Cholera Transmission in Bangladesh: Social Networks and Neighborhoods

Sophia H. Giebultowicz

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Geography.

Chapel Hill
2010

Approved by:

Michael Emch

Melinda Meade

John Florin

Abstract

Sophia Giebultowicz: Cholera Transmission in Bangladesh: Social Networks and Neighborhoods
(Under the direction of Michael Emch)

Transmission of infectious pathogens across networks is well-documented, yet remains primarily focused on diseases spread by sexual contact. Such analytical tools, however, may also facilitate understanding of how other types of health outcomes are related to physical and social contacts. This research examines the relationship between cholera incidence and the social network that links households in rural Bangladesh. Using twenty-one years of longitudinal demographic and health data, clustering of similar disease rates in the network was measured and compared to spatial autocorrelation of cholera at the neighborhood level. Results indicate that rates are significantly concentrated amongst households within the same local environment, and that social clustering is only evident during certain years examined. These outcomes suggest that intervention efforts should place priority on identifying local-level environmental factors, but also consider the potential of networks as they assist transmission, as well as their role in interactions within a defined neighborhood.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Abbreviations

ICDDR,B: International Center for Diarrheal Disease Research, Bangladesh

HDSS: Health and Demographic Surveillance System

Chapter 1

Introduction

Cholera is an acute diarrheal illness that remains endemic to many less-developed countries. Individuals infected with the disease are often asymptomatic, but those that develop the actual illness exhibit either mild or severe diarrhea that may lead to dehydration and possibly death (World Health Organization, 2008). Cholera is primarily spread by the fecal-oral route, from an infected human reservoir. However, evidence also shows that the *V. Cholerae* pathogen inhabits seas, ponds, and other aquatic environments (Colwell et al, 1977; Colwell et al, 1990), from where it is capable of introduction into vulnerable populations. Two routes of transmission based on these reservoirs are identified by Miller et al (1985). *Primary transmission* occurs via the local estuarine environments where *V. cholerae* is able to survive, spreading to the individual through some form of contact with water or, alternatively, consumption of shellfish or aquatic plants contaminated in their local habitat. *Secondary transmission*, in turn, refers to the diffusion of cholera from an infected individual to susceptibles in the population.

The objectives of this research are to measure the extent of secondary transmission in a cholera-endemic region of Bangladesh. Social networks are constructed and used to represent interactions of individuals with those to whom they are related and thus more likely to interact with. The degree of clustering of cholera along these networks may then be interpreted as the "effect" of cholera in a housing unit on the occurrence of cholera in another, socially-affiliated household. While certain environmental factors are controlled for, there are likely unidentified spatial processes that may produce common outcomes for socially-affiliated units located close in space. This "spatial error" component is thus accounted for in the analysis as well.

Improving understanding of cholera transmission, via primary or secondary routes, is critical for prevention purposes. Rather than identifying single factors associated with cholera, this research uses a holistic perspective to assess the overall risk of disease diffusion through social networks and the various interactions they encompass. Research on networks and health has largely omitted diarrheal diseases such as cholera, despite the possibility of new and informative insights on the nature of illness as related to both physical transmission and behavioral risks related to social networks and interactions. The study area of Matlab, located in rural Bangladesh, provides an opportunity to examine cholera transmission not only as related to environment, but also population-level variables. Furthermore, using longitudinal demographic data at the individual level makes it possible to construct a social network and identify any forms of clustering amongst contacts in Matlab. The specific research questions are: 1) To what extent does cholera cluster in social networks in Matlab, as compared to spatial autocorrelation? 2) When controlling for known and unknown environmental variables, is social clustering significant? To address these questions, two separate analyses were conducted. This thesis is structured as follows: an introductory chapter, a second chapter describing the first analysis, a bridge chapter, a fourth chapter that outlines the second chapter, and a final conclusion chapter. The first analysis uses twenty-one years of data (1983-2003) to compare clustering of cholera both in social networks and space across time. The second analysis uses the same data to measure clustering of cholera in the network controlling for both known environmental risk factors as well as spatial error, or unknown variables.

Background

Cholera

Caused by the bacterium *Vibrio cholerae*, cholera remains present in many parts of the world in its current seventh pandemic. As warmer temperatures facilitate vibrio growth, the disease often exhibits annual cycles in endemic areas, with high variability of disease rates across years due to changing climate and environmental factors (Sack et al, 2004). In Bangladesh, where *v. cholerae*

exists naturally as part of the aquatic environment, cholera has two peaks, during the warmer pre-monsoon season in April and later, a larger outbreak from September through December, following the monsoons (Faruque at al, 2005; Jensen et al, 2006; Siddique et al, 1992). The most common symptom of cholera is watery diarrhea, accompanied perhaps by vomiting and eventually severe dehydration. The significant amounts of fluid lost in the process may cause death within the first hours of onset; without treatment the case fatality rate is about 50%. However, simple and effective oral rehydration treatment, when available and administered properly, prevents the majority of deaths.

Forms of Cholera Transmission

Though spread from person to person by the fecal-oral route, *V. cholerae* is able to naturally survive in aquatic environments (Colwell et al, 1977; Colwell et al, 1990). Research shows that aquatic flora such as algae and phytoplankton provide a reservoir for the bacteria (Islam et al, 1990; Islam et al,1993), and that microscopic creatures including copepods and zooplankton are also suitable for the survival of the pathogen (Nalin et al, 1977; Nalin et al, 1996; Huq et al, 1983). Cholera bacteria are thus able to inhabit brackish, coastal, and fresh waters for significant periods of time due to the presence of these reservoirs. Environmental and climate-related factors may then affect pathogen survival and reproduction rates, as well as likelihood of introduction into human populations. In cholera-endemic areas, for example, the disease displays distinct seasonal trends which are correlated with warmer water temperatures and phytoplankton blooms (Lobitz et al, 2000; Epstein et al, 1993; Colwell, 1996).

The relationship between primary and secondary transmission is documented in a model of transmission supported by researchers (Miller at al, 1985; Craig, 1988; Franco et al, 1997). Initially, seasonal and environmental triggers encourage growth of *v. cholerae* at various local sites. The higher volume of the pathogen thus increases the opportunity for contact with the area population. Once cholera is introduced through a number of index cases, the disease continues to spread, primarily facilitated by human contamination of food, water, and the environment. In terms of geographic

distribution, primary transmission cases are expected to scatter around environmental reservoirs, while secondary transmission is more likely to occur in spatial clusters (Miller et al, 1985).

Broadly speaking, secondary transmission occurs most often when an individual infected with cholera contaminates a source in which the pathogen can survive, followed by consumption of the contaminated entity by susceptibles. Traditionally, water is considered the primary "source" for this type of transmission; certain foods, however, are also able to harbor the cholera pathogen, through contact with either contaminated water or an infected person (Sack et al, 2004). Both are thus considered vehicles of transmission. While direct person-to-person spread of cholera is significantly less common, it functions as another form of diffusion.

Various routes of secondary transmission are documented throughout the literature. Individuals may encounter *v. cholerae* by means of a public source, such as through bathing or consuming water from lakes, rivers, or other bodies thought to be previously contaminated by human activity (Birmingham et al., 1997; Acosta et al., 2001; Shaprio et al, 1999; Swerdlow et al., 1997; Spira et al., 1980). Contamination of designated public water supplies may also lead to epidemics. This might affect a single common source, such as a well (Tauxe et al., 1988) or common pump, as in John Snow's famous study of a cholera epidemic in London (Snow, 1855). However, the poor quality of filtration systems in many developing countries often translates to cholera transmission via consumption of water from municipal sources (Ramakrishna et al., 1996). This has especially been the case in many Latin American settings, including Colombia (Cardenas et al., 1992), Ecuador (Weber et al., 1994), and Peru (Swerdlow et al., 1992; Ries et al., 1992).

Drinking water or eating food outside of the home may also indicate contamination of a public source (Quick et al., 1995; Gunn et al., 1981). Elevated risk is also documented in individuals who had consumed beverages and/or food from street vendors, evident in a number of settings (Weber et al., 1994; Hutin et al., 2003; Ries et al., 1992; Koo et al., 1996). Indeed, commercial distribution of food is an additional source of secondary transmission. In the Bangladeshi capital of Dhaka, cases of cholera were linked to eating at restaurants and to food aid distribution centers (Khan et al., 1983),

and in Bangkok a group of tourists developed cholera after consuming contaminated yellow rice at a buffet (Boyce et al., 1995).

Secondary transmission also occurs at the interfamilial and household-level. Living within the same household as an infected individual can present risk, as in cholera-endemic Bangladesh, where, during a 1982 cholera epidemic, the newly emerged classical strain of the disease was more likely to affect members of the same family living together (Shahid et al., 1984). In a Peruvian setting, significantly more cases of diarrhea occurred among relatives of patients with cholera than among those of non-cholera patients, also suggesting interfamilial transmission (Fuduka et al., 1995). Concentration of cholera within families and/or households has also been identified in Gaza, the Philippines, and in Calcutta (Lasch et al., 1984; Tamayo et al., 1965; Sengupta et al., 1994). Certain studies identify the specific household-level risk factor, such as consuming water from a storage container in which hands had been placed (Swerdlow et al., 1992), eating from a common cooking pot (Shahid et al., 1984), and eating leftovers that were contaminated within the household setting (St. Louis et al., 1990). Eating food prepared by a member of the household who recently was ill or in contact with another sick individual is also potentially related to additional infections within the household (Estrada-Garcia et al., 1996; Holmberg et al., 1984).

Outside of the immediate household, food as a vehicle for the spread of cholera has also been implicated. Sharing food with individuals exhibiting diarrheal symptoms was considered a risk factor in a Kenyan epidemic (Shaprio et al., 1999), and in rare occurrence within the United States, two cholera cases were diagnosed as a result of consuming fruit prepared by an asymptomatic companion (Ackers et al., 1997). Contaminated food or beverages prepared and served at social events are equally likely to cause disease outbreaks. Attendance at funerals (Shapiro et al., 1999; Gunnlaugsson et al., 1998; St. Louis et al., 1990), weddings (Lasch et al., 1994), and parties (Swerdlow et al., 1992) are all linked to numerous cholera cases in Africa, Gaza, and Peru.

Though neither within households nor at social events, direct person-to-person transmission of cholera is specifically referred to in studies by both Goh et al. (1999) and Mhalu et al. (1984). Both

occurred at hospitals, emphasizing the potential of sick individuals to spread the disease to those within close proximity. This is an important consideration for caretakers, relatives and other contacts in household settings. Furthermore, Emch (1999) identified many individuals living within a small household as associated with cholera incidence, indicating the risk of crowded conditions.

It is also important to note that various behaviors may decrease the risk of cholera in individuals and their households. Boiling drinking water or disinfecting it by another means are examples of such protective measure (Mintz et al., 1994; Weber et al., 1994; Deb et al., 1986), as is the presence of soap and the action of regular hand-washing (St. Louis et al., 1990; Weber et al., 1994; Hutin et al., 2003; Ries et al., 1995). Storing household water in narrow-mouthed containers has also been proved effective in reducing contamination and decreasing cholera (Deb et al., 1986). Avoiding the use of nearby and potentially contaminated surface water for household purposes may also be protective, especially if neighbors are known to be sick (Spira et al., 1980; Hughes et al., 1981).

Other Risk Factors for Cholera

Various studies conducted in Matlab, the study area, have investigated risk factors associated with cholera. Emch (1999) identified multiple households using latrines, living within a flood-controlled area, small household area, large household population, and high population density as variables positively related to cholera incidence. Ali et al (2002a) also found high population density to be a risk factor, as well as proximity to surface water bodies and poor educational level. Additionally, living in a flood-controlled area has shown a relationship to cholera incidence (Ali et al, 2002b). Breast-feeding has been shown to have a protective effect in children (Clemens et al, 1990; Glass and Black, 1992), while Glass and Black (1992) found children between the ages of 2 and 15 and women of child-bearing age to be at highest risk.

Social Networks and Health

Social network analysis is commonly used to map and measure relationships between social entities, or nodes in a network (Wasserman & Faust, 1994; Hanneman, 2001). A social network is a set of actors connected by their relationships to one another; these are usually based on kinship, friendship, or other forms of interaction such as sexual contact or information-sharing (Klovdahl et al, 2001). One method of investigating the process of disease diffusion is to identify the social networks that exist in a given population and analyze the relationships between actors in order to explain how a disease may be transferred across these linkages. Research has shown that social networks can affect health outcomes and that they are a valuable epidemiological tool for analyzing the diffusion of non-randomly distributed diseases (e.g., not transmitted by a vector) in a population. Network methods allow identification of the major sources of infection in a connected group, as well as the most susceptible actors.

