J. Appl. Prob. **35,** 651–661 (1998)
Printed in Israel

HETEROGENEITY IN EPIDEMIC MODELS AND ITS EFFECT ON THE SPREAD OF INFECTION

HÅKAN ANDERSSON,* *Stockholm University*
TOM BRITTON,** *Uppsala University*

Abstract

We first study an epidemic amongst a population consisting of individuals with the same infectivity but with varying susceptibilities to the disease. The asymptotic final epidemic size is compared with the corresponding size for a homogeneous population. Then we group a heterogeneous population into households, assuming very high infectivity within households, and investigate how the global infection pressure is affected by rearranging individuals between the households. In both situations considered, it turns out that whether or not homogenizing the individuals or households will result in an increased spread of infection actually depends on the infectiousness of the disease.

Keywords: Epidemic model; heterogeneous population; total size; household population

AMS 1991 Subject Classification: Primary 92D30

1. Introduction

This paper studies to some extent how the total size of an epidemic may be affected by introducing variability among the individuals. The underlying epidemic process (a special case of the model described in Ball and Clancy [5]) is a generalization of the so-called general epidemic model, and we are mainly concerned with two different situations. First, we assume that all the individuals are equally infectious but may have different susceptibility. The asymptotics for the total epidemic size is studied under the constraint that the population mean of the susceptibilities is kept fixed. Second, we group a multitype population into households, where we assume very high infectivity within households, and again draw some conclusions about the total size for different household formations. Here homogeneity refers to the case where all households have the same infectivity and the same susceptibility. As we shall see, this model may be treated as an epidemic in a multitype population *without* households. Our main goal is to quantify the following very loose statement.

If the disease is very contagious then homogenizing the population increases the size of the epidemic, while for a less infectious disease the largest epidemic arises in a heterogeneous setup.

It must be kept in mind that the results derived in Section 3, hinting at the above phenomenon, rely on the concept of *individual stochastic thresholds*, which is a purely mathematical construction without a clear immunological interpretation. On the other hand, the corresponding

Received 28 August 1996; revision received 10 January 1997.
* Postal address: Mathematical Statistics, Department of Mathematics, Stockholm University, S-106 91 Stockholm, Sweden. Email address: hakan@matematik.su.se.
** Postal address: Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden.
T.B. supported by The Bank of Sweden Tercentenary Foundation.

results for the household case treated in Sections 4 and 5 are less controversial – heterogeneous situations are created here by simply swapping or moving individuals between households. Yet another convincing example will be given in Andersson [1].

Let us mention a few other papers that investigate this kind of question from a theoretical point of view. Ball [2] and Becker and Marschner [9] announce interesting comparison results for some classical epidemic models. Related problems for the carrier-borne epidemic are discussed by Becker [8], Lefèvre and Malice [13] and Ball [3]. Marschner [14] considers epidemic models where the population is stratified into classes, the individuals having a preference for either within- or between-class contact. These models are compared with corresponding homogeneous ones. O'Neill [15] treats a model for a fatal disease, where individuals are members in high-risk or low-risk activity groups. Further, members of the high-risk group may change their behaviour by entering the low-risk group. This model is compared with the corresponding homogeneous one, considered for example by Ball and O'Neill [6]. Clancy [12] compares the distribution of the total size for a standard multigroup model with the corresponding distribution for a model (see Ball [4]) where infectives move from group to group and infect only within their current group.

Note carefully that our result partly contradicts the general opinion that the total size of an epidemic is always reduced when heterogeneities are introduced in the population (see for example [2], [3], [8], [13], [15]). We feel therefore that it is very important to gain further insight in how to compare epidemic models with each other.

2. The underlying epidemic model

Consider a closed population, each individual in the population being of one of ρ types. In the beginning we have ν_i susceptibles and η_i infectives of type i, $1 \le i \le \rho$. Let $\nu = \nu_1 + \cdots + \nu_\rho$ denote the total number of initial susceptibles, and also write $\pi_i = \nu_i/\nu$, $1 \le i \le \rho$. Following Ball and Clancy [5], we give each individual a label (i, r), where i denotes the type of the individual and $r = -(\eta_i - 1), -(\eta_i - 2), \ldots, -1, 0, 1, \ldots, \nu_i$ indicates the number of the individual in a type-specific list. The first η_i members of that list refer to the initial infectives and the subsequent members to the initial susceptibles. We assume that to each individual (i, r) is attached an infective period $I^{(i,r)}$ with distribution F_i and expected value μ_i, say. In case (i, r) becomes infected, she remains infective for a period $I^{(i,r)}$ and during that time makes contact with type j individuals according to a Poisson process with intensity $\lambda_{ij}\pi_j$, $1 \le j \le \rho$. Each time point of this process indicates a contact with an individual chosen at random among the ν_j initial susceptibles of type j. If the individual so contacted is still susceptible she becomes infected (and infective). All the infective periods and all the Poisson processes are assumed to be mutually independent.

