Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

A deterministic epidemic model taking account of repeated contacts between the same individuals

O. Diekmann, M.C.M. de Jong and J.A.J. Metz

Department of Analysis, Algebra and Geometry

AM-R9517 1995

Report AM-R9517
ISSN 0924-2953

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

A Deterministic Epidemic Model taking Account of Repeated Contacts between the same Individuals

O. Diekmann[1,3], M.C.M. de Jong[2] and J.A.J. Metz[3,1]

1. CWI, P.O. Box 94079, 1090 GB Amsterdam

2. ID-DLO, P.O. Box 65, 8200 AB Lelystad

3. EEW, section Theoretical Biology, University of Leiden, P.O. Box 9516, 2300 RA Leiden

The Netherlands

Abstract

We introduce a certain population contact structure and derive, in three different ways, the final size equation for a quite general superimposed epidemic process. The contact structure is characterized by the following two properties: i) each individual contacts exactly k other individuals; ii) these k acquaintances are a random sample of the (infinite) population.

AMS Subject Classification (1991): 92D30
Keywords & Phrases: contact structure, basic reproduction ratio, final size

1. Introduction

The transmission of an infectious agent is often predicated upon some form of contact between two hosts. A submodel for the host contact process is, therefore, an important constituent of a model for the spread of an infectious disease.

The law of mass action asserts that contact rates are proportional to the densities of the types of individuals involved, where it is allowed that the population is stratified according to spatial –or social position, the latter capturing a large variety of context dependent distinctions. With finite rates, the expected number of contacts of an individual in a given finite interval of time, is finite. In the deterministic limit the subpopulation sizes are, by assumption, infinite and hence the probability that two particular individuals have contact equals zero. As a consequence, the probability that two individuals have contact twice is zero as well.

Such properties of the contact model contradict our daily experience of (human and animal) social relationships (and therefore motivated ad hoc modifications , e.g. van Druten [12]). So it seems relevant to study the effect of repeated contacts with a (possibly large, either fixed or changeable) group of "acquaintances" on the spread of an infectious disease.

The acquaintance relation defines a network of connections among the members of the population and the modelling now entails a specification of this network. In spatial lattice models (Durrett [7]) one assumes a very regular contact network. In random graph models (Blanchard, Bolz & Krüger [2] Barbour & Mollison [1]) one only prescribes certain statistical properties of the network. In pair formation models (Dietz & Hadeler [6], Diekmann, Dietz & Heesterbeek [5], one assumes that contacts are restricted to partners that form an isolated pair for an extended period of time. Yet another class of models stratifies the population according to households.

Here we shall consider a class of models that combines features of spatial lattice– and (the limit of) random graph models. We picture an individual as surrounded by a circle of k acquaintances. But these acquaintances do not contact each other. In fact we assume that the probability that two

acquaintances have other acquaintances in common equals, in the infinite population, zero. Admittedly this again contradicts our daily experience and we think that, ultimately, models should be constructed in which the correlation between acquaintances of acquaintances is neither as rigid as it is in spatial lattices nor as absent as we postulate it here.

It is common practice in stochastic epidemic modelling to use "contact" and "infect" as synonymous. This reflects that one does not first model a contact process and then superimposes transmission, but that one does both things at one go. There is a certain economy in doing so. But we plead for the more laborious road, in order to unravel the influences of, respectively, contact structure and infectivity.

In the impressive paper [10] J.A. MARTIN-LÖF analyses generalised Reed-Frost processes that, on first sight, should contain the present model as a special case. Indeed, the rules concerning the active part of transmitting the infective agent are very flexible. But concerning the passive part of being a potential victim, there are no rules at all. In contrast, in the model presented here an individual can possibly receive the agent from exactly k acquaintances, no more no less. This has an effect on the final size equation.

Debatable as our assumptions still may be from a modelling point of view, they stand out by the attractive feature that they allow us not only to compute the basic reproduction ratio R_0, which is straightforward, but also to derive a simple final size equation. To present this derivation is the main aim of the present paper.

We shall use the expression for R_0 and the equation for the final size to investigate how the epidemic depends on the size k of the circle of acquaintances. For $k = 2$ we recover the epidemic on the integer lattice on the line, and for $k \to \infty$ we are back to KERMACK & MCKENDRICK [8].

In a follow-up publication [4] we shall retain a mass action contact process as a second model of transmission, infectious matter being partitionned between the two modes. Thereby we obtain a one-parameter "bridge" (a homotopy) between pure mass action and pure "acquaintance" transmission, which allows for a biological interpretation. Thus we will be able to make comparisons and to quantify the effect of repeated contacts further.

There may be many ways to construct a probabilistic interaction structure among N individuals such that our assumptions are satisfied in the limit for N tending to infinity (along a subsequence, perhaps, to avoid combinatorial obstructions). We shall ignore this issue.