There are two basic types of networks that exist when discussing health outcomes, known as non-transmission and transmission networks. The former consist of emotional relationships and interactions associated with various psycho-social effects on health. These linkages may provide social support associated with positive health or health outcomes; an example of this is the number of studies showing lower mortality in individuals with more social ties (Bell et al, 1999, Berkman et al, 1979; House et al, 1982; Blazer, 1982), as well as additional research showing increased risk of certain chronic illnesses for individuals experiencing a greater degree of social isolation (Kinney et al, 2003; Eng et al, 2002). Networks may also influence behaviors related to health, such as contraceptive use and prenatal behavior in women (Valente et al, 1997; Kincaid et al, 2000; Harley et al, 2006).

Transmission networks, in contrast, are those along which a parasitic agent, infection, or other physical component is transferred. These rely far more on physical interaction. An excellent example of a transmission network is one based on sexual contact, along which sexually transmitted infections diffuse from person to person. This is well-documented in the spread of HIV/AIDS (Jaffe et al, 1983; Auerbach et al, 1984; Bell et al, 1999) as well as Chlamydia and gonorrhea (Wylie et al, 2005; Ghani

et al, 1997; Aral et al, 1999). Infectious diseases not transmitted via sexual contact have also been examined using network methods, but this is a more complex matter as modes of transmission can vary widely. The majority of research on non-sexually transmitted infectious diseases and networks has looked at tuberculosis incidence as a health outcome (Fitzpatrick et al, 2001; Klovdahl et al, 2001; Cook et al, 2007); other studies have modeled smallpox and flu outbreaks based on networks and contact structure for the purposes of developing effective intervention strategies, such as vaccines (Wallinga et al, 1999).

Social networks do not exist in a vacuum; rather, they function within a geographic environment. Certain research has therefore incorporated spatial distance and other geographical variables into the study of networks and health. Much of this has been done using sexual networks. Wylie et al (2005) identified distinct geographical clusters of different Chlamydia strains corresponding with smaller sexual networks in Manitoba, with two smaller networks connected by mobile individuals, and Zenilman et al (1999) identified distinct sexual networks based on geographic proximity in Baltimore. Rothenberg et al (2005) showed geographic and social clustering in a group of urban individuals at risk of HIV/AIDS, and Wylie et al (2007) found a dense network of hotels in Winnipeg where Intravenous Drug Use (IDU) occurred, with individuals acting as spatial connections between hotels and agents of disease transmission. In the case of non-STIs, Klovdahl et al (2001) found social settings such as bars and clubs to be points of tuberculosis transmission in a network where actors consisted of both patients and places they frequented in the Houston metropolitan area.

Theoretical Framework

This research is guided by the theoretical framework of disease ecology, within the broader subdiscipline of medical geography. This particular perspective is concerned with "the ways human behavior, in its cultural and socioeconomic context, interacts with environmental conditions to produce or prevent disease (Meade and Earickson, 2005)." Unlike the concepts of specific etiology and germ theory, the disease ecology perspective sees health outcomes as far more than direct

pathogen-host interactions. Rather, it encourages careful examination of the social, economic, behavioral, cultural, environmental and biological context in which disease occurs. This must be considered, in addition to the characteristics of the illness itself, in order to fully understand health events and outcomes in a population.

The work of individuals such as Jacques May (1958) and Rene Dubos (1965) was critical in introducing the concepts of modern disease ecology, especially in relation to medical geography. May (1954) argued that human activity, combined with the characteristics of the surrounding environment, could be used to explain the distribution of disease and disease foci. Furthermore, he stressed that the occurrence of disease required the interaction of agent and host in time and space. A key concept of his work was the identification of the various cultural buffers playing a part in the human-disease relationship, such as the link between house type and clothing to the occurrence of malaria, or the higher risk of malnutrition in certain populations based on primary crops and cooking preferences (Meade, 1977; Mayer, 1996). However, as Meade (1977) argues, May's descriptions were essentially static and failed to identify the underlying theoretical processes of disease ecology. Meade thus developed the perspective further, re-defining health as adaptability rather than a biological entity and using it as a measure of the interactions among population (e.g. genetics, demographic composition), environment (including all forms of "insults" to the human form produced by physical, chemical, and psychosocial conditions, as well as institutional environment), and culture (norms and practices, behavior, and perceptions/understanding of the surrounding world). The three vertices of this "triangle of human ecology" are inherently related to one another, as well as to health and disease outcomes; the interactions between them provide a foundation for integrating and analyzing the factors that contribute to ecologies of disease.

An additional theoretical component of this research is the concept of a "cultural vector," as introduced originally by Ewald (1988) for the purpose of describing pathogen virulence and evolution. In contrast to arthropod vectors, such as mosquitoes, a cultural vector is defined as "a set of characteristics that allow transmission from immobilized hosts to susceptibles when at least one of the

characteristics is some aspect of human culture (Ewald, 1988).” When applied to waterborne diseases, examples of cultural vectors include materials contaminated by the infected host, the individual removing the contaminated materials, the sewage systems or other water bodies that move the contaminated water to a public source, and any persons or equipment that might then deliver the contaminated water to susceptibles (Ewald, 1994). The definition of a cultural vector is used here to include non-immobilized hosts, due to their arguable potential for contaminating materials as well. Additionally, food is considered a vehicle of transmission, whether contaminated originally by the host, or acting as the “equipment” that delivers the pathogen to susceptibles through some form of contact with contaminated water.

The framework and concepts described above are used to structure and answer the research questions here. Specifically, rather than simply assuming infection occurs as a result of direct contact with the cholera pathogen, this work firstly questions the behavioral factors that are also involved in the processes of diffusion. Various components of domestic and social life, which fall under the general classification of cultural norms and traditions in the lives of the population, are potentially related to cholera risk. This includes elements such as household sanitation, participation in social events, and caring for the sick. Secondary transmission is aided primarily by cultural vectors. For example, an infected individual may contaminate some type of “material,” i.e. food or water, facilitating its spread to other members of the household. Or, a caretaker may also transfer the pathogen to food or water served to friends and family after interaction with the sick individual. Equipment such as wide-mouthed household storage containers for water may effectively deliver cholera from host to susceptible. Protective behaviors among family members and social contacts, however, may also affect risk in the opposite way; that is, disinfecting water and hand-washing will function as buffers to transmission through cultural vectors. The use of social networks, while not specifically pinpointing the behaviors that occur within and beyond the household

In addition, this research considers the critical component of environment when evaluating cholera risk and transmission. Environmental factors that are potentially associated with cholera risk

are considered, as well as the effect of those variables correlated in space that remain unaccounted for in the models used. Furthermore, the kinship structure and relations of the study population are used to develop a social network that represents the daily interactions that may encompass risk. The scope of this research thus goes beyond simply pathogen-host relationships, integrating vital components of human and disease ecology to further understand cholera transmission.

References

Ackers, M., Pagaduan, R., Hart, G., Greene, K.D., Abbott, S., Mintz, E., and Tauxe, R.V. 1997. Cholera and sliced fruit: Probable secondary transmission from an asymptomatic carrier in the United States. *International Journal of Infectious Diseases*, 1(4): 212-214.

Acosta, J.C., Galindo, C.M., Kimario, J., Senkoro, K., Urassa, H., Casals, C., Corachán, M., Eseko, N., Tanner, M., Mshinda, H., Lwilla, F., Vila, J., and Alonso, P.L. 2001. Cholera outbreak in southern Tanzania: risk factors and patterns of transmission. *Emerging Infecrious Diseases (Supplement 3):* 583–587.

Ali, M., Emch, M., Donnay, J. P., Yunus, M., & Sack, R. B. 2002a. Identifying environmental risk factors for endemic cholera: A raster GIS approach. *Health & Place,* 8(3), 201-210.

Ali, M., Emch, M., Donnay, J. P., Yunus, M., & Sack, R. B. 2002b. The spatial epidemiology of cholera in an endemic area of Bangladesh. *Social Science & Medicine,* 55(6), 1015-1024.

Auerbach, D.M., W.W. Darrow, H.W. Jaffe, and J.W. Curran. 1984. Cluster of Cases of the Acquired Immune Deficiency Syndrome: Patients Linked by Sexual Contact. *The American Journal of Medicine,* 76(3):487-92.

Bell, D.C., Atkinson, J.S., and Carlson, J.W.. 1999. Centrality measures for disease transmission networks. *Social Networks*, 21(1): 1-21.

Berkman, L.F., and Syme, S.L. 1979. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *American Journal of Epidemiology*, 109(2):186-204.

Birmingham, M.E., Lee, L.A., Ndayimirije, N., Nkurikiye, S., Hersh, B.S., and Wells, J.G. 1997. Epidemic cholera in Burundi: patterns of transmission in the Great Rift Valley Lake region. *The Lancet*, 349: 981–85.

Blazer, D.G. 1982. Social support and mortality in an elderly community population. *American Journal of Epidemiology*. 115(5): 684-694.

Boyce, T.G., Mintz, E.D., Greene, K.D., Wells, J.G., Hockin, J.C., Morgan, D., and Tauxe, R.V. 1995. *Vibrio cholerae* O139 Bengal infections among tourists to southeast Asia: An intercontinental foodborne outbreak. *Journal of Infectious Diseases*, 172(5): 1401–1404.

Cardenas, V., Saad, C., Varona, M., and Linero, M. 1993. Waterborne cholera in Riohacha, Colombia, 1992. *Bulletin of the Pan-American Health Organization*, 27(4): 313-30.

Clemens, J.D., Sack, D.A., Harris, J.R., Khan, M.R., Chakraborty, J., Chowdhury, S., Rao, M.R., Van Loon, F.P., Stanton, B.F., and Yunus, M. 1990. Breast feeding and the risk of severe cholera in rural Bangladeshi children. *American Journal of Epidemiology*. 131(3):400-11.

Colwell, R.R. 1996. Global Climate and Infectious Disease: The Cholera Paradigm. *Science*, 274(5295): 2025-2031.

Colwell, RR, Kaper, J, and Joseph, SW. 1977. Vibrio cholerae, Vibrio parahaemolyticus, and other vibrios: occurrence and distribution in Chesapeake Bay. *Science*. 198:394.

Colwell, RR, Tamplin, ML, Brayton, PR, Gauzens, AL, Tall, BD, Herrington, D, Levine, MM, Hall, S, Huq, A, and Sack, DA. 1990. Environmental aspects of Vibrio cholerae in transmission of cholera. In: *Advances in research on cholera and related diarrhoea*, 7th ed Tokyo, KTK Scientific Publishers. 327–343.

Cook, V.J., Sun, S.J., Tapia, J., Muth, S.Q., Arguello, D.F., Lewis, B.L., Rothenberg, R.B., and McElroy, P.D. 2007. Transmission network analysis in Tuberculosis contact investigations. *Journal of Infectious Diseases* 196: 1517-27.

Craig, M. 1988. Time-space clustering of vibrio cholerae 01 in Matlab, Bangladesh, 1970-1982. *Social Science and Medicine*, 26 (1): 5-13.

Deb, B.C., Sircar, B.K., Sengupta, P.G., De, S.P., Mondal, S.K., Gupta, D.N., Saha, N.C., Ghosh, S., Mitra, U., and Pal, S.C. 1986. Studies on interventions to prevent eltor cholera transmission in urban slums. *Bulletin of the World Health Organization*, 64(1): 127-131.

Dubos, R. J. 1965. *Man adapting*. New Haven: Yale University Press.

Dutt AK, Alwi S, Velauthan T. 1971. A shellfish-borne cholera outbreak in Malaysia. *Trans R Soc Trop Med Hyg*; 65:815-8.

Emch, M. 1999. Diarrheal disease risk in Matlab, Bangladesh. *Social Science & Medicine,* 49(4): 519-530.

Eng, P.M., Rimm, E.B., Fitzmaurice, G., and Kawachi, I. 2002. Social Ties and Change in Social Ties in Relation to Subsequent Total and Cause-specific Mortality and Coronary Heart Disease Incidence in Men. *American Journal of Epidemiology*. 155(8): 700-709.

Epstein, P R. 1993. Algal blooms in the spread and persistence of cholera. *Biological Systems* 31(2-3): 209-21.

Estrada-Garcia, T, and Mintz, ED. 1996. Cholera: Foodborne transmission and its prevention. *European Journal of Epidemiology* 12(5): 461-469.

