

Research Article

A Network Model for Dispersion of Communicable Diseases

Ling Bian

*Department of Geography
University at Buffalo, State University
of New York*

David Liebner

*School of Medicine and Biomedical
Sciences
University at Buffalo, State University
of New York*

Abstract

The spread of communicable diseases through a population is an intrinsic spatial and temporal process. This paper presents an individual-based analytical framework for modeling the spatial and temporal heterogeneity in the disease transmission. The framework specifies a network model structure and six associated parameters. These parameters describe the properties of nodes, the properties of links, and the topology of the network. Through this model structure and associated parameters, this framework allows the representation of discrete individuals, individualized interactions, and interaction patterns in a network of human contact. The explicit representation of the spatial distribution and mobility of individuals in particular facilitates the modeling of spatial heterogeneity in the disease transmission.

1 Introduction

The spread of communicable diseases (those that are transmitted from individual to individual) through a population is an intrinsic spatial and temporal process. The temporal progress of an epidemic has been modeled through a family of population-based models. The simplest form of these models divides a population into three mutually exclusive segments. These are those who are susceptible to the disease (S), those who are infectious and can spread the disease to susceptibles (I), and those who are recovered from a previous infection and can no longer spread or catch the disease (R). These so-called SIR models use mathematical equations to describe the dynamics, in terms of size, of the three segments through the course of an epidemic (Kermack and McKendrick 1927, Anderson and May 1991). These models are explicitly temporal, but non-spatial. The spatial version of the SIR model is a dynamic wave model. In this wave model, the

Address for correspondence: Ling Bian, Department of Geography, University at Buffalo, State University of New York, Amherst, NY 14261, USA. E-mail: lbian@buffalo.edu

infectious population segment is on the crest of the wave, while the susceptible segment is in front of the crest, leaving the recovered behind. Diseases spread in a population when the wave travels through the space (Cliff and Ord 1981, Cliff et al. 1986, Rhodes and Anderson 1997). These models, including their spatial version, have been used in numerous applications throughout the history of modern epidemiology.

An epidemic is usually expressed as a curve, reporting daily infections during an outbreak. By adjusting the value of parameters, these models can be fitted to an observed curve. The parameters are usually presented as an averaged value over a population. The observations of recent outbreaks of severe acute respiratory syndrome (SARS), anthrax, and ebola raised serious questions on how these parameters should be estimated. At the core of the issue is the heterogeneity in disease transmission (Arita et al. 2003, Dye and Gay 2003, Francesconi et al. 2003, Mbogo et al. 2003, Meyers et al. 2003, Galvani 2004, Kretzschmar et al. 2004). It is argued that an averaged value over a population may not lead to a correct understanding of how diseases spread (Dye and Gay 2003). The potential bio-terror threats, such as smallpox, further demand appropriate approaches to modeling the observed heterogeneity (Rhodes and Anderson 1997, McKenzie 2004, Walden and Kaplan 2004). This issue is of great interest and significance to researchers, health policy makers, and the public (McKenzie 2004).

In the past decade, individual-based modeling began to gain momentum in many disciplines (note that individual-based modeling is often equated to or considered as part of agent-based modeling in certain disciplines). As an alternative to population-based models, individual-based models present a different modeling concept and can compensate for many shortcomings inherent in population-based models. This paper presents an individual-based analytical framework for epidemiological modeling. This framework allows the representation of heterogeneous transmission of diseases by modeling discrete individuals and the interactions between them in a network of human contact. The spatial distribution and mobility of these individuals facilitate the modeling of spatial heterogeneity in the disease transmission. This paper focuses on the model structure and its associated parameters to define the analytical framework.

The remainder of the paper begins with the basics of population-based models in order to contrast the individual-based analytical framework. The sections that follow discuss an individual-based conceptual framework, a model structure, and the parameters associated with the model structure. A simulated influenza epidemic is used throughout the discussion to illustrate the effects of individual parameters. The spatial and temporal heterogeneity in the spread of disease under two scenarios are demonstrated toward the end of the paper.

2 Population-Based Models

The three-population-segments SIR models are based on the assumption that at any given time, a number of individuals are removed from the susceptible segment and added to the infectious segment, and in the meantime, a number of individuals are moved from the infectious to the recovered segment. Assuming that these changes are continuous, differential equations are typically used to express the changes:

$$dS/dt = -\beta SI, \quad dI/dt = \beta SI - gI, \quad \text{and} \quad dR/dt = gI \quad (1)$$

where S , I , and R denote the susceptible, infectious, and recovered individuals, respectively; β is the infection coefficient, the product of the average number of contacts

within a given time period and the probability of infection for a contact between susceptible and infectious individuals; and g is the recovery rate. This differential equation model structure uses a small number of simple parameters. Adjusting β and g , a SIR model can estimate the size of the infectious segment through an epidemic. In a single epidemic, the size of the infectious segment rises after an epidemic begins and declines after reaching its peak (Kermack and McKendrick 1927, Anderson and May 1991).

The SIR model is the simplest form of the family of population-based models. The more complex models may take additional considerations into account. For example, the model may allow births, deaths, and migration into and out of the population. Each segment can be further divided. For example, the infectious segment could be broken down into infected and infectious segments. The former includes individuals who are infected, but not yet infectious because they are in the latent period, whereas those in the latter segment can infect others. A single epidemic forecast can be extended to multiple epidemics. A variety of complexities have been introduced to the basic framework of SIR models by adding more parameters into the equations (Anderson and May 1991). The basic modeling framework, however, has remained the same throughout the history of modern epidemiology. These models do not explicitly address causal factors in the spread of diseases. Instead, they directly estimate the size of affected population segments. The deterministic form of these models is advantageous for its simplicity in structure and mathematics.

In recent years, the SIR model and its derivatives have drawn increasing criticism for failing to produce realistic and useful results, especially for complex disease systems (Holmes 1997, Koopman and Lynch 1999). Recent reports on the spread of some of the most dangerous communicable diseases, such as SARS, anthrax, and ebola, further reveal the limitations of the population-based modeling framework. By adjusting the value of parameters, the population-based models can fit a curve to an observed epidemic, but results can be misleading. These population-wide homogeneous parameters cannot help decipher how diseases spread and whether the decline of an epidemic is a result of intervention measures or heterogeneity in infection (Dye and Gay 2003). Many reports indicate that it is the heterogeneity in transmission that often leads to the sustaining or decline of an epidemic, instead of intervention measures (Arita et al. 2003, Dye and Gay 2003, Francesconi et al. 2003, Mbogo et al. 2003, Meyers et al. 2003, Galvani 2004, Kretzschmar et al. 2004).