2. The Model Ingredients

We shall, as in METZ [11], allow for variability in infectious output. We distinguish individuals with a variable ξ taking values in a set Ω. The population composition (or, equivalently the probability that an individual draws a certain future infectivity upon becoming infected) is described by a probability measure m on Ω. As a function of time τ elapsed since becoming infected, the infectious output of an individual of type ξ is given by $a(\tau, \xi)$. This infectious output is uniformly distributed among k acquaintances. We normalise "infectious output" such that it equals the probability per unit of time of transmission. Hence the ("survival") probability that a specific acquaintance is not yet infected by an infectious individual of type ξ at "disease-age" τ is given by

$$\mathcal{P}(\tau, \xi) = e^{-\frac{1}{k}\int_0^\tau a(\sigma,\xi)d\sigma} \tag{2.1}$$

and the expected value of this quantity equals

$$\mathcal{F}(\tau) = \int_\Omega e^{-\frac{1}{k}\int_0^\tau a(\sigma,\xi)d\sigma} m(d\xi). \tag{2.2}$$

The overall probability of transmission Q is now defined by

$$Q = 1 - \mathcal{F}(\infty). \tag{2.3}$$

The most familiar example is where infectious output is a constant, say α, during an exponentially distributed (say with parameter γ) period. Then $\Omega = \mathbb{R}_+$, $a(\tau, \xi) = \alpha$ for $0 \leq \tau \leq \xi$ and zero otherwise, and $m(d\xi) = \gamma e^{-\gamma\xi} d\xi$. From this specification we find

$$Q = \frac{\alpha/k}{\gamma + \alpha/k} \tag{2.4}$$

as one can also deduce directly by thinking in terms of competing events. When we introduce a latent period preceding the infectious period, we have to multiply the expression with the reduction factor accounting for the possibility that the individual never "enters" the infectious period (since, for instance, it dies).

As a last example, consider $\Omega = [2, 3]$, m the uniform distribution on this interval and $a(\tau; \xi) = \alpha$ for $1 \leq \tau \leq \xi$ and zero otherwise. Then

$$Q = 1 - \frac{k}{\alpha}[e^{-\frac{\alpha}{k}} - e^{-2\frac{\alpha}{k}}] \tag{2.5}$$

3. The Basic Reproduction Ratio R_0

Any infected individual necessarily has at least one infectious or immune acquaintance, to wit the one that transmitted the infectious agent to it. In the initial phase of the epidemic this will be the only one. Hence the expected number R_0 of secondary cases per primary case in the initial phase equals

$$R_0 = (k-1)Q. \tag{3.1}$$

So the threshold condition for an epidemic outbreak (which, as always, reads $R_0 > 1$) cannot be satisfied for $k = 2$. This corresponds exactly to the well-known fact that the epidemic on a one-dimensional lattice with nearest neighbour transmission necessarily stops (i.e. only minor outbreaks are possible, and the infectious agent cannot propagate to infinity).

At the other extreme, for $k \to \infty$, we find

$$\lim_{k\to\infty} R_0 = \int_0^\infty A(\tau) d\tau \tag{3.2}$$

where

$$A(\tau) := \int_\Omega a(\tau, \xi) m(d\xi) \tag{3.3}$$

is the expected infectious output at disease-age τ (note that the linearity in the limit $k \to \infty$ allows us to interchange taking the expectation with the other operations). So here we find back another well-known result, the expression for R_0 for uniform transmission in an infinite population (i.e. for mass action).

In Figure 1 we show R_0 as a function of k for the case of an exponentially distributed infectious period, i.e. for Q given by (2.4), with $\alpha = 10\gamma$ and $\alpha = 2\gamma$.

We may also pose the following question: given a limit for $k \to \infty$, which is above threshold, what is the minimal value of k for which R_0 is still above threshold? If we put $R_0 = (k-1)Q = 1$ and $Q = \frac{\alpha/k}{\gamma+\alpha/k}$ we find $k = \frac{2\alpha}{\alpha-\gamma}$.

4. The probability of a minor outbreak

We can use a branching process approximation to give a slightly more detailed description of the initial phase of an epidemic. Given ξ and given that the infective has $k-1$ susceptible acquaintances, the probability that it will infect j acquaintances equals

$$\binom{k-1}{j}(1 - q(\infty, \xi))^{k-1-j} q(\infty, \xi)^j$$

where

$$q(\infty,\xi) := 1 - e^{-\frac{1}{k}\int_0^\infty a(\sigma,\xi)d\sigma} \tag{4.1}$$

This motivates us to define, for $0 \le j \le k-1$,

$$p_j = \binom{k-1}{j} \int_\Omega (1-q(\infty,\xi))^{k-1-j} q(\infty,\xi)^j m(d\xi) \tag{4.2}$$

and $p_j = 0$ otherwise, and to consider the branching process in which an individual gets j offspring with probability p_j. The probability π that this process goes extinct when we start it off with one individual can be found as the solution of the consistency condition

$$\pi = g(\pi) \tag{4.3}$$

where

$$g(z) := \sum_{j=0}^{\infty} p_j z^j. \tag{4.4}$$

The conditions $g(0) > 0$ and $g'(1) > 1$ guarantee that $0 < \pi < 1$. Note that for the present definition of p_j

$$g'(1) = (k-1)Q = R_0. \tag{4.5}$$

The description fits, with the appropriate interpretation of offspring, the epidemic process except for one detail. If the initial infective is infected from the outside, it still has k susceptible acquaintances. Hence the probability $\tilde{\pi}$ that, given such an initial condition, the infective agent goes extinct before a substantial epidemic develops equals

$$\begin{aligned} \tilde{\pi} &= \sum_{j=0}^{k} \binom{k}{j} \int_\Omega (1-q(\infty,\xi))^{k-j} q(\infty,\xi)^j m(d\xi)\pi^j \\ &= \int_\Omega (1-q(\infty,\xi)+\pi q(\infty,\xi))^k m(d\xi) \end{aligned} \tag{4.6}$$

where, for $R_0 > 1, \pi$ is the unique root in $(0,1)$ of equation (4.3) with g defined by (4.4) and (4.2).