Consider now a sequence of epidemic processes indexed by the population size ν. We are going to study an asymptotic situation where $\nu \to \infty$ in such a way that all proportions $\pi_i^{(\nu)}$ converge to strictly positive limits. Moreover, the numbers $\eta_i^{(\nu)}$ are kept constant independent of ν. Define $\tau_i^{(\nu)}$ to be the relative proportion of the initially susceptible type i individuals that become infected during the course of the epidemic. In [5] it is shown that the distribution of the vector $\tau^{(\nu)} = (\tau_1^{(\nu)}, \ldots, \tau_\rho^{(\nu)})$ converges as $\nu \to \infty$ to a distribution concentrated on the at most two solutions of the following non-linear system of equations:

$$1 - \tau_j = \exp\left\{-\sum_{i=1}^{m} \pi_i \lambda_{ij} \mu_i \tau_i\right\}, \qquad 1 \le j \le \rho, \tag{2.1}$$

where now of course the numbers π_i denote the limiting values. It is also shown that the largest eigenvalue R_0 of the matrix $(\pi_i \lambda_{ij} \mu_i)$ reflects the qualitative behaviour of the process in the sense that if $R_0 \leq 1$ then (2.1) has only the solution $\tau = \boldsymbol{0}$, whereas if $R_0 > 1$ then besides the trivial solution there always exists a strictly positive root. In this case we say that the epidemic is above threshold, and we say that a major epidemic has occurred if the final proportion infected is close to the positive root. In our analysis of the asymptotic total size, unless otherwise stated we always refer to the *largest* root of (2.1). At the end of Section 3, the probability of a major epidemic will be discussed. The quantity R_0 is called the *basic reproduction number*. From now on we will skip the index ν in order not to overburden notation. We also suppress the limits on the summation indices whenever possible.

In the sequel we will often assume that the contact rate $\lambda_{ij}\pi_j$ from type i individuals to type j individuals splits up into a product; i.e. $\lambda_{ij} = \alpha_i \beta_j$ for all i, j. The parameter α_i reflects how infectious and social the i-individuals are during their infective stage, and β_j measures how social and liable to infection susceptible j-individuals are. One can show that in this case $R_0 = \sum_i \pi_i \alpha_i \beta_i \mu_i$ (cf. [9]).

The following equivalent construction of the epidemic model described above is similar to that of Sellke [16]. We provide each individual (i, r) with an exponentially distributed random variable $Q^{(i,r)}$ having intensity β_i. The variables are mutually independent, and they are also independent of all the variables and Poisson processes defined above. The variable $Q^{(i,r)}$ is called the *individual stochastic threshold* for the individual (i, r). The initial infectives (i, r) are infective for a time $I^{(i,r)}$ and are then removed. At any time $t \geq 0$ the susceptible with label (j, r) accumulates 'exposure to infection' at rate $\sum_i \alpha_i Y_i(t)/\nu$, where $Y_i(t)$ denotes the number of infective type i individuals at time t. As soon as this exposure to infection reaches the threshold $Q^{(j,r)}$, the individual (j, r) becomes infected, and then stays infective for a time $I^{(j,r)}$. This construction is seen to be equivalent to our epidemic process (cf. [5]). The interpretation of $Q^{(i,r)}$ as a measure of the susceptibility of individual (i, r) (low threshold = high susceptibility) is crucial in what follows.

3. Individuals with varying susceptibilities

In what follows we want to compare a homogeneous population with a multitype population with varying susceptibility. Given a homogeneous ($\rho = 1$) epidemic process, i.e. given parameters (ν, η, F, λ), it is unfortunately not at all clear what the 'correct' equivalent multitype processes to compare with should look like. In fact, there are usually several answers depending on the phenomenon one wishes to focus on. In this section we assume that all the infectious periods are identically distributed (in particular $\mu_i = \mu$ for all i) and also that $\lambda_{ij} = \alpha_i \beta_j$ where α_i are equal to α for all i, so all that differs are the susceptibility rates. Below we will suggest one way to construct comparable heterogeneous processes.