Ewald, P.W. 1988. Cultural vectors, virulence, and the emergence of evolutionary epidemiology. *Oxford Survey of Evolutionary Biology*, 5: 215-245.

Ewald, P.W. 1994. Evolution *of Infectious Disease*. Oxford: Oxford University Press.

Faruque, S. M., Roy, S.K., Alim, A.R., Siddique, A.K., and Albert, M.J. 1995. Molecular epidemiology of toxigenic Vibrio cholerae in Bangladesh studied by numerical analysis of rRNA gene restriction patterns. *Journal of Clinical Microbiology*, 33(11): 2833–2838.

Fitzpatrick, L.K., Hardacker, J.A., Heirendt, W., Agerton, T., Streicher, A., Melnyk, H., Ridzon, R., Valway, S., and Oronato, I. 2001. A preventable outbreak of tuberculosis investigated through an intricate social network. *Clinical Infectious Diseases,* 33(11):1801-1806.

Franco, A.A., Fix, A.D., Prada, A., Paredes, E., Palomino, J.C., Wright, A.C., Johnson, J.A., McCarter, R., Guerra, H., Morris, J.G. Jr. 1997. Cholera in Lima, Peru, correlates with prior isolation of Vibrio cholerae from the environment. *American Journal of Epidemiology*, 146(12): 1067-1075.

Fuduka, J.M, Augusta, Y., Chaparro, L., Campos, M., and Chea, E. 1995. Clinical characteristics and risk factors for *Vibfio cholerae* infection in children. *The Journal of Pediatrics*, 126(6): 882-886.

Franco, A.A., Fix, A.D., Prada, A., Paredes, E., Palomino, J.C., Wright, A.C., Johnson, J.A., McCarter, R., Guerra, H., Morris, J.G. Jr. 1997. Cholera in Lima, Peru, correlates with prior isolation of Vibrio cholerae from the environment. *American Journal of Epidemiology*, 146(12): 1067-1075.

Hutin, Y., Luby, S., and Paquet, C. 2003. A large cholera outbreak in Kano City, Nigeria: the importance of hand-washing with soap and the danger of street-vended water. *Journal of Water and Health*, 1(1) : 45-52.

Glass, R.I., Black, R.E., 1992. The epidemiology of cholera. In: Barua, D., Greenough, W.B. (Eds.), *Cholera*. Plenum Medical Book Company, New York, p. 129.

Goh, K.T., Teo, S.H., Lam, T.S., and Ling, M.K.. 1999. Person-to-person transmission of cholera in a psychiatric hospital. *Journal of Infection,* 20:193-200.

Gunn, R.A., Kimball, A.M., Mathew, P.P., Dutta, S.R., and Rifaat, A.H.M. 1981. Cholera in Bahrain: epidemiological characteristics of an outbreak. *Bulletin of the World Health Organization* (WHO), 59(1): 61-66.

Gunnlaugsson, G., Einarsdottir, J., Angulo, F.J., Mentambanar, S.A., Passa, A., and Tauxe, R.V. 1998. Funerals during the 1994 cholera epidemic in Guinea-Bissau, West Africa: The need for disinfection of bodies of persons dying of cholera. *Epidemiology and Infection*, 120(1): 7-15.

Hanneman, R.A. 2001. *Introduction to social network methods*. Riverside, California: Department of Sociology, University of California at Riverside.

Harley, K., and Eskenazi, B. 2006. Time in the United States, Social Support, and Health Behaviors During Pregnancy Among Women of Mexican Descent. *Social Science and Medicine* 62(12): 3048-3061.

Holmberg SD, Harris JR, Kay DE. 1984. Foodborne transmission of cholera in Micronesian households. *The Lancet*, (i): 325-328.

House, J.S., Robbins, C., and Metzner, H.L. 1982. Association of social relationships and activities with mortality: Prospective evidence from the Tecumseh community health study. *American Journal of Epidemiology*. 116(1):123-140.

Hughes, J.M., Boyce, J.M., Levine, R.J., Khan, M., Aziz, K.M., Huq, M.I., Curlin, G.T. 1981. Epidemiology of El Tor cholera in rural Bangladesh: importance of surface water in transmission. Scientific Report no. 52, International Centre for Diarrhoeal Disease Research, Bangladesh. 21 p.

Hunter, J.M. 1966. River blindness in Nangodi, Northern Nigeria: a hypothesis of cyclical advance and retreat. *Geographical Review* 56, 398-416.

Huq, A., Small, E.B., West, P.A., Huq, M.I., Rahman, R., and Colwell, R.R. 1983. Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Applied Environmental Microbiology*, 45 (1): 275–283.

Islam, M.S., Drasar, B.S., and Bradley, D.J. 1990. Long-term persistence of toxigenic *Vibrio cholerae* 01 in the mucilaginous sheath of a blue-green alga,*Anabaena variabilis*. *Journal of Tropical Medicine and Hygiene*, 93 (2): 133–139.

Islam, M.S., Drasar, B.S., and Sack, R.B. 1993. The aquatic environment as a reservoir of *Vibrio cholerae*: a review. *Journal of Diarrhoeal Disease Research*, 11 (4): 197–206.

Jaffe, H.W., Choi, K., Thomas, P., Haverkos, H.W., Auerbach, D.M., Guinan, M.E., Rogers, M.F., Spira, T.J., Darrow, W.W., Kramer, M.A., Friedman, S.M., Monroe, J.M., Friedman, A.E., Laubenstein, L.J., Marmor, M., Safai, B., Dritz, S.K., Fannin, S.L., Orkwis, J.P., Kelter, A., Rushing, W.R., Thacker, S.B., and Curran, J.W. 1983. National case-control study of Kaposi's sarcoma and Pneumocystis carinii pneumonia in homosexual men: part 1, epidemiologic results. *Annals of Internal Medicine*, 99(2):145-51.

Khan, M.U., Shahidullan. M., Ahmed, W.U., Purification, D., and Khan, M.A. 1983. The eltor cholera epidemic in Dhaka in 1974 and 1975. *Bulletin of the WHO*, 61(4): 653-9.

Koo, D., Aragon, A., Moscoso, V., Gudiel, M., Bietti, L., Carrillo, N., Chojoj, J., Gordillo, B., Cano, F., Cameron, D.N., Wells, J.G., Bean, N.H., and Tauxe. R.V. 1996. Epidemic Cholera in Guatemala, 1993: Transmission of a Newly Introduced Epidemic Strain by Street Vendors. *Epidemiology and Infection*, 116(2): 121-126.

Lasch, E.E., Abed, Y., Marcus, O., Shbeir, M., El Alem, A., Ali Hassan, N. 1984. Cholera in Gaza in 1981: epidemiological characteristics of an outbreak. *Tranactions of the Royal Society of Tropical Medicine and Hygiene*, 78(4):554-7.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A.S.G., and Colwell, R.R. 2000. Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proceedings of the National Academy of Sciences*, 97(4): 1438–1443.

Kincaid, D.L.. 2000. Social networks, ideation, and contraceptive behavior in Bangladesh: a longitudinal analysis. *Social Science and Medicine*, 50: 215-231.

Kinney, A.Y., Bloor, L.E., Dudley, W.N., Millikan, R.C., Marshall, E., Martin, C., and Sandler, R. 2003. Roles of Religious Involvement and Social Support in the Risk of Colon Cancer among Blacks and Whites. *American Journal of Epidemiology*, 158(11): 1097-1107.

Klovdahl, A.S. 2001. Networks and tuberculosis: an undetected community outbreak involving public places. *Social Science & Medicine*, 52(5): 681-694.

May, J.M. 1958. *The Ecology of Human Disease*. MD Publications, Inc., New York.

May, J.M. 1954. Cultural aspects of tropical medicine. *American Journal of Tropical Medicine and Hygiene*, 3: 422-30.

Mayer, J.D. 1996. The political ecology of disease as one new focus for medical geography. *Progress in Human Geography*, 20(4): 441-456.

Meade, M. S. 1977. Medical geography as human ecology: The dimension of population movement. *The Geographical Review*, 67(4), 379-393.

Meade, Melinda S. and Earickson, Robert J. 2005. *Medical Geography*. Guilford Press.
Pyle, Gerald S. 1979. *Applied Medical Geography*. V.H. Winston and Sons, Washington, D.C.

Mhalu et al. 1984. Hospital Outbreaks of Cholera Transmitted Through Close Person-to-Person Contact. *The Lancet*, 14;2(8394):82-4.

Mintz, E.D., T. Popovic, and P.A. Blake. 1994. Transmission of Vibrio Cholerae 01. In *Vibrio Cholerae and Cholera: Molecular to Global Perspectives*. (Wachsmuth, K.I., Blake, P.A. and Olsik, O., Eds.), pp. 321-344. American Society for Microbiology Press, Washington, DC.

Miller, C.J., Feachem, R.G., and Drasar, B.S. 1985. Cholera epidemiology in developed and developing countries: new thoughts on transmission, seasonality, and control. *Lancet*, 8423: 261-3.

Nalin, DR. 1976. Cholera, copepods, and chitinase. *Lancet*, 2 (7992): 958.

Nalin, D.R., Daya, V., Reid, A., Levine, M.M., and Cisneros, L. 1979. Absorption and growth of *Vibrio cholerae* on chitin. *Infection and Immunology*, 25 (2): 76-770.

Pyle, Gerald S. 1983. "Three Decades of Medical Geography in the United States." In *Geographical Aspects of Health*, ed. N.D. McGlashan and J.R. Blunden. Academic Press.

Quick et al, 1995. Epidemic Cholera in Rural El Salvador: Risk Factors in a Region Covered by a Cholera Prevention Campaign. *Epidemiology and Infection*, 114(2): 249-255.

Rabbani, GH, and Greenough WB 3d. 1999. Food as a Vehicle of Transmission of Cholera. *Journal of Diarrhoeal Diseases Research*, 17(1):1-9.

Ramakrishna, B. S., Kang, G., Rajan, D. P., Mathan, M., Mathan, V. I. 1996. Isolation of Vibrio cholerae O139 from the drinking water supply during an epidemic of cholera. *Tropical Medicine & International Health*, 1(6):854-858.

Ries, A.A., Vugia, D.J., Beingolea, L., Palacios, A., Vasquez, E., Wells, J.G., Baca, N.G., Swerdlow, D.L., and Pollack, M. 1992. Cholera in Piura, Peru: A modern urban epidemic. *Journal of Infectious Diseases,* 166(6): 1429-433.

Sack, D.A., Sack, R.B., Nair, G.B., Siddique, A.K. 2004. Cholera. *The Lancet*, 363: 223-33.

Shapiro, R.L., Otieno, M.R., Adcock, P.M., Phillips-Howard, P.A., Hawley, W.A., Kumar, L., Waiyaki, P., Nahlen, B.L., and Slutsker, L. 1999. Transmission of epidemic Vibrio cholerae O1 in rural western Kenya associated with drinking water from Lake Victoria: an environmental reservoir for cholera? *American Journal of Tropical Medicine and Hygiene,* 60(2): 271-276.

St. Louis, M., J.D. Porter, A. Helal, K. Drame, N. Hargrett-Bean, J.G. Wells, and R.V. Tauxe. 1990. Epidemic Cholera in West Africa: The Role of Food-Handling and High-Risk Foods. *American Journal of Epidemiology*, 131(4): 719-728

Sengupta et al. 1994. Epidemiology of *Vibrio cholerae O139 with Special Reference* to Intrafamilial Transmission in Calcutta, *Journal of lnfection,* 31: 45-47.

Shahid, N.S., Samadi, A.R., Khan, M.U., and Huq, Mi.I. 1984. Classical vs El Tor cholera: a prospective family study of a concurrent outbreak. *Journal of Diarrheal Disease Research*, 2(2): 73-78.

Spira et al. 1980. Microbiological surveillance of intra-neighborhood El Tor cholera transmission in rural Bangladesh. *Bulletin of the World Health Organization* (WHO), 33(5):627-36.

Snow, J. 1855. *On the mode of communication of cholera*. 2nd ed. London, England: J. Churchill.

Swerdlow et al. 1992. Waterborne transmission of epidemic cholera in Trujillo, Peru: lessons for a continent at risk. *The Lancet*, 340(8810):28-33.

Swerdlow, D.L. and M. Isaacson. 1994. The epidemiology of cholera in Africa. In: *Vibrio cholerae and Cholera: Molecular to Global Perspectives* (Wachsmuth, K.I., Blake, P.A. and Olsik, O., Eds.), pp. 297-307. American Society for Microbiology Press, Washington, DC.