The limitations of population-based models are rooted in the assumptions underlying them. Of many assumptions embedded in the population-based models, a number of basic ones regarding the population are most relevant to this discussion. These include the following five assumptions. First, all individuals are assumed to be identical except for their infection status, i.e. either susceptible, infectious, or recovered. Second, all individuals have interactions with all other individuals, i.e. the so-called free mixing or homogeneous mixing. Third, all individuals interact with each other at an equal rate. Fourth, regarding the spatial location of the individuals, the classic population-based models are non-spatial. The spatial version of the classic models assumes a uniform distribution. Fifth and lastly, these models consider individuals to be immobile. These homogeneity assumptions inherently limit the usefulness of population-based models in explaining the observed heterogeneity in disease infection, spatially as well as temporally. Alternative modeling concepts, such as individual-based modeling, have attracted attention in the past decade. The analytical framework discussed in this paper represents this recent shift (DeAngelis and Gross 1992, Judson 1994).

3 Conceptual Framework for Individual-Based Models

The analytical framework presented in this paper is built on an individual-based conceptual framework for epidemiological modeling (Bian 2004). This conceptual framework explicitly considers discrete individuals, their properties, and interactions between them for modeling the spread of diseases in a spatially explicit fashion. Based on this conceptual framework, the analytical framework discussed in this paper further defines the model structure and its associated parameters. This section briefly describes the conceptual framework and associated assumptions.

The conceptual framework detailed in Bian (2004) is developed based on several assumptions. First, individuals may differ in properties, such as age or occupation, in addition to infection status (Assumption 1). Second, individuals interact with a finite number of other individuals within a given period of time (Assumption 2). Third, the number of interactions varies by individuals (Assumption 3). Fourth, individuals are spatially distributed (Assumption 4). Fifth and lastly, individuals are mobile (Assumption 5). These assumptions are established to explicitly account for the heterogeneity in individuals, the individualized interactions between them, and subsequent heterogeneity in disease transmission, spatially and temporally. Specifically, these assumptions are intended for modeling disease transmission in an urban environment on a daily basis.

Built on these heterogeneity assumptions, the conceptual framework is composed of the following elements, properties, and relationships. First, individuals are used as the modeling unit, each with certain properties, such as health status and behavior, including interaction with others. These properties and behavior may change through time and location. Second, the temporal dynamics of disease development are treated as discrete periods. These periods indicate an individual's susceptibility to infection or the ability to spread diseases. The length of these periods may affect the spatial and temporal heterogeneity of disease transmission. These periods are used as one of the properties of an individual to describe their health status.

Third, individuals have an explicit nighttime location at homes and a daytime location at workplaces. Most individuals travel between the two locations on a daily basis. The movement of individuals is then represented as individual lifelines, following the concept of time geography (Hägerstrand 1970, Pred 1977, Lenntorp 1978, Löytönen 1998, Kwan 1999, Miller 2005). These lifelines intersect at homes and workplaces, forming a local network. The travel between homes and workplaces by these individuals links the local networks into a population-level network. Through this two-level (local and population levels) and two-population (nighttime and daytime populations) network, diseases may spread by local infections and long-distance dispersions. The local infections occur through interactions at home or workplaces, while long distance dispersions occur when individuals travel. Fourth, the interactions may vary, for example, in the number of contacts within and outside a local network, the probability of being infected once in contact with an infectious individual, and the length of a health status period.

To summarize, the conceptual framework emphasizes discrete individuals, individualized interactions, and changes in health status and interaction of these individuals with location and time. Collectively, these differences and changes portray a heterogeneous spread of disease through the two-level and two-population network. This conceptual framework is used to guide the design of the analytical framework discussed in the following sections.

4 Analytical Framework

Based on the conceptual framework that identifies the basic elements and relationships, the analytical framework further defines the model structure and its associated parameters. The model structure takes the form of a network that consists of nodes and links. Each individual is represented as a node and the interaction between any two individuals is a link. The properties of nodes, the properties of links, and the topology of the network are specified by a set of parameters.

In recent years, the concept of network has re-emerged as an actively pursued research topic (Watts and Strogatz 1998, Keeling 1999, Albert et al. 2000). These mathematics-based studies provide insights into how network topology affects the performance of a network and how the topology of a network is characterized. Epidemiological studies, on the other hand, offer ample anecdotal evidence supporting a network model structure for describing the spread of diseases (Jernigan et al. 2002, Traeger et al. 2002, Arita et al. 2003, Francesconi et al. 2003, Griffith et al. 2003, Meyers et al. 2003, Galvani 2004, Hsueh et al. 2004, Lau et al. 2004, Zhuang et al. 2004). Both fields of studies contribute to the establishment of this analytical framework.

In order to explicitly account for the spatial heterogeneity in disease transmission, the analytical framework places two identical sets of individuals into two corresponding spaces. These are the nighttime population in the home space and the daytime population in the workplace space. Within each space, each individual has an explicit location, represented by x and y coordinates. In the home space, individuals are assigned to families. Family members are located close to each other, forming a local network. Individuals are nodes and the connections between individuals are links. Similarly, individuals in the workplace space are assigned to workplaces, and co-workers are located close to each other to form another set of local networks. The local networks in the respective spaces are linked together through individuals who have locations in both spaces. This spatially explicit design is necessary to reveal the spatial heterogeneity in the disease transmission.

A total of six parameters are used to describe this network. These are: (1) length of latent period, (2) length of infectious period, (3) disease infection rate, (4) number of connections of an individual, (5) degree of inter-connection between family members and co-workers, and (6) ratio between workplace connections to family connections. Of the six parameters, the first two, length of latent period and length of infectious period, are properties of individuals, or properties of nodes in the context of a network. The third parameter, infection rate, depicts individualized interaction between an infectious and a susceptible individual, or a property of a link. The last three parameters, number of connections, degree of inter-connections, and ratio between family connections and workplace connections, characterize connections between individuals, or the topology of the network.

For discussion purposes, a simulation of an influenza epidemic involving 5,000 individuals is used to illustrate the analytical framework. Influenza is chosen for the simulation because it is common and readily communicable between individuals. In order for the discussion to focus on the six parameters, the simulation simply assumes that the population is closed. That is, there are no births, deaths, or migrations during the epidemic, and the recovered individuals are immune. For actual working models, the individual-based framework can easily accommodate an open population. The 5,000 individuals are grouped into families of an average size of three and workplaces of an

average size of 11. The latter is the number of people with whom an individual is in direct contact on a daily basis; thus, it may appear to be smaller than the size of an actual workplace. A single infection is introduced to the 5,000 initially susceptible individuals. The epidemic is described by a curve that reports the number of daily infections through the course of the epidemic, i.e. the epidemic curve. Three descriptive terms are used to characterize an epidemic curve. These are the peak time of an epidemic, the number of infections at the peak time, and the total number of infections.

