

Epilogue
New Directions in Migration
Health and Medicine

Chapter 13

Mapping and Modeling Disease Risk Among Mobile Populations

HENRY G. MWAMBI, PH.D. AND KHANGELANI ZUMA, PH.D.

Human Mobility and Epidemiology

Human mobility patterns play an important role in the spread of many infectious diseases and in designing control strategies for them. Given that epidemiology is the study of the occurrence of disease in a person and at a time and place, this implies that mobility patterns and their distribution within and between countries or regions are important in explaining the dissemination of existing, emerging, or re-emerging diseases. Mobility patterns have been responsible for the introduction of infectious agents into areas where they never before existed. A well-known example is the spread of Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS), which first emerged in the early 1980s. With the passage of time, HIV began to occur nearly on the entire face of the earth with varying proportions. A more recent example is the Severe Acute Respiratory Syndrome (SARS) an atypical pneumonia and part-time sexually transmitted infection (STI) spread through both sexual and casual contact that first appeared in November 2002 in China. First reported in Asia in February 2003, SARS was spread via international travelers in a few months to more than two dozen countries in Asia, North America, South America, and Europe before the global outbreak of 2003 was contained—but by the last case was recorded, there were a total of 8,437 known cases of the disease, with 813 deaths.

Mathematical and statistical models have been put to good use in understanding the spread of infectious pathogens and how best to contain them (Anderson, Fraser, Ghani, Donnelly, Riley, Ferguson, Leung, Lam, & Hedley, 2004; Donnelly, 2004; Donnelly, Ghani, Leung, Hedley, Fraser, Riley, Abu-Raddad, Ho, Thach, Chau, Chan, Lam, Tse, Tsang, Liu, Kong, Lau, Ferguson, & Anderson 2003). The spread and persistence of highly pathogenic diseases, such as SARS, pose challenges to present day epidemiologists because their transmission is aided by key factors such as human mobility patterns including air travel, and the continued growth and overcrowding in big cities, as seen in many parts of the world, particularly in developing countries, as well as other risk factors. Considering the foregoing, epidemic

outbreaks of new infectious agents are likely to become more common than ever, therefore, epidemiologists need to be equipped with quantitative tools and skills in order to counter such challenges. Data-oriented mathematical and statistical methods are crucial for the success of these efforts. It has been argued that international collaboration in the analysis of epidemiological and contact-network databases can provide further insight into the spread of emerging or re-emerging infections such as SARS and avian influenza (Donnelly et al., 2003).

Models incorporating geographic mobility among regions in the diffusion of infectious diseases are not new. Sattenspiel and Dietz (1995) developed a model for the spread of infectious diseases among discrete geographic regions that incorporated a mobility process describing how contact occurs between individuals from different regions. Their general mobility formulation included a range of mobility patterns from complete isolation of all regions to permanent migration between them. In addition, the authors showed how to incorporate mobility processes into the basic Susceptible-Infective-Recovered (SIR) epidemic model, which was applied to describe the 1984 measles epidemic on the Caribbean island of Dominica. For STIs including HIV/AIDS, a number of socioeconomic factors are definitely responsible for the recent rise in their incidence and prevalence but clearly most of the facilitating factors are linked to human mobility and migration (Lurie, Williams, Zuma, Mkaya-Mwamburi, Garnett, Sturm, Sweat, Gittelsohn, & Abdool Karim, 2003a; Lurie, Williams, Zuma, Mkaya-Mwamburi, Garnett, Sturm, Sweat, Gittelsohn, & Abdool Karim, 2003b; Zuma, Gouws, Williams, & Lurie 2005). Examples include increasing urbanization, migrant worker effect, expansion of the tourist travel sector, and new attitudes towards sexual behavior in modern societies. Generally, human mobility patterns and social networks in relation to the spread of infectious diseases are multilevel in nature. By design, it is expected that there will be more movements of humans within cities, towns and regions in the same country than movements between two or more countries. As one goes down to a finer geographic resolution, social networks and contacts become tighter. However, in developed countries where transport systems are highly efficient, communication between areas within a country and between countries in a modeling sense is more probable. This means that the transportation of an infectious agent between countries with such highly connected systems will be more efficient.

Mathematical Modeling and Infectious Diseases

The earliest known mathematical model for infectious diseases is attributed to Daniel Bernoulli (1760) for the transmission of smallpox. After the work by Bernoulli there was a period of no further development in the area, possibly due to the lack of understanding of the mechanisms driving the

infectious processes of diseases. Advances in biology and bacteriology reversed this situation and a major development followed with the first model for malaria, an indirectly transmitted disease, developed by Sir Ronald Ross (1909, 1911). In his model, Ross described the process whereby a human host acquires malaria from an infectious mosquito bite or conversely, a disease-free mosquito gets infected by biting an infected human host. Ross introduced a very important concept in disease control by suggesting that the eradication of malaria could be possible if the mosquito population could be reduced to a certain finite threshold number, say N^{crit} , suggesting that the complete eradication of mosquitoes is not necessary. This observation led to the threshold phenomenon in epidemiology. Following this, Kermack and McKendrick (1927) developed an epidemic model in which the infectivity of an individual depends on the time when the individual becomes infective or what essentially could be interpreted as an age of infection model. In this case, an epidemic is defined as a sudden outbreak of a disease, which infects a sizeable portion of the population in a region before it disappears. The Kermack-McKendrick SIR epidemic model is formulated as a two-dimensional system of ordinary differential equations representing transitions (in continuous time) from the Susceptible class (S) to the Infective class (I) and then from the Infective to the Recovered class (R). Since this period in history, there has been an increasingly rapid transition of biology and medicine from qualitative (descriptive) to quantitative (predictive) sciences in the form of mathematical and statistical models to explain the growth of human (and animal) populations and the spread of infectious diseases as well as the design of control strategies. A significant contribution in this field is the work of Anderson and May (1991) as well as a number of other model variants including contact rates, quarantine, and isolation which followed the recent SARS epidemic, and which can be grouped as infection models and analyzed using the Kermack and McKendrick approach.

