generation interval contraction and epidemic data analysis, Math. Biosci. 213 (2008) 71.
[7] N. Keyfitz, Introduction to the Mathematics of Populations, revised ed., Addison-Wesley, 1977.
[8] N. Keyfitz, H. Caswell, Applied mathematical demography, third ed., Statistics for Biology and Health, Springer, 2005.
[9] P. Haccou et al., Branching processes: variation growth and extinction of populations, Cambridge Studies in Adaptive Dynamics, Cambridge University Press, 2005.
[10] C.R. Rao, Statistics and Truth. Putting Chance to Work, second ed., World Scientific, 1997.
[11] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2, John Wiley & Sons, 1971.
[12] R.M. Anderson, R.M. May, Infectious Diseases of Humans. Dynamics and Control, Oxford University Press, 1971.
[13] S. Cauchemez et al., A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data, Stat. Med. 23 (2004) 3469.
[14] N. Becker, Analysis of Infectious Disease Data, Chapman & Hall, 1989.
[15] R.E. Hope Simpson, The period of transmission in certain epidemic diseases: an observational method for its discovery, Lancet 17 (1948) 755.