In the simulation, alternate values are given to each parameter, while holding other parameters constant, in order to illustrate the effect of the parameter. The default values for these indices are set as follows. The length of latent period = 1 day, and the length of infectious period = 3 days. The infection rate is set to 0.1, out of a range of 0 to 1, for an individual at either home or a workplace. For the combined rate at home and workplace during a 24-hour day, the infection rate = $0.1 + 0.1 - 0.1 * 0.1 = 0.19$. The first two 0.1 terms represent the independent infection rate at home and workplace, respectively. The term $0.1 * 0.1$ represents the joint probability of infection. That is, an individual is infected in either home or workplace and subsequently, in 12 hours, infected again in the other place. This situation is deemed unlikely and subtracted from the combined rate. The average number of connections of an individual = 12. This is the sum of the average number of family contacts of two (with an average family size of three) and the average number of workplace contacts of 10 (with an average workplace size of 11). The degree of inter-connection = 0.5 out of a range of 0 to 1. This parameter is represented by a covariance that is explained later in section 5. Lastly, the ratio between workplace and family connections = 3:1. A total of 500 realizations are performed using the Monte Carlo method for each parameter value. The average value of the 500 realizations is used to construct each epidemic curve.

The six parameters are explained for their role in a network, their importance in epidemiological modeling, their effect on an epidemic, and how to estimate their values in a working model. The descriptions are presented in the following order: node properties, link properties, and network topology.

1. Length of latent period – In a network, nodes may have properties. Length of latent period that describes the health status of an individual is used to represent one such property. The latent period begins when an individual is infected and ends when the individual becomes infectious. The length of latent period is one of the most fundamental factors that affect the progress of an epidemic, especially for individualized modeling (Rhodes and Anderson 1997, Dye and Gay 2003, Galvani 2004, McKenzie 2004, Walden and Kaplan 2004). By representing a property of individuals, this parameter is an implementation of the individual difference assumption (Assumption 1) outlined above in the conceptual framework.

Figure 1 demonstrates the effect of the length of latent period on an epidemic by varying the length from 0.5 to 5 days. A shorter latent period leads to an earlier and higher peak in the epidemic (Figure 1a) and a greater total number of infections during the course of an epidemic (Figure 1b). This is because a short latent period may allow a rapid increase in the number of infected individuals early in the epidemic. When the susceptible pool becomes exhausted soon after the epidemic begins, the total number of affected individuals declines quickly. For a working model, the selection of the length of latent period depends on the clinical information for specific diseases. Individual difference in the length of latent period can be easily implemented when treated as one of the properties of an individual.

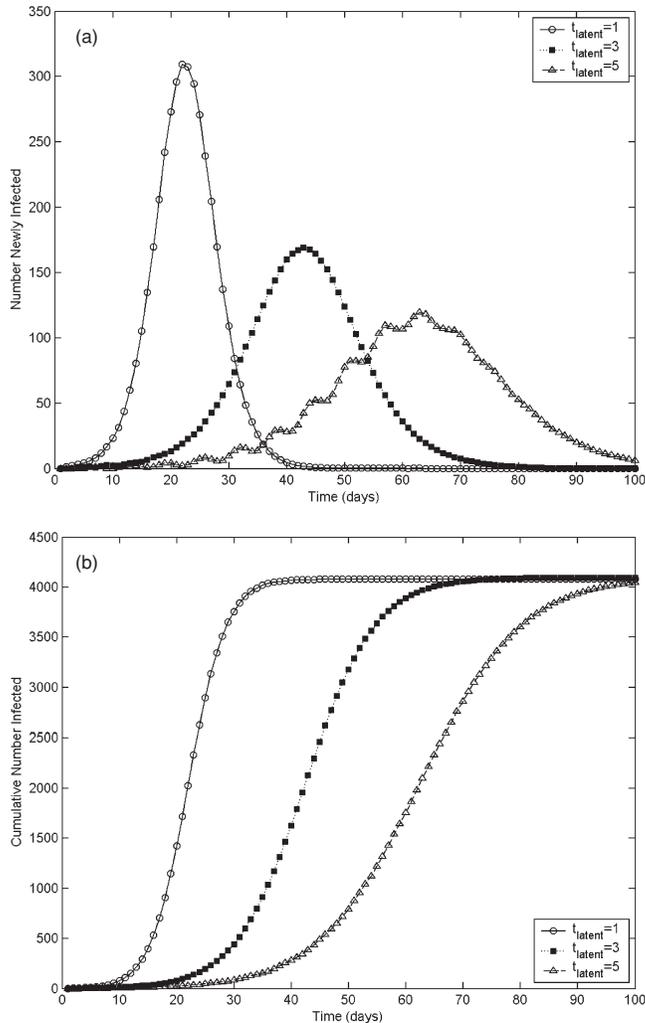


Figure 1 A demonstration of the effect of the length of latent period on an epidemic: (a) the number of daily new infections throughout an epidemic in response to the length of latent period (t_{latent}) of 1, 3, and 5 days; and (b) the cumulative number of infections throughout the epidemic in response to the three lengths of latent period

2. Length of infectious period – Much like the length of a latent period discussed above, the length of an infectious period is another parameter used to represent the properties of a node. This parameter describes the length of a period in which an individual can infect other individuals. It is also a fundamental factor that influences the spread of disease for individualized modeling (Rhodes and Anderson 1997, Dye and Gay 2003, Galvani 2004, McKenzie 2004, Walden and Kaplan 2004). As one of the properties of an individual, this parameter is also an implementation of the individual difference assumption (Assumption 1). Figure 2 displays the effect of the length of an infectious period by varying the length from 1 to 5 days. A longer infectious period

