Partially mixed household epidemiological model with clustered resistant individuals

David E. Hiebeler* and Amanda Keck Criner
Department of Mathematics and Statistics, 333 Neville Hall, University of Maine, Orono, Maine 04469-5752, USA
(Received 24 May 2006; revised manuscript received 30 August 2006; published 21 February 2007)

We study the dynamics of the spread of an infectious disease within a population partitioned into households, and stratified into resistant and nonresistant individuals. Variability in the level of resistance between households increases the initial rate of spread of the infection, as well as the infection level at the endemic equilibrium. This phenomenon is seen even when all individuals in the population are equally likely to be resistant, and can also be predicted by including spatial clustering of resistant individuals within an improved mean-field approximation.

DOI: 10.1103/PhysRevE.75.022901 PACS number(s): 87.19.Xx, 87.23.Ge, 89.65.−s

Growing evidence from both empirical and theoretical studies shows that the spatial structure of hosts affects the dynamics of infectious diseases among both human and wildlife populations [1–5], as well as the spread of malicious software among computers or other information within social networks [6–18]. The spatial arrangement of individuals, and the frequency and spatial scale of interactions between individuals may affect the dynamics just as much as fundamental parameters such as contact and recovery rates [19–21]. Similarly, in ecological systems, the presence of spatially clustered uninhabitable sites and the spatial scale of dispersal strongly affect population dynamics [22,23]. In epidemiology, the importance of host connectivity has been observed in a variety of topologies, including lattices, small-world and scale-free networks [8,24–27]. In this paper, we examine the dynamics of an infectious disease in a population with localized hierarchical structure. Because real biological and social networks include variability in susceptibility, our model also includes resistant individuals which are distributed in a clustered fashion.

Consider a continuous-time SIS-R epidemiological model applied to a fixed population of N individuals partitioned into n_2 groups referred to as households, although the groups may represent cities, dormitories within a university, or other sectors of a population. Each household contains n_1 individuals. Some individuals in the population are geographically fixed and resistant to the infectious disease being considered, while other individuals are highly mobile, and susceptible to infection. One example of such a structured population would be the case where households represent cities; the population is divided into a relatively wealthy class which is resistant to infection due to better living conditions and health care, and a class of relatively impoverished migrant workers who often move from one region to another and who are susceptible to infection.

Denote the proportion of individuals in household k who are resistant by R_k. Individuals may also be susceptible (S_k) or infectious (I_k), and note that $S_k+I_k+R_k=1$. Each infectious individual contacts and attempts to infect others at rate $\phi>0$, with the amount of mixing between households determined by the parameter α. Each contact is with an individual chosen at random from among the entire population with probability α, and otherwise with an individual chosen at random from within the same household. When an infectious individual contacts a susceptible individual, the latter becomes infectious; contacting another infectious individual or a resistant individual has no effect. Finally, each infectious individual independently recovers to the susceptible state at rate $\mu>0$. This is therefore essentially an SIS model, but with the presence of fixed resistants and partitioning into households; this is a generalization of a previous household epidemiological model [26] combined with the spatially distributed unsuitable habitat (in the form of resistant individuals) from prior ecological models [22,23]. This model differs from previous network or lattice-based systems, such as those in Refs. [3,8,9], because here every individual is connected with every other individual (i.e., it is a fully connected graph) but the probabilities or rates of attempting infections along different connections are not equal.

Taking the expectation over stochastic realizations of the system, as will be done implicitly throughout development of the model, yields the continuous-state differential equation for I_k

$$\frac{dI_k}{dt} = E[I]\phi\alpha S_k + I_k\phi(1-\alpha)S_k - \mu I_k, \tag{1}$$

where the three terms represent between-household infection, within-household infection, and recovery, respectively. Note that the term $E[I]$ is a population mean taken over all households, $E[I]=\frac{1}{n_2}\sum_{k=1}^{n_2} I_k$, representing the probability that a randomly chosen individual from the entire population is infectious. The differential equation for S_k is simply $dS_k/dt = -dI_k/dt$, and $dR_k/dt=0$ because resistant individuals are fixed.