Under the above restriction, the fundamental equation (2.1) becomes

$$1 - \tau_j = \exp\left\{-\beta_j \alpha \mu \sum_i \pi_i \tau_i\right\}, \qquad 1 \leq j \leq \rho,$$

hence multiplying by π_j, summing over j and writing $\tau = \sum_i \pi_i \tau_i$ yields

$$1 - \tau = \sum_j \pi_j \exp\{-\beta_j \alpha \mu \tau\}. \tag{3.1}$$

Obviously, τ is the overall total size of the epidemic. In Ball [2] this quantity is compared with the corresponding total size, τ_* say, obtained by running a homogeneous process with the weighted mean of the old contact rates as the new contact rate:

$$\beta = \sum_i \pi_i \beta_i.$$

Ball shows that $\tau_* \geq \tau$ with equality if and only if all the numbers β_i are equal. In fact, the results in [2] are remarkably precise; in short they state that the homogeneous counterpart gives a stochastically larger accumulated number of infectives for each time point t and for all population sizes ν. We may gain some insight into this phenomenon by looking at expected thresholds (cf. Section 2). Write $Q^{(i,r)}$ for the individual stochastic thresholds of the heterogeneous model and $Q_*^{(r)}$ for the thresholds in the homogeneous case, and define

$$\Phi = \frac{1}{\nu} E\left(\sum_i \sum_{r=1}^{\nu_i} Q^{(i,r)}\right) \quad \text{and} \quad \Phi_* = \frac{1}{\nu} E\left(\sum_{r=1}^{\nu} Q_*^{(r)}\right)$$

to be the corresponding population means of the expected thresholds. Since all the Q-variables are exponentially distributed, we have

$$\Phi = \sum_i \frac{\pi_i}{\beta_i} \geq \left(\sum_i \pi_i \beta_i\right)^{-1} = \frac{1}{\beta} = \Phi_*$$

because the arithmetic mean always dominates the harmonic mean, thus the homogeneous population as a whole is actually made more susceptible to the disease than the heterogeneous counterpart. The rate β is chosen so that *initially* the two epidemics spread at the same rate, but this breaks down as soon as a positive fraction of the individuals have become infected.

The following problem now suggests itself. Given a contact rate $\beta > 0$, compare the resulting homogeneous epidemic process with the multitude of multitype epidemics having susceptibilities $\beta_1, \ldots, \beta_\rho$ and relative proportions $\pi_1, \ldots, \pi_\rho$ satisfying

$$\frac{1}{\beta} = \sum_i \frac{\pi_i}{\beta_i}, \tag{3.2}$$

because all these multitype epidemics will then have the same population mean of the expected thresholds as the homogeneous epidemic. We have the following result.

Observation 1. *Define $\beta_c = 2/(\alpha\mu(1 - e^{-2}))$. If $\beta \geq \beta_c$ then the homogeneous epidemic process, given by $\hat{\beta} = (\beta, \ldots, \beta)$, yields the largest total size. If $\beta < \beta_c$ then the largest outbreak is obtained in a heterogeneous case.*

To prove this, we fix β and $\pi_1, \ldots, \pi_\rho$ and define for each given $\beta = (\beta_1, \ldots, \beta_\rho)$ obeying (3.2) the function

$$f(\beta, t) = \sum_i \pi_i \exp\{-\beta_i \alpha\mu t\}, \qquad t \geq 0.$$

The total size $\tau(\beta)$ is then the largest root of the equation $1 - t = f(\beta, t)$ (see Equation (3.1)). In particular, in the homogeneous case we have, writing $\tau = \tau(\hat{\beta})$,