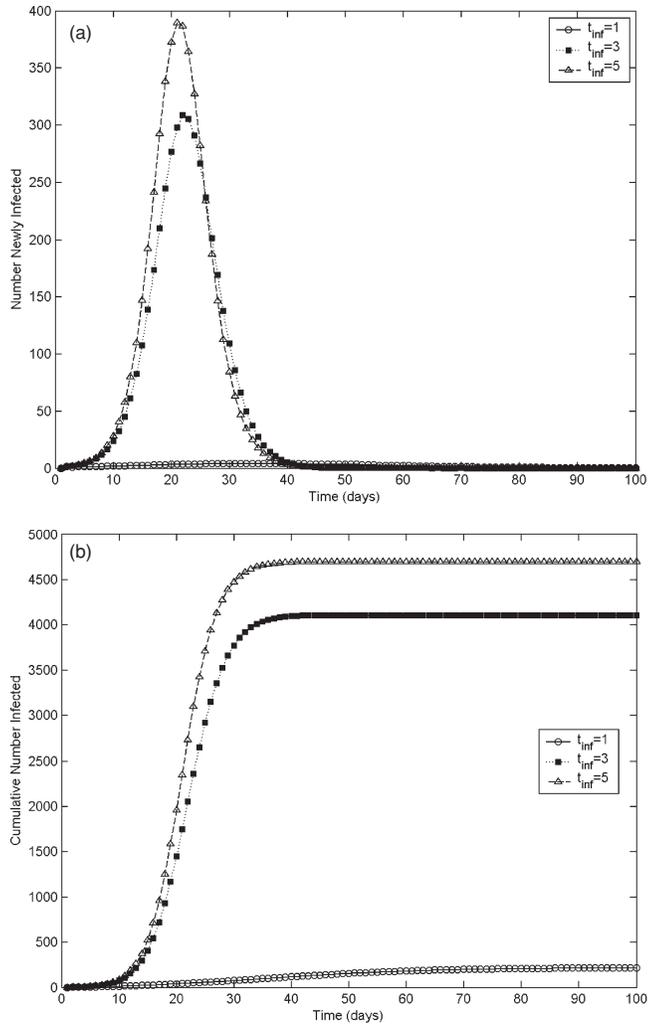


Figure 2 A demonstration of the effect of the length of infectious period on an epidemic: (a) the daily new infections throughout an epidemic in response to the length of infectious period (t_{inf}) of 1, 3, and 5 days; and (b) the cumulative number of infections throughout the epidemic in response to the three lengths of infectious period

results in a slightly earlier peak, a higher peak (Figure 2a), and a greater number of individuals affected during the epidemic (Figure 2b). Similar to the length of latent period, the selection of the length for an infectious period depends on the clinical information on specific diseases. Note that, for certain diseases, all infected individuals will most likely become infectious as assumed in this simulation.

3. Infection rate – Links may have both quantity and direction. Disease infection rate, along with infection direction, is used to represent these properties of link. In the context of epidemiology, disease infection is a directional event. Communicable diseases

are always transmitted from the infectious to the susceptible individuals. This direction determines the temporal sequence of disease transmission. An infection begins with an infectious individual and spreads to other individuals in a local network, subsequently to other local networks, and eventually to the entire population.

The infection rate represents the probability for a susceptible individual to be infected once in contact with an infectious individual. This is one of the most commonly used parameters in population-based models (Anderson and May 1991, Ferguson et al. 1997, Keeling 1999). In the individual-based modeling framework, this is another implementation of the individual difference assumption (Assumption 1). The infection rate dictates many characteristics of an epidemic. Figure 3 shows the effect of the infection rate on an epidemic by varying the rate from 0 (0%) to 1 (100%). An infection rate of 1 or close to 1 causes a rapid rise in the number of daily infections, a large number of infections at the peak time, and a large total number of infections throughout the epidemic. An infection rate of 0 or close to 0 does not initiate an epidemic.

Similar to the length of latent and infectious periods, the selection of the infection rate for a working model depends on the clinical information for specific diseases. With the individual-based framework, different infection rates can be assigned among individuals. For example, individuals of certain ages (e.g. children and elderly) and occupations are most vulnerable to infection and may have higher infection rates than others. Age and occupation can be treated as properties of individuals in a working model. These properties can be estimated from nighttime population data, for example, the census data in the U.S.

4. Number of connections of an individual – The number of connections of an individual, when expressed as number of links of a node, is one of the most basic parameters to describe the topology of a network (Watts and Strogatz 1998, Keeling 1999, Albert et al. 2000). This parameter can be used to determine how nodes are linked into a network and, consequently, how a disease spreads through the network. In combination with infection rate, the number of connections of an individual is one of the most important parameters used in population-based epidemiological models (Anderson and May 1991, Keeling 1999, Ferguson et al. 1997).

Within the individual-based modeling framework, the number of direct connections is used to describe the sum of the number of family members and the number of co-workers associated with an individual. Heterogeneity in the number of connections has been attributed to the heterogeneity in the spread of communicable diseases (Dye and Gay 2003, Francesconi et al. 2003, Meyers et al. 2003). This parameter thus implements the assumptions of a finite and diverse number of connections that an individual may have (Assumptions 2 and 3). Figure 4 illustrates the effect of the number of connections of an individual on an epidemic, with the value ranging between 1 to 25. The greater the number of direct connections there is, the earlier the epidemic peaks with a great number of infections at the peak time. Subsequently, a greater number of individuals are affected during the epidemic.

For a working model, the estimation of the sum of the number of family and workplace connections depends on the demographic make up and employment structure of a community. The number of connections at home can be estimated from the family size reported in census data. The estimation of the number of connections at workplaces requires two pieces of information, an individual's workplace and the size of the workplace. Although there is no standardized data like census for the daytime population in the U.S., the workplace size can be estimated from other data sources. It is challenging,

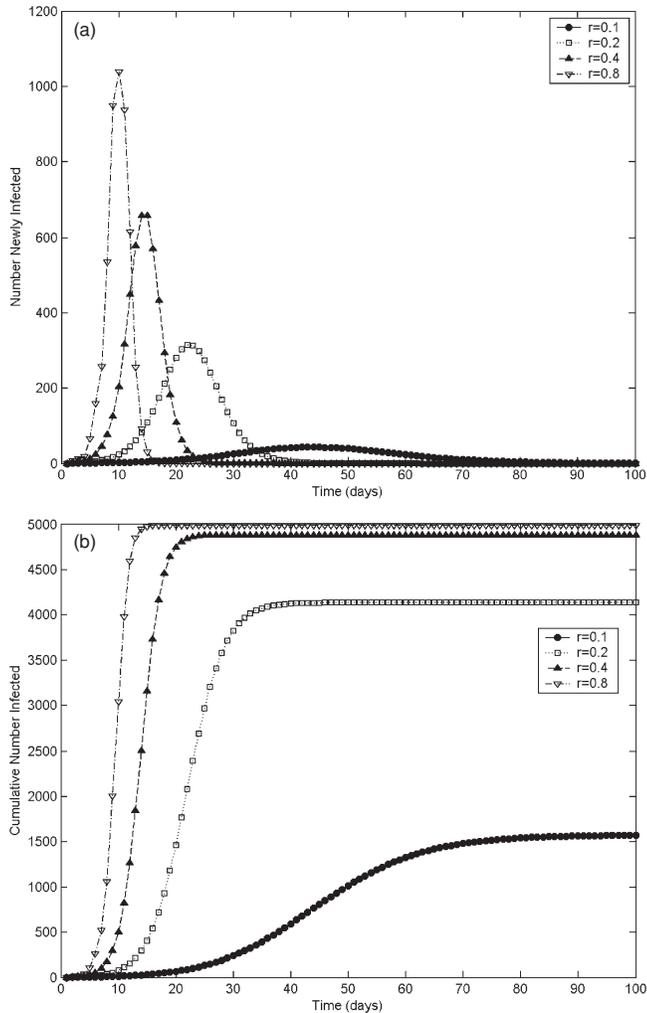


Figure 3 A demonstration of the effect of the infection rate on an epidemic: (a) the daily new infections throughout an epidemic in response to the infection rate (r) of 0.1, 0.2, 0.4, and 0.8; and (b) the cumulative number of infections throughout the epidemic in response to the four infection rates

however, to identify the workplace for an individual. The census data provide, at an aggregated level, the occupation type and the time required to travel to work for a population. This information, though at the aggregated level, can help estimate the workplace for an individual and subsequently the number of connections at the workplace. First, the home location and occupation of an individual can be estimated using TIGER data. Possible workplace locations of the occupation can be obtained from many sources. The workplace that meets the required time for an individual to travel to work can be assigned to the individual.