The Basic SIR Model

In this model the assumption is made that an individual can only occupy one of three possible disease states or classes. Firstly, an individual is uninfected and susceptible to the disease; secondly, if the individual comes into contact with the infectious agent and becomes infected then his/her state will change from that of susceptible to infected; and finally, the individual may develop some immunity to the disease and change from infective to the recovered or removed class. It should be noted, however, that this structure is in most cases a major simplification because in reality the disease stages may be more detailed. In order to briefly explain the formulation of the SIR epidemic model, the enticing approach developed by Daley and Ghani (1999) is followed. Let $x(t)$, $y(t)$ and $z(t)$ respectively denote the number of individuals in

the susceptible, infected, and recovered classes of the disease at time t respectively. The total number of individuals is $N = x(t) + y(t) + z(t)$ and assumed to be constant as in its original formulation (Kermack & McKendrick, 1927). The variables $x(t)$, $y(t)$ and $z(t)$ are taken as continuous deterministic variables, hence they can be modeled using a system of differential equations given by:

$$\frac{dx}{dt} = -\beta xy \quad (13.1)$$

$$\frac{dy}{dt} = \beta xy - \gamma y \quad (13.2)$$

$$\frac{dz}{dt} = \gamma y \quad (13.3)$$

The initial state of the system is $(x(0), y(0), z(0)) = (x_0 > 0, y_0 > 0, 0)$ assuming there is at least one infected person in the population otherwise the population will remain uninfected. In its basic form, the model assumes a closed population with no migration in or out of that population. The parameter β denotes the rate of infection per individual per unit time or what may be called the hazard of infection. Note that susceptibles get infected at a rate that is proportional to the product of susceptible and infected individuals. This concept is known as “the law of mass action” governing the spread of infectious diseases assuming homogeneous mixing between those susceptible to and infected with the disease. The parameter γ is the recovery or the removal rate of an infected individual. Written in this form the model assumes that the duration of infection follows an exponential distribution with a mean given by $1/\gamma$. An equation relating $x(t)$ and $z(t)$ can be obtained (Daley & Ghani, 1999) by first dividing the equation for $x(t)$ by the equation for $z(t)$ to yield:

$$\frac{dx}{dz} = -\frac{\beta}{\gamma} x = -\frac{x}{\rho} \text{ where } \rho = \frac{\gamma}{\beta} \quad (13.4)$$

Then the solution to equation (13.4) is easily obtained as

$$x = x_0 e^{-\frac{z}{\rho}} \quad (13.5)$$

where ρ can be defined as a measure of the relative removal rate. This therefore means that

$$y = N - x_0 e^{-\frac{z}{\rho}} - z \quad (13.6)$$

hence

$$\frac{dz}{dt} = \gamma(N - x_0 e^{-\frac{z}{\rho}} - z) \quad (13.7)$$

The above equation has a parametric solution given by

$$\gamma t = \int_0^z \frac{dv}{N - v - x_0 e^{-\frac{v}{\rho}}}, \quad 0 \leq t \leq \infty \quad (13.8)$$

There are two main results which can be inferred from system (13.1)-(13.3). The first one is the criticality condition which comes from equation (13.2). Since at the start of the epidemic, this equation can be re-expressed as

$$\frac{dy}{dt} = \beta y(x_0 - \rho) \tag{13.9}$$

it follows that if the epidemic is ever to grow, then we require that

$$\left. \frac{dy}{dt} \right|_{t=0} > 0 \text{ or } x_0 > \rho$$

(i.e., the initial number of susceptibles must exceed a threshold value equal to ρ). The second important result from the work of Kermack and McKendrick (1927) is derived as follows. From equation (13.8), as $t \rightarrow \infty$, $z(t)$ approaches its limiting value $z_\infty < N$. Thus in the limit

$$N - z_\infty - x_0 e^{-\frac{z_\infty}{\rho}} = 0$$

Now suppose that x_0 is close to N ; then z_∞ is the approximate solution of

$$0 = N - N e^{-z_\infty/\rho} \approx N - z_\infty - N(1 - z_\infty/\rho + z_\infty^2/2\rho^2)$$

Thus if say $N = \rho + v$ with $v \ll \rho$ then

$$z_\infty \approx \frac{2v}{1 + \frac{v}{\rho}} \approx 2v$$

Thus if $x_0 \approx N = \rho + v$ with $v > 0$, then since $y_\infty = 0$, $x_\infty \approx \rho - v$ since $x_\infty + z_\infty = N$. According to this result, some susceptibles will ultimately survive the epidemic free from infection. An important fundamental property related to the first result noted above is that there is a basic reproductive number R_0 of the disease, determining whether the disease will die out without spreading or whether there will be an epidemic. R_0 is defined as the number of expected secondary infections caused by a single infective introduced into a wholly susceptible population of size N over its entire infectious duration. In this case, since the mean infective period is $1/\gamma$, then $R_0 = \beta N/\gamma$, from the final size equation it is possible to calculate the fraction x_∞/N of the population that escapes the epidemic.

A more realistic model is one that includes a latent or exposed (incubating) class whose members progress to the infectious class at a rate of say σ to which could also be included a disease-induced mortality at rate α . If the time-scales of the disease process are much faster than the demographic processes, then the basic reproductive number of the disease becomes $R_0 = \beta N/(\gamma + \alpha)$. For a complete analysis of an age of infection model closely related to the original Kermack-McKendrick model, of which system (13.1)-(13.3) is a special case, the reader is referred to the work of Brauer (2005) in which general contact rates are allowed.

Extensions to Incorporate Reactions to an Epidemic

An actual epidemic model differs considerably from the idealized model system (13.1)-(13.3) and its extension to the Susceptible-Exposed-Infected and Infectious-Recovered (SEIR) model with the SARS epidemic as a notable recent example follows. Some key differences are:

1. Various vaccination strategies are possible, such as the vaccination of health care workers and other first-line responders to the epidemic, vaccination of individuals who have been in contact with diagnosed infectious individuals, or vaccination of members of the population who are in close proximity to the diagnosed infectious individuals.
2. Those diagnosed as infected can be hospitalized, both for treatment and as a means of isolation from the rest of the population.
3. Contact tracing (Mueller, Kretzschmar, & Dietz, 2000) may be used to identify people at risk of becoming infective, and consequently so that they may be quarantined (or instructed to remain at home and to avoid contacts) and monitored so that they may be isolated immediately if and when they become infective.
4. Sometimes isolation may be imperfect, as in the case of in-hospital transmission of infection, which can be a major problem.

In hospitals, transmission can account for many new cases as was the case with SARS. This is an important heterogeneity in disease transmission which must be accounted for whenever there is any risk of transmission. Such details were accounted for in models for SARS (Anderson et al., 2004; Donnelly et al., 2003) and approaches used for the SARS epidemic can also be relevant to other epidemics. The SARS outbreak attracted renewed effort to epidemic modeling, which is of great value in coping with future disease outbreaks. Thus if a vaccine is available for a disease that threatens to be an epidemic outbreak such as avian influenza, a vaccinated class that is protected at least partially ought to be included in the model development. But for an epidemic outbreak, where no vaccine protection is available, isolation and quarantine are the two main control measures available. Thus one can formulate a model for an epidemic once such control measures are in place, as in the work of Brauer (2005), who assumes that the epidemic has just started and so the number of infectious individuals is still small and almost all members of the population are still susceptible. In this approach, a class of quarantined individuals (Q) and another class of isolated (J) members are introduced. A general model with six compartments called the Susceptible-Exposed-Quarantined-Infected-Isolated-Recovered (SEQIJR) model is formulated in order to capture the course of the epidemic with no vaccine but with some control measures in place. The control reproductive number R_C is then defined as the number of secondary infections caused by a single infective in a population consisting essentially only of susceptible individuals with control measures in place. In order to derive the expression for R_C the following assumptions are necessary:

1. Exposed members may be infective with infectivity reduced by a factor θ_E , where $0 \leq \theta_E < 1$.
2. Exposed members who are not quarantined become infectious at rate σ_1 .
3. Exposed members are quarantined at rate α_1 per unit time. Although quarantine is not perfect, it should be assumed that it reduces the contact rate by, say θ_Q .
4. Infective individuals are diagnosed at rate α_2 per unit time and isolated. Quarantined members are monitored and isolated immediately on showing disease symptoms at rate σ_2 .
5. There is a possibility of transmission of disease by isolated members, with an infectivity factor of θ_J .
6. Infectious individuals who are not isolated leave the infective class at rate γ_1 with a fraction f_1 recovering, while isolated members leave the isolated class at rate γ_2 with a fraction f_2 recovering.