$$1 - \tau = \exp\{-\beta\alpha\mu\tau\}. \tag{3.3}$$

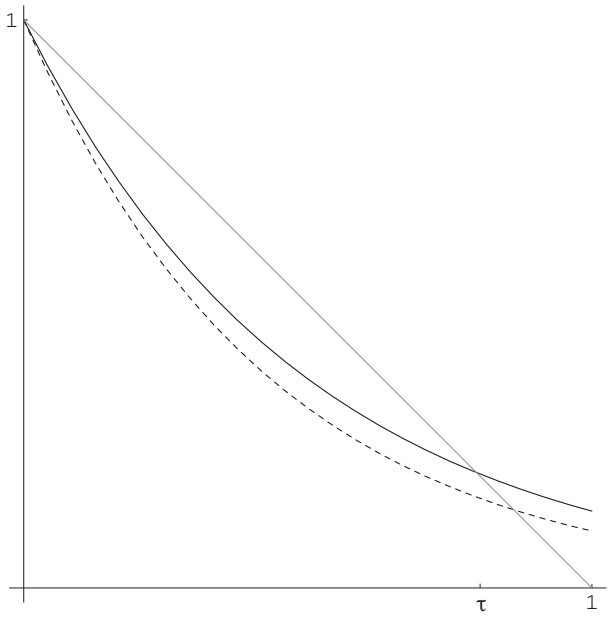

FIGURE 1: Perturbation of $\hat{\beta}$ and its impact on the crossing point τ.

In order to make an analysis of the behaviour of $\tau(\beta)$ for small departures $\beta = \hat{\beta} + \delta$ from $\hat{\beta}$, we note that if $f(\beta, \tau) < f(\hat{\beta}, \tau)$ then the root $\tau(\beta)$ obviously satisfies $\tau(\beta) > \tau$ (see Figure 1), the statement with reversed inequality signs being equally valid. Denote by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ the inner product $\sum_i u_i v_i$ of the vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^\rho$. Also, let $|\boldsymbol{u}|$ denote the norm $\sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$ of $\boldsymbol{u}$. Since

$$\sum_i \frac{\pi_i}{\beta_i} = \frac{1}{\beta} - \frac{1}{\beta^2}\langle \pi, \delta \rangle + \frac{1}{\beta^3}\langle \pi, \delta^2 \rangle + o(|\delta|^2),$$

we must have $\langle \pi, \delta^2 \rangle - \beta \langle \pi, \delta \rangle = o(|\delta|^2)$ in order to stay sufficiently close to the surface S prescribed by (3.2) (here δ^2 denotes the vector $(\delta_1^2, \ldots, \delta_\rho^2)$). Taylor expansion then yields

$$\begin{aligned} f(\beta, \tau) &= f(\hat{\beta}, \tau) + \alpha\mu\tau \exp\{-\beta\alpha\mu\tau\}(-\langle \pi, \delta \rangle + \tfrac{1}{2}\alpha\mu\tau \langle \pi, \delta^2 \rangle) + o(|\delta|^2) \\ &= f(\hat{\beta}, \tau) + \alpha\mu\tau \exp\{-\beta\alpha\mu\tau\}\langle \pi, \delta^2 \rangle(\tfrac{1}{2}\alpha\mu\tau - 1/\beta) + o(|\delta|^2). \end{aligned}$$

We see that the sign of $\tau(\alpha\mu\tau/2 - 1/\beta)$ determines how the total size changes for small departures from $\hat{\beta}$. According to (3.3) we have $\alpha\mu\tau/2 - 1/\beta = 0$ exactly when $1 - 2/(\beta\alpha\mu) = \mathrm{e}^{-2}$, i.e. when $\beta = \beta_c$. We can sum up the results as follows.

If $0 < \beta \leq 1/(\alpha\mu)$ then $\tau = 0$ (see (3.3)), and it easy to see that we can find β with $\tau(\beta) > 0$. Indeed, the basic reproduction number R_0 can easily be pushed above 1, e.g. by choosing one of the components β_i to be very large. If $1/(\alpha\mu) < \beta < \beta_c$ then $\tau > 0$ and perturbations of $\hat{\beta}$ always lead to an increasing total size. Finally, if $\beta \geq \beta_c$ then we have a *local* maximum of $\tau(\beta)$ at the point $\hat{\beta}$. In fact, $\tau(\beta)$ has its *global* maximum at $\hat{\beta}$ which we now show. Assume for contradiction that the maximum is not global, i.e. there exists a point on the surface S with even larger total size. Pick a smooth curve γ on S with these two extreme points as endpoints. Somewhere along this curve, at the point β^* say, the total size takes the minimal value $\tau^* = \tau(\beta^*)$. But this is impossible, since Taylor expansion shows that along

TABLE 1: Maximal total size for different values of β

β	(β_1, β_2)	τ
0.50	(14.5, 0.25)	0.57
1.00	(9.5, 0.53)	0.64
1.50	(6.40, 0.85)	0.73
2.00	(4.14, 1.32)	0.81
2.31	(2.31, 2.31)	0.86
2.50	(2.50, 2.50)	0.89

the tangent $\boldsymbol{\nu}$ to γ at β^*,

$$f(\beta^* + \varepsilon\boldsymbol{\nu}, \tau^*) > f(\beta^*, \tau^*)$$

if ε is small enough, giving even smaller values of the total size in a neighbourhood of β^*.