To construct an analytically tractable model, take the expectation over households of Eq. (1), which gives

$$\frac{dE[I]}{dt} = \phi(\alpha E[I]E[S] + (1-\alpha)E[IS]) - \mu E[I]. \tag{2}$$

The standard mean-field approximation for models of this type assumes that all individuals are well-mixed among the population and there are no spatial correlations in any state variables, which implies $E[IS]\approx E[I]E[S]$. This simplifies the above differential equation into the following form, drop-

*Electronic address: hiebeler@math.umaine.edu

ping the explicit "$E[\cdot]$" notation for simplicity:

$$\frac{dI}{dt} = \phi\alpha IS + (1-\alpha)\phi IS - \mu I = I(\phi S - \mu)$$

$$= I(\phi(1-R-I) - \mu).$$

Notice that α cancels from the model, as the mean-field approximation neglects all spatial information; assuming complete mixing is equivalent to assuming $\alpha=1$. This model is a continuous-time logistic equation which may be written in the standard form $dx/dt=rx(1-x/K)$. The initial rate of spread of the infection will be $r=\phi(1-R)-\mu$, the long-term (equilibrium) proportion of infectious individuals will be $K=1-R-\mu/\phi$, and therefore proportion $\tilde{K}=1-\mu/[\phi(1-R)]$ of nonresistant individuals will be infectious at equilibrium, assuming $\mu/\phi<1-R$. If the latter condition is violated, the model predicts that the infection will fail to spread, and the system will reach a steady-state distribution in which all nonresistants will be susceptible. The basic reproductive number for this version of the model is $R_0=\phi(1-R)/\mu$, and the condition for persistence of the infection does correspond to $R_0>1$.

Now define Q as the probability that if a randomly chosen nonresistant individual selects an individual within the same household (a "housemate"), the selected housemate will also be nonresistant. Q is a measure of the clustering among nonresistant individuals. Following [26], this clustering Q is given by

$$Q = \frac{E[(1-R)^2]}{E[1-R]},$$

where the expectations are taken over households.

Next, we will develop an improved mean-field approximation by including the (fixed) spatial structure of resistant individuals; it can be interpreted as a local-dispersal as opposed to infinite-dispersal mean-field approximation [28,29]. To approximate $E[IS]$ in Eq. (2), observe that because of the definition of S_k and I_k, $E[IS]$ can be interpreted as the probability that among two individuals chosen independently at random from the same randomly selected household, the first is infectious and the second is susceptible. This may be broken into the product of the marginal probability that the first individual is infectious, and the conditional probability that the second is susceptible given that the first is infectious. The latter may be further broken down into the product of the probabilities that the second individual is nonresistant, and that it is susceptible given that it is nonresistant. The above expansions, together with the assumption that susceptible and infectious individuals are well mixed, yields the relation $E[IS]=E[I]QE[S]/(1-E[R])$, which when substituted into Eq. (2) yields

$$\frac{dI}{dt} = I\left(\phi(1-R-I)\left[\alpha + \frac{(1-\alpha)Q}{1-R}\right] - \mu\right). \tag{3}$$

Observe that this improved mean-field approximation now includes α, as well as information about both the overall density and clustering of individuals, both resistant and nonresistant.

The improved mean-field approximation given by Eq. (3) is still in the form of a continuous-time logistic equation, where the initial rate of spread of the infection is given by $r=\phi\alpha(1-R)+\phi(1-\alpha)Q-\mu$ and the equilibrium proportion of nonresistant individuals who are infectious at equilibrium is $\tilde{K}=1-\mu[\phi\alpha(1-R)+\phi(1-\alpha)Q]^{-1}$. Both of these quantities increase as ϕ increases and as μ decreases, as expected. But now the spatial clustering as well as the overall level of resistance in the population also affect the initial rate of spread and the long-term level of infection. Because infections are more likely to initially arise in households with large numbers of nonresistants (assuming all susceptibles are equally likely to be the initial source of the infection), increasing either the clustering Q of nonresistants or the amount of intrahousehold contacts $1-\alpha$ increases r. This effect persists even to the steady-state distribution, i.e., increasing clustering Q or intrahousehold contacts $1-\alpha$ also increases the proportion $\tilde{K}$ of nonresistants infected at the steady-state, as long-distance (interhousehold) contacts are more likely to encounter resistant individuals. The results involving α can be derived mathematically from the model using the fact that Hölder's inequality implies $Q \geqslant E[1-R]$, which in turn implies $\partial r/\partial\alpha \leqslant 0$ and $\partial\tilde{K}/\partial\alpha \leqslant 0$ (with equality only if $Q=E[1-R]$, which occurs when all households have the same number of resistant individuals). The basic reproductive number for the improved model is $R_0=\phi[\alpha(1-R)+(1-\alpha)Q]/\mu$, and the endemic equilibrium is biologically feasible (satisfies $\tilde{K}>0$) when $R_0>1$.