For $k \to \infty$ the distinction between $\tilde{\pi}$ and π becomes negligible and, by the familiar convergence of the binomial distribution to the Poisson distribution, we find the limiting equation

$$\pi = \int_\Omega e^{(\pi-1)\int_0^\infty a(\sigma,\xi)d\xi} m(d\xi). \tag{4.7}$$

A simple computation reveals that in the special case of the exponentially distributed infectious period the solution is given by

$$\pi = \frac{\gamma}{\alpha} = R_0^{-1}. \tag{4.8}$$

5. The final size equation

The epidemic in a demographically closed population will inevitably come to an end (unless we assume that immunity is only temporary). Let us denote by s_∞ the fraction of the population that escapes from ever being infected. Then $1 - s_\infty$ is the (relative) final size of the epidemic. In this section we shall employ heuristic probabilistic arguments to derive an equation from which s_∞ can be computed. The deterministic justification of our assertions is postponed to the next two sections.

Imagine a large graph in which every vertex is linked to exactly k other vertices, while short cycles and separate components are avoided by randomness. Consider an edge linking A and B. We ask the question: will B become infected by A, should A itself become infected? Without knowledge on A, the probability that the answer is yes simply equals Q.

Imagine a large epidemic outbreak has taken place. We consider an arbitrary individual and ask with what probability it was *not* affected. In order to decide whether or not it was infected, we look backwards, i.e. we follow edges with probability Q, to see whether we arrive at an infected individual in this manner. By definition s_∞ equals the probability that we won't.

To let the size of the graph go to infinity has two effects:

- the process of following edges backwards is described by a branching process with

$$p_j = \binom{k-1}{j}(1-Q)^{k-1-j}Q^j \tag{5.1}$$

  for $0 \leq j \leq k-1$ and $p_j = 0$ otherwise, except for the first generation which has

$$p_j = \binom{k}{j}(1-Q)^{k-j}Q^j \tag{5.2}$$

  for $0 \leq j \leq k$ and $p_j = 0$ otherwise (note that by looking backwards we eliminated the dependence that complicated (4.2))
- s_∞ equals the extinction property of this branching process (here the argument is that either the branching process goes extinct or it explodes and therefore necessarily connects to the "giant (outbreak) component" (Bollobás [3]), which we assume to be unique).

So the claim is that

$$s_\infty = (1-Q+\pi Q)^k \tag{5.3}$$

where, for $R_0 > 1$, π is the unique solution in (0,1) of the equation

$$\pi = (1-Q+\pi Q)^{k-1}. \tag{5.4}$$

In the next two sections we shall substantiate this claim.

For $k \to \infty$ the distinction between s_∞ and π disappears and in the limit we recover the familiar Kermack-McKendrick final size equation

$$s_\infty = e^{(s_\infty - 1)\int_0^\infty A(\tau)d\tau} \tag{5.5}$$

In Figure 2 we depict the final size as a function of k for Q given by (2.4) with $\alpha/\gamma = 10$ and $\alpha/\gamma = 2$.

6. The deterministic generation process

A remarkable feature of epidemic processes is that, even though they are nonlinear, one can study certain aspects, such as the final size, at will in the context of a discrete time generation formulation or in continuous time. The final size will be the same, despite the fact that there will be different answers to the question "who is held responsible for infecting whom?". this applies equally to stochastic and deterministic models and was, as far as we know, first noted by Ludwig [9].

We shall index the generations by n. In each generation, a susceptible possibly has acquaintances which are susceptible, infected or immune. The number that are infected determines the probability that the susceptible becomes infected, but also the addition to the number of immune acquaintances in case the susceptible is not infected. So we have to divide the population into subpopulations according to the status of the individual concerned *and* the status of its circle of acquaintances. We adopt the convention that a subpopulation indexed by n refers to the size of that subpopulation immediately

before infection takes place in generation n. Moreover, size refers to relative size, i.e. our variables are proportions of the total population.

The following observation is crucial for the model formulation. Consider one specific individual and focus on one of its k acquaintances. We want to know with what probability this acquaintance belongs to a certain subpopulation. Without further knowledge the "random=uniform" principle tells us that this probability equals the subpopulation size. Suppose, however, that we know about a fraction of the totality of all acquaintances of the subpopulation that the specific individual we consider is ruled out. Then we have to multiply the subpopulation size by one minus this fraction to find the probability.

