FREQUENCY DEPENDENT RISK OF INFECTION AND THE SPREAD OF INFECTIOUS DISEASES

BU-1503 -M **September, 2000**

Juan P. Aparicio
Angel F. Capurro
and
Carlos Castillo-Chavez

Keywords: epidemic models, social networks, dynamical systems, tuberculosis.

Abstract: Disease evolution is not only intimately connected to changes in social dynamics but highly driven by them. Disease emergence and re-emergence is tightly linked to shifts on contact structures. Contact sub-structures and, consequently, local networks become epidemiologically active only when infectious individuals are present. In this article, we study the impact of networks on disease evolution. Our work is based on the idea that the "birth" of an infectious individual gives rise to a local epidemiologically active network or generalized household, or an epidemiologically-active cluster, of individuals who are likely to have intimate contacts with the source case. Each source case is also a potential source of secondary infections, in the general population, via casual contacts and this is also considered in this paper. Hence, our focus is on the development and analysis of models of disease dynamics in networks that include local dynamics. Our work is applicable to slow diseases which are transmitted via casual and via (local) intimate contacts.

BU 1503-M

Frequency Dependent Risk of Infection and the Spread of Infectious Diseases

Juan P. Aparicio[1,2], Angel F. Capurro[2,3] and Carlos Castillo-Chavez[2,4]

[1] *Department of Biometrics & Mathematical and Theoretical Biology Institute, Cornell University, 431 Warren Hall, Ithaca, NY 14853-7801, USA*

[2] *Departamento de Investigación, Universidad de Belgrano-CONICET Zabala 1851, Piso 12, 1426 Buenos Aires, Argentina.*

[3] *Laboratorio de Ecología, Universidad Nacional de Luján-CONICET, Ruta 5 y 7, 6700 Luján, Argentina.*

[4] *Department of Theoretical and Applied Mechanics Cornell University, 317 Kimball Hall, Ithaca, NY 14853-7801, USA*

September 21, 2000

Abstract

Disease evolution is not only intimately connected to changes in social dynamics but highly driven by them. Disease emergence and re-emergence is tightly linked to shifts on contact structures. Contact sub-structures and, consequently, local networks become epidemiologically active only when infectious individuals are present. In this article, we study the impact of networks on disease evolution. Our work is based on the idea that the "birth" of an infectious individual gives rise to a local epidemiologically active network or generalized

household or an epidemiologically-active cluster, of individuals who are likely to have intimate contacts with the source case. Each source case is also a potential source of secondary infections, in the general population, via casual contacts and this is also considered in this paper. Hence, our focus is on the development and analysis of models of disease dynamics in networks that include local dynamics. Our work is applicable to slow diseases which are transmitted via casual and via (local) intimate contacts.

key words: epidemic models, social networks, tuberculosis, dynamical systems.

1 Introduction

A useful classification of contacts can only be given in the context of a particular disease. Here, we study the transmission dynamics of communicable diseases like tuberculosis that, in general, have long periods of latency followed by short periods of infectiousness, recovery without permanent immunity, and/or death. The "birth" of an infectious individual gives rise to a local network or an epidemiologically active generalized household or active cluster. This local network is composed of individuals who are extremely likely to have intimate contacts with the source case. Of course, each source case is also a potential source of secondary infections, in the general population, via casual contacts. Therefore, the development and study of models of

disease dynamics in networks where casual and intimate contacts are common is of importance. In this note, we carry out such a study in the simplest possible setting. Our goals here are to highlight the importance of contact structures based on the quality of contacts. This view is useful in the study of the transmission dynamics of slow communicable diseases like tuberculosis (Aparicio *et al.*, 2000a).

First, we consider a homogeneous population and assume that an infectious individual, the source case, has contacts with susceptible individuals in two settings. The source case either has contacts within the local network, cluster or generalized household which he has just activated (by becoming infectious) or within the "world". The local network consists of his close associates while the world is comprised of the rest of the network.