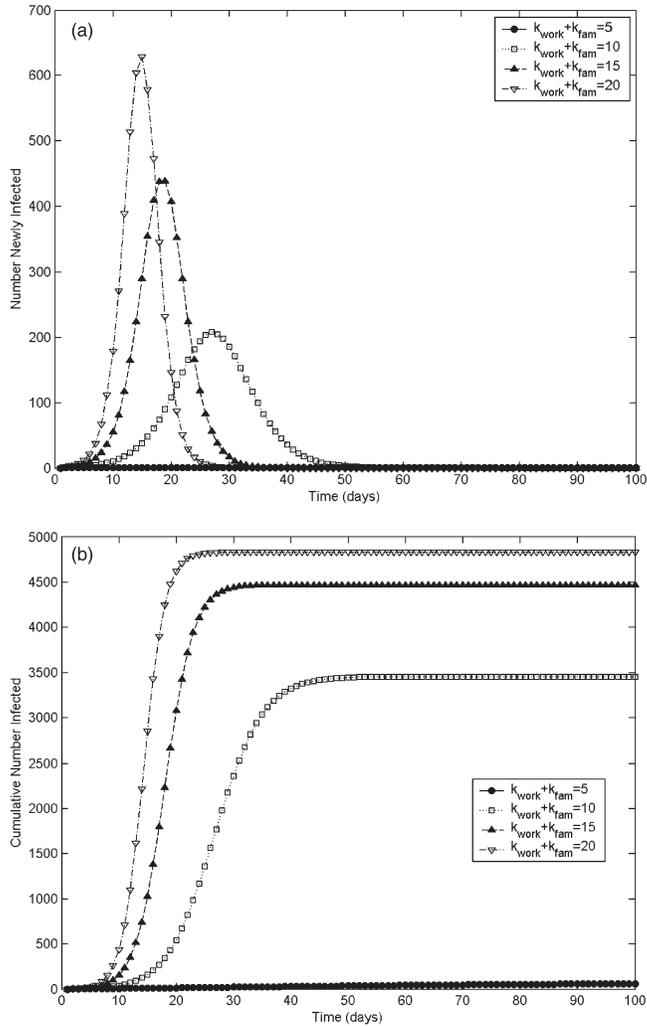


Figure 4 A demonstration of the effect of the number of connections on an epidemic: (a) the daily new infections throughout an epidemic in response to the number of connections ($k_{work} + k_{fam}$) of 5, 10, 15, and 20; and (b) the cumulative number of infections throughout the epidemic in response to the four connection numbers

5. Degree of inter-connection between family members and co-workers – Like the number of links for a node, the degree of inter-connection is one of the most basic parameters used to describe the topology of a network (Watts and Strogatz 1998, Keeling 1999, Albert et al. 2000). This parameter describes how many nodes that are linked to a given node are also linked between themselves. In the context of disease infection, this is one of the parameters that affects the progress of an epidemic, especially in the context of individual-based modeling (Keeling 1999). Heterogeneity in the degree of inter-connections can determine which parts of a society may be

affected (Arita et al. 2003). This parameter directly implements the individual spatial distribution and mobility assumptions (Assumptions 4 and 5). It is used as an indicator of social, in addition to demographic and employment structures, of a community.

Various mathematical indices have been explored to express this parameter (Watts and Strogatz 1998, Keeling 1999, Albert et al. 2000). These indices mostly involve relationships between no more than three nodes and are not spatially explicit. This paper presents an alternative index to describe the inter-connection between family members and co-workers related to an individual. The index is expressed as a covariance of an individual's location in the home space and in the workplace space for x and y coordinates, respectively. A low covariance value indicates that an individual's home location differs from their workplace location. Thus, this individual interacts with different groups of individuals between daytime and nighttime, implying a low degree of inter-connection between family members and co-workers. A high value of covariance indicates that an individual interacts with a similar group of individuals in daytime and nighttime, implying a high degree of inter-connection. With the covariance varying between 0 and 1, Figure 5 exhibits the effect of inter-connection on an epidemic. A lower covariance raises the daily number of infections rapidly earlier in an epidemic, resulting in a greater number of total infections throughout an epidemic. This is because lower degrees of inter-connection between family members and co-workers allow an individual to interact with a greater number of individuals on a daily basis, consequently creating a rapid spread of diseases through the population.

For a working model, the estimation of the degree of inter-connection requires knowledge of the social structure of a community, in addition to demographic and employment structures. This presents a greater challenge than the estimation for the number of connections described above. This is because inter-connections require a more accurate estimation on how the nighttime population is linked to the daytime population. Current census data can be used for the intended estimation. Additional research is needed to support more accurate estimations.

6. Ratio between workplace and family connections – The ratio between workplace and family connections associated with an individual is another index that is presented in this paper to describe the topology of a network. In the context of individual-based epidemiological modeling, this parameter contrasts the local networks in the home space with those in the workplace space. The parameter is used as another indicator of social, demographic and employment structures, and an implementation of individual spatial distribution and mobility assumptions (Assumptions 4 and 5). Figure 6 illustrates the effect of this parameter by varying the ratio value between 1, 3, and 5. The effect can be demonstrated by a higher ratio value. This high ratio value produces an epidemic with a later peak, a lower peak, and fewer total infections. A large workplace to family ratio means a large size for workplaces and a small size for families. In this situation, the small family size results in inefficient spread of diseases through family connections, and the disease spread throughout the population is primarily dependent on the workplace connections. This situation makes the entire network inefficient and makes it difficult for diseases to spread.

The quantities of nodes, the quantity and direction of links, and the topology of the network collectively characterize the exposure of individuals to diseases in a network of human contact. According to these characteristics, various spatial and temporal patterns in disease transmission can be modeled.