Swerdlow, D.L., Malenga, G., Begkoyian, G., Nyangulu, D., Toole, M., Waldman, R.J., Puhr, D.N.D., and Tauxe, R.V. 1997. Epidemic cholera among refugees in Malawi, Africa: treatment and transmission. *Epidemiology and Infection*, 118: 207-214.

Tamayo et al. 1965. Studies of Cholera El Tor in the Philippines. *Bulletin of the World Health Organization (WHO)*;33(5):645-649.

Tauxe et al. 1988. Epidemic Cholera in Mali: High Mortality and Multiple Routes of Transmission in a Famine Area. *Epidemiology and Infection*, 100(2): 279-289.

Tauxe, R., Seminario, L., Tapita, R. and Libel, M. 1994. The Latin American epidemic. In: *Vibrio cholerae and Cholera: Molecular to Global Perspectives* (Wachsmuth, K.I., Blake, P.A. and Olsik, O., Eds.), pp. 321-344. American Society for Microbiology Press, Washington, DC.

Valente, T.W., Watkins, S.C., Jato, M.N., A. Van Der Straten, and L.M. Tsitsol. 1997. Social Network Associations with Contraceptive Use Among Cameroonian Women in Voluntary Associations. *Social Science and Medicine*, 45(5): 677-687.

Wallinga, J., Edmunds, W.J., and Kretzschmar, M. 1999. Perspective: Human contact patterns and the spread of airborne infectious diseases. *Trends in Microbiology*, 7 (9): 372-377.

Wasserman, S. and Faust, K. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Weber et al. 1994. Epidemic Cholera in Ecuador: Multidrug-Resistance and Transmission by Water and Seafood. *Epidemiology and Infection*, 112(1): 1-11.

World Health Organization (WHO). 2008. Cholera Fact Sheet. http://www.who.int/.

Wylie, J.L., Cabral, T., and Jolly, A.M. 2005. Identification of Networks of Sexually Transmitted Infection: A Molecular, Geographic, and Social Network Analysis. *Journal of Infectious Diseases*, 191(6): 899-906.

Wylie, J.L., Shah, L., and Jolly, A. 2007. Incorporating geographic settings into a social network analysis of injection drug use and bloodborne pathogen prevalence. *Health and Place*, 13(3):617-28.

Zenilman, JM, Ellish, N, Fresia, A, Glass, G. 1999. The geography of sexual partnerships in Baltimore: applications of core theory dynamics using a geographic information system. *Sexually Transmitted Diseases*, 26(2):75-81.

Chapter 2

Spatial and Social Clustering of Cholera in Matlab, Bangladesh

Abstract: Cholera is a diarrheal disease that remains endemic in Bangladesh, where poor sanitation and environmental factors are known to be risk factors for the disease. Using twenty years of data from the Matlab Demographic and Health Surveillance System, this study examines the clustering of cholera in a kinship-based social network. Social clustering is compared to spatial clustering of the disease at multiple geographic scales using the Moran's *I* spatial autocorrelation statistics. The results show that while spatial clustering of cholera is consistent over time at multiple neighborhood scales, there is seldom social clustering of cholera. A combination of spatial and social network analysis tools can be useful for understanding disease transmission.

Introduction

Infectious disease transmission can be affected by one's interactions with others and therefore their social networks. Social network analysis is a tool commonly used to measure relationships between people within a network and to understand connections such as kinship, information-sharing, or sexual contact (Wasserman & Faust, 1994; Hanneman, 2001; Klovdahl et al, 2001). These methods have proven to be a valuable epidemiological tool for examining the diffusion of certain diseases, particularly those spread by non-random means through a population. Much of the previous research on networks and health has focused on HIV/AIDS and other sexually transmitted infections (Jaffe et al, 1983; Auerbach et al, 1984; Bell et al, 1999; Wylie et al, 2005; Ghani et al, 1997; Aral et al, 1999) and, less so, on contagious illnesses such as tuberculosis (Fitzpatrick et al, 2001; Klovdahl et al, 2001; Cook et al, 2007) or health outcomes associated with social processes such as obesity (Christakis et al, 2007). This study uses a similar approach with a disease not often examined in the social networks literature; namely, it seeks to measure the effect of networks on the transmission of cholera, a diarrheal illness commonly spread through contaminated food or water. In rural Bangladesh, analysis of potential risk is based on social connections to other individuals with the illness. Furthermore, it is compared to risk attributed to local neighborhood-level risk factors known to increase the likelihood of infection (Emch, 1999; Ali et al, 2002a; Ali et al, 2002b).

To compare the effects of individual cholera risk from a local environmental- and neighborhood-level as opposed to transmission from family and social contacts, both spatial and social clustering of cholera in rural Bangladesh are examined over a twenty-one year period. The specific research questions are as follows: 1) To what extent does cholera cluster at the neighborhood level? 2) Is there significant social clustering of cholera, or a higher likelihood of illness in individuals related to others with cholera? Comparing social and spatial clustering can further develop understanding of the various pathways of transmission risk in the study area. Environmental factors are known to play a significant role in cholera occurrence, yet the nature of person-to-person

transmission is not well understood, especially through a social networks approach. This paper presents the following hypotheses: 1) There are spatial clusters of cholera incidence in local neighborhood environments, but also that risk is higher for those who are socially connected to others who are diagnosed with the disease; 2) The nature of both the spatial and social relationships may not be consistent over time, and vary based on external factors such as flooding in a particular year, or local efforts to reduce transmission through education and improved sanitation.

Background

The study area is Matlab, Bangladesh which is approximately 50 km southeast of Dhaka, at the confluence of the Meghna and Ganges rivers (Figure 1). The population is about 200,000 and people reside in clusters of patrilineally-related households called *baris*. A *bari* may contain anywhere from one to a dozen households, usually with an average of five or six (Ali et al, 2005). Since 1966, the International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B) has administered a surveillance system in Matlab, monitoring the population and recording all demographic and health-related events (D'Souza, 1981). This database offers a unique opportunity to examine the relationships among various environmental and population variables and related health outcomes.

Diarrheal diseases such as cholera are a significant cause of morbidity and remain endemic to the region, primarily due to environmental/seasonal factors and poor sanitation (Black et al, 1981; Emch, 1999; Emch et al, 2002). Research shows that aquatic flora such as algae and phytoplankton provide a reservoir for the cholera-causing V*ibrio cholerae* bacterial pathogen (Islam et al, 1990; Islam et al, 1993), and that microscopic organisms including copepods and zooplankton are also related to pathogen concentrations (Nalin et al, 1977; Nalin et al, 1996; Huq et al, 1983). *V. cholerae* are thus able to inhabit brackish, coastal, and fresh waters for significant periods of time due to the presence of these reservoirs. Environmental and climate-related factors may then affect pathogen survival and reproduction rates, as well as likelihood of introduction into human populations. In cholera-endemic areas, for example, the disease displays distinct seasonal trends that are correlated

with warmer water temperatures and phytoplankton blooms (Lobitz et al, 2000; Epstein et al, 1993; Colwell, 1996).

There thus remains a high risk of contaminated surface water from rivers and canals, which is often used by the population for drinking, bathing, and daily household tasks because of the high iron content of water from area tubewells (Emch, 1999; Hoque et al, 1996). Furthermore, the temporary latrine structures that many households use and the lack of a public sewage system also causes contamination of surface water and potentially exposes the population to various bacteria, including those causing cholera (Emch, 1999). An additional means of exposure occurs through the introduction of *V. cholerae* into drinking water or food by household members or other related persons.

Cholera transmission occurs through the fecal-oral route, yet the pathogen is also able to survive naturally in aquatic environments (Colwell et al, 1977; Colwell et al, 1990). Two forms of transmission are described by researchers, specifically known as primary and secondary (Miller at al, 1985; Craig, 1988; Franco et al, 1997). The former is the result of direct contact with the pathogen in its environmental reservoir; this often occurs as seasonal events encourage growth of the bacteria in the environment, raising the risk of contact with the local population. The latter form is the result of subsequent transmission from infected human hosts to other susceptible individuals through fecal contamination of food, water, or the local sanitation environment.

Secondary transmission occurs in various ways and is documented in previous research. Cholera can spread by bathing in or drinking from water bodies contaminated by human activity (Birmingham et al., 1997; Acosta et al., 2001; Shaprio et al, 1999; Swerdlow et al., 1997; Spira et al., 1980), or through formal public water supplies such as wells, pumps, or municipal systems (Tauxe et al., 1988; Snow, 1855; Ramakrishna et al., 1996; Cardenas et al., 1992; Weber et al., 1994; Swerdlow et al., 1992; Ries et al., 1992). Transmission is also linked to eating food from a restaurant or street vendor (Weber et al., 1994; Ries et al., 1992; Koo et al., 1996). However, infection can also occur at the interfamilial or household level, or outside of the home at social events. In such cases, food or

water also act as vehicles of transmission, but contamination occurs through members of the family or household, or through social contacts (Quick et al., 1995; Gunn et al., 1981).

This research focuses on secondary transmission, in particular when related to social and physical contact with others who have cholera. A social network was constructed for the population of Matlab and the connections used to represent individual-level interactions, examining disease risk based on those relationships. This method may be useful when asking questions related to person-to-person transmission and risk. However, the level of risk at the spatial level is also identified, as primary transmission dynamics and environmental factors within one's neighborhood may play a significant role in cholera occurrence. Local surface water used for consumption, for example, may either naturally support the cholera pathogen or be contaminated by an individual carrying the bacteria. Specific risk factors previously identified in Matlab include multiple households sharing a common latrine, small household area, high population density, and proximity to surface water, amongst other variables (Emch, 1999; Ali et al, 2002a).

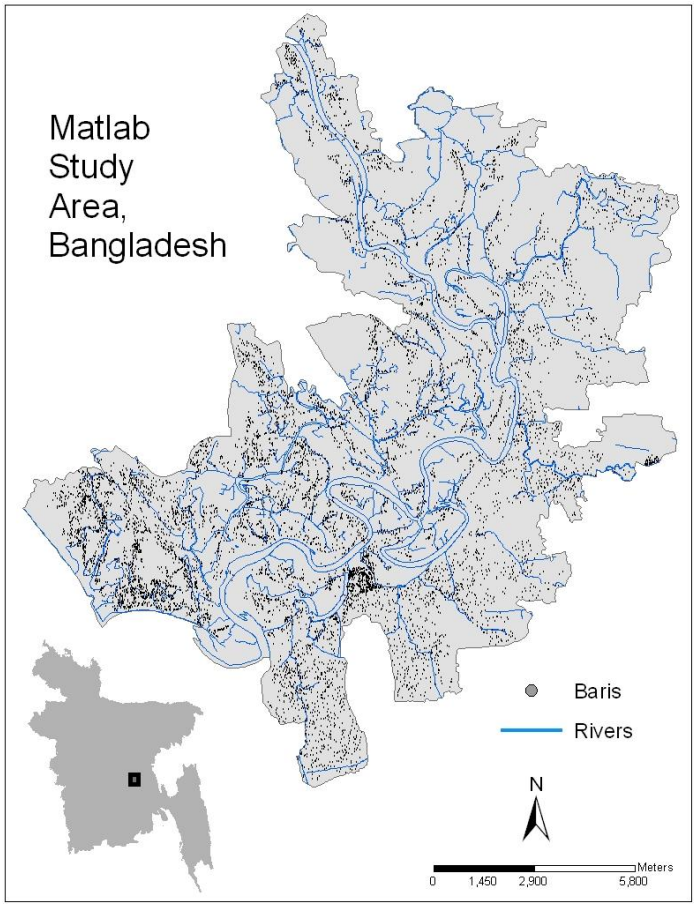

Figure 2.1: Matlab study area in Bangladesh, showing rivers and distribution of *bari*s

Data

This study used a combination of health, demographic, and geographic data to examine the social and spatial clustering of cholera in Matlab. Since 1966, the ICDDR,B has administered a Health and Demographic Surveillance System (HDSS) in the study area. Each resident, upon entry into the study through either birth or in-migration, is assigned a unique identification number within the database known as a Registration ID (RID). The individual is linked through this ID to a village, *bari*, and household. As a person may live in one *bari* initially but then relocate to another, every *bari* of residence for an individual is recorded in the DSS database, including dates of in- and out-migration. Community health workers visit each *bari* in Matlab twice a month and record data on births, deaths, and migrations. Individuals who are sick are referred to the ICDDR,B hospital to which

transportation for the patient is provided at no cost. Data on laboratory-diagnosed cases of diarrheal diseases are recorded at the health facility and then linked to demographic information for individuals, their *baris*, and households.