Without going into further details, an expression for R_C is readily constructed (see full paper by Brauer, 2005):

$$R_C = \frac{\theta_E N\beta}{\Lambda_1} + \frac{N\beta\sigma_1}{\Lambda_1\Lambda_2} + \frac{\theta_Q N\beta\alpha_1}{\Lambda_1\sigma_2} + \frac{\theta_J N\beta\sigma_1\alpha_2}{\gamma_2\Lambda_1\Lambda_2} + \frac{\theta_J N\beta\alpha_1}{\gamma_2\Lambda_1} \quad (13.10)$$

where $\Lambda_1 = \alpha_1 + \sigma_1$ and $\Lambda_2 = \alpha_1 + \gamma_1$ giving the overall rate of leaving the exposed and infectious classes respectively. The epidemiological interpretation of each term in R_C is as follows. The mean duration in the Exposed class is $1/\Lambda_1$ with contact rate modified to $\theta_E\beta$, giving a contribution of $\theta_E N\beta/\Lambda_1$. A fraction σ_1/Λ_1 goes from the Exposed class to the Infectious class, with contact rate β and mean duration of $1/\Lambda_2$, giving a contribution of $N\beta\sigma_1/\Lambda_1\Lambda_2$. Next, a fraction α_1/Λ_1 goes from the Exposed class to the Quarantined class, with contact rate $\theta_Q\beta$ and mean duration $1/\sigma_2$, giving a contribution to R_C of $\theta_Q N\beta\alpha_1/\Lambda_1\sigma_2$. A fraction $\sigma_1\alpha_2/\Lambda_1\Lambda_2$ move from the Exposed to the Infectious class then to the Isolated class (J), with contact rate $\theta_J\beta$ and a mean duration of $1/\gamma_2$, giving a contribution of $\theta_J N\beta\sigma_1\alpha_2/\gamma_2\Lambda_1\Lambda_2$. Lastly a fraction α_1/Λ_1 move from E to Q to J with contact rate $\theta_J\beta$ and a mean duration of $1/\gamma_2$ giving a contribution of $\theta_J N\beta\alpha_1/\Lambda_1\gamma_2$. Adding all these contributions together leads to the expression of R_C .

Modeling Epidemics on Social Networks

Localized spread of infectious diseases has been successfully captured through various spatial models such as those developed by Mollison (1977), Durrett and Levin (1994), and Grenfell and associates (2001), among others. Such approaches recognize that the predominantly local nature of disease transmission leads to high degree of spatial heterogeneity and hence the population is not well mixed. As an alternative, the use of social network analysis has increasingly become an important tool to further our understanding of the spread of epidemics, particularly when proximity in space is no longer the determining risk factor for transmission. The approach is most useful in the development of

more effective targeted control and treatment strategies. A wide range of communicable human diseases can be considered as spreading through a network of possible transmission routes. The implied network structure of a particular disease is vital in determining its dynamics, mixing pattern, and spread. The structure becomes particularly crucial when the average number of connections per individual is small as is the case for many STIs including HIV/AIDS. The use of social network models in epidemics is ideally an extension of the traditional population-level, pathogen-focused analysis of epidemics to one that is focused more closely on the host. Attention is no longer only on an individual's risk behavior, but focused on that individual's risk environment—which can accordingly increase or decrease an individual's risk behavior depending on the infectious status and behaviors of the people with whom s/he typically interacts. Thus, routinely collected contact tracing data is a critical source of information in the construction of disease networks for specific areas of occurrence; however in many parts of the world such data are either partially available or non-existent. This problem is further magnified in developing countries whose health budgets are highly constrained. Nonetheless it is recommended that centralized computer-based data systems be kept at all cost to cumulatively capture the contact tracing information gathered by public-health workers and physicians.

Social network models have already been applied to study infectious diseases in the study of STI transmission patterns (e.g., Wylie & Jolly, 2001). A mathematical model to address the question of heterogeneity in the rates of partner change and sexual mixing patterns was developed to investigate ethnic inequalities in the incidence of STIs in south-east London (Turner, Garnett, Ghani, Sterne, & Low, 2004). A more general analysis of (sexual) network models is found in a paper by Eames and Keeling (2002), who developed an intuitive mathematical framework to deal with the heterogeneities implicit within contact networks and those that arise because of the infection process. The researchers demonstrated how such models can be used to estimate parameters of epidemiological importance, and how they can be extended to examine the effectiveness of various control strategies, particularly screening and contact tracing.

In general, a network model for an infectious disease focuses on one of the fundamental issues of epidemiology—such as, who can acquire infection from whom (Lurie et al., 2003b). Such networks represent an individual within a population as a *node*, with connecting *edges* denoting relationships that could lead to the transmission of the disease. For many of the common airborne diseases it may be difficult to define which contacts form an edge, but for STIs, edges are more precisely defined as and correspond to sexual partnerships, thus networks are disease dependent. The initial spread and long-term behavior of any infectious disease are determined by both its epidemiological characteristics and the graph theoretical properties of the network. Some of these properties include an average number of neighbors, degree of clustering and the path length between nodes (Ghani & Garnett, 1998; Kretzschmar & Morris, 1996; Morris & Kretzschmar, 1995). One of the key features of an infection occurring within the constraints of a network is

the rapid build-up of correlations in the infectious status of connected individuals. This aggregation will have the net effect of reducing the average number of susceptible partners per infected individual and consequently slow the spread of an epidemic. Note that standard epidemiological models such as the SIR model ignore this important correlation structure. In practice, however, the chains of transmission detected are seldom more than a few individuals long. This necessitates the need for modeling approaches that are capable of utilizing the available detailed information about the network, but does not require the complete network to be reconstructed. This goal was achieved by modeling partnerships as dynamic variables, developing a set of differential equations for the various types of connected pairs within the network in the study of STIs (Eames & Keeling, 2002). The beauty of this approach is that the models can easily be parameterized through the use of readily attainable contact tracing data, yet retain a high degree of generality.