Table 1 gives the maximal value of the total size and the corresponding optimal point along the curve described by (3.2) for a range of parameter values β. Here $\rho = 2$, $\alpha = \mu = 1$ and $\pi_1 = \pi_2 = 1/2$. We have included the value $\beta_c \approx 2.31$. Note how the pairs (β_1, β_2), giving the maximal value of τ, approach the diagonal as β increases. Also note that the value of the total size for $\beta = \beta_c$ is extremely high – for all realistic values the optimum is to be found at a point off the diagonal. (Obviously, interchanging the coordinates of (β_1, β_2) also gives an optimal point.)

Let us make one final remark. It is important to consider not only the final epidemic size in case of a large outbreak, but also the *probability* of a large outbreak along the surface S. Fix β, $\{\pi_i\}$ and $\{\beta_i\}$ as before. From Becker and Marschner [9] we may extract the following equation for the probabilities θ_i of a large epidemic given a single initial infective of type i, $1 \le i \le \rho$:

$$1 - \theta_i = E\left[\exp\left\{-\alpha I \sum_j \pi_j \beta_j \theta_j\right\}\right], \qquad 1 \le i \le \rho,$$

where the infectious period I has distribution F. Now, it is natural to assume that the probability of the initial infective being a type i individual is proportional both to the number of type i individuals and to the susceptibility β_i, i.e. the unconditional explosion probability is given by $\theta = \sum_i \pi_i \beta_i \theta_i / \langle \pi, \beta \rangle$. It follows that

$$1 - \theta = E[\exp\{-\alpha I \langle \pi, \beta \rangle \theta\}].$$

We have previously observed that $\beta = (\sum_i \pi_i / \beta_i)^{-1} \le \langle \pi, \beta \rangle$. This implies that the probability of a large outbreak is *always* minimal in the homogeneous setup, even in the case of a very infectious disease where, as we have seen, the total size in case of an outbreak is *maximized* for the homogeneous case.

4. An epidemic model for a multitype population with households

The formation of households is obviously a very important feature to take under consideration when studying the spread of an epidemic among a human population. Only recently, however (by Ball *et al.* [7]), has a thorough probabilistic analysis of an epidemic model with households, assuming homogeneous individuals, been performed. Britton [11] also treats the

case with households of homogeneous individuals and constructs a statistical test to detect if the disease transmits at a higher rate within households. Becker and Dietz [10] study the other extreme, meaning that they assume that *all* household members become infected once the disease has entered the household – an assumption we adopt in this and the next section. Here we extend the model of Section 2 to an epidemic in a population built up of households of individuals, assuming very high infectivity within the households. By treating the households as 'macro individuals' it is shown that this model falls under a multitype epidemic model *without households* defined by Ball and Clancy [5].

Start with the model defined in Section 2 and adopt the notation of that section. On top of this we assume that each individual belongs to exactly one household and that once someone in a household is infected she immediately infects all other household members (see the end of Section 5 for comments on this assumption). Once a household is infected each of the individuals has contact with individuals outside the household according to the rules defined in Section 2. Beside different individuals we now also have different types of households. Assume there are r different household structures labelled $1, \dots, r$, and let k_i^h denote the number of i-individuals in h-households (households will be indexed by g and h as opposed to individuals where i and j are used). Let n_h denote the number of h-households and $p_h = n_h/n$ the corresponding proportion, where $n = \sum_h n_h$ is the total number of households. Further, we define $\gamma = \nu/n$ to be the average household size. The asymptotics we will consider is where γ is finite and n tends to infinity. For this notation to be consistent with the notation of Section 2 it follows that $\nu_i = \sum_h k_i^h n_h$.