Let $S_j(n)$ denote the subpopulation size of susceptibles which had $k-j$ of their acquaintances infected in the generations up to and including $n-2$. In other words, j of their acquaintances are either susceptible or infected in generation $n-1$ and then infectious in generation n, while $k-j$ of their acquaintances are "removed". Similarly, let $I_j(n)$ denote the subpopulation size of those that are infected in generation $n-1$, and which have $k-j$ removed acquaintances. Now consider an acquaintance of a susceptible, for which it is known that it was not infected in the generations up to and including $n-2$. Let $p(n)$ denote the probability that this acquaintance belongs to the subpopulation of those infected in generation $n-1$. Without the knowledge that the acquaintance was not infected in earlier generations, this probability would simply be

$$\frac{1}{k}\sum_{j=1}^{k-1} jI_j(n).$$

But the conditional probability that takes our a priori knowledge into account is obtained by multiplying this quantity with the inverse of

$$\frac{1}{k}\left(\sum_{j=1}^{k-1} jI_j(n) + \sum_{j=1}^{k} jS_j(n)\right).$$

Thus we arrive at the expression

$$p(n) = \frac{\sum_{j=1}^{k-1} jI_j(n)}{\sum_{j=1}^{k-1} jI_j(n) + \sum_{j=1}^{k} jS_j(n)}. \tag{6.1}$$

Consider a member of the $S_l(n)$ subpopulation. The probability that $l-j$ of its acquaintances are infectious equals

$$\binom{l}{l-j} p(n)^{l-j}(1-p(n))^j.$$

Given such a configuration, the probability that the susceptible becomes infected in generation n equals

$$1-(1-Q)^{l-j},$$

while with probability

$$(1-Q)^{l-j}$$

the susceptible remains susceptible. Consequently,

$$S_j(n+1) = \sum_{l=j}^{k} \binom{l}{l-j} p(n)^{l-j}(1-p(n))^j (1-Q)^{l-j} S_l(n) \tag{6.2}$$

and

$$I_j(n+1) = \sum_{l=j+1}^{k} \binom{l}{l-j} p(n)^{l-j}(1-p(n))^j(1-(1-Q)^{l-j})S_l(n). \tag{6.3}$$

Note that in the last equation necessarily $j \leq k-1$, since an infective needs to have had at least one infected acquaintance. Note also that the future is fully determined by the values of $p(n)$ and $S_j(n)$ and that one can consider $I_j(n)$ as auxiliary variables.

Together the equations (6.1)-(6.3) describe the deterministic dynamics on a generation basis. The initial condition consists of a value for $p(0)$ and each of $S_j(0)$, $0 \leq j \leq k$, such that $\sum_{j=0}^{k} S_j(0) \leq 1$. (A slightly more precise, but also more cumbersome, initial condition would specify $I_j(0)$ and $S_j(0)$ such that $\sum_{j=0}^{k}(I_j(0)+S_j(0)) = 1$ and then compute $p(0)$ from (6.1) with $n = 0$.)

The analysis of the difference equations is greatly facilitated by the introduction of the generating functions

$$G_n(z) = \sum_{j=0}^{k} S_j(n) z^j \tag{6.4}$$

and

$$H_n(z) = \sum_{j=0}^{k-1} I_j(n) z^j \tag{6.5}$$

LEMMA 6.1. The equations (6.1)-(6.3) are equivalent with the system

$$G_{n+1}(z) = G_n((1-Q)p(n) + z(1-p(n)), \tag{6.6}$$

$$p(n+1) = 1 - \frac{G_n'(1-Qp(n))}{G_n'(1)}. \tag{6.7}$$

PROOF.

$$G_{n+1}(z) = \sum_{j=0}^{k} \sum_{l=j}^{k} \binom{l}{l-j} p(n)^{l-j}(1-p(n))^j(1-Q)^{l-j} z^j S_l(n)$$

$$= \sum_{l=0}^{k} \sum_{j=0}^{l} \binom{l}{l-j} p(n)^{l-j}(1-p(n))^j(1-Q)^{l-j} z^{j-l} z^l S_l(n)$$

$$= \sum_{l=0}^{k} \left(\frac{(1-Q)p(n)}{z} + 1 - p(n) \right)^l z^l S_l(n) = G_n((1-Q)p(n) + z(1-p(n)).$$

Manipulating in a similar manner we find the identity

$$H_{n+1}(z) + G_{n+1}(z) = G_n(p(n) + z(1-p(n)).$$

From (6.1) we obtain at once that

$$p(n) = \frac{H_n'(1)}{H_n'(1) + G_n'(1)}.$$

Hence

$$p(n+1) = \frac{H_{n+1}'(1)}{H_{n+1}'(1) + G_{n+1}'(1)} = \frac{G_n'(1)(1-p(n)) - G_{n+1}'(1)}{G_n'(1)(1-p(n))} = 1 - \frac{G_n'(1-Qp(n))}{G_n'(1)}.$$

Thus we derived (6.6) and (6.7) from (6.1)-(6.3). Conversely, (6.6) and (6.7) constitute a closed system of difference equations which define $p(n)$ and $G_n(z)$ once we specify $p(0)$ and $G_0(z)$. Equation (6.6) guarantees that $G_n(z)$ is a polynomial of degree k provided $G_0(z)$ is such a polynomial. We now use equation (6.4) the other way around, i.e. to define $S_j(n)$ in terms of $G_n(z)$. Subsequently we use (6.3) to define $I_j(n)$. Finally, to verify (6.1) and (6.2) amounts to reading chains of identities derived above in the other direction, and sorting our according to powers of z. □