Standard models for the dynamics of diseases or the mean-field models (Anderson & May, 1991) classify individuals according to their infection history by keeping track of the densities of susceptible, infected, and immune hosts. In general, these models consider the proportion of individuals in each class and ignore the underlying network or spatial structure. For STIs and childhood respiratory diseases (e.g., the respiratory syncytial virus [RSV]) there is generally little or no immunity, so individuals return to the susceptible state upon recovery hence a suitable model for such infections is the SIS. The simplest correlation dynamics equations keep track of states of neighboring pairs of hosts on the lattice (Van Baalen, 2005). The general pair-wise network model for the Susceptible-Infective-Susceptible (SIS) model dynamics, following Keeling (1997), is outlined as follows. First label individuals as S or I and superscripts to denote their number of partners. Thus $[I^n]$ denotes the number of infected individuals with n partners and $[S^n I^m]$ the number of partnerships between a susceptible with n partners and an infected with m partners. It is through such partnerships between susceptible and infected individuals that infection can be transmitted. Considering the dynamics of infectious individuals, two basic events can occur: either the infectious individual recovers, assumed to occur at rate ν or a susceptible individual gets infected by an infectious individual assumed to occur at rate γ . This leads to the following equation for the number of infected individuals with n partners:

$$\frac{d[I^n]}{dt} = -\nu[I^n] + \gamma \sum_m [S^n I^m] \tag{13.11}$$

The second term on the right hand side gives the total number of infected partners of all S^n , each of whom transmits infection at rate γ . By making the standard assumption of ignoring partnerships but using contact data to estimate mixing between classes, this number can be estimated by:

$$[S^n I^m] \approx \frac{[S^n]}{[n]} \times \frac{[I^m]}{[m]} \times [nm] \tag{13.12}$$

where $[n]$ denotes the number of individuals with n partners (with a similar meaning for $[m]$) and $[nm]$ the number of partnerships with n and m partners. Although this mean-field approximation includes risk-structured heterogeneities, it ignores the correlations in infection status that emerges between connected individuals. To model the dynamics of pairs, such as the $[S^n I^m]$ pair, equations of the form given below can be constructed.

$$\begin{aligned} \frac{d[S^n I^m]}{dt} = & \gamma \sum_p ([S^n S^m I^p] - [I^p S^n I^m]) - \gamma [S^n I^m] \\ & - \nu [S^n I^m] + \nu [I^n I^m] \end{aligned} \quad (13.13)$$

The terms in the above equation refer to the creation of the $[S^n I^m]$ pair caused by infection of an S^m within an $[S^n S^m]$ pair, loss of the pair caused by infection of the S^n from outside or within the partnership, loss of the pair because of the recovery of the infected individual, and the last term which is specific to an SIS process represents the creation of the pair due to an $[I^n]$ recovery (which will be absent from the SIR model). In a similar manner, one can construct equations for all types of pairs by considering all possible events that can lead to that pair. This process could in theory be extended to model triplets, such as $[S^n S^m I^p]$, in terms of quadruples and so on, but the system rapidly becomes more complicated and in addition, the amount of data available to characterize the triplets is limited. This problem can be alleviated by making use of the *moment closure approximation* (Dushoff, 1999; Van Baalen, 2005), which allows the estimation of the number of triplets in terms of pairs, which closes the system, enabling one to calculate the behavior of individuals and pairs.

The impact of network models can be explained when important epidemiological quantities are considered, such as the basic reproduction ratio, R_0 , already defined as the average number of secondary cases produced by an average infectious individual in a wholly susceptible population. It is calculated as a measure of initial growth of an infinitesimal infection in an otherwise susceptible population. For a structured population, however, the growth rate may depend on which class of individuals is infected. It might therefore be necessary to allow the level of infection to equilibrate between classes (so that high-risk individuals are more likely to be infected) before calculating R_0 . For network models, R_0 should be calculated only once early spatial correlations (which develop within a couple of generations) have formed (for a detailed analysis, refer to Eames and Keeling, 2002 and Keeling, 1999). For an SIR version of the pair-wise network model, the basic reproduction ratio is given by

$$R_0 = \frac{\gamma(\lambda - 1)}{\nu} \quad (13.14)$$

where λ is the dominant eigenvalue of the matrix M given by:

$$M_{mn} = \frac{[nm](m-1)}{m \times [m]} \quad (13.15)$$

The matrix M is therefore a useful means of quantifying the connectedness of contact networks. Thus the strong correlations between the infection statuses of neighboring individuals play two roles. First, the negative correlation between susceptible and infectious individuals acts to dampen the epidemic spread and therefore reduces R_0 . Second, in standard mean-field models, which ignore partnerships and correlations, R_0 is the same for both the SIS and SIR models. However in a pair-wise SIR structure, infectious individuals have a high proportion of recovered individuals as their neighbors, which will limit further spread of the disease. This limitation does not exist in the SIS formulation, and hence epidemic growth is more rapid. Thus equation (13.14) above offers only a lower bound for the SIS disease process. The importance of taking partnerships into account was demonstrated by using simulation studies where the mean-field model consistently overestimated initial spread of an infection over a range of values of the dimensionless infection parameter here given by γ/ν (Eames & Keeling, 2003). The key aim of modeling epidemics and analysis is to help in the design of control and treatment strategies. Clearly a control strategy focused on high-risk individuals (those with high numbers of contacts), taking advantage of the heterogeneities present in the network of partnerships, is likely to be more successful than that applied homogeneously across the population. More importantly at the verge of eradicating the disease, the high-risk classes can act as both reservoir and possible invasion routes for new infections. Hence when an intervention is about to achieve success it becomes increasingly important to target those individuals most central to disease spread.

Migrant Worker Effect: A Real Application

Introduction

HIV was identified as the cause of AIDS two years after its identification as a disease. Today, HIV affects all countries of the globe, making it and its disease consequences the most significant emerging infection of the late 20th century (Nicoll & Gill, 1999). To date, epidemiological factors determining the geographical spread of STIs/HIV are still not completely understood.

The geographical spread of STIs/HIV is determined by an interaction of factors related to demography, socioeconomics, and sexual behavior. HIV, like other infections that spread from person to person, follows the movement of people (Decosas & Adrien, 1997; Decosas, Kane, Anarfi, Sodji, & Wagner, 1995; Mabey & Mayaud, 1997; Quinn, 1994). The predominant socioeconomic factor (particularly in developing regions) is the rural-urban *labor migration* of young sexually active men leaving their sexual partners behind (Decosas, et al., 1995; Pison, Le Guenno, Lagarde, Enel, & Seck, 1993). Mobile people are at higher risk of STIs/HIV than those in stable living arrangements (Lagarde, Pison, & Enel, 1996; Pison, et al., 1993), primarily

because the conditions of migration bring, for instance men into heterosexual contact with commercial sex workers and other women at high risk of STIs/HIV (Jochelson, Mothibeli, & Leger, 1991). The consequent sexual networking between urban and rural areas determines the diffusion rate of STIs/HIV into local societies (Fleming & Wasserheit, 1999). Furthermore, the women left behind sometimes have to exchange sex for favors as a survival strategy (Evian, 1993). The stark reality of the impact of STIs/HIV on society requires deeper understanding of factors determining the spread of STIs/HIV and further understanding of the relationship between STIs and HIV.