One noteworthy feature of this model is that the response of r and $\tilde{K}$ to changes in parameters deviate from predictions of the ordinary mean-field approximation even when the resistants are distributed according to what would typically be considered a null model, i.e., every individual in the population is resistant independently with a fixed probability $E[R]$. In that case, the number of resistants (and nonresistants) in each household follows a binomial distribution, perhaps approximated by a Poisson if the number of households is large and frequency of resistance is low. Because $Q>E[1-R]$ in this case, both r and $\tilde{K}$ are elevated above what their levels would be in a population where all households have the same number of resistants, in which $E[(1-R)^2]=(E[1-R])^2$ and therefore $Q=E[1-R]$.

The above conclusions regarding the initial rate of spread r of the disease are still valid even if the assumption of mixing among susceptible and infectious individuals is relaxed. Removing this assumption changes the transient dynamics and steady-state distribution of the system, but not the initial short-term behavior. As long as infections are equally likely to initially appear among all susceptibles with equal probability, the behavior of the system as a perturbation near the disease-free equilibrium will be dominated by the distribution of resistant hosts. This differs from lattice models, where spatially localized clustering immediately dominates the dynamics [30]. In the model studied here, assuming the number of individuals per household n_1 is large, because individuals within a household all interact equally each household behaves like a standard epidemiological system with mass action dynamics with additional infection at-

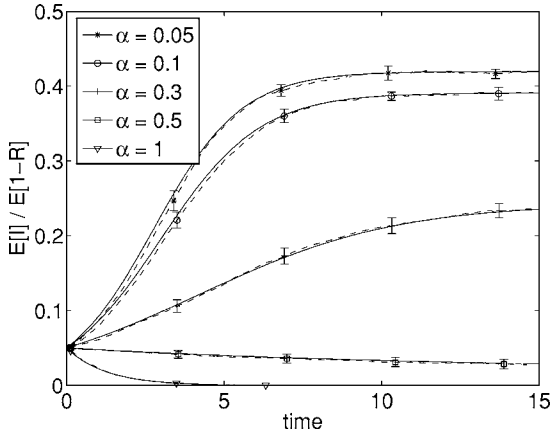

FIG. 1. Mean results of stochastic simulations (solid lines) together with predictions from the ODE model (dashed lines), showing the proportion of nonresistant individuals infected $E[I]/E[1-R]$ over time. Parameters were $\phi=2$, $\mu=1$, 300 households each containing 300 individuals, $E[1-R]=0.1$, $E[(1-R)^2]=0.09$ giving $Q=0.9$, with 5% of susceptible hosts initially infectious. Error bars on simulation curves show ±1 standard deviation among 20 replicate simulations. The value of α increases from 0.05 to 1 moving from the upper curve to the lower curve.

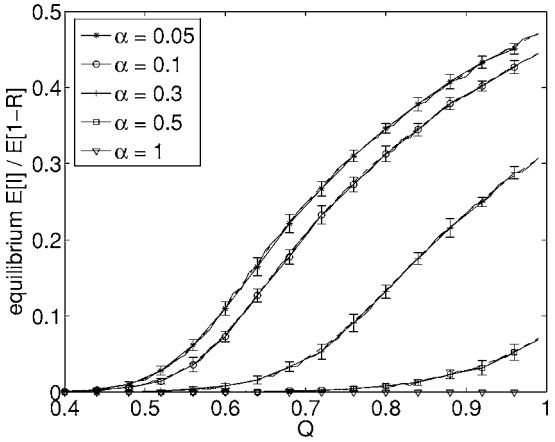

FIG. 2. Mean results of stochastic simulations (solid lines) together with predictions from the ODE model (dashed lines), showing the equilibrium proportion of nonresistant individuals infected $E[I]/E[1-R]$ as a function of clustering Q. Parameters were: $\phi=2$, $\mu=1$, 300 households each containing 300 individuals, $E[1-R]=0.1$, with 5% of susceptible hosts initially infectious. Error bars on simulation curves show ±1 standard deviation among 20 replicate simulations. The value of α increases from 0.05 to 1 moving from the upper curve to the lower curve.

tempts from the other households. That is, we have many coupled systems, each of which has internal dynamics corresponding to a well-mixed system. Because of the large neighborhood size, the improved mean-field approximation accurately predicts the initial growth rate r; the large neighborhood size together with the assumption of mixing of susceptible and infectious individuals allows the approximation to accurately predict the steady-state distribution $\tilde{K}$. Both of these results have been confirmed via stochastic simulations of the system, as displayed in Figs. 1 and 2. The largest difference in $E[I]$ between the ODE model developed here and stochastic simulations as shown in Fig. 1 had magnitude less than 5.2×10^{-4}; the largest difference between simulations and predictions in Fig. 2 had magnitude less than 7.6×10^{-4}. As the number of individuals per household becomes larger, these differences will decrease, as stochastic fluctuations play a less important role and the continuum approximation made within each household becomes more accurate. Note however that the ultimate fate of the system, which must be fixation to the disease free equilibrium (DFE) where $E[I]=0$, will be determined by noise, which has been neglected from this model. But for populations which are not too small and for the parameter values explored here, the system generally persists for very long times at a fixed quasistationary equilibrium, conditioned on the fact that it has not gone to fixation to the DFE [31].