LEMMA 6.2. The solution of (6.6) is given by

$$G_n(z) = G_0((1-Q)(1-B(n)) + B(n)z) \tag{6.8}$$

where $B(0) = 1$ and

$$B(n+1) = (1-p(n))B(n). \tag{6.9}$$

PROOF. Suppose (6.8) holds for a particular n. Then

$$\begin{aligned} G_{n+1}(z) &= G_n((1-Q)p(n) + z(1-p(n)) \\ &= G_0((1-Q)(1-B(n)) + B(n)((1-Q)p(n) + z(1-p(n)))) \\ &= G_0((1-Q)(1-(1-p(n))B(n)) + (1-p(n))B(n)z) \\ &= G_0((1-Q)(1-B(n+1)) + B(n+1)z) \end{aligned}$$

or, in words, (6.8) holds for $n+1$. Since obviously (6.8) is correct for $n = 0$, the proof is complete. □

The expression (6.8) allows us to rewrite (6.7) as

$$p(n+1) = 1 - \frac{G_0'(1-Q(1-B(n)+B(n)p(n)))}{G_0'(1-Q(1-B(n)))} \tag{6.10}$$

or, equivalently, as

$$p(n+1) = 1 - \frac{G_0'(1-Q(1-B(n+1)))}{G_0'(1-Q(1-B(n)))}. \tag{6.11}$$

LEMMA 6.3. The sequence $B(n)$ is positive, decreasing and satisfies

$$B(n+1) = (1-p(0))\frac{G_0'(1-Q+QB(n))}{G_0'(1)}. \tag{6.12}$$

PROOF. We derive two expressions for the product $\Pi_{l=1}^{n}(1-p(l))$, basing ourselves on, respectively, (6.9) and (6.11). By (6.9) this product equals

$$\frac{B(n+1)}{B(1)} = \frac{B(n+1)}{1-p(0)}$$

and by (6.11) it also equals

$$\frac{G_0'(1-Q+QB(n))}{G_0'(1)}.$$

Equating the two expressions we find (6.12). Since G_0 is a polynomial with nonnegative coefficients, both G_0' and G_0'' are positive. It then follows at once from (6.12) that $B(n)$ is positive and decreasing. □

COROLLARY 6.4. *The limit $B(\infty)$ exists and satisfies*

$$B(\infty) = (1 - p(0))\frac{G_0'(1 - Q + QB(\infty))}{G_0'(1)}. \tag{6.13}$$

THEOREM 6.5. *The distribution of susceptibles at the end of the epidemic, with respect to the number of their acquaintances that are susceptible as well, has generating function*

$$G_\infty(z) = G_0((1 - Q)(1 - B(\infty)) + B(\infty)z) \tag{6.14}$$

where $B(\infty)$ is the unique root in (0,1) of equation (6.3). In particular,

$$s_\infty = G_\infty(1) = G_0(1 - Q + QB(\infty)). \tag{6.15}$$

PROOF. The right hand side of (6.13) is positive for $B(\infty) = 0$, less than one for $B(\infty) = 1$ and has a positive first and second derivative with respect to $B(\infty)$. Hence there exists a unique root in (0,1). The convergence of $B(n)$ can be "lifted" to the convergence of the generating function $G_n(z)$ by means of (6.8). □

We now specialise to the case of a negligibly small initial infection. That is, we let both $p(0) \downarrow 0$ and assume that $S_j(0) = 0$ for $j < k$ while $S_k(0) \uparrow 1$. The limiting initial generating function is simply

$$G_0(z) = z^k \tag{6.16}$$

and consequently (6.13) reduces to

$$B(\infty) = (1 - Q + QB(\infty))^{k-1}. \tag{6.17}$$

(Note that now $B(\infty) = 1$ is always a root, but that, as a graphical argument shows, we converge to the root in (0,1) when letting $p(0) \downarrow 0$, provided such a root exists, i.e. provided $R_0 = (k-1)Q > 1$.) Comparing (6.17) with (5.4) we find that for this limiting situation $B(\infty)$ equals π, the extinction probability of the branching process. So s_∞ computed from (6.15) as

$$s_\infty = (1 - Q + QB(\infty))^k \tag{6.18}$$

agrees exactly with the formula (5.3) and our first deterministic justification of the claim (5.3) is achieved. Moreover, we now know that

$$S_j(\infty) = \binom{k}{j}(1 - Q)^{k-j}(1 - B(\infty))^{k-j}B(\infty)^j. \tag{6.19}$$

7. THE DETERMINISTIC CONTINUOUS TIME FORMULATION

So far we could work with the overall probability of transmission Q, but now we will need the full survival function $\mathcal{F}$ introduced in (2.2). And we shall employ another survival function. Let, for a susceptible, $B(t)$ denote the probability that an acquaintance is susceptible at time t, given that it was susceptible at time zero (note carefully that the individual whose acquaintance we consider is, by assumption, susceptible in the entire interval from zero to t). Of course B is not known; it is our task to determine it. But for a while we shall pretend that we know B.