Migration and HIV Diffusion Risks

In a study of the effects of migration in the transmission dynamics of HIV, migrant men from two adjacent health districts in South Africa's northern province of Kwa-Zulu/Natal were recruited at two primary migration destinations, Richards Bay (an industrial area) and Carletonville (a mining town). Migrant men were eligible to participate in the study if they were from Hlabisa or Nongoma districts, if they had at least one regular partner living in at least one of the two districts, and if they had been a migrant for at least the last six months (study methodology and results have been reported elsewhere, see Lurie et al., 2003a, Lurie et al., 2003b; Zuma et al., 2005; Zuma & Lurie, 2005). Migrant men gave information to locate their rural partners who were then invited to participate in the study. Non-migrant men and their partners living within a one kilometre radius from a migrant couple's home were asked to participate in the study. A detailed questionnaire was administered and urine and blood were collected for STI/HIV testing respectively.

A total of 168 couples were recruited into the study, of whom 98 (58.3%) were couples in which the male partner was a migrant, and 70 (41.7%) in which the male partner was not a migrant. The overall prevalence of HIV was 19.9% with 24.4% of men and 15.5% of women infected and among 69.6% of the couples, none of the partners was infected with HIV. Migrant couples were as likely as non-migrant couples to have neither partner infected with HIV (65.3% versus 75.7%) that is, to be HIV-concordant¹. In 9.5% of the couples, both partners were infected with HIV, but this did not differ significantly by the migration status of the male partner (Lurie et al., 2003b). In 20.8% of the couples, one of the partners was infected with HIV. Migrant couples were 2.5 times more likely than non-migrant couples to be HIV-discordant² (26.5% versus 12.8%). Of the 35 discordant couples, the man was HIV-positive in 25 (71%) of the cases and the woman in the remaining 10 (29%) cases. The proportion of men who were infected in the migrant discordant couples was essentially the same as in non-migrant HIV discordant couples (Lurie et al., 2003b).

¹ Both sexual partners have the same HIV status

² Only one sexual partner is infected with HIV in a couple

The Mathematical Model

In order to estimate the relative risk (RR) of infection for migrant and non-migrant men and women from their spouses and from partners outside the relationship, a set of parameters need to be defined (for greater detail of the model and results, see Lurie et al., 2003b). For a man and a woman in a sexual partnership, the man may be infected from outside the relationship with probability α , the woman may be infected from outside the relationship with probability β . The man may also be infected by his wife with probability γ (if she is already infected) and the woman may be infected by her husband with probability δ (if he is already infected). If the probabilities of infection are known, then the probabilities of each of the four concordance possibilities can be calculated. Combining probabilities gives:

$$\begin{aligned} P_{nm} &= (1 - \alpha)(1 - \beta) \\ P_{pn} &= \alpha(1 - \beta)(1 - \delta) \\ P_{np} &= \beta(1 - \alpha)(1 - \gamma) \\ P_{pp} &= \alpha\beta + \alpha\gamma + \beta\gamma - \alpha\beta(\gamma + \delta) \end{aligned}$$

where the first subscript indicates the HIV status of the man (positive or negative) and the second indicates that of the woman. The parameters are varied in order to maximize the likelihood of the fit of the estimated probabilities to the observed probabilities assuming binomial errors. Since there are four parameters and only three independent observations, an appropriate value for the ratio of the likelihood that an infected man infects his wife to the likelihood that an infected woman infects her husband, is assumed to be δ/γ .

Fitting this mathematical model to the above described data shows that both men and women are more likely to be infected by partners outside the relationship than to be infected by their spouses, irrespective of the migration status of the man. Migrant men are 26 times more likely to be infected by partners outside the relationship than from inside the relationship, whereas, women whose partners are migrants are 2.1 times more likely to be infected from outside the relationship than from inside. The same is true for non-migrant couples but with smaller RR of 10.5 for non-migrant men and 0.8 for their partners.

The impact of migration on the transmission dynamics of HIV can be better understood by comparing the RRs of infection for migrants as against non-migrants from outside versus inside their primary relationship for both men and women. Both men and women are likely to be infected from outside the primary relationship 1.44 and 1.53 respectively; however, they are less likely to be infected by their spouse if they are part of the migrant couple (Lurie et al., 2003b). The model assumes that within a spousal relationship, male-to-female HIV transmission is twice as likely as female-to-male transmission. Changing the relative transmissibility from men to women in either direction changes the RR estimates by less than 1.5% in all cases.

Discussion

It has long been assumed that the primary direction of spread of HIV has been from returning migrant men, who become infected while away at work, to their rural partners upon their return home. If this were the case, the male would be the HIV infected partner in most of the discordant couples; however, in nearly one-third of the discordant couples the female was the infected partner. Although this confirms the importance of migration as a risk factor for infection in both men and women, it changes the understanding of the way in which migration enhances risk.

The analysis in this chapter has focused on the man as a migrant and a woman as a non-migrant. In recent years, female circular migration has increased in South Africa as well as other places. The impact that migration has on the health of female migrants has not been investigated as extensively as it has been for men and most studies have concentrated on the migration of men and the risk that this entails for them and their non-migrant female partners (Decosas et al., 1995; Jochelson et al., 1991; Pison et al., 1993). Fewer have explored HIV infection risk factors among migrant women (Brewer et al., 1998; Zuma, Gouws, Williams, & Lurie, 2003) but have demonstrated that migrant women are also at high risk of HIV infection during their migration periods. There is a need to take drastic steps to address the social and economic pressures that migrant men, migrant women, and partners of migrant men face in the process of migration—such as the encouragement of industrial decentralization and regional developments to reduce the need for migration as well as to improve the conditions of migration.

Mapping Disease Risk

GIS, Mapping, and Clustering

Disease mapping and geographical information systems (GIS)³ are becoming necessary as new technologies to improve decision-making processes in disease surveillance and control activities. These tools provide health professionals the ability to quickly analyze spatial relationships and disease risk factors in order to facilitate policy planning and implementation. The technique is used to visualize spatial patterns in the geographical distribution of disease, usually for explorative and descriptive purposes, to gain important clues about the etiology of a disease, and to provide information for further studies. As a result, disease mapping has become a valuable approach to hypothesis generation in explorative epidemiology.

Because of its growing usefulness, the development of methods for disease mapping has received great attention. The mapping and GIS program

³GIS is a surveillance system that uses remote sensing to predict disease spread.

of the World Health Organization (WHO) has spearheaded a global partnership in the promotion and implementation of GIS to support decision-making for a wide range of infectious diseases and as early as 1997, the WHO held a workshop in Rome on “Disease Mapping and Risk Assessment for Public Health Decision-Making.” The workshop concluded with the general belief that geographical analysis of the distribution of risk factors can be useful in prioritizing preventive measures. Disease mapping was identified as useful for health service provision and targeting interventions if avoidable risk factors are known. It was however agreed that, no methodology of choice can be recommended in general and that analytical methods should be selected on the basis of the structure of the data to be analyzed and of the hypotheses to be investigated. In most circumstances, it might be helpful to envisage a first level of descriptive analysis, to be followed by more specific and problem-dependent analyses involving parameter estimation and hypothesis testing (Lawson, Biggeri, Boehning, Lesaffre, Viel, & Bertollini, 1997).