Contacts (with susceptibles) of an infectious individual are classified by their level of intimacy. The quality of the contact is closely connected to its frequency and its duration. We do not define them precisely but rather by their context. Intimate or *generalized household* contacts tend to occur in families, schools, and at work. *Casual contacts* are those that are not intimate, they occur in the "world" and at random. Hence, as soon as an individual becomes infectious then he/she defines two networks of significantly different sizes: a generalized household, a small local network of individuals who have been put at a higher risk of infection by their personal connection

to the source case and, global network, a 'place' where infections are generated by "accident". Risk from the local network persists only through the epidemiological life of the source case. It is likely that the number of casual contacts of an infectious individual is a lot greater than his/her average number of intimate contacts. It is also likely that the probability of disease transmission per contact is significantly lower for casual than for intimate contacts.

Traditionally, contacts between susceptibles and infectious individuals are modeled from the point of view of the susceptible, that is, the rate of infection is modeled by a term of the form $\beta S\frac{I}{N}$. It is assumed that the rate of contact between susceptible and infectious individuals is proportional to the infectious prevalence I/N, or, equivalent, to the susceptible frequency S/N. Since typically all contacts are identical (same duration and intensity) then this distinction is irrelevant. That is, the contact process can be interpreted from the point of view of the infectious or the susceptible individuals. On the other hand, since we focus on diseases with low probability of transmission per casual contact and higher probability of transmission per intimate contacts then contacts must be differentiated by duration and strength or intensity. Furthermore, since intimate contacts occur mostly with individuals with whom one has frequent interactions then a modeling approach based on the networks of the (infectious) source case seems the most appropriate as

the process is no longer symmetric.

Hence, the focus is on modeling diseases with two-types of frequency dependent contacts: contacts within an epidemiologically-active generalized household or small local network and, contacts within the "world", a huge network, with a lower probability of transmission per contact.

Shifts in social dynamics driven by cultural changes (urbanization) or population dynamics (demography) have generated a dynamic social landscape that fosters disease evolution. Increases in contact rates due to urbanization and mass-transportation have altered the frequency of contacts and possibly their quality. Models that incorporate contact-quality at multiple levels need to be developed and studied. Here we focus exclusively on the impact of local and global networks (see Aparicio *et al.*, 2000b; and this volume for additional approaches and extensions).

2 Modeling cluster transmission

We consider slow diseases with a long latency period and assume that transmission is *only* possible within a generalized household or an epidemiologically-active cluster. We consider four epidemiological classes: susceptible, latent, infectious and recovered. We assume that the recruitment rate Λ is constant in order to exclude the role of demography. We let μ denote the inverse of life expectancy or the per-capita disease induced mortality rate and r the

recovery rate. The total removal rate of infectious individuals, denoted by γ, is therefore given by $\gamma \equiv r + \mu_d + \mu$. We keep track of whether or not an individual is in an epidemiologically-active cluster or not. We let $N_c(t)$, denote the population size of individuals in epidemiologically-active generalized households while $N_{nc}(t)$ denotes the population of those who are not. First we neglect the effects of casual infections, that is, we assume that only susceptible persons belonging to the N_c-population may acquire infection. To simplify the notation, non-infectious populations are identified via subscripts. Subscript 1 is used to identify individuals in generalized-households (epidemiologically-active clusters) while subscript 2 is used to identify those who are not. The period of infectiousness, $1/\gamma$, is assumed to be significantly shorter than life expectancy (as is the case in tuberculosis). This last assumption, justifies our decision to neglect births, deaths, and disease progression within the N_c-population. It is further assumed that cluster size is constant and equal to n; that all infectious individuals, or active cases are equally infectious; and, that all generalized household contacts are equivalent. In other words, a constant risk of infection per capita and per unit of time (β) is assumed for all of susceptible individuals who are members of generalized-households. The per capita disease progression rate is denoted by k and, consequently the fraction of infected people who develop the disease is given by $f = k/(k + \mu)$. To highlight the role of epidemiologically-active clusters

on disease transmission, we highlight the mechanics of the infection process.