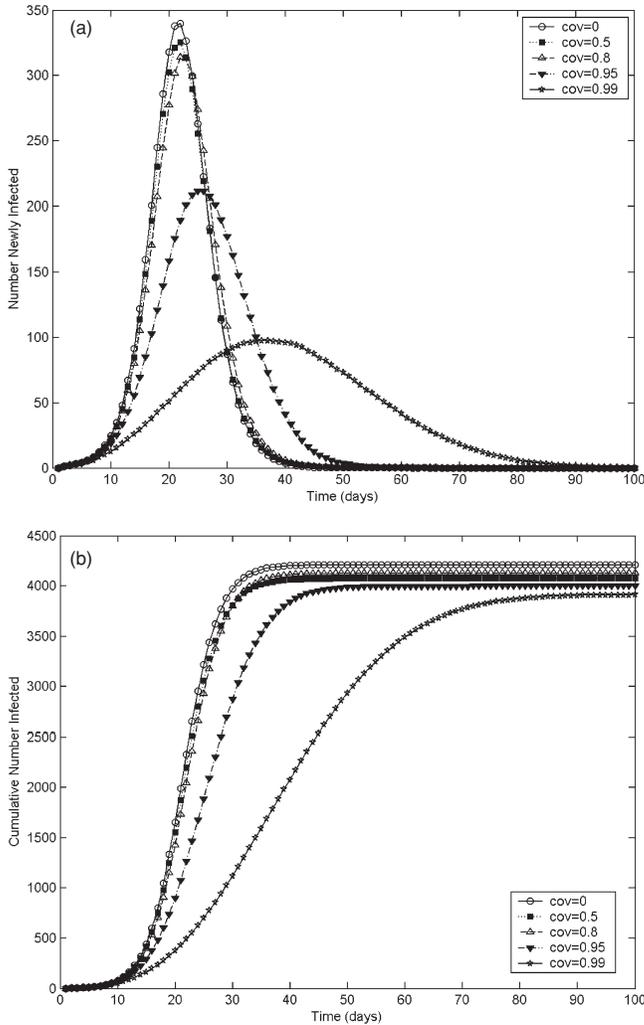


Figure 5 A demonstration of the effect of the degree of inter-connection on an epidemic: (a) the daily new infections throughout an epidemic in response to the degree of inter-connection (cov) of 0, 0.5, 0.8, 0.95, and 0.99; and (b) the cumulative number of infections throughout the epidemic in response to the five degrees of inter-connection

5 Spatial and Temporal Heterogeneity

By explicitly considering the causal factors in disease transmission, such as individuals, individualized interactions and the interaction patterns, the individual-based models have the flexibility to model the heterogeneity in disease transmission. This capability allows the modeling of not only heterogeneous spatial and temporal patterns, but also patterns under different scenarios. Each scenario is represented by a combination of values of the six parameters. The spatial and temporal heterogeneity in disease transmission that results from two such scenarios is illustrated below.

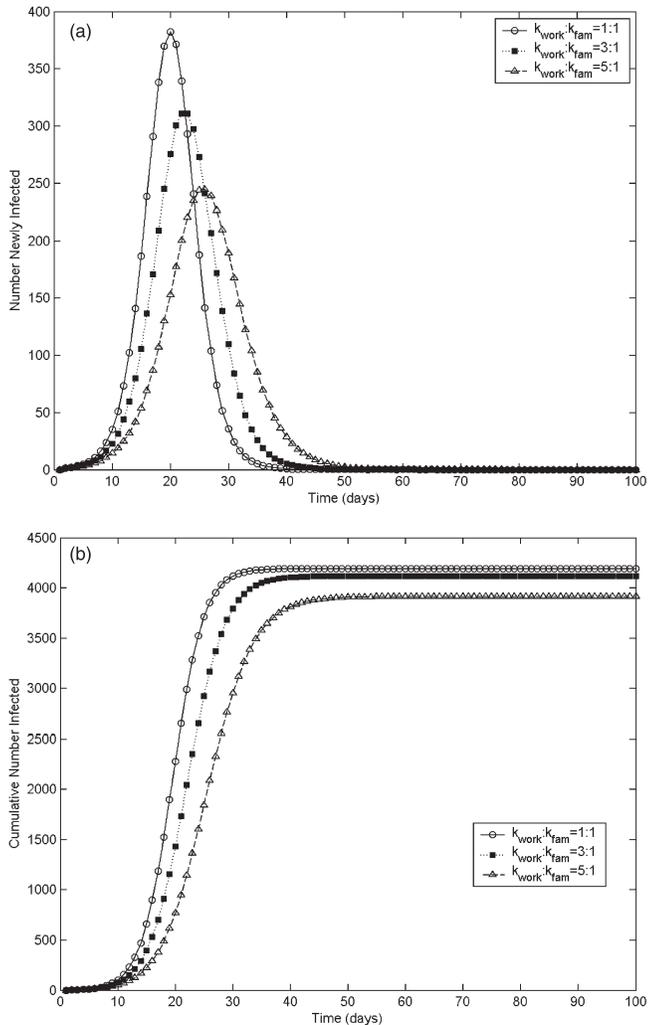


Figure 6 A demonstration of the effect of the ratio between workplace and family connections on an epidemic: (a) the daily new infections throughout an epidemic in response to the ratio ($k_{work}:k_{fam}$) of 1:1, 3:1, and 5:1; and (b) the cumulative number of infections throughout the epidemic in response to the three ratios

One of the six parameters, the degree of inter-connections between family members and co-workers, is used for the illustration. This parameter in particular represents the demographic, employment and social structures of a community, as discussed above. The spatial and temporal patterns resulting from variations in this parameter are displayed in Figure 7. To make the illustration legible, only 1,000 individuals are simulated. Each of the 1,000 individuals has a unique “pseudo” physical location in a home space (Figure 7a) and in a workplace space (Figure 7b). In a true physical space, all family members or co-workers would have occupied an identical location. In addition to its physical meaning, the distance between individuals also represents social distance.