Next assume that p is the hazard rate corresponding to B, i.e.

$$B(t) = e^{-\int_0^t p(\sigma)d\sigma}. \tag{7.1}$$

In a sense (specified below) p measures the force of infection in the population (but p does not equal, for a randomly selected susceptible, the probability per unit of time to become infected and hence it is not the force of infection in the traditional sense of the word).

We denote by $S_j(t, \sigma_k, \ldots, \sigma_{j+1})$ the density, at time t, of susceptibles with j susceptible acquaintances and $k - j$ acquaintances that were infected at, respectively, the times $0 < \sigma_k < \ldots < \sigma_{j+1} \leq t$.

LEMMA 7.1.

$$S_j(t,\sigma_k,\ldots,\sigma_{j+1}) = \frac{k!}{j!}p(\sigma_k)B(\sigma_k)\ldots p(\sigma_{j+1})B(\sigma_{j+1})(B(t))^j\mathcal{F}(t-\sigma_k)\ldots\mathcal{F}(t-\sigma_{j+1})\bar{S}_k(0) \qquad (7.2)$$

where $\bar{S}_k(0)$ denotes the size of the population, at time zero, of susceptibles with all acquaintances susceptible.

PROOF. The probability per unit of time that the number of susceptible acquaintances drops from l to $l-1$ equals, by definition, $lp(t)$. This explains the factor $kp(\sigma_k)(k-1)p(\sigma_{k-1})\ldots(j+1)p(\sigma_{j+1}) = \frac{k!}{j!}p(\sigma_k)\ldots p(\sigma_{j+1})$. Moreover, the k acquaintances have to "survive" to, respectively, $\sigma_k,\ldots,\sigma_{j+1},t,\ldots,t$, which explains the factor $B(\sigma_k)\ldots B(\sigma_{j+1})(B(t))^j$. Finally, the individual that we consider should remain susceptible itself, which explains the factor $\mathcal{F}(t-\sigma_k)\ldots\mathcal{F}(t-\sigma_{j+1})$. □

When the initial population composition includes susceptibles with infective acquaintances, one has to specify the expected future infectivity of these acquaintances. Next one combines this information with the expected changes after $t = 0$ to derive analogues of formula (7.2). We refrain from an elaboration of such formulas.

As the times at which the acquaintances were infected are not of primary importance to us, we now define

$$\bar{S}_j(t) := \int_0^t \int_{\sigma_k}^t \ldots \int_{\sigma_{j+1}}^t S_j(t,\sigma_k,\ldots,\sigma_{j+1})d\sigma_{j+1}\ldots d\sigma_k. \qquad (7.3)$$

Using $p(\sigma)B(\sigma)d\sigma = B(d\sigma)$ we find.

COROLLARY 7.2.

$$\bar{S}_j(t) = \frac{k!}{j!}(B(t))^j \int_0^t \ldots \int_{\sigma_{j+3}}^t \mathcal{F}(t-\sigma_{j+2}) \int_{\sigma_{j+2}}^t \mathcal{F}(t-\sigma_{j+1})B(d\sigma_{j+1})B(d\sigma_{j+2})\ldots B(d\sigma_k)\bar{S}_k(0). \qquad (7.4)$$

And by taking the limit $t\to\infty$, while noting that $\mathcal{F}(\infty) = 1-Q$, we deduce

COROLLARY 7.3.

$$\bar{S}_j(\infty) = \binom{k}{j}(1-Q)^{k-j}B(\infty)^j(1-B(\infty))^{k-j}\bar{S}_k(0) \qquad (7.5)$$

which coincides with (6.19) upon taking $\bar{S}_k(0) = 1$. (But now we have a clear interpretation of $B(\infty)$, which in fact allows us to infer (7.5) directly, without going through the intermediate steps (7.2) and (7.4).) Finally, the explicit formula (7.5) has the immediate

COROLLARY 7.4.

$$\sum_{j=0}^k \bar{S}_j(\infty) = (B(\infty)+(1-Q)(1-B(\infty)))^k\bar{S}_k(0) \qquad (7.6)$$

$$\sum_{j=1}^k j\bar{S}_j(\infty) = kB(\infty)(B(\infty)+(1-Q)(1-B(\infty)))^{k-1}\bar{S}_k(0). \qquad (7.7)$$

We now set out for a determination of $p(t)$. Let $\tilde{S}_j(t)$ denote the size of the subpopulation of susceptibles with exactly j susceptible acquaintances, at time t. Then the total number of susceptible acquaintances of susceptibles is given by

$$N(t) = \sum_{j=1}^{k} j\tilde{S}_j(t). \tag{7.8}$$

This number decreases for two reasons:

i) a susceptible becomes infected itself

ii) one of the acquaintances of a susceptible becomes infected.

As the difference is only one of perspective, the rates of i) and ii) are equal at the population level (note that in our way of bookkeeping every link is counted twice; so also a change in the disposition of a link should be counted twice; for the change under consideration, we do that once under i) and once under ii); finally note that in an infinitesimal time interval it cannot happen that both partners in the connection are infected). For every susceptible acquaintance the probability per unit of time of becoming infected is, by definition, $p(t)$. Therefore

$$\frac{dN}{dt} = -2pN. \tag{7.9}$$

Note that this relation gives a more precise meaning to the interpretation of p as a kind of force of infection.