Numerous disease-mapping methods exist from the simple to the complicated (Lawson et al., 1997). While many of the earlier methods adopted a frequentist approach, Bayesian approaches based on Markov Chain Monte-Carlo (MCMC) methods have been gaining importance. In one of the earliest applications of the latter method, an empirical Bayes approach was used to shrink the Standardized Mortality Ratio (SMR) towards a local or global mean (Besag, York, & Mollié, 1991; Clayton & Bernardinelli, 1992). In the paper by Besag and associates (1991) the method was generalized to allow for different spatial heterogeneity. In their model Clayton and Bernardinelli (1992) discuss a Markov Random Field (MRF) approach as representing spatially structured heterogeneity. A nonparametric Bayesian approach was later proposed for the detection of clusters of elevated (or lowered) risk for the identification of unknown risk factors regarding the disease (Knorr-Held & Raer, 2000).

Mapping Mobile Populations

The first two case studies illustrate the use of GIS techniques in disease mapping and control. The third case study is on modeling disease risk in space and time, where data are both longitudinal in time and spatial in nature.

(1) Using Remote Sensing and GIS to Identify Villages in Uganda at High Risk for Sleeping Sickness

GIS and remote sensing were used to identify villages at high risk for sleeping sickness (also known as human Africa trypanosomiasis, caused by *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense*), as defined by reported incidence (Odiit, Bessel, Fevre, Robinson, Kinoti, Coleman, Welburn, McDermott, & Woolhouse, 2006). Sleeping sickness is a

vector disease spread by the riverine tsetse fly species *Glossina fuscipes fuscipes*; therefore, tsetse fly densities and infection rates are major entomological determinants of sleeping sickness which is regarded as a re-emerging disease (WHO, 1986).

Landsat Enhanced Thematic Mapper (ETM) satellite⁴ data were classified to obtain a map of land cover, and Normalized Difference Vegetation Index (NDVI) and Landsat band-5⁵ were derived as unclassified measures of vegetation density and soil moisture, respectively. GIS functions were used to determine the areas of land cover types and mean NDVI and band-5 values within 1.5 km radii of 389 villages where sleeping sickness incidence had been estimated. Analysis was carried out using backward logistic regression, and proximity to swampland and low population density were found to be predictive factors of reported sleeping sickness presence, with distance to a sleeping sickness hospital as an important confounding variable. The study area comprised the Tororo district in eastern Uganda (Odiit et al., 2006). A sample of 389 villages out of a total of 884 census villages was selected with each village covering an area of approximately 0.5 to 5km², in an area that has two distinct wet (September-November and March-May) and dry (June-August and December-February) seasons. Increasing land pressure has forced people to encroach on marginal habitats to expand the area under cultivation. In eastern and southern Africa, where sleeping sickness occurs, reservoir hosts are a major contributing factor to its persistence. The human population in the area is split between rural mixed farmers, growing subsistence crops and rearing small holdings of cattle, and those living in the region's urban centers. The movement of infected cattle from endemic areas has been implicated in the re-emergence of sleeping sickness in areas where it was not known to be endemic (Fevre et al., 2001) but where tsetse flies are prevalent. Such movements of cattle bring the sleeping sickness agent into contact with the causal vector, which in turn transmits it to humans.

In this study logistic regression was used to first assess the statistical significance of satellite-derived variables, distance to a sleeping sickness hospital and population densities. For logistic regression, a binary response was defined as taking a value of 1 if sleeping sickness was present in a village and 0 otherwise, then using the presence and absence of disease as a predictor variable, backward logistics regression analysis of variables with associations of significance was carried out. After this, backward-stepwise logistic regression was performed to find the most parsimonious model of sleeping sickness risk (Greenland & Maldonado, 1994). This is a typical example in which human and animal mobility critically affects the spread of an infectious disease.

⁴A series of satellites by the USGS and NASA.

⁵Band designated to discriminate moisture content of the soil and vegetation and also able to penetrate thin clouds.

(2) Bayesian and GIS Mapping of Childhood Mortality in Burkina Faso

Investigators used GIS ArcView and an empirical Bayes smoothing technique to map the annual childhood mortality rates for each of 39 villages in the Nouna Demographic Surveillance Area (DSA) in Burkina Faso, West Africa (Sankoh, Berke, Simboro, & Becher, 2002). The study was restricted to children under the age of five years and was carried out between 1993 and 1998. The annual population of children younger than five years per village ranged from 15 to 454, showing a wide range of village size. In summary, annual mortality rates for each village in the study area was calculated using mid-year populations of children under five as the denominator. Two mapping techniques were implemented: first, the GIS software ArcView was used to map the crude mortality rates, and then the data were smoothed by the method of empirical Bayes (shrinkage) estimation⁶. The geostatistical method of Krigging was then administered to spatially interpolate the data for successive years. As an output of the above analysis, a semivariogram (the spatial dependence structure) of the mortality rates was estimated. The method of Krigging was used to produce isopleth maps showing the risk of children living in a certain place in the study region to die in a given year (Sankoh, et al., 2002). The maps showed no clear spatial trend pattern but the authors found that there was a tendency of villages in the northeastern region to produce higher incidence or risk values, which confirmed clustering of disease reported earlier by Sankoh and associates (2001). It is important to note that disease mapping was used primarily as an explorative tool to provide a general insight as opposed to precise estimates of incidence or spatial trends (Kafadar, 1999).

In general, the Bayesian smoothing technique is used to address the issue of heterogeneity in the population at risk and it is therefore a useful tool to use in explorative mapping of disease and mortality. In this study the method was helpful for visual identification of clustering in the northeastern side of the study region.

(3) Modeling Risk from a Disease in Time and Space

Both models for longitudinal and spatial data were combined in a hierarchical Bayesian framework, with particular emphasis on the role of time- and space-varying covariate effects (Knorr-Held & Besag, 1997). Data analysis was implemented via Markov chain Monte Carlo methods and the methodology was applied to the Ohio lung cancer data covering the period of 1968 to 1988. The state of Ohio is located in northeastern United States and is divided into 88 counties. The database consisted of the population size and the number of deaths from lung cancer, stratified by age, gender, and race (white or non-white), for each year between 1968 and 1988 and for each

⁶Other smoothing techniques include the loess, kernel smoothing, or head-banging to name a few (Elliot et al., 2000).

county. Two approaches that adjust for unmeasured spatial covariates, particularly tobacco consumption, were used; the first included the use of random effects model to account for unobserved heterogeneity and the second involved the addition of a simple urbanization measure as a surrogate for smoking behavior. The Ohio data set has been of particular interest because of the suggestion that a nuclear plant located in the southwest of the state may have caused increased levels of lung cancer. The authors, however, concluded that Bayesian smoothing may not be the most appropriate tool for a focused analysis of this nature, and that the Ohio dataset does not provide enough information for any proper conclusion to be drawn. Thus the authors' main interest in the data was to illustrate the use of Bayesian mapping in time and space.