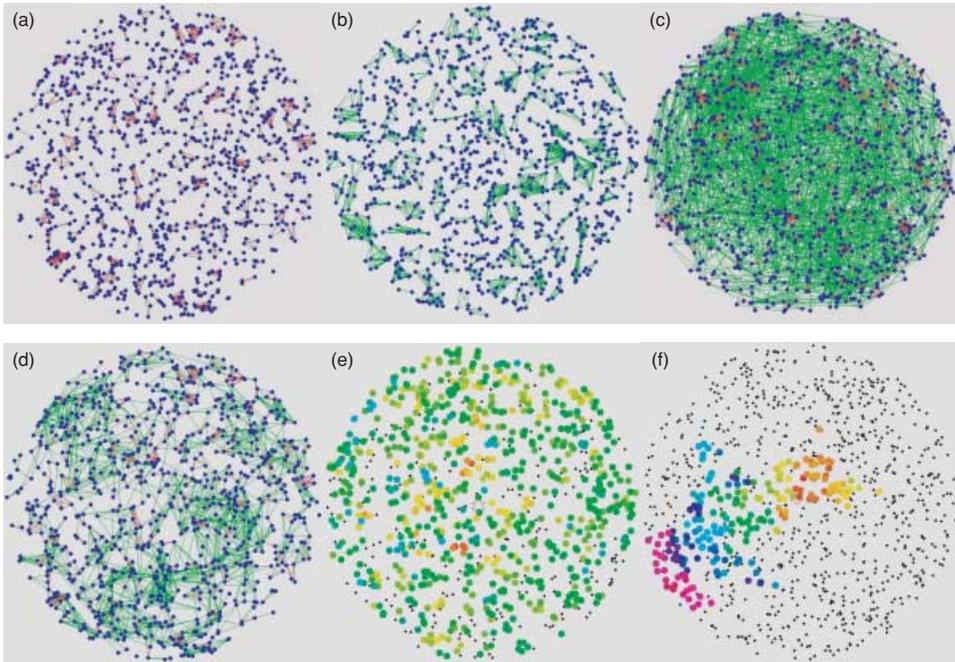


Figure 7 An illustration of the 1,000 simulated individuals in the home space (a) and workplace space (b). The red links indicate family members and green links indicate co-workers. The inter-connection pattern between family members and co-workers is displayed in the home space, represented as covariance = 0.8 in (c) and covariance = 0.99 in (d). The resultant spatial and temporal disease transmission patterns are displayed in (e) and (f), respectively. The “rainbow” color scheme represents the number of days for an individual to be infected, with orange-red indicating 1 day and magenta-red indicating 70 days.

Those individuals who are close to each other are linked into families or workplaces. The connection is occasional and weak between those individuals who are not linked by explicit family or workplace ties.

Figures 7c and 7d show the 1,000 individuals in their home space with both family connections (red links) and workplace connections (green links) between individuals. Two patterns of inter-connections are simulated. These are represented by covariances between individuals’ home locations and workplace locations for the x and y coordinates, respectively. Figure 7c shows a lesser degree of inter-connection (covariance = 0.8) than that in Figure 7d. A lower degree of inter-connection allows both a greater number of and more diverse connections for individuals. In contrast, Figure 7d shows a higher inter-connection between family members and co-workers (covariance = 0.99), resulting in more limited connections outside a home.

In the simulation, the first infectious individual is introduced to the center of the home space. Figures 7e and 7f display two simulation results showing spatial and temporal patterns of disease transmission corresponding to the two inter-connection patterns, respectively. In response to a lower degree of inter-connection (displayed in Figure 7c), the spread of disease appears to be rather random in terms of how many days it takes

for an individual to be infected and where infected individuals are located (Figure 7e). This is because the connections between individuals are both numerous and diverse. In response to a higher degree of inter-connection (displayed in Figure 7d), the disease spreads in a distinct pattern, spatially and temporally, according to the social distance away from the initial infectious individual (Figure 7f). This is because with a high degree of inter-connection, family members are mostly co-workers to each other. This results in a large number of relatively isolated groups with a strong link within a group and a weak link between groups. The disease must spread gradually across space through the distance-dependent between-group links, most likely in certain directions. In a working model, similar spatial and temporal patterns may appear. That is, a low degree of inter-connection between family members and co-workers may create a random infection pattern in both the residential and workplace areas. In contrast, a high degree of inter-connection may yield spatially dependent infection patterns.

Other parameters may yield similarly diverse patterns under different scenarios. These results may differ from those that would have been forecast by the classic or the spatial version of population-based models. Based on homogeneity assumptions, the population-based models would have produced spatially isotropic and temporally dependent patterns. In this sense, individual-based models provide realistic insights into how diseases spread, who might be affected, and what control measures would be effective.

The combined effect of the six parameters can be evaluated in a six-dimension parameter space where each dimension represents one parameter. Different communities can find their locations in this space according to their demographic, employment, and social characteristics. The vulnerability of these communities to certain diseases can then be evaluated. For example, for a given disease, the community with a low number of family and workplace connections, a high degree of inter-connections, and a high ratio of workplace-to-family connections would be in a “low risk” location in the six-dimension space. In contrast, a community with a high number of connections, a low degree of inter-connection, and a low ratio of workplace-to-family connections (close to 1) would be in a “high risk” location.

6 Discussion and Conclusions

The assumptions of discrete individuals, individualized interactions, the explicit spatial distribution of these individuals, and their mobility promise a realistic representation of heterogeneity in the spread of diseases. By directly considering causal factors in disease transmission at the individual level, the network model structure and its associated parameters can effectively portray this heterogeneity. The explicit representation of spatial distribution and mobility of individuals in particular supports the modeling of spatial heterogeneity. The individual-based analytical framework discussed in this paper can incorporate additional or different parameters to represent various conditions and scenarios. This framework can be extended to represent disease transmission outside home and workplaces. An example would be transmission between individuals of the “weekend” population at locations typical of evening and weekend activities.

The modeling of spatial and temporal heterogeneity in the spread of diseases offers a firm foundation to devise strategies for disease eradication and prevention. Vaccination design, for example, may depend on the identification of those individuals who are

vulnerable to infection, those family and workplace compositions that facilitate connections, and those communities whose demographic, employment, and social structures foster rapid spread of diseases. Based on this knowledge, resources and efforts can be directed toward the most critical nodes, links, and parts of a network of human contact.

As an alternative to population-based models, the individual-based modeling concept presents a number of unique challenges. First, the increased realism in the individual-based representation demands realistic estimation of many aspects of disease transmission, perhaps to a greater degree than population-based models. Second, because of the large number of diverse individuals involved in the modeling, the estimation of parameters may have to rely on stochastic approaches rather than the deterministic approach used in the population-based models. Third, the estimation of parameters for individuals ideally requires data at the corresponding level. Presently, due to a lack of such data (McKenzie 2004), the estimation has to draw from data at the aggregated level, such as census data, or from anecdotal reports. Fourth and lastly, individual-based models are most appropriate for forecasts in areas of small spatial extents or small populations, such as cities, communities and neighborhoods, or areas where individual heterogeneity and mobility are assumed. Population-based models are best suited for modeling pandemics over large spatial extents or in areas that are assumed to be homogeneous internally.

It is argued that epidemiological models can only produce crude approximations to reality. The key is not to look for the absolute numeric predictions, but for differences in outcomes between different scenarios or between different models (McKenzie 2004). In this sense, the modeling frameworks, population-based and individual-based alike, offer a means to systematically evaluate disease intervention strategies, determine the most important issues in decision-making, and identify critical gaps in current knowledge.

Acknowledgements

This research was funded in part by the National Institute of Environmental Health Sciences, National Institutes of Health under Award No. R01 ES09816–01. The authors wish to thank the anonymous reviewers for their valuable comments.