In this new model we may treat a household as an individual since all individuals become infected at the same time. However, this 'individual' does not make infective contacts at a constant rate during its infective stage, as was assumed in the model of Section 2. The reason for this is that, unless all infective periods are degenerate with equal constant values, after some time some individuals in the household will remain infective whereas others are removed. The rate of infective contacts decreases each time someone in the household is removed, until all individuals in the household are removed – when this happens the whole household is removed. Fortunately, Ball and Clancy [5] study a model in which infective individuals may change state. All that is required of the random process governing the state-changes is that it is independent of the remaining population – an assumption which is fulfilled in the present case since all infective periods were defined to be independent of everything else. If the model did not assume immediate infection of all household members, then this would imply that individuals could get infected from someone outside the household while the disease was spreading within the household. But then the random process changing infective state would be *dependent* on the remaining population and the results in [5] could not be applied.

An important quantity is the expected number ℓ_{gh} of h-households that one g-household would infect if all h-households were susceptible. The following relation holds:

$$\ell_{gh} = E\left(\sum_i \left((I^{(i,1)} + \cdots + I^{(i,k_i^g)}) \sum_j \pi_j \lambda_{ij} \frac{n_h k_j^h}{\nu_j} \right)\right). \tag{4.1}$$

A g-household can infect an h-household by any type of individual in the g-household having contact with any type of individual from the h-household. This explains the summation over i and j. The factor $(I^{(i,1)} + \cdots + I^{(i,k_i^g)})$ is the cumulative infective period of all i-type individuals in a g-household. While infective, an i-individual has contact with j-individuals at rate $\pi_j \lambda_{ij}$, but only the fraction $n_h k_j^h / \nu_j$ of these individuals belong to h-households. This

explains Equation (4.1). After some simple algebra the expression becomes

$$\ell_{gh} = \frac{p_h}{\gamma} \sum_{i,j} k_i^g k_j^h \lambda_{ij} \mu_i. \tag{4.2}$$

Under the assumption (cf. Section 2) that the contact rates between individuals split up into products $\lambda_{ij} = \alpha_i \beta_j$, Equation (4.2) can be written $\ell_{gh} = (p_h/\gamma) a_g b_h$, where $a_g = \sum_i k_i^g \alpha_i \mu_i$ is ν times the average infection pressure caused by a g-household and $b_h = \sum_j k_j^h \beta_j$ measures the susceptibility of an h-household.

5. Adding individuals to households

In this section we compare the contribution of 'infection pressure' from different household formations. The comparison is performed by swapping or moving individuals between households and we derive the necessary criteria on the households in order to get a larger/smaller contribution to the pressure by a swap. It turns out that whether or not a swap will cause a larger contribution often depends on the remaining population and in particular the accumulated infection pressure it has caused.

Assume as before that the contact rates between individuals satisfy $\lambda_{ij} = \alpha_i \beta_j$ and that the population consists of a large number of households. We concentrate on two of the households, one of type g and one of type h, and treat the remaining population as a black box. Assume two 'new' individuals, one i- and one j-individual, have to join the two households – one in each household. Which of the two possible ways to do this will, on the average, result in a larger contribution to the global infection pressure?

To address the question posed we will consider the random graph corresponding to the epidemic. It is well known that a suitably defined directed random graph can model the final outcome of more or less any epidemic model – the existing directed components indicate transmission if the host gets infected. By neglecting the arrows to and from household g and h we may study the epidemic that would appear if g and h did not belong to the population. Assume the remaining population has caused the infection pressure A, where the infection pressure caused by an i-individual with an infective period of length $I^{(i)}$ is defined as $\alpha_i I^{(i)}/\nu$. Given A, we now compute the *expectation* of the additional infection pressure, X say, obtained by entering the two households into the population, that is no longer neglecting the arrows to and from the two households. The reason for choosing *expectation* as the quantity of interest would perhaps be better motivated if many such pairs of households were entered simultaneously. Let I denote the case where i joins household g and consequently that j joins household h. Conditioning upon which of the households that get infected by the pressure A, we have

$$\begin{aligned}
E_A^{\mathrm{I}}(X) &= E_A^{\mathrm{I}}(X \mid \text{neither}) P_A^{\mathrm{I}}(\text{neither}) + E_A^{\mathrm{I}}(X \mid g) P_A^{\mathrm{I}}(g) \\
&\quad + E_A^{\mathrm{I}}(X \mid h) P_A^{\mathrm{I}}(h) + E_A^{\mathrm{I}}(X \mid g \text{ and } h) P_A^{\mathrm{I}}(g \text{ and } h) \\
&= 0 + \frac{a_g + \alpha_i}{\nu} (1 - \exp\{-(b_g + \beta_i)A\}) \exp\{-(b_h + \beta_j)A\} \\
&\quad + \frac{a_h + \alpha_j}{\nu} \exp\{-(b_g + \beta_i)A\} (1 - \exp\{-(b_h + \beta_j)A\}) \\
&\quad + \frac{a_g + a_h + \alpha_i + \alpha_j}{\nu} (1 - \exp\{-(b_g + \beta_i)A\}) (1 - \exp\{-(b_h + \beta_j)A\}) \\
&\quad + o(\nu^{-1}).
\end{aligned}$$