It is important to note that disease occurrence data in time and space is in the form of counts that nominally follow binomial or Poisson distributions. The key property with data of this type is that the outcomes are naturally correlated in time and space. The challenge is then to incorporate this correlation structure in any model one adopts to model the process. The generalized linear model (GLM) therefore is the starting point in an attempt to model such data statistically because the GLM neatly synthesizes likelihood-based approaches to regression analysis for a variety of outcome measures (McCullagh & Nelder, 1989). Extensions of the GLM involve models with random terms in the linear predictor giving rise to generalized linear mixed models (GLMMs). These models are useful for modeling the dependence among outcome variables inherent in longitudinal or repeated measures designs and for producing shrinkage estimates in multi-parameter problems, such as the construction of maps of small area disease rates (Clayton & Kaldor, 1987). For more details about these models including inference methods in GLMMs the reader is referred to the work of Breslow (1993).

Conclusions and Future Research

This chapter has presented the SIR model as the basic mean-field epidemic model (Anderson & May, 1991; Daley & Gani, 1999) upon which further extensions and modifications can be implemented to capture more complex disease processes, in order to introduce basic concepts associated with disease modeling in order to enhance understanding and analysis of such processes. The extension of the above basic model has been particularly directed towards the understanding of the recent SARS epidemic and other general properties which have been described more elegantly by Brauer (2005). Further, the chapter has emphasized the need to develop statistical methods to enable the estimation of key parameters of a disease process and to enable the evaluation of the significance of some key factors driving epidemics such as HIV. As an example of the effect of population mobility on disease transmission, migrant worker effects and the spread of HIV/AIDS in Africa have been discussed.

The chapter has highlighted the concept of social networks as a means of enhancing present understanding of the local dynamics of infectious diseases, particularly when proximity of individuals in space is no longer the determinant factor of whether or not an individual infects or gets infected. Spatial explicit models have been used (e.g., Durrett & Levin, 1994) when spatial heterogeneity plays a significant role in the spread of a disease. The discussion has been focused on sexual network models because partnerships are more easily defined with these infections. The spread of STIs, especially HIV/AIDS depends very much on patterns of sexual contact prevalent in a given population. In some societies serial monogamy may be the norm such that having more than one partner at the same time is an exception, while in other societies polygamy is the norm or at least widely accepted. Understanding of such differences is therefore important to understanding the spread of such infections and to help in designing control and intervention strategies. This chapter has also presented GIS and disease mapping techniques, which are becoming increasingly more applicable by improving decision-making process in disease surveillance control activities. The technique is useful as a tool to visualize spatial patterns in the geographic distribution of disease, for explorative and descriptive purposes as well as to provide information for further studies. Many earlier methods for disease mapping methods adopted the frequentist approaches, but currently Bayesian inference methods (parametric and non-parametric) are gaining popularity because of the advent of powerful computational methods such as the MCMC methods of parameter estimation and inference generation. The combination of standard statistical and Bayesian modeling approaches in order to understand the spread of the highly pathogenic diseases such as HIV/AIDS, Malaria, TB, and childhood diseases in Africa are among the future research interests of the authors of this chapter. Finally, it is our goal to enhance capacity in the continent in the field of disease modeling and mapping to inform policy on the most optimal control strategies that have high efficiency and of minimal cost relative to some existing methods.

References

- Anderson, R.M. & May, R.M. (1991). *Infectious Diseases of Humans*. Oxford, UK: Oxford University Press.
- Anderson, R.M., Fraser, C., Ghani, A.C., Donnelly, C.A., Riley, S., Ferguson, N.M., Leung, G.M., Lam, T.H., & Hedley, A.J. (2004). Epidemiology, transmission dynamics and control of SARS: The 2002-2003 epidemic. *Philosophical Transactions of the Royal Society, London B*, 359, 1091-1105.
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43, 1-59.
- Brauer, F. (2005). The Kermack-Mackendrick epidemic model revisited. *Mathematical Biosciences*, 198, 119-131.
- Breslow, N.E. & Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association*, 88(421), 9-25.

- Brewer, T.H., Hasbun, J., Ryan, C.A., Hawes, S.E., Martinez, S., Sanchez, J., Butler de Lister, M., Constanzo, J., Lopez, J., & Holmes, K.K. (1998). Migration, ethnicity and environment: HIV risk factors for women on the sugar cane plantations of the Dominican Republic. *AIDS*, 12(14), 1879-1887.
- Clayton, D.G. & Bernardinelli, L. (1992). Bayesian methods for mapping disease risks. In (J. Cuzick & P. Elliot, Eds.) *Small Area Studies in Geographical and Environmental Epidemiology*, pp. 205–220, Oxford, UK: Oxford University Press.
- Clayton, D.G. & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43, 671-681.
- Daley, D.J. & Gani, J. (1999). *Epidemic Modelling. An Introduction*. Cambridge Studies in mathematical Biology. Cambridge, UK: Cambridge University Press.
- Decosas, J., Kane, F., Anarfi, J.K., Sodji, K.D., & Wagner, H.U. (1995). Migration and AIDS. *Lancet*, 346(8978), 826-828.
- Decosas, J. & Adrien, A. (1997). Migration and HIV. *AIDS*, 11(Suppl A), S77-S84.
- Donnelly, C.A., Ghani, A.C., Leung, G.M., Hedley, A.J., Fraser, C., Riley, S., Abu-Raddad, L.J., Ho, L.M., Thach, T.Q., Chau, P., Chan, K.P., Lam, T.H., Tse, L.Y., Tsang, T., Liu, S.H., Kong, J.H., Lau, E.M., Ferguson, N.M., & Anderson, R.M. (2003). Epidemiological determinants of causal agents of severe acute respiratory syndrome in Hong Kong. *Lancet*, 361, 1761-1766.
- Donnelly, C.A. (2004). Infectious Disease Epidemiology and Surveillance. Workshop on the epidemiological Methods and Clinical Trials for Preventive Health, 7-9 July 2004, Sydney, Australia.
- Durret, R. & Levin, S.A. (1994). Stochastic spatial models: a user's guide to ecological applications. *Philosophical Transactions of the Royal Society, London B*, 343, 329-350.
- Dushoff, J. (1999). Host heterogeneity and disease endemicity. A moment-based approach. *Theoretical Population Biology*, 56, 325-335.
- Eames, K.T. & Keeling, M.J. (2002). Modelling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences*, 99(20), 13330-13335.
- Elliot, P., Wakefield, J.C., Best, N.G., & Briggs, D.J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford, UK: Oxford University Press.
- Evian, C. (1993). The socio-economic determinants of the AIDS epidemic in South Africa—a cycle of poverty. *South Africa Medical Journal*, 83, 653-656.
- Fevre, E.M., Coleman, P.G., Odiit, M., Magona, J.W., Weburn, S.C., & Woolhouse, M.E.J. (2001). The origins of a new *Trypanosoma brucei rhodensiense* sleeping sickness outbreak in eastern Uganda. *Lancet*, 358, 625-628.
- Fleming, D.T. & Wasserheit, J.N. (1999). From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sexually Transmitted Infections*, 75(1), 3-17.
- Ghani, A.C. & Garnett, G.P. (1998). Measuring sexual partner networks for transmission of sexually transmitted diseases. *Journal of the Royal Statistics*, A161, 227-238.
- Greeland, S. & Maldonado, G. (1994). The interpretation of multiplicative-model parameters as standardized parameters. *Statistics in Medicine*, 13, 989-999.
- Jochelson, K., Mothibeli, M., & Leger, J.P. (1991). Human immunodeficiency virus and migrant labor in South Africa. *International Journal of Health Services*, 21(1), 157-173.
- Kafadar, K. (1999). Simultaneous smoothing and adjusting mortality rates in U.S. counties: Melanoma in white females and white males. *Statistics in Medicine*, 18, 3167-3188.