The population database for this study contains information on all residents of Matlab between January 1st, 1983 and December 31st, 2003, including the RID, sex, date of birth, all *bari*s of residence and associated dates of in- and out-migration. In addition, individual-level information on all laboratory-diagnosed cholera cases during the same time period (*n*=8,765) is used, which also includes the RID, sex, date of birth, residence at time of diagnosis, and date of diagnosis. For purposes of spatial analysis, a GIS database of Matlab was used (Emch et al, 2002). The Matlab GIS includes a variety of geographic features including the spatial location of each *bari* and the hospital location.

Methods

Social networks were constructed and used as a means of documenting individuals related to one another and therefore more likely to engage in some form of interaction, either within or outside of the household. Integrating disease data with network information allowed identification of clustering of cholera within the network. This was a way to model secondary transmission, and evaluate its effects in a cholera-endemic setting. Clustering of cholera in space at various neighborhood levels was also examined in order to compare the effects of the local environment.

A kinship-based social network was developed based on longitudinal and individual-level population data collected by the ICDDR,B. As described above, this dataset contains information on the exact dates a given individual resided in a certain *bari*; therefore, it is possible to "track" an individual from *bari* to *bari* over time in the Matlab study area. The assumption guiding this network is that when an individual moves, he or she maintains interaction with the previous *bari* of residence due to existing relations. Though the original migration is directional, the resulting interaction between the two *bari*s is mutual; therefore the social connections are non-directional. These inter-*bari*

migrations are primarily based on kinship, i.e. marriage into a different family. Specifically, the actors in this network are people with some kinship-based relationship that will foster movement between physical residences, and thus the relationships measured are based on family connections. Naturally, this type of network will not capture all social interactions in the lives of the Matlab population because it primarily includes only family members. However, while kinship is certainly not the only measure of representing how individuals are connected to one another, it is appropriate for the setting of this study given that kinship networks are an important part of social interactions in both rural communities (Guest and Chamratrithirong, 1992) and in lower socio-economic settings (Hollinger and Haller, 1990). A kinship network is one type of indicator of who individuals are interacting with on a regular basis, due to familial visits and customs. While individuals in the study area certainly interact with others to whom they are not related, many of the more prolonged social interactions such as visits between households and shared meals are likely to include kin. Using the migrations of related individuals between *bari*s as evidence of a social connection is more precise than simply assuming all related individuals interact with one another.

Creation of the network was based on individual-level migrations linking *bari*s, which are the "nodes," or units of analysis in the network. Each individual-level migration from *bari* x to *bari* y creates a social linkage between those two *baris*; each linkage of this type is called a dyad. A complete list of all dyads, or an edgelist, can be represented in graphical or matrix form. In a social adjacency matrix, 1 represents the presence of a single, non-directional social connection between two *bari*s and 0 represents no social connection.

Multiple social networks were first created, constructed from the data for each year; i.e., the linkages between *bari*s for 1983 were not considered when constructing the social network for *bari*s in 1984, and so forth. This was to account for uncertainty regarding how long an active social linkage based on a migration may last. Given a cumulative network, or one that considers all social connections created from the beginning of the dataset, all were considered equal when in reality they may have decreased in strength over time. For the purposes of this analysis, each year was first used

independently in order to take note of trends. Secondly, to account for connections created between *bari*s and that may have in fact remained over time, a cumulative network was created. Beginning with 1983, connections created in each year to remained throughout the remainder of the study period. These accumulated network connections and all known cholera cases were then used to predict clustering of cholera rates across *bari*s in 2003.

To determine neighborhood-level clustering of cholera, for each *bari* all other *bari*s located within a 500, 1000, and 2000 meter buffer were identified and recorded using ArcGIS software. This was used to make three distance-band spatial matrices of all *bari*s, where 1 represented a common neighborhood between two *bari*s and 0 represented no common neighborhood. Three different buffers, or "neighborhoods," were used in order to compare spatial clustering at various scales. The total number of *bari*s evaluated in both the social and spatial analysis was 8,873. The dependent variable of interest was the rate of cholera in a *bari* during a specific year, aggregated from all individual recorded cases located in the disease database. For the entire 21-year study period, there were 8,765 cases of cholera in Matlab. Individual-level cholera cases were assigned to a *bari* for each year using the unique RID of the individual diagnosed with cholera at a treatment center. For every *bari*, the total number of cases that year was divided by the total population of the *bari* to produce a cholera rate. For each year, there was thus an *n* x1 vector of *bari*-level dependent cholera values. An additional vector was created containing the value of change in the cholera rate for each *bari*, or the difference between the rate of the current year of interest and the previous year.

For each year, the four 8,873 x 8,873 matrices, one of social adjacency and three representing the different shared spatial neighborhoods, were row-standardized into weights matrices. This gave both social affiliates and spatial neighbors equal "weight" in terms of their influence on a certain *bari*. The matrices could then each be multiplied by either the *n* x 1 vector of cholera rates per *bari* or the vector of change in cholera rate, generating a lag operator which represents the average rate of cholera or cholera change in neighboring *bari*s, or those either socially-affiliated (social lag) or spatially connected (spatial lag).

For the analysis, the global Moran's *I* statistic was used to identify clustering in both the social and spatial clustering. Typically used as a measure of spatial autocorrelation, the Moran's *I* can also be applied to detect clustering within other types of networks representing elements such as language or cultural variables (Dow, 2007). In this case, the result provides a global measure of the degree of clustering of *bari*s with similar cholera rates, in a traditional spatial application but also within the social network. The term "global" refers to average clustering within the entire social network or geographic area, as opposed to a "local" measure, which is used in spatial analysis to identify the relationship between a specific value and the average of the neighboring values. The Moran's I produces a coefficient ranging from -1 (indicating perfect dispersion) and 1 (indicating perfect correlation), with a value of zero implying a random spatial pattern. The measure was used in this study as an indicator of overall clustering of similar cholera rates. Clustering of the change in cholera rates across *bari*s was also tested, i.e. whether those *bari*s connected socially and spatially saw similar changes in rates of cholera from the previous year. Z-scores for significance were Monte Carlo simulation-derived using 10,000 runs, under a null hypothesis of no network autocorrelation either in geographic or social space. The test was run for each of the twenty-one years using the social connectivity matrix as well as the three various spatial distance matrices. Additionally, the Moran's I was run for the cumulative social network consisting of connections created between 1983 and 2002, using this to test for clustering in 2003. Each separate analysis produced both the coefficient representing the extent of clustering and a z-score for significance for each year of the data. The cumulative network produced a single coefficient and z-score for 2003. We used Stata 9 and MATLAB 7.7.0 for the analyses.

Results

Table 1 displays cholera rates per 1,000 in Matlab during the twenty-one years used in this analysis. Overall, rates fluctuated across the data, peaking in 1993 with 1,147 cases. As of 2000, the number of cases decreased and the rate remained less than 1 per 1,000. The social network used to

examine clustering of these cholera cases in the population was based on migrations between *baris*, with a connection in the network created by an individual moving from one *bari* to another as recorded in the ICDDR,B database. For every year, each *bari* in the database therefore had a variable associated with it indicating how many non-directional ties were created during that period. The average number of connections was less than one across all years, indicating that *bari*s typically saw few in-and-out migrations.

Year	# Cholera Cases	Total Population	Rate per 1,000
1983	825	181,224	4.55
1984	452	182,779	2.47
1985	467	183,354	2.55
1986	788	185,857	4.24
1987	414	189,649	2.18
1988	342	192,362	1.78
1989	63	194,833	0.32
1990	47	196,777	0.24
1991	229	199,145	1.15
1992	722	199,450	3.62
1993	1,147	207,332	5.53
1994	739	207,848	3.56
1995	371	208,459	1.78
1996	267	210,279	1.27
1997	509	211,157	2.41
1998	591	213,073	2.77
1999	255	214,344	1.19
2000	120	214,370	0.56
2001	99	203,922	0.49
2002	167	205,709	0.81
2003	151	206,953	0.73

Table 2.1: Number of cholera cases, total population, and rate per 1,000 by year

The extent to which cholera rates were spatially clustered at three different neighborhood scales was then measured, as well as clustering of values representing the difference between the rate in the current and previous year. Results indicated that overall, there is indeed similarity in the cholera rates of *bari*s located within common neighborhoods. Tables 2-4 show z-scores and

associated probabilities of spatial autocorrelation at the three neighborhood levels. Z-scores of more than 1.96 indicate a very strong probability of clustering in space. Significant spatial clustering ($p < 0.05$) of both cholera rates and yearly change in cholera rates for *bari*s occurred at the 500-meter neighborhood for all years of data with the exception of 1990, at which it was significant at the $p < 0.1$ level (Figure 2 and Table 2). The results were similar for spatial clustering at both the 1000-meter and 2000-meter neighborhood (Figures 3 and 4, Tables 3 and 4); at 1000 meters, all years showed significant clustering of cholera rates except for 1990, and all years showed significant clustering of change in cholera rates. At 2000 meters, clustering for both variables was significant at $p < 0.05$ for all years, again with the exception of 1990, at which both were significant at $p < 0.1$ (Figure 4 and Table 4).

Year	**z-score and probability: clustering of cholera rates**		**z-score and probability: clustering of rates of change**	
1983	12.71	p<0.0001	12.53	p<0.0001
1984	8.62	p<0.0001	11.34	p<0.0001
1985	12.94	p<0.0001	9.98	p<0.0001
1986	12.30	p<0.0001	10.15	p<0.0001
1987	3.56	p=0.0004	5.92	p<0.0001
1988	6.88	p<0.0001	4.26	p<0.0001
1989	2.34	p=0.0192	6.23	p<0.0001
1990	0.91	p=0.3639	1.93	p=0.0536
1991	4.21	p<0.0001	3.36	p=0.0008
1992	3.60	p=0.0003	3.77	p=0.0002
1993	13.89	p<0.0001	4.07	p<0.0001
1994	9.24	p<0.0001	10.81	p<0.0001
1995	2.25	p=0.0246	4.52	p<0.0001
1996	5.70	p<0.0001	2.38	p=0.0175
1997	9.47	p<0.0001	9.40	p<0.0001
1998	3.28	p=0.0010	6.09	p<0.0001
1999	6.93	p<0.0001	3.62	p=0.0003
2000	5.61	p<0.0001	8.26	p<0.0001
2001	3.49	p=0.0005	6.16	p<0.0001
2002	3.41	p=0.0006	6.58	p<0.0001
2003	8.74	p<0.0001	5.10	p<0.0001

Table 2.2: Significance of spatial clustering of cholera rates and change in cholera rates, 500-meter neighborhoods.

Year	z-score and probability: clustering of cholera rates		z-score and probability: clustering of rates of change	
1983	8.61	p<0.0001	8.24	p<0.0001
1984	8.53	p<0.0001	6.99	p<0.0001
1985	13.14	p<0.0001	9.17	p<0.0001
1986	17.29	p<0.0001	11.69	p<0.0001
1987	5.97	p<0.0001	6.81	p<0.0001
1988	8.02	p<0.0001	5.56	p<0.0001
1989	4.01	p<0.0001	6.51	p<0.0001
1990	1.58	p=0.1146	3.16	p=0.0016
1991	4.37	p<0.0001	3.37	p=0.0008
1992	6.20	p<0.0001	5.97	p<0.0001
1993	18.15	p<0.0001	5.25	p<0.0001
1994	9.63	p<0.0001	10.17	p<0.0001
1995	2.66	p=0.0079	4.45	p<0.0001
1996	6.06	p<0.0001	2.28	p=0.0227
1997	11.28	p<0.0001	7.60	p<0.0001
1998	4.16	p<0.0001	5.84	p<0.0001
1999	8.82	p<0.0001	4.62	p<0.0001
2000	6.61	p<0.0001	9.45	p<0.0001
2001	3.61	p=0.0003	6.87	p<0.0001
2002	3.76	p=0.0002	5.91	p<0.0001
2003	8.42	p<0.0001	4.40	p<0.0001

Table 2.3: Significance of spatial clustering of cholera rates and change in cholera rates, 1000-meter neighborhoods.