As seen in a related model without mixing and without resistant individuals [26], decreasing the amount α of interhousehold contacts slows the initial spread of an infectious disease, as infectious individuals become clustered and are therefore less likely to encounter susceptible individuals to infect. This is in stark contrast to the results observed here, where decreasing α actually increases the initial rate of spread of the infection as well as the final equilibrium infection level, as seen in Fig. 1. Figure 2 shows that this effect becomes stronger as the clustering Q of nonresistants increases. In a system without mixing of susceptible and infectious individuals but still containing spatially clustered resistant individuals and both intra- and interhousehold contacts ($0<\alpha<1$), the two opposing influences of clustered infectious and resistant individuals will both play a role in the dynamics; either effect may play a more dominant role, depending on the amount of clustering of resistants and the extent of interhousehold contact among individuals. Accurately predicting the transient dynamics of such a system would require an expanded model including equations describing the dynamics of higher moments, e.g., as in Ref. [26]. As shown here, for a model with mixing of susceptible and infectious individuals but fixed resistant individuals, clustering of the fixed portion of the population can have a dramatic influence on the dynamics of epidemics on household-structured populations, yet the dynamics can still be accurately predicted by mathematical approximations when this clustering is included within the mean-field approximation framework.

This work was conducted with financial support from the University of Maine Office of the Vice President for Research.

[1] B. M. Bolker, Bull. Math. Biol. **61**, 849 (1999).
[2] M. Boots and A. Sasaki, Ecol. Lett. **3**, 181 (2000).
[3] A. Grabowski and R. A. Kosiński, Phys. Rev. E **70**, 031908 (2004).
[4] M. J. Tildesley, N. J. Savill, D. J. Shaw, R. Deardon, S. P. Brooks, M. E. Woolhouse, B. T. Grenfell, and M. J. Keeling, Nature (London) **440**, 83 (2006).
[5] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, Science **312**, 447 (2006).
[6] H. Hinrichsen, Adv. Phys. **49**, 815 (2000).
[7] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001).
[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[9] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).
[10] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[11] M. E. J. Newman, I. Jensen, and R. M. Ziff, Phys. Rev. E **65**, 021904 (2002).
[12] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 0351011(R) (2002).
[13] J. M. Read and M. J. Keeling, Proc. R. Soc. London, Ser. B **270**, 699 (2003).
[14] J. Balthrop, S. Forrest, M. Newman, and M. M. Williamson, Science **304**, 527 (2004).
[15] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, J. Theor. Biol. **235**, 275 (2005).
[16] D.-U. Hwang, S. Boccaletti, Y. Moreno, and R. López-Ruiz, Math. Biosci. Eng. **2**, 317 (2005).
[17] M. Keeling, Theor Popul. Biol. **67**, 1 (2005).
[18] J. Saramäki and K. Kaski, J. Theor. Biol. **234**, 413 (2005).
[19] H. Andersson and T. Britton, J. Appl. Probab. **35**, 651 (1998).
[20] F. Ball, D. Mollison, and G. Scalia-Tomba, Ann. Appl. Probab. **7**, 46 (1997).
[21] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds, Proc. Natl. Acad. Sci. U.S.A. **102**, 11157 (2005).
[22] D. Hiebeler, Ecology **81**, 1629 (2000).
[23] D. Hiebeler, Theor Popul. Biol. **66**, 205 (2004).
[24] M. A. M. de Aguiar, E. M. Rauch, and Y. Bar-Yam, Phys. Rev. E **67**, 047102 (2003).
[25] D. Hiebeler, Lect. Notes Comput. Sci. **3515**, 360 (2005).
[26] D. Hiebeler, Bull. Math. Biol. **68**, 1315 (2006).
[27] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[28] D. Hiebeler, J. Theor. Biol. **187**, 307 (1997).
[29] D. Hiebeler, J. Math. Biol. (to be published).
[30] H. Matsuda, N. Ogita, A. Sasaki, and K. Sato, Prog. Theor. Phys. **88**, 1035 (1992).
[31] L. J. Allen, *Stochastic Processes with Applications to Biology* (Pearson Prentice Hall, Upper Saddle River, NJ, 2003).