Whenever an infected individual (belonging to the E_2 class) becomes infectious, that is, when he/she moves into the I-class, his/her generalized-household becomes epidemiologically active. A group of people of size n is instantaneously moved from the N_{nc}-population into the N_c-population. How are they be re-distributed among the N_c-epidemiological classes? They are transferred in proportion to their frequencies in the N_{nc}-population, namely, in proportion to S_2/N_2, E_2/N_2, and R_2/N_2–. From the definition of epidemiologically-active cluster, it becomes clear that kE_2 new (epidemiologically-active) generalized households are "born" per unit of time. Hence, the total recruitment rates into the corresponding N_c-susceptible, latent and recovered or treated epidemiological classes are $\frac{S_2}{N_2}nkE_2$, $\frac{E_2}{N_2}nkE_2$ and $\frac{R_2}{N_2}nkE_2$. The recovery or death of the active case makes the generalized household inactive and this occurs at the rate γI. Whenever an epidemiologically active cluster becomes inactive, individuals are returned to their corresponding epidemiological classes in the N_{nc}-population. Since the "death" rate of the epidemiologically-active generalized-household is γI, and the average generalized-household size is n then, $n\gamma I$ individuals must be moved from N_c- into the N_{nc}-population per unit of time. The relation $N_1 = nI$, which holds by definition, implies that $n\gamma IS_1/N_1 = \gamma S_1$. Hence, γS_1 susceptible individuals must be returned per unit of time to the N_{nc}-population, per

unit of time. Rates of return for other epidemiological classes are computed in the same way. If we further assume that those treated will never get the disease again (permanent immunity) then we arrive at the following epidemiological model where the mode of transmission is exclusively via generalized households (intimate contacts):

$$\frac{dS_1}{dt} = -(\beta+\gamma)S_1 + \frac{S_2}{N_2}nkE_2, \tag{1}$$

$$\frac{dE_1}{dt} = \beta S_1 - \gamma E_1 + \frac{E_2}{N_2}nkE_2, \tag{2}$$

$$\frac{dR_1}{dt} = -\gamma R_1 + \frac{R_2}{N_2}nkE_2, \tag{3}$$

$$\frac{dI}{dt} = kE_2 - \gamma I, \tag{4}$$

$$\frac{dS_2}{dt} = \Lambda - bS_2 + \gamma S_1 - \frac{S_2}{N_2}nkE_2, \tag{5}$$

$$\frac{dE_2}{dt} = \gamma E_1 - (\mu+k)E_2 - \frac{E_2}{N_2}nkE_2. \tag{6}$$

$$\frac{dR_2}{dt} = rI + \gamma R_1 - \mu R_2 - \frac{R_2}{N_2}nkE_2. \tag{7}$$

For model 1-7, the *Basic Reproductive Number*, defined as the number of secondary cases generated by a typical infectious individual in a population where every one is susceptible, is given by

$$\mathcal{R}_0 = \frac{\beta n}{(\beta+\gamma)}\frac{k}{(\mu+k)} \equiv Q_0 f, \tag{8}$$

where $Q_0 = \frac{\beta n}{\beta+\gamma}$ is the number of secondary infections caused by one infectious individual in a fully susceptible population; and $f \equiv \frac{k}{\mu+k}$, is the fraction of infected individuals who develop the disease (become infectious). From Expression (8) we see that $\mathcal{R}_0$ is always bounded by cluster size (n) regardless of the level of infectiousness or the length of the infectious period. Hence, mean cluster size plays a fundamental role in determining the severity of an epidemic where the main mode of transmission is via generalized households. The elasticity of $\mathcal{R}_0$ with respect to a parameter λ is $\mid \frac{1}{\mathcal{R}_0}\frac{\partial \mathcal{R}_0}{\partial \lambda} \mid$. It is straightforward to see that cluster size is the most significant parameter.

2.1 Modeling short latency periods

A ('source') case produces secondary infections and some of them will progress into the infectious class ('secondary' cases). It is likely that some proportion of individuals in an epidemiologically-active generalized household were in fact members of a prior epidemiologically-active generalized household. In other words, generalized household overlap is likely and its likelihood must be linked to duration of the latency period. The shorter the latent period the higher the probability of generalized household overlap.

A simple way of modeling generalized cluster overlap is by assuming that the proportion of susceptible individuals in a newly established epidemiologically-active generalized household is given by a weighted average of the pro-

portion of suceptibles from the N_{nc}- and N_c- populations. This simplistic approach that captures cluster overlap in a phenomenological way can be used to model partial overlap between epidemiologically-active generalized-households just by replacing the term $\frac{S_2}{N_2} nkE_2$ with the weighted average $\left((1-\rho)\frac{S_2}{N_2} + \rho\frac{S_1}{N_1}\right) nkE_2$, with $0 < \rho < 1$. Whenever $\rho = 0$, we recover the case where there is no cluster overlap and that corresponds to diseases with extremely long latency (non-infectious) periods followed by very short infectious periods.