Year	z-score and probability: clustering of cholera rates		z-score and probability: clustering of rates of change	
1983	9.91	p<0.0001	9.50	p<0.0001
1984	9.2	p<0.0001	4.41	p<0.0001
1985	15.08	p<0.0001	6.69	p<0.0001
1986	26.33	p<0.0001	14.16	p<0.0001
1987	9.65	p<0.0001	6.42	p<0.0001
1988	5.69	p<0.0001	5.14	p<0.0001
1989	3.05	p=0.0023	3.39	p=0.0007
1990	1.81	p=0.0708	1.8	p=0.0727
1991	5.88	p=0.0023	3.2	p=0.0014
1992	8.37	p<0.0001	6.6	p<0.0001
1993	25.62	p<0.0001	5.19	p<0.0001
1994	9.68	p<0.0001	10.39	p<0.0001
1995	4.19	p<0.0001	4.47	p<0.0001

1996	5.47	p<0.0001	2.64	p<0.0001
1997	10.74	p<0.0001	4.81	p<0.0001
1998	5.60	p<0.0001	6.51	p<0.0001
1999	8.56	p<0.0001	4.85	p<0.0001
2000	4.06	p<0.0001	7.06	p<0.0001
2001	2.8	p<0.0001	4.37	p<0.0001
2002	4.1	p<0.0001	4.96	p<0.0001
2003	7.39	p<0.0001	3.32	p<0.0001

Table 2.4: Significance of spatial clustering of cholera rates and change in cholera rates, 2000-meter neighborhoods.

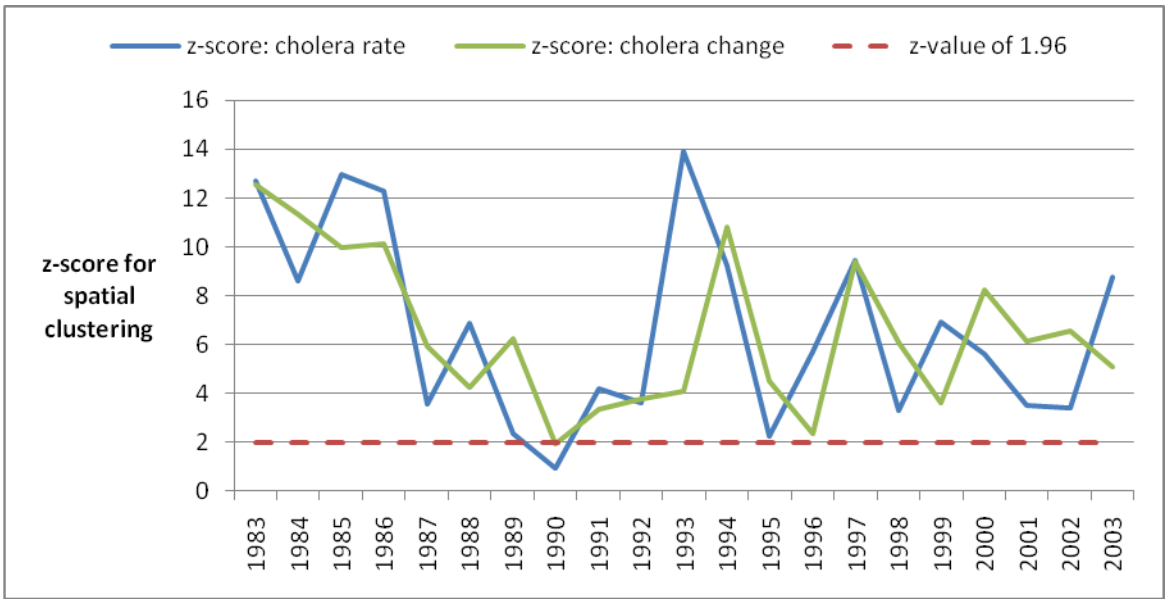

Figure 2.2: Significance of spatial clustering of cholera rates and change in cholera rates, 500-meter neighborhoods.

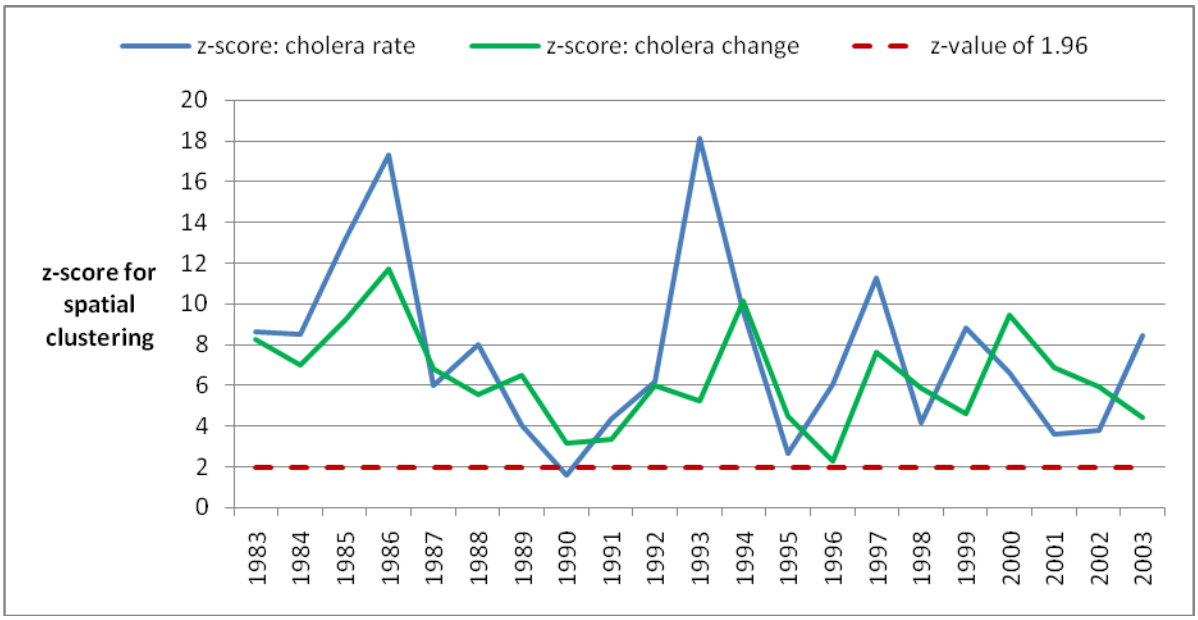

Figure 2.3: Significance of spatial clustering of cholera rates and change in cholera rates, 1000-meter neighborhoods.

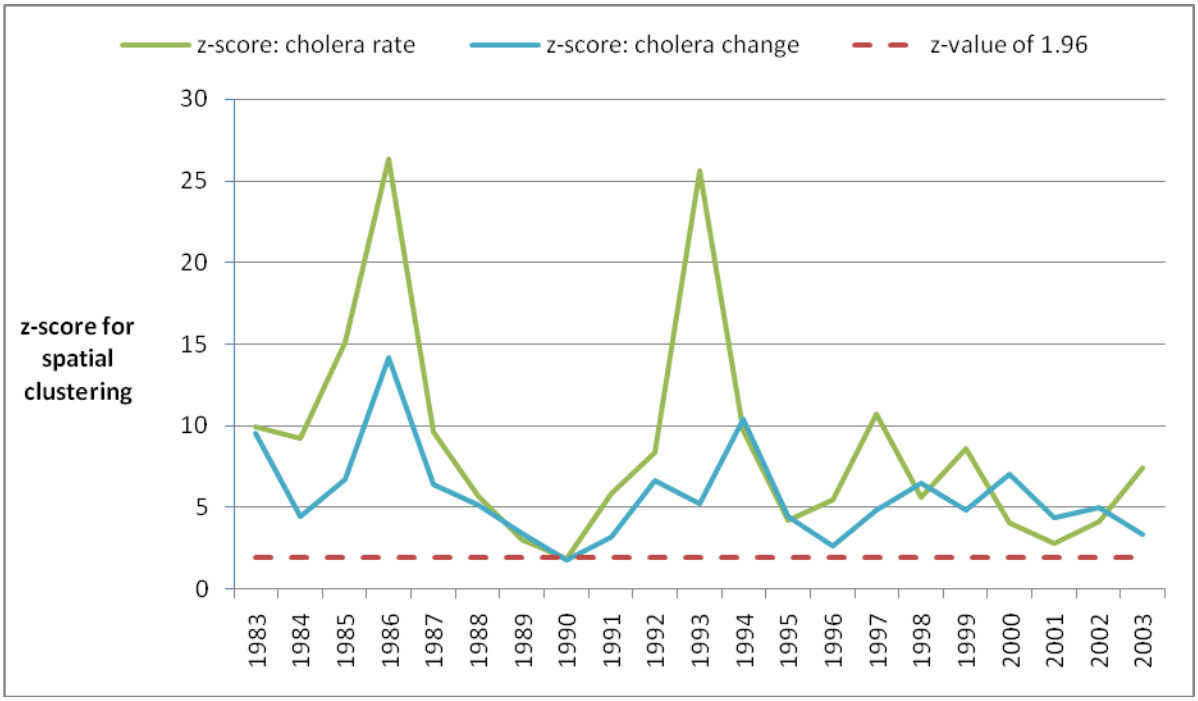

Figure 2.4: Significance of spatial clustering of cholera rates and change in cholera rates, 2000-meter neighborhoods.

The global Moran's I was then used to assess the degree of clustering of cholera across *baris* connected by the kinship-based social network created, as well as clustering of yearly change in cholera rates across socially connected *baris*. Regarding trends across years, results here differed

significantly from the measures of spatial clustering at the various neighborhood levels (Table 5 and Figure 5). Namely, evidence for social clustering of cholera rates amongst connected *bari*s was found during only three different years of the data. Clustering of change in cholera rates occurred only in one year. Significant ($p < 0.05$) values for cholera in 1989, 1993, and 2000, and for change in cholera rates in 1989. Change in cholera rates was also significant at the 0.1 level in 2000.

Year	**z-score and probability: clustering of cholera rates**		**z-score and probability: clustering of rates of change**	
1983	0.27	p=0.397	0.26	p=0.397
1984	1.55	p=0.061	0.23	p=0.411
1985	1.51	p=0.066	-1.27	p=0.897
1986	0.04	p=0.485	0.13	p=0.449
1987	0.19	p=0.423	0.47	p=0.319
1988	0.02	p=0.493	0.00	p=0.5
1989	8.49	p<0.0001	2.84	p=0.002
1990	0.02	p=0.493	0.04	p=0.482
1991	-0.26	p=0.604	-0.19	p=0.426
1992	-0.43	p=0.666	-0.42	p=0.338
1993	14.61	p<0.0001	-0.22	p=0.411
1994	-0.04	p=0.518	0.66	p=0.254
1995	0.06	p=0.475	-0.50	p=0.310
1996	-0.11	p=0.455	0.02	p=0.491
1997	0.56	p=0.289	0.43	p=0.335
1998	-0.05	p=0.482	-0.30	p=0.384
1999	-0.04	p=0.484	-0.02	p=0.508
2000	5.31	p<0.0001	1.95	p=0.026
2001	0.13	p=0.450	0.05	p=0.481

2002	-0.42	p=0.338	-0.01	p=0.506
2003	-0.01	p=0.495	0.05	p=0.481

Table 2.5: significance of social clustering of cholera rates and change in cholera rates

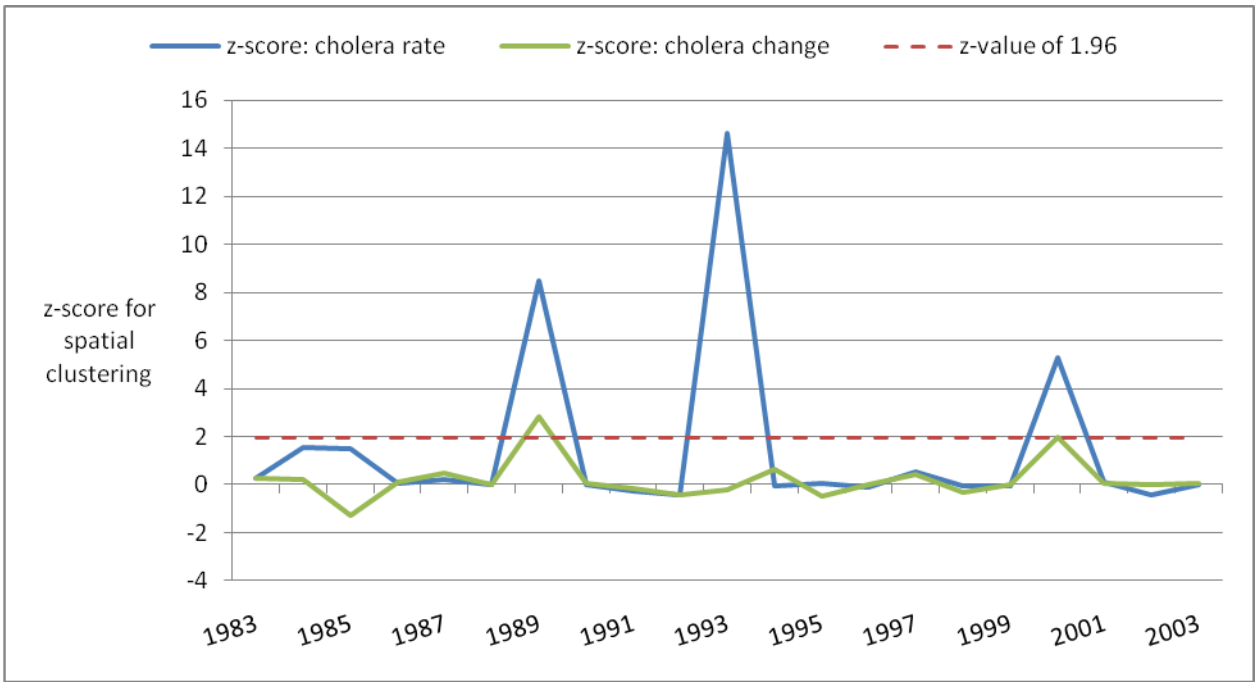

Figure 2.5: Significance of social clustering of cholera rates and change in cholera rates.