- Keeling, M. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society, London, B*, 266, 859-869.
- Keeling, M. (1997). Modelling the persistence of measles. *Trends in Microbiology*, 5, 513-518.
- Kermack, W.O. & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society, A*, 115, 700-721.
- Knorr-Held, L. & Besag, J. (1997). *Modeling risk from a disease in time and space*. Technical Report Series, NRCSE-TRS No. 005, University of Munich, Germany.
- Knorr-Held, L. & Raer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56, 13-21.
- Kretzschmar, M. & Morris, M. (1996). Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences*, 133, 165-195.
- Lagarde E., Pison, G., & Enel, C. (1996). A study of sexual behavior change in rural Senegal. *Journal of Acquired Immune Deficiency Syndromes*, 11, 282-287.
- Lawson, A., Biggeri, A., Boehning, D., Lesaffre, E., Viel, J.F., & Bertollini, R. (1999). *Disease Mapping and Risk Assessment for Public Health*. Chichester, UK: Wiley.
- Lurie, M.N., Williams, B.G., Zuma, K., Mkaya-Mwamburi, D., Garnett, G.P., Sturm, A.W., Sweat, M.D., Gittelsohn, J., & Abdool Karim, S.S. (2003a). The impact of migration on HIV-1 transmission in South Africa: a study of migrant and nonmigrant men and their partners. *Sexually Transmitted Diseases*, 30(2), 149-156.
- Lurie, M., Williams, B.G., Zuma, K., Mkaya-Mwamburi, D., Garnett, G.P., Sweat, M.D., Gittelsohn, J., & Abdool Karim, S.S. (2003b). Who infects whom? HIV-1 concordance and discordance among migrant and non-migrant couples in South Africa. *AIDS*, 17(15), 2245-2252.
- Mabey, D. & Mayaud, P. (1997). Sexually transmitted diseases in mobile populations. *Genitourinary Medicine*, 73(1), 18-22.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models* (2nd Ed.). London, UK: Chapman and Hall.
- Mueller, J., Kretzschmar, M., & Dietz, K. (2000). Contact tracing in deterministic and stochastic models. *Mathematical Biosciences*, 164, 61-75.
- Morris, M. & Kretzschmar, M. (1995). Concurrent partnerships and transmission dynamics in networks. *Social Networks*, 17, 299-318.
- Mollison, D. (1977). Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society, B*, 39, 283-326.
- Nicoll, A. & Gill, O.N. (1999). The global impact of HIV infection and disease. *Communicable Disease and Public Health*, 2(2), 85-95.
- Odiit, M., Bessel, P.R., Fevre, E.M., Robinson, T., Kinoti, J., Coleman, P.G., Welburn, S.C., McDermott, J., & Woolhouse, M.E.J. (2006). Using remote sensing and geographic information systems to identify villages at high risk *rhodesiense* sleeping sickness in Uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100, 354-362.
- Pison, G., Le Guenno, B., Lagarde, E., Enel, C., & Seck, C. (1993). Seasonal migration: A risk factor for HIV infection in rural Senegal. *Journal of Acquired Immune Deficiency Syndromes*, 6, 196-200.
- Quinn, T.C. (1994). Population migration and the spread of types 1 and 2 human immunodeficiency viruses. *Proceedings of the National Academy of Sciences*, 91, 2407-2414.
- Ross, R. (1911). *The Prevention of Malaria*, (2nd Ed.). London, UK: Murray.

- Sattenspiel, L. & Dietz, K. (1995). A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences*, 128, 71-91.
- Sankoh, O.A., Ye, Y., Sauerborn, R., Mueller, O., & Becher, H. (2001). Clustering of childhood mortality in rural Burkina Faso. *International Journal of Epidemiology*, 20, 485-492.
- Sankoh, O.A., Berke, O., Simboro, S., & Becher, H. (2002). Bayesian and GIS mapping of childhood mortality in rural Burkina Faso. SFB 544 *Control of Tropical Infectious Diseases*, Discussion Paper 03/2002, Uni-Heidelberg, Germany.
- Turner, K.M.E., Garnett, G.P., Ghani, A.C., Sterne, J.A.C., & Low, N. (2004). Investigating ethnic inequalities in the incidence of sexually transmitted infections: mathematical modeling study. *Sexually Transmitted Infections*, 80, 379-385.
- Van Baalen, M. (2005). Contact networks and the evolution of virulence. In: *Adaptive Dynamics of Infections Diseases: In Pursuit of Virulence Management*. Diekmann, U., Metz, J.A.J., Sabelis M.W. and Sigmund, K. (Eds), pp. 85-103. Cambridge, UK: Cambridge University Press.
- WHO (1986). Epidemiology and control of human African trypanosomiasis. Report of WHO expert committee on sleeping sickness. Geneva: World Health Organization, Technical Report Series No. 739.
- WHO (1997). Disease mapping and risk assessment for public health decision-making. Report on a WHO workshop, Rome, Italy, 2-4 October, 1997. EUR/ICP/EHRO 02 03 01 (In Lawson et al. 1999: pp. 453-468).
- Wylie, J.L. & Jolly, A. (2001). Patterns of Chlamydia and Gonorrhoea infection in sexual networks in Manitoba, Canada. *Sexually Transmitted Diseases*, 28(1), 14-24.
- Zuma, K., Gouws, E., Williams, B.G., & Lurie, M. (2003). Risk factors for HIV infection among women in Carletonville, South Africa: Migration, demography and sexually transmitted diseases. *International Journal of STD & AIDS*, 14, 814-817.
- Zuma, K., Lurie, M., Williams, B.G., Mkaya-Mwamburi, D., Garnet, G.P., & Sturm, A.W. (2005). Risk factors of sexually transmitted infections among migrant and non-migrant sexual partnerships from rural South Africa. *Epidemiology and Infections*, 133, 421-428.
- Zuma, K. & Lurie, M. (2005). Application and comparison of methods for analyzing correlated interval censored data from sexual partnerships. *Journal of Data Science*, 3, 241-256.