2.2 Modeling casual infections

No disease will be transmitted exclusively in generalized households. In fact, there is growing evidence that even diseases with low probability of transmission per contact, like tuberculosis, can be transmitted often via casual contacts. Tuberculosis transmission in public places like bars and airplanes has been documented (see Rafalli *et al.*, 1996, for a recent review; see also Cobelens *et al.*, 2000). We choose to model the rate of casual infections in the classical way, that is, via a term of the form $\beta^* I \frac{S}{N}$. However, for diseases like TB, with *low prevalence of active cases* further approximations are possible. Since, the prevalence of active cases is $I/N \ll 1$ then $N_1 \ll N_2$. Hence, we choose to approximate S/N by S_2/N_2. Using this approximation

our model with intimate and casual contacts becomes:

$$\frac{dS_1}{dt} = -(\beta+\gamma)S_1 + \left((1-\rho)\frac{S_2}{N_2} + \rho\frac{S_1}{N_1}\right) nkE_2, \tag{9}$$

$$\frac{dE_1}{dt} = \beta S_1 - \gamma E_1 + \left((1-\rho)\frac{E_2}{N_2} + \rho\frac{E_1}{N_1}\right) nkE_2, \tag{10}$$

$$\frac{dR_1}{dt} = -\gamma R_1 + \left((1-\rho)\frac{R_2}{N_2} + \rho\frac{R_1}{N_1}\right) nkE_2, \tag{11}$$

$$\frac{dI}{dt} = kE_2 - \gamma I, \tag{12}$$

$$\frac{dS_2}{dt} = \Lambda - bS_2 + \gamma S_1 - \left((1-\rho)\frac{S_2}{N_2} + \rho\frac{S_1}{N_1}\right) nkE_2 - \beta^* I\frac{S_2}{N_2}, \tag{13}$$

$$\frac{dE_2}{dt} = \gamma E_1 - (\mu+k)E_2 - \left((1-\rho)\frac{E_2}{N_2} + \rho\frac{E_1}{N_1}\right) nkE_2 + \beta^* I\frac{S_2}{N_2}, \tag{14}$$

$$\frac{dR_2}{dt} = rI + \gamma R_1 - \mu R_2 - \left((1-\rho)\frac{R_2}{N_2} + \rho\frac{R_1}{N_1}\right) nkE_2, \tag{15}$$

where now $N_i \equiv S_i + E_i + R_i$. The resulting Basic Reproductive Number for this model is

$$\mathcal{R}_0 = \left(\frac{\beta n(1-\rho)}{\beta + (1-\rho)\gamma} + \frac{\beta^*}{\gamma}\right)\frac{k}{(k+\mu)}. \tag{16}$$

The infectious period determines the time-scale of the disease dynamics in the N_c-population while life expectancy is the characteristic time-scale for the dynamics (including disease dynamics) in the N_{nc}-population.

Whenever, the infectious period is a lot shorter than the life-expectancy (as it would be the case for tuberculosis), we approximate the variables S_1,

E_1, R_1 and I by their quasi-equilibrium values which are obtained by setting Equations (9-12) equal to zero. They are:

$$S_1(t) \cong \frac{(1-\rho)}{\beta + (1-\rho)\gamma} \frac{S_2(t)}{N_2(t)} nkE_2(t), \tag{17}$$

$$E_1(t) \cong \frac{1}{\gamma}\left(\frac{\beta}{\beta + (1-\rho)\gamma}\frac{S_2(t)}{N_2(t)} + \frac{E_2(t)}{N_2(t)}\right) nkE_2(t), \tag{18}$$

$$R_1(t) \cong \left(\frac{1}{\gamma} + \frac{R_2(t)}{N_2(t)}\right) nkE_2(t), \tag{19}$$

$$I(t) \cong \frac{k}{\gamma} E_2(t). \tag{20}$$

Whenever this approximation is valid the system reduces (9-15) to a 3-dimensional system for the slow variables associated with the disease dynamics in the N_{nc} -population. The reduced system is