Lastly, the social network containing all cumulative connections created between 1983 and 2002, combined with disease incidence, was used to predict clustering of cholera rates in 2003. The results showed no significant values for that year, indicating that the cholera rates in *bari*s during 2003 were not affected by the accumulated network connections. That is, having been connected to *bari*s with cholera in past years, as opposed to only the present year, was not related to incidence in the last year of the data.

Discussion and Conclusions

This study sought to identify the extent to which social connectivity poses individual cholera risk. Previous studies suggest that neighborhood-level factors are important predictors of cholera incidence, affecting populations through both primary and secondary transmission dynamics. In order

to measure the impact of secondary transmission, especially through contact between households linked by kinship, connected *bari*s were identified to create a social network under the assumption that individuals residing within those *bari*s were more likely to come into contact with one another. This paper is an example of how the global Moran's I statistic can be used to assess spatial as well as social clustering of a health outcome for comparison purposes.

The results of this analysis support the hypothesis that there is clustering of similar cholera rates across *bari*s within the same neighborhoods during the entire study period. This is likely due to environmental and demographic risk factors in these locations, such as surface water proximity, high population density, and shared latrine use. The same was generally true for clustering of differences in cholera rates from the previous year, providing further evidence that local effects matter. The fact that similar disease rates are seen at three spatial scales ranging from 500 to 2000 meters indicates that common environmental and demographic effects are important in both smaller and larger neighborhoods.

In addition to spatial clustering, this research examined the extent to which social connections to households with cholera predict disease occurrence in *bari*s. It appeared that clustering of similar cholera rates in the social network was less common. There were, however, three separate years in which there was highly significant social clustering. Clustering in change of cholera rates from the previous year also occurred in two of the same time periods. During two of these years, average rates of cholera across *bari*s were not exceptionally high, though 1993 did see the highest rate of cholera during the study period, which may explain the higher level of clustering.

While social network analysis is frequently used in epidemiological studies, no other study has examined cholera and social networks. These results show that while cholera almost always is clustered in space, it is sometimes clustered socially as well. There are two main limitations to this study. First, the network used represents kinship connections. This did not include all other social interactions that occured in the lives of the population of Matlab, which may play a role in risk as well. Secondly, it is important to consider that spatial clustering may then have been the result of

some of these non-familial interactions, as individuals are more likely to come into contact with those who live in closer proximity. Therefore social factors may in fact play a role in the spatial clustering. However, the comparison between social clustering and spatial clustering suggests that the local environment is much more important than social connections in cholera transmission.

Improving understanding of disease transmission dynamics is critical for public health. While improvements in sanitation, socioeconomic status, and education have decreased rates of diarrheal disease in Bangladesh and other countries in the developing world, it remains a priority to identify specific pathways of transmission and thus develop effective intervention methods. A combination of social network and spatial analysis can help improve understanding of transmission dynamics.

References

Acosta, J.C., Galindo, C.M., Kimario, J., Senkoro, K., Urassa, H., Casals, C., Corachán, M., Eseko, N., Tanner, M., Mshinda, H., Lwilla, F., Vila, J., and Alonso, P.L. 2001. Cholera outbreak in southern Tanzania: risk factors and patterns of transmission. *Emerging Infecrious Diseases (Supplement 3):* 583–587.

Ali, M., Emch, M., Donnay, J. P., Yunus, M., & Sack, R. B. 2002a. Identifying environmental risk factors for endemic cholera: A raster GIS approach. *Health & Place,* 8(3), 201-210.

Ali, M., Emch, M., Donnay, J. P., Yunus, M., & Sack, R. B. 2002b. The spatial epidemiology of cholera in an endemic area of Bangladesh. *Social Science & Medicine,* 55(6), 1015-1024.

Ali, M. Emch, M. von Seidlein, M. Yunus, D.A. Sack, M. Rao, J. Holmgren and J.D. Clemens. 2005. Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis, *Lancet* 366 (9479) (2005), pp. 44–49.

Aral et al. 1999. Sexual Mixing Patterns in the Spread of Gonococcal and Chlamydial Infections. *American Journal of Public Health* 89 (6): 825-832.

Auerbach, D.M., Darrow, W.W., Jaffe, H.W., and Curran, J.W. 1984. Cluster of Cases of the Acquired Immune Deficiency Syndrome: Patients Linked by Sexual Contact. *The American Journal of Medicine,* 76(3):487-92.

Bell, D.C., Atkinson, J.S., and Carlson, J.W. 1999. Centrality measures for disease transmission networks. *Social Networks,* 21(1): 1-21.

Birmingham, M.E., Lee, L.A., Ndayimirije, N., Nkurikiye, S., Hersh, B.S., and Wells, J.G. 1997. Epidemic cholera in Burundi: patterns of transmission in the Great Rift Valley Lake region. *The Lancet,* 349: 981–85.

Black, R.E., Merson, M.H., Huq, I., Alim, A.R.M.A., Yunus, M.D., 1981. Incidence and severity of rotavirus and Escherichia coli diarrhoea in rural Bangladesh. *The Lancet*, 1:141-143.

Christakis, N.A., and Fowler, J.H. 2007.The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine,* 357: 370-379.

Colwell, R.R. 1996. Global Climate and Infectious Disease: The Cholera Paradigm. *Science,* 274(5295): 2025-2031.

Colwell, RR, Kaper, J, and Joseph, SW. 1977. Vibrio cholerae, Vibrio parahaemolyticus, and other vibrios: occurrence and distribution in Chesapeake Bay. *Science*. 198:394.

Colwell, RR, Tamplin, ML, Brayton, PR, Gauzens, AL, Tall, BD, Herrington, D, Levine, MM, cholera. In: *Advances in research on cholera and related diarrhoea*, 7th ed Tokyo, KTK Scientific Publishers. 1990. pp. 327–343.

Cook, V.J., Sun, S.J., Tapia, J., Muth, S.Q., Arguello, D.F., Lewis, B.L., Rothenberg, R.B., and McElroy, P.D. 2007. Transmission network analysis in Tuberculosis contact investigations. *Journal of Infectious Diseases,* 196: 1517-27.

Craig, M. 1988. Time-space clustering of vibrio cholerae 01 in Matlab, Bangladesh, 1970-1982. *Social Science and Medicine*, 26 (1): 5-13.

D'Souza, S., 1981. A population laboratory for studying disease processes and mortality: the demographic surveillance system in Matlab, Bangladesh. Special Publication No. 13, ICDDR,B, Dhaka, Bangladesh.

Dow, M.M. 2007. Galton's problem as multiple network autocorrelation effects: cultural trait transmission and ecological constraint. *Cross-Cultural Research*, 41(4): 336-363.

Emch, M. 1999. Diarrheal disease risk in Matlab, Bangladesh. *Social Science & Medicine,* 49(4): 519-530.

Emch, M., and Ali, M. 2001. Spatial and temporal patterns of diarrheal disease in Matlab, Bangladesh. *Environment and Planning*, 33: 339-350.

Epstein, P R. 1993. Algal blooms in the spread and persistence of cholera. *Biological Systems* 31(2-3): 209-21.

Franco, A.A., Fix, A.D., Prada, A., Paredes, E., Palomino, J.C., Wright, A.C., Johnson, J.A., McCarter, R., Guerra, H., Morris, J.G. Jr. 1997. Cholera in Lima, Peru, correlates with prior isolation of Vibrio cholerae from the environment. *American Journal of Epidemiology*, 146(12): 1067-1075.

Guest, P. and Chamratrithirong, A. 1992. The social context of fertility decline in Thailand, in: Goldscheider C, ed., Fertility Transitions, Family Structure, and Population Policy, Boulder, CO, USA: Westview Press.

Hanneman, R.A. 2001. *Introduction to social network methods*. Riverside, California: Department of Sociology, University of California at Riverside.

Hollinger, F., and Haller, M. 1990. Kinship and social networks in modern societies: a cross-cultural comparison among seven nations. *European Sociological Review*, 6: 103-124.

Hoque, B.A., Ahmed, S.A., Chakraborty, J., Chowdhury, UK, Sack, R.B., 1996. Water quality and health situation analysis of Meghna–Dhonagoda project. ICDDR,B Report, pp. 1–47.

Huq, A., Small, E.B., West, P.A., Huq, M.I., Rahman, R., and Colwell, R.R. 1983. Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Applied Environmental Microbiology*, 45 (1): 275–283.

Islam, M.S., Drasar, B.S., and Bradley, D.J. 1990. Long-term persistence of toxigenic *Vibrio cholerae* 01 in the mucilaginous sheath of a blue-green alga,*Anabaena variabilis*. *Journal of Tropical Medicine and Hygiene*, 93 (2): 133–139.

Islam, M.S., Drasar, B.S., and Sack, R.B. 1993. The aquatic environment as a reservoir of *Vibrio cholerae*: a review. *Journal of Diarrhoeal Disease Research*, 11 (4): 197–206.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A.S.G., and Colwell, R.R. 2000. Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proceedings of the National Academy of Sciences* 97(4): 1438–1443.

Klovdahl, A.S. 2001. Networks and tuberculosis: an undetected community outbreak involving public places. *Social Science & Medicine*, 52(5): 681-694.

Miller, C.J., Feachem, R.G., and Drasar, B.S. 1985. Cholera epidemiology in developed and developing countries: new thoughts on transmission, seasonality, and control. *Lancet*, 8423: 261-3.

Nalin, DR. 1976. Cholera, copepods, and chitinase. *Lancet*, 2 (7992): 958.

Nalin, D.R., Daya, V., Reid, A., Levine, M.M., and Cisneros, L. 1979. Absorption and growth of *Vibrio cholerae* on chitin. *Infection and Immunology*, 25 (2): 76-770.

Shapiro, R.L., Otieno, M.R., Adcock, P.M., Phillips-Howard, P.A., Hawley, W.A., Kumar, L., Waiyaki, P., Nahlen, B.L., and Slutsker, L. 1999. Transmission of epidemic Vibrio cholerae O1 in rural western Kenya associated with drinking water from Lake Victoria: an environmental reservoir for cholera? *American Journal of Tropical Medicine and Hygiene*, 60(2): 271-276.

Siddique AK, Zaman K, Baqui AH, et al. 1992. Cholera epidemics in Bangladesh: 1985–1991. *Journal of Diarrhoeal Disease Research*, 10: 79–86.

Spira et al. 1980. Microbiological surveillance of intra-neighborhood El Tor cholera transmission in rural Bangladesh. *Bulletin of the World Health Organization* (WHO), 33(5):627-36.

Swerdlow et al. 1992. Waterborne transmission of epidemic cholera in Trujillo, Peru: lessons for a continent at risk. *The Lancet*, 340 (8810): 28-33.

Swerdlow, D.L. and M. Isaacson. 1994. The epidemiology of cholera in Africa. In: *Vibrio cholerae and Cholera: Molecular to Global Perspectives* (Wachsmuth, K.I., Blake, P.A. and Olsik, O., Eds.), pp. 297-307. American Society for Microbiology Press, Washington, DC.