Combining (7.9) and (7.1) we obtain.

COROLLARY 7.5.

$$B(t) = \left(\frac{N(t)}{N(0)}\right)^{1/2}. \tag{7.10}$$

Before outlining how to pose and solve the initial value problem, we shall give, as promised, our second deterministic justification of the claim (5.3). To do so, we specialise to the case where, at time zero, susceptibles have only susceptible acquaintances. Then $\tilde{S}_j(t) = \bar{S}_j(t)$, and $N(0) = k\bar{S}_k(0)$ and (7.7) implies

$$\frac{N(\infty)}{N(0)} = B(\infty)(1 - Q + QB(\infty))^{k-1} \tag{7.11}$$

which, in combination with the limit of (7.10), yields the by now familiar (cf. (5.4) and (6.7)) equation

$$B(\infty) = (1 - Q + QB(\infty))^{k-1}. \tag{7.12}$$

Combining this equation with (7.6) we find (5.3).

For good measure we add a few remarks on the initial value problem. The general procedure is as follows. First one derives, as in Lemma 7.1, expressions for the various subpopulation sizes in terms of $B(t)$. Next these are used to express $N(t)$ in terms of past values of $B(t)$. Substitution of the result into (7.10) yields a nonlinear Stieltjes renewal equation for $B(t)$. We then define B as the unique solution of this equation, which turns all formulas into explicit expressions and justifies all our inferences.

8. CONCLUDING REMARKS

When individuals systematically contact the same individuals, the probability that an infectious individual has contact with an already infected individual is higher than it is otherwise. Thus the effectivity of the infectious output gets reduced.

This phenomenon manifests itself in spatial contact models, but there it is confounded by the phenomenon that neighbours have, as a rule, several common neighbours. To disentangle the influences of these two effects we have introduced a contact structure characterized by two properties:

(i) each individual has contact with precisely k other individuals,

(ii) these k so-called acquaintances form a random sample of the (infinite) population.

It is *not* a straightforward exercise to formulate a model for the spread of an infectious agent in a population with such a presumed contact structure (we openly admit to have made several false attempts before hitting upon the, according to our conviction, correct answer). In Section 7 we presented the continuous time formulation that finally resulted from our deliberations and in Section 6 we presented a, somewhat artificial, generation formulation.

The most important outcome of our analysis, besides an explicit expression for the basic reproduction ratio R_0, is a rather simple equation that determines the final size of the epidemic in a virgin population. A graphical representation of these quantities is presented in Figures 1–3 and from these one can deduce the quantitative aspects of the reduction in effectivity of infectious output due to repetition of contacts.

The present model concerns a homogeneous population: each individual has exactly the same properties. A more complicated structure arises when individuals are allowed to differ in the number of acquaintances. We may consider such a situation in future work.

A question that formed part of our original motivation is the following: if most contacts are with a rather limited number of fixed acquaintances, but some are with random members of the total population, how much does the outcome differ from the case where all contacts are with random members. This question is addressed in our companion paper [4].

Acknowledgement
The initial ideas of this paper were conceived during the authors visits to the Isaac Newton Institute in Cambridge, during the special half year on Epidemic Models organized by Bryan Grenfell, Denis Mollison and Valery Isham. We thank Thomas Hantke for help with the production of the figures.

References

1. Barbour, A. and Mollison, D. (1990). Epidemics and random graphs. In: Gabriel, J.-P., Lefèvre, C. and Picard, P. (eds.). *Stochastic Processes in Epidemic Theory*, Lect. Notes in Biomath. **86**, 86–89.
2. Blanchard, Ph., Bolz, G.F. and Krüger, T. (1990). Modelling AIDS-epidemics or any venereal disease on random graphs. In: Gabriel, J.-P., Lefèvre, C. and Picard, P. (eds.). *Stochastic Processes in Epidemic Theory*, Lect. Notes in Biomath. **86**, 104–117.
3. Bollobás, B. (1985). *Random Graphs*, Academic Press, London etc..
4. De Jong, M.C.M., Diekmann, O. and Metz, J.A.J. (in preparation).
5. Diekmann, O., Dietz, K. and Heesterbeek, J.A.P. (1991). The basic reproduction ratio for sexually transmitted diseases, part 1: theoretical considerations. Math. Biosc. **107**, 325–339.
6. Dietz, K. and Hadeler, K.P. (1988). Epidemiological models for sexually transmitted diseases, J. Math. Biol. **26**, 1–25.
7. Durrett, R. (1995). Spatial epidemic models. In: Mollison, D. (ed.). *Epidemic Models: Their Structure and Relation to Data*, Cambridge Univ. Press, 187–201.
8. Kermack, W.O. and McKendrick, A.G. (1927). Contributions to the mathematical theory of epidemics, part 1. Proc. Roy. Soc. Edinb. **A 115**, 700–721 (reprinted in Bull. Math. Biol. **53** (1991) 33–55.
9. Ludwig, D. (1975). Final size distributions for epidemics, Math. Biosc. **23**, 33–46.
10. Martin-Löf, A. (1986). Symmetric sampling procedures, general epidemic processes and their threshold limit theorems, J. Appl. Prob. **23**, 265-282.