$$\frac{dS_2}{dt} \cong \Lambda - \mu S_2 - (\mu + k)\mathcal{R}_0 \frac{S_2 E_2}{N_2}, \tag{21}$$

$$\frac{dE_2}{dt} \cong -(\mu + k)E_2 + (\mu + k)\mathcal{R}_0 \frac{S_2 E_2}{N_2}. \tag{22}$$

$$\frac{dR_2}{dt} \cong \frac{rk}{\gamma} E_2 - \mu R_2, \tag{23}$$

where $\mathcal{R}_0$ is still given by expression (16). Using approximation (20), the rate of infection term becomes $\frac{\gamma}{k}(\mu + k)\mathcal{R}_0 \frac{S_2 I}{N_2} = \gamma Q_0 \frac{S_2 I}{N_2}$. Hence, Model (1-7) can be approximated by a classical epidemic model with a transmission coefficient given by $\beta_{eff} \equiv \gamma Q_0$. However, the number of secondary infections produced by every infectious individual $Q_0 \frac{S}{N}$ does not depend in a multiplicative way

on the infectious period $1/\gamma$. Effectively, we have that

$$Q_0 = \frac{\beta n(1-\rho)}{\beta + (1-\rho)\gamma} + \frac{\beta^*}{\gamma}. \tag{24}$$

The last term in Expression (24) gives the contribution of casual infections to secondary cases and depends linearly on the infectious period while the first term (intimate contacts) is bounded by cluster size. The use of a constant rate (or risk) of infection per susceptible individuals in epidemiologically-active generalized-households (β) gives rise to a nonlinear dependence of $\mathcal{R}_0$ on β. Since it is assumed that risk is exponentially-distributed with parameter β in a cluster then risk becomes an epidemiologically-relevant and measurable parameter. Note that the use of a classical modeling approach for transmission via casual contacts renders the parameter β^* non-measurable- a typical problem of classical models. In fact, despite its epidemiological derivation and justification our lack of approaches to estimate contact rates reduces the value of β^*. In fact, β^* becomes mostly a fitting parameter.

In our case, however, we have that $\beta_{eff} \equiv \gamma Q_0$. It is not necessary to know the value of the infectious period $1/\gamma$ because the rate of infection is approximately $(k+\mu)Q_0\frac{SE}{N}$. Although Q_0 is still a function of the infectious period (see expression (24)), standard epidemiological surveys provide estimates of Q_0 using a direct classification of contacts for the source case. Therefore our formulation uses parameters closer to the those obtained from epidemiological studies.

2.3 Discussion

Our approach allows for the exploration of various sociological settings including the impact of urbanization, mass transportation, and global travel on contact rates and, consequently, on casual infections. For slow diseases like tuberculosis, cluster overlap, here modeled by ρ, is minimal. However, for diseases like influenza or measles cluster overlap is critical. Models like those described here are not appropriate for the study of the role of clusters on the transmission of "fast" and highly infectious diseases. In fact, the results of Watts and Strogatz (1998) support, albeit indirectly, the use of classical models.

In classical network models for the spread of infectious diseases each person is modeled as a node and the links among them are those with potential infective contacts. For the study of disease transmission in small communities an approach based on a full characterization of the entire network of contacts is possible. For large communities, such a global network analysis is not only impractical but unnecessary. Watts and Strogatz (1998) have shown, in some sense, that the spread of an infectious disease in a 'small-world' network is quite similar to the spread of a disease in a randomly connected network, provided that all the contacts are identical and the number of contacts (cluster size) remains constant (with some minimal cluster overlap). Local network analysis which is based on our knowledge of the links of sampled individuals

(see Morris, 1995, for a short review) is close to our work. From the point of view of epidemiological field studies (data) local network approaches appear to be significantly more appropriate than global ones.

In small-world networks, randomly connected networks, and some intermediate networks (considered by Watts & Strogatz) individuals are assumed to have contacts with a fixed number of individuals (nodes). Hence, the network is seen as an inter-connected collection of (almost independent) clusters where transmission is possible only within the cluster. A small-world network can be viewed as a highly clustered network where the probability of contact between two individuals is very high whenever they share a contact (connection to a common node).