The remainder on the right-hand side comes from the small chance that household g infects household h and vice versa. Denote the second alternative by II, i.e. i belongs to household h and j to g. The corresponding expected additional force $E_A^{\mathrm{II}}(X)$ is computed similarly. Now define $\Delta(A) = \lim_{\nu\to\infty} \nu(E_A^{\mathrm{I}}(X) - E_A^{\mathrm{II}}(X))$. Simple algebra yields

$$\begin{aligned}\Delta(A) = {} & (a_g + \alpha_j)\exp\{-(b_g + \beta_j)A\} + (a_h + \alpha_i)\exp\{-(b_h + \beta_i)A\} \\ & - (a_g + \alpha_i)\exp\{-(b_g + \beta_i)A\} - (a_h + \alpha_j)\exp\{-(b_h + \beta_j)A\}. \end{aligned} \tag{5.1}$$

We show next that the sign of Δ can change at most once for positive A. This is equivalent to proving that $f(A)$ has at most one strictly positive root, where

$$f(A) = c_1\exp\{-d_1A\} + c_2\exp\{-d_2A\} - c_3\exp\{-d_3A\} - c_4\exp\{-d_4A\},$$

such that $c_1 + c_2 = c_3 + c_4$ and $d_1 + d_2 = d_3 + d_4$, all coefficients being positive. Assume without loss of generality that $d_4 \geq d_i$ for all i which implies that $d_3 \leq d_i$ for all i. Define

$$\begin{aligned} g(A) &= \exp\{d_3A\}(d_4f(A) + f'(A)) \\ &= c_1(d_4 - d_1)\exp\{-(d_1 - d_3)A\} + c_2(d_4 - d_2)\exp\{-(d_2 - d_3)A\} - c_3(d_4 - d_3). \end{aligned} \tag{5.2}$$

From the first line of (5.2) we see that g and f' have the same sign at the roots of f. From the second line of (5.2) we note that g is decreasing in A due to the assumptions on d_i. This implies that g changes sign at most once. But f' changes sign between adjacent roots of f, so these two observations imply that f has at most two roots and, since 0 is a root, consequently at most one strictly positive root.

Since the function Δ has at most one strictly positive root it is of particular interest to compute $\Delta(A)$ for small positive and large positive A. The critical value A_c where Δ changes sign, if it ever does, can only be solved numerically. It is not hard to show that

$$\Delta(A) = [(a_g - a_h)(\beta_i - \beta_j) + (b_g - b_h)(\alpha_i - \alpha_j)]A + o(A), \tag{5.3}$$

so the sign of the first order term determines which alternative gives the largest average additional infection pressure when A is small. For large A, the sign of $\Delta(A)$ is identical to the sign of

$$\min\{b_g + \beta_i,\ b_h + \beta_j\} - \min\{b_g + \beta_j,\ b_h + \beta_i\}, \tag{5.4}$$

except when either (5.4) equals 0 or at least one of the households *and* one of the individuals has no infectivity at all.

Assume that (5.4) and the first order term in (5.3) are both non-zero. Let us also assume that

$$a_g \geq a_h,\ b_g \geq b_h; \qquad \alpha_i \geq \alpha_j,\ \beta_i \geq \beta_j.$$

We believe that this situation is most common in applications. In particular, this is always true if $\alpha_i = c\beta_i$, implying that infectivity and susceptibility depend only on social activity. Another example is where we have homogeneous individuals belonging to households of varying sizes, and where individual j is fictitious, that is $\alpha_j = \beta_j = 0$ (this means that individual i shall join either of two households of different size). By using (5.3) and (5.4) we easily conclude that $\Delta(A)$ is positive for small A and negative for large A (see Figure 2 for a particular case). We have thus shown the following.