Swerdlow, D.L., Malenga, G., Begkoyian, G., Nyangulu, D., Toole, M., Waldman, R.J., Puhr, D.N.D., and Tauxe, R.V. 1997. Epidemic cholera among refugees in Malawi, Africa: treatment and transmission. *Epidemiology and Infection*, 118: 207-214.

Wasserman, S. and Faust, K. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Chapter 3

Bridge

The results of the Moran's *I* as run for cholera clustering in both spatial proximity and social networks indicate that the disease does concentrate in space consistently over the twenty-one years examined. Only during three years, however, was there social clustering of cholera in the kinship-based social network created. The same general trend occurred when the difference between cholera rates in the previous and current year were measured for spatial and social autocorrelation. The evidence for concentration of cholera in geographic space suggests that risk-related environmental factors are important, while social networks are significantly less predictive of cholera occurrence. Furthermore, clustering in the social network may be caused by households connected to one another being located near each other as well, thus experiencing the same environmentally-related risk. The second analysis presented here controls for factors that are in the local environment, as well as those that remain unidentified, since data on numerous variables is not always available. This will then produce a coefficient that represents the "social effect," as well as values indicating the spatial error. Accounting for possible environmental variables that may increase risk elaborates on the social effect, which may be confounded by neighborhoods. The two analyses are thus complimentary and illustrate how spatial and social network interactions are related to one another in cholera transmission dynamics.

Chapter 4

The Importance of Spatial versus Social Networks on Cholera Transmission

Abstract: Social network analysis is an emerging and effective epidemiological tool for analyzing diffusion of infectious diseases. While past research has primarily focused on sexually-transmitted infections, the role of networks in transmission of other pathogens remains somewhat unexplored. The objectives of this research are to investigate social networks as they relate to transmission of cholera, a diarrheal disease, and compare this risk to neighborhood-level effects. Twenty-one years of demographic data from the Matlab, Bangladesh study area are used to create a social network of households. Using GIS software, households sharing a neighborhood are also identified. Laboratory-diagnosed cholera case data is then integrated and the variables run in a global Moran's I to test for separate social and spatial clustering. In addition, a spatial error-social effects model tests for clustering of rates of cholera in socially-linked households while simultaneously accounting for spatial factors. Results show 1) consistent spatial clustering at the neighborhood level but separate social clustering for only three years of data; 2) five years of evident social clustering while accounting for both known and unknown environmental variables. These findings suggest that transmission of cholera through the environment is significant in an endemic region, with social networks perhaps accounting for transmission as well, but not as consistently. This knowledge is important for effective intervention efforts preventing diffusion in vulnerable populations.

Introduction

Identifying and analyzing the connections between actors in social networks is an emerging means across disciplines of understanding the importance of relationships. A network may be defined by many different types of associations, including kinship, friendship, institutional linkages, or

physical contact (Wasserman & Faust, 1994; Hanneman, 2001; Klovdahl et al, 2001). In epidemiological and other medically-related studies, networks are a useful tool for analyzing the effects of social support and peer influence on health and health behaviors (Kinney et al, 2003; Valente et al, 1997), modeling vaccination strategies (Halloran et al, 2002; Miller et al, 2007), and for understanding processes of infectious disease diffusion. However, the impact of networks on the transmission of many diseases remains unknown, especially those that spread by more various and complex means (Klovdahl et al, 2001). Understanding these diffusion dynamics is essential to developing effective intervention strategies, such as vaccination and other public health measures.

This study focuses on the effects of social contacts on risk of cholera, a diarrheal disease that remains endemic to many parts of the developing world. Specifically, we question how much individual risk exists based on one's degree of social connectedness to other people with the disease. No research has yet looked at the transmission of cholera through social networks; however, understanding this process is particularly significant due to continued interest amongst researchers on the exact dynamics by which cholera spreads to, and within, human populations. While environmental factors facilitate transmission, behavioral factors related to social interactions, such as food-sharing, may lead to disease diffusion as well. For the purpose of gaining insight as to the dynamics of primary versus secondary transmission, risk within social networks is examined here and compared to neighborhood-level cholera clustering.

Background

Cholera Transmission

Primarily a waterborne disease, cholera has two documented forms of transmission. The pathogen responsible, *vibrio cholerae*, is known to naturally occur in coastal and brackish waters, which provide an environmental reservoir (Colwell et al, 1977; Colwell et al, 1990; Islam et al, 1990; Islam et al, 1993). Humans may thus become infected through direct contact with the bacteria, designated as primary transmission. As seasonal events encourage growth of the bacteria in the

environment, risk of introduction into the human population may increase. Alternatively, cholera may occur as the result of interaction with another individual carrying the bacteria, either through contaminated food or water as a vehicle for the pathogen, or through direct person-to-person contact. This is referred to as secondary transmission (Miller et al, 1985; Craig, 1988; Franco et al, 1997).

The latter form occurs in a number of ways. Cholera may spread from bathing in or drinking from water bodies contaminated by humans carrying the pathogen (Birmingham et al., 1997; Acosta et al., 2001; Shaprio et al, 1999; Swerdlow et al., 1997; Spira et al., 1980), or through formal public water supplies such as wells, pumps, or municipal systems that the pathogen has been introduced to by humans (Tauxe et al., 1988; Snow, 1855; Ramakrishna et al., 1996; Cardenas et al., 1992; Weber et al., 1994; Swerdlow et al., 1992; Ries et al., 1992). Transmission is also linked to consuming food or drink from restaurants and street vendors (Weber et al., 1994; Ries et al., 1992; Koo et al., 1996). However, infection can also occur within the household through family members (Shahid et al., 1984; Fuduka et al., 1995; Lasch et al., 1984; Tamayo et al., 1965; Sengupta et al., 1994), or outside of the home at social events (Quick et al., 1995; Gunn et al., 1981). In such cases, food or water also act as vehicles of transmission, but contamination occurs through members of the family or household, or through social contacts.

Social Networks and Disease

Throughout the literature on networks and health, HIV/AIDS and sexually transmitted infections are perhaps the most frequently examined disease outcomes (Jaffe et al, 1983; Auerbach et al, 1984; Bell et al, 1999; Wylie et al, 2005; Ghani et al, 1997; Aral et al, 1999). In these studies, transmission is tracked through sexual networks or individuals sharing needles for the purposes of intravenous drug use. Other contagious illnesses are also the subject of past research (Fitzpatrick et al, 2001; Klovdahl et al, 2001; Cook et al, 2007). However, the means by which diseases spread across networks where connections are not sexual, or transmitted through a defined vehicle such as needles, are more complex. Klovdahl et al (2001) traced transmission of tuberculosis amongst individuals in

Houston to social locations, such as bars and clubs, using interview and survey methods. Fitzpatrick et al (2001) used similar methods to track a TB outbreak in Indiana. Tuberculosis is a respiratory illness that can be transmitted through a single infectious aerosol droplet, placing people at high risk if in close proximity to a symptomatic carrier; therefore social networks are a useful method for identifying diffusion patterns.

Bates et al (2007) considered networks when examining diarrheal disease risk in rural Ecuador. In their analysis, data on spatial household proximity and household degree centrality, or the number of non-directional social connections to other households, was collected. For nine different communities, mean density and degree values were compared to community diarrheal disease incidence. While higher spatial density was negatively related to diarrheal disease incidence, higher social centrality displayed a positive association, suggesting higher risk due to crowding but lower risk associated with social cohesion. The authors speculate that the latter finding is indicative of social networks not facilitating transmission.

The model used in this study considers similarity in rates of cholera amongst households that are located within similar neighborhoods. However, in contrast to Bates et al (2007), the component of this research that considers social networks measures prevalence at the household level as related to cholera rates in socially connected households. Social clustering is measured while simultaneously controlling for both known and unknown environmental and spatial variables. This further clarifies the extent to which clustering that appears is actually a result of transmission related to social and behavioral factor, as opposed to common underlying neighborhood dynamics.

Study Area and Data

The study region is located in Matlab, Bangladesh, which lies southwest of Dhaka, the capital (Figure 1). Primarily an agricultural area, it is approximately 184 square kilometers with a population of about 200,000 that resides in clusters of patrilineally-related households called *baris*. A *bari* typically contains anywhere from one to a dozen households, averaging at five or six (Ali et al, 2005).

The International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B) has administered a surveillance system in Matlab since 1966, monitoring the population and recording all demographic and health-related events (D'Souza, 1981). Such a database makes it possible to examine the relationships among various environmental, population, and health-related variables. Diarrheal diseases are especially of interest, as they remain a significant cause of morbidity in rural Bangladesh (Black et al, 1981; Baqui et al, 1992; Emch, 1999).

The health and demographic data, in combination with geographic data, was used to analyze the social and spatial clustering of cholera in Matlab. Within the demographic database, each resident is identified by a unique identification number known as a Registration ID (RID), which is assigned upon entry into the study area through either birth or in-migration. The individual is linked through this ID to a village, *bari*, and household. Since a person may not remain in the same one *bari* while living in Matlab, every *bari* of residence for an individual is recorded in the DSS database, including dates of in- and out-migration. Community health workers visit each *bari* in Matlab every two weeks for the purposes of recording data on births, deaths, and migration, as well as any information regarding illnesses amongst members of the *bari*. Individuals who are sick at the time of the visit are then referred to either the main ICDDR, B hospital or to one of the additional three treatment centers in the area, to which transportation for the patient is provided at no cost. Laboratory-diagnosed cases of diarrheal diseases are then entered into a database at the health facility and then linked to demographic information for individuals, their *baris*, and households.

The population database used includes information on all residents of Matlab for the period between January 1st, 1983 and December 31st, 2003, including the RID, sex, date of birth, all *bari*s of residence and associated dates of in- and out-migration. This was integrated with individual-level data on all laboratory-diagnosed cholera cases from the same time period (n=8,765), which also includes the RID, sex, date of birth, residence at time of diagnosis, and date of diagnosis. A GIS database of Matlab created using aerial photography was used for the spatial component of the analysis (Emch et

al, 2002). The Matlab GIS includes geographic features such as rivers, ponds, roads, health center sites, and the spatial location of each *bari*.

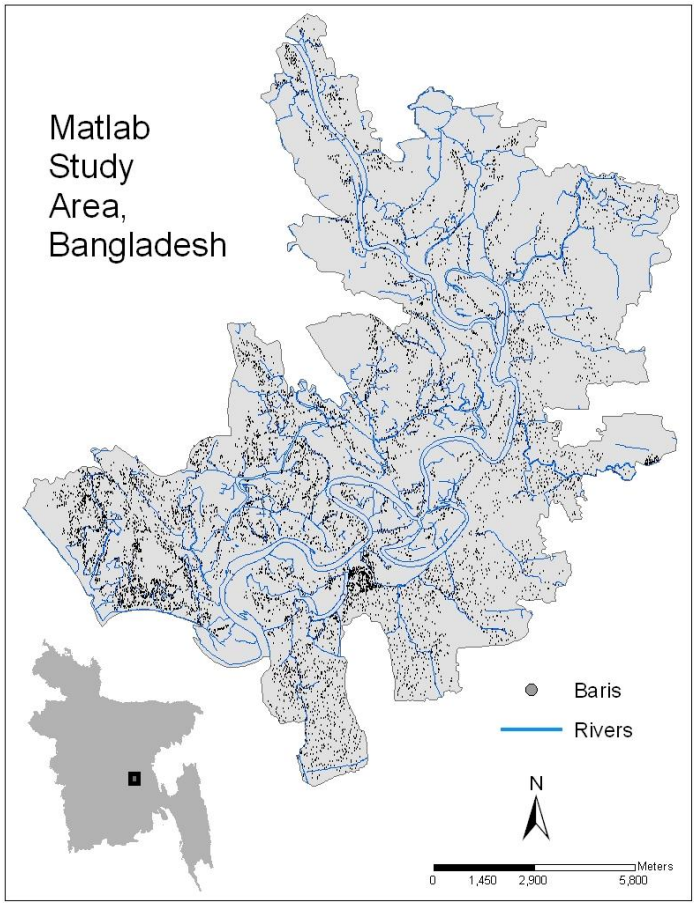

Figure 4.1: Matlab study area in Bangladesh, showing rivers and distribution of *bari*s

Methods

As the objective was to analyze cholera transmission in the social and domestic sphere, demographic and migration data were used to construct a social network and thus identify with whom individuals were more likely to interact. This primarily kinship-based network used longitudinal and individual-level population data from the ICDDR,B database, which contains the exact dates a given individual resided in a certain *bari*. Therefore, one is able to follow an individual from *bari* to *bari* over the course of his or her time in the Matlab study area. Migrations between *baris* are primarily

kinship-based