References

- Albert R, Jeong H, and Barabási A 2000 Error and attack tolerance of complex networks. *Nature* 406: 378–82
- Anderson R M and May R M 1991 *Infectious Diseases of Humans: Dynamics and Control*. New York, Oxford University Press
- Arita I, Kojima K, and Nakane M 2003 Transmission of Severe Acute Respiratory Syndrome. *Emerging Infectious Diseases* 9: 1183–4
- Bian L 2004 A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B* 31: 381–95
- Cliff A D and Ord J K 1981 *Spatial Process: Models and Applications*. London, Pion
- Cliff A D, Haggett P, and Ord J K 1986 *Spatial Aspects of Influenza Epidemics*. London, Pion
- DeAngelis D L and Gross L J (eds) 1992 *Individual-based Models and Approaches in Ecology: Populations, Communities, and Ecosystems*. New York, Chapman and Hall
- Dye C and Gay N 2003 Modeling the SARS epidemic. *Science* 300: 1884–5

- Ferguson N M, May R M, and Anderson R M 1997 Measles: Persistence and synchronicity in disease dynamics. In Tilman D and Kareiva P (eds) *Spatial Ecology*. Princeton, NJ, Princeton University Press: 137–57
- Francesconi P, Yoti Z, Declich S, Onok P A, Fabiani M, Olango J, Andraghetti R, Rollin P E, Opira C, Greco D, and Salmaso S 2003 Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda. *Emerging Infectious Diseases* 9: 1420–37
- Galvani A 2004 Emerging infections: What have we learned from SARS? *Emerging Infectious Diseases* 10: 1351–2
- Griffith K S, Mead P, Armstrong G L, Painter J, Kelley K A, Hoffmaster A R, Mayo D, Barden D, Ridzon R, Parashar U, Teshale E H, Williams J, Noviello S, Perz J F, Mast E E, Swerdlow D L, and Hadler J L 2003 Bioterrorism-related inhalational anthrax in an elderly woman, Connecticut, 2001. *Emerging Infectious Diseases* 9: 681–8
- Hägerstrand T 1970 What about people in regional science? *Papers of the Regional Science Association* 24: 7–21
- Holmes E E 1997 Basic epidemiological concepts in a spatial context. In Tilman D and Kareiva P (eds) *Spatial Ecology*. Princeton, NJ, Princeton University Press: 111–36
- Hsueh P, Chen P, Hsiao C, Yeh S, Cheng W, Wang J, Chiang B, Chang S, Chang F, Wong W, Kao C, and Yang P 2004 Patient data, early SARS epidemic, Taiwan. *Emerging Infectious Diseases* 10: 489–93
- Jernigan D B, Raghunathan P L, Bell B P, Brechner R, Bresnitz E A, Butler J C, Cetron M, Cohen M, Doyle T, Fischer M, Greene C, Griffith K S, Guarner J, Hadler J L, Hayslett J L, Meyer R, Petersen L R, Phillips M, Pinner R, Popovic T, Quinn C P, Reefhuis J, Reissman D, Rosenstein N, Schuchat A, Shieh W, Siegal L, Swerdlow D L, Tenover F C, Traeger M, Ward J W, Weisfuse I, Wiersma S, Yeskey K, Zaki S, Ashford D A, Perkins B A, Ostroff S, Hughes J, Fleming D, Koplan J P, and Gerberding J L 2002 Investigation of bioterrorism-related anthrax, United States, 2001: Epidemiologic findings. *Emerging Infectious Diseases* 8: 1019–28
- Judson O P 1994 The rise of the individual-based model in ecology. *TREE* 9: 9–14
- Keeling M J 1999 The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London B* 266: 859–67
- Kermack W O and McKendrick A G 1927 A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A* 115: 700–21
- Koopman J and Lynch J 1999 Individual causal models and population system models in epidemiology. *American Journal of Public Health* 89: 1170–4
- Kretzschmar M, van den Hof S, Wallinga J, and van Wijngaarden J 2004 Ring vaccination and smallpox control. *Emerging Infectious Diseases* 10: 832–41
- Kwan M-P 1999 Gender and individual access to urban opportunities: A study using space-time measures. *Professional Geographer* 51: 210–27
- Lau J T F, Tsui H, Lau M, and Yang X 2004 SARS transmission, risk factors, and prevention in Hong Kong. *Emerging Infectious Diseases* 10: 587–92
- Lenntorp B 1978 A time-geographic simulation model of individual activity programmes. In Carlstein T, Parkes D, and Thrift N (eds) *Human Activity and Time Geography*. New York, John Wiley and Sons: 162–80
- Löytönen M 1998 Time geography, health and GIS. In Gatrell A C and Löytönen M (eds) *GIS and Health*. London, Taylor and Francis: 97–110
- MacDonald G 1957 *The Epidemiology and Control of Malaria*. Oxford, Oxford University Press
- Mbogo C M, Mwangangi J M, Nzovu J, Gu W, Yan G, Gunter J T, Swalm C, Keating J, Regens J L, Shililu J I, Githure J I, and Beier J C 2003 Spatial and temporal heterogeneity of *Anopheles* mosquitoes and *Plasmodium falciparum* transmission along the Kenya coast. *American Journal of Tropical Medicine and Hygiene* 68: 734–42
- McKenzie F E 2004 Smallpox models as policy tools. *Emerging Infectious Diseases* 10: 2044–7
- Meyers L A, Newman M E J, Martin M, and Schrag S 2003 Applying network theory to epidemics: Control measures for *Mycoplasma pneumoniae* outbreaks. *Emerging Infectious Diseases* 9: 204–10
- Miller H J 2005 A measurement theory for time geography. *Geographical Analysis* 37: 17–45
- Pred A 1977 The choreography of existence: Comments on Hagerstrand's time geography and its usefulness. *Economic Geography* 53: 207–21

- Rhodes C J and Anderson R M 1997 Epidemic threshold and vaccination in a lattice model of disease spread. *Theoretical Population Biology* 52: 101–18
- Traeger M S, Wiersma S T, Rosenstein N E, Malecki J M, Shepard C W, Raghunathan P L, Pillai S P, Popovic T, Quinn C P, Meyer R F, Zaki S R, Kumar S, Bruce S M, Sejvar J J, Dull P M, Tierney B C, Jones J D, and Perkins B A 2002 First case of bioterrorism-related inhalational anthrax in the United States, Palm Beach County, Florida, 2001. *Emerging Infectious Diseases* 8: 1029–34
- Walden J and Kaplan E H 2004 Estimating time and size of bioterror attack. *Emerging Infectious Diseases* 10: 1202–5
- Watts D J and Strogatz S H 1998 Collective dynamics of ‘small-world’ networks. *Nature* 393: 440–2
- Zhuang S, Ning F, Zhou W, He X, Lin C, and Chin D P 2004 Superspreading SARS events, Beijing, 2003. *Emerging Infectious Diseases* 10: 256–60