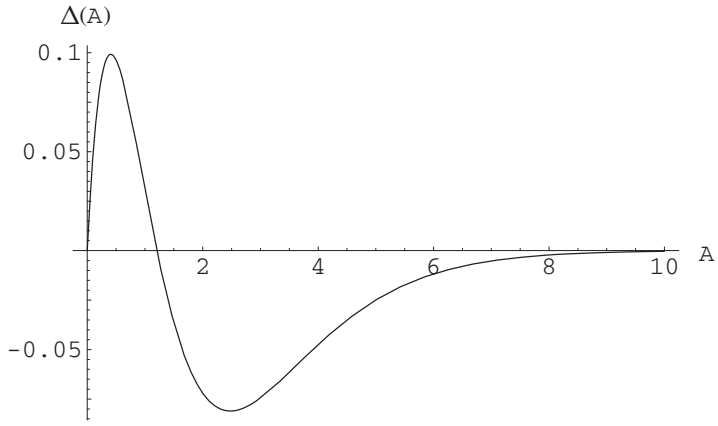

FIGURE 2: Δ as a function of A when $a_g = 10$, $a_h = 8$, $b_g = 1$, $b_h = 0.8$, $\alpha_i = 4$, $\alpha_j = 3$, $\beta_i = 0.4$, $\beta_j = 0.2$.

Observation 2. *Consider an infectious disease with high infectivity within households. Let the infection pressure be A. Then, under the conditions above, there exists* A_c *such that if* $A > A_c$ *then the largest contribution to the infection pressure is obtained when households are homogenized, whereas if* $A < A_c$ *the largest additional pressure is obtained when households are heterogenized.*

We have, for technical reasons, assumed that once someone in a household is infected this individual infects *all* other household members. This is of course not true in real life. However, for infectious diseases with high infectivity upon contact such as measles, it is approximately true if we consider only individuals who are susceptible when the epidemic starts. Further, the household model can be derived as the limit of a sequence of epidemic models with within-household infection rates increasing to infinity. For this reason the statement above should hold also when the infection rate within the household is not extremely high.

References

[1] ANDERSSON, H. (1998). Limit theorems for a random graph epidemic model. To appear in *Ann. Appl. Prob.*

[2] BALL, F. (1985). Deterministic and stochastic epidemics with several kinds of susceptibles. *Adv. Appl. Prob.* **17**, 1–22.

[3] BALL, F. (1990). A new look at Downton's carrier-borne epidemic model. In *Stochastic Processes in Epidemic Theory*, Eds J.-P. Gabriel, C. Lefèvre and P. Picard. (*Lecture Notes in Biomathematics* **86**.) Springer, Berlin, pp. 71–85.

[4] BALL, F. (1991). Dynamic population epidemic models. *Math. Biosci.* **107**, 299–324.

[5] BALL, F. AND CLANCY, D. (1993). The final size and severity of a generalised stochastic multitype epidemic model. *Adv. Appl. Prob.* **25**, 721–736.

[6] BALL, F. AND O'NEILL, P. (1993). A modification of the general stochastic epidemic motivated by AIDS modelling. *Adv. Appl. Prob.* **25**, 39–62.

[7] BALL, F., MOLLISON, D. AND SCALIA-TOMBA, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7**, 46–89.

[8] BECKER, N. G. (1973). Carrier-borne epidemics in a community consisting of different groups. *J. Appl. Prob.* **10**, 491–501.

[9] BECKER, N. G. AND MARSCHNER, I. (1990). The effect of heterogeneity on the spread of disease. In *Stochastic Processes in Epidemic Theory*, Eds J.-P. Gabriel, C. Lefèvre and P. Picard. (*Lecture Notes in Biomathematics* **86**.) Springer, Berlin, pp. 90–103.

[10] Becker, N. G. and Dietz, K. (1995). The effect of household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127**, 207–219.
[11] Britton, T. (1997). Tests to detect clustering of infected individuals within families. *Biometrics* **53**, 98–109.
[12] Clancy, D. (1994). Some comparison results for multitype epidemic models. *J. Appl. Prob.* **31**, 9–21.
[13] Lefèvre, C. and Malice, M.-P. (1988). Comparisons for carrier-borne epidemics in heterogeneous and homogeneous populations. *J. Appl. Prob.* **25**, 663–674.
[14] Marschner, I. (1992). The effect of preferential mixing on the growth of an epidemic. *Math. Biosci.* **109**, 39–67.
[15] O'Neill, P. (1995). Epidemic models featuring behaviour change. *Adv. Appl. Prob.* **27**, 960–979.
[16] Sellke, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Prob.* **20**, 390–394.