11. Metz, J.A.J. (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.

12. Van Druten, J.A.M., de Boo, Th., Reintjes, A.G.M., Jager, J.C., Heisterkamp, S.H., Coutinho, R.A., Bos, J.M. and Ruitenberg, E.J. (1988). Reconstruction and prediction of spread of HIV infection in populations of homosexual men. In: Jager, J.C. and Ruitenberg, E.J. (eds.). *Statistical Analysis and Mathematical Modelling of AIDS*, Oxford Univ. Press, 52–76.

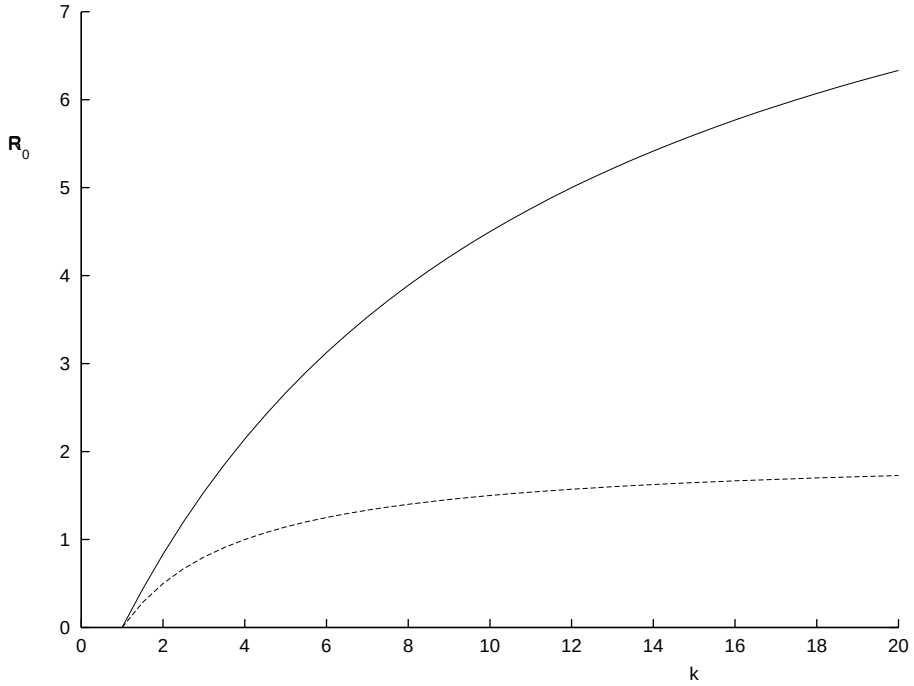

Figure 1. The reproduction ratio R_0 as a function of the number of acquaintances k(—– $\alpha/\gamma = 10$; – – – $\alpha/\gamma = 2$). Note that the acquaintances structure always leads to a reduction of R_0 (the limit for $k \to \infty$ equals α/γ, i.e. the value of R_0 with random contacts).

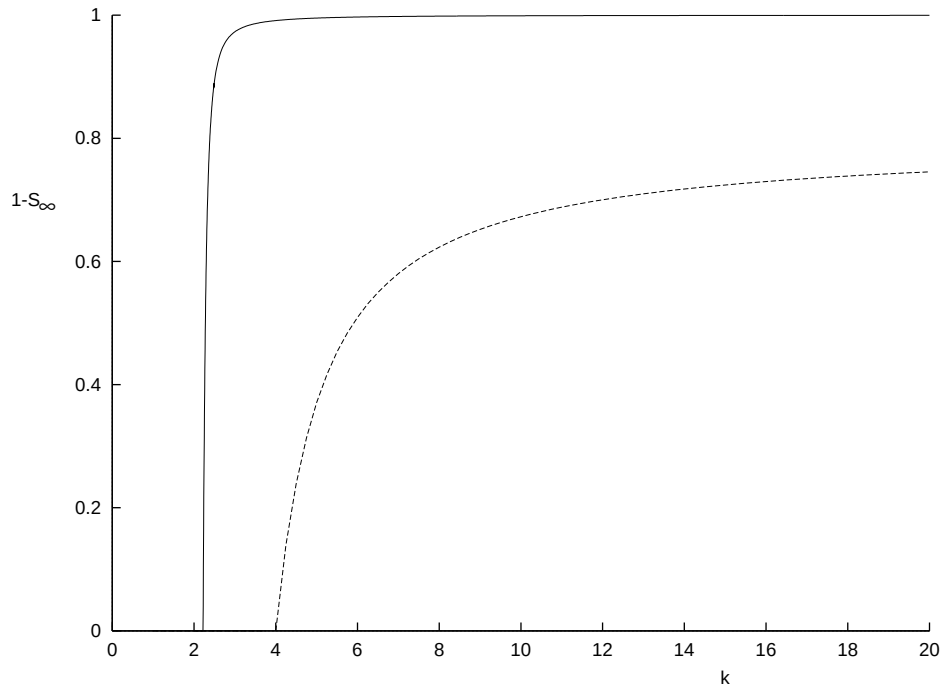

Figure 2. The final size $1 - s_\infty$ as a function of the number of acquaintances k. Note that the shift in the location of the critical value depends on the value of α/γ.