Although this network structure is very different from that of a randomly connected network, its structure does not seem to affect significantly the spread of a disease. In other words, randomly connected and small world networks predict qualitatively similar epidemics.

Our work shows that the critical parameter for disease spread is mean cluster size. It seems reasonable to expect that every human community is, at least in some sense, well represented by a small world network. If this is so then our work (when combined with the results of Watts & Strogatz' (1998)) shows that disease spread is highly dependent on the distribution of generalized household (local networks) sizes.

Research on local network analysis has focused mostly on efforts to characterize the mixing of heterogeneous populations (see, for example, Anderson *et al.*, 1990; Blythe *et al.*, 1991; Castillo-Chavez & Blythe, 1989; Jacquez *et al*, 1989; Hyman & Stanley, 1988; Sattenspiel, 1987; Hethcote & Yorke, 1984). In most of the works just cited, links are identical within homogeneously mixing sub-populations. Our approach considers a homogeneously mixing population but discriminates in the type of links within subpopulations–in the simplest possible way. Sustained contact are different from casual ones. Sustained contacts are more infectious, on the average, than casual contacts. At the population level, casual contacts may still have a significant impact on the spread of a disease. In fact, casual contacts enhance the effect of the "long-distance" rare interactions assumed in small-world networks. Casual contacts reduce the differences between randomly connected and small-world networks. Our work support this view through the incorporation of contact quality (edges between nodes are not identical).

Networks here (fixed in size as in Watts & Strogatz, 1998) allow for dynamic links between nodes. When an individual becomes infectious all the links between the source case and its local network become active (intimate contacts in generalized household) as well as some randomly distributed links outside the local network (casual infections). In other words, individuals are only temporary members of a network.

References

Anderson, R.M., Gupta, S, & Ng, W. (1990). The significance of sexual partner contact networks for the transmission dynamics of HIV. *JAIDS* **3**, 417-429.

Aparicio, J.P., Capurro, A.F. & Castillo-Chavez, C. (2000a). Transmission and dynamics of tuberculosis in generalized households. *J. theor. biol.*, in press.

Aparicio J. P., Capurro A. F., & Castillo-Chavez C. (2000b) On the fall and rise of tuberculosis. Department of Biometrics, Cornell University, Technical Report Series, BU-1477-M.

Blythe, S. Castillo-Chavez, C., Palmer, J. & Cheng, M. (1991). Towards a unifed theory of mixing and pair formation, *Math. Biosc.* **107**, 349-407.

Castillo-Chavez, C. & Blythe, S. (1989). Mixing framework for social/sexual behavior. In: *Mathematical and statistical approaches to AIDS epidemiology* (Castillo-Chavez ed.). Springer, Berlin, 275-288.

Colbelens, F.G.J., van Deutekmon, H., Draayer-Jansen, I.W.E., Scepp-Beelen, A.C.H.M., van Gerven, P.J.H.J., van Kessel, R.P.M., & Mensen, M.E.A. (2000). Risk of infection with *Mycobacterium tuberculosis* in

travellers to areas of high tuberculosis endemicity. The Lancet, **356**, 461–465.

Hethcote, H. W. & York J. A. (1984). Gonorrhea Transmission and Control. Lectures Notes in Biomathematics 56, Springer, New York.

Jacquez, J.A., Simon, C.P., & Koopman, J. (1989). Structured mixing: heterogeneous mixing by the definition of activity groups. In: *Mathematical and statistical approaches to AIDS epidemiology* (Castillo-Chavez ed.). Springer, Berlin, 275-288.

Hyman, J. & Stanley, E. (1988). Using mathematical models to understand the AIDS epidemics. *Math. Bios.* **90**, 415-473.

Morris, M. (1995). Data driven network models for the spread of disease. In *Epidemic Models: their structure and relation to data*, (Mollison D. ed.), Cambridge University Press, 302-322.

Raffalli, J., Sepkowitz, K. A. & Armstrong, D. (1996). Community-Based Outbreaks of Tuberculosis. *Arch. Intern. Med.*, **156**, 1053-1060.

Satenspield, L. (1987). Population structure and the spread of disease. *Human Biology*, **59**, 411-438.

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks, *Nature*, **393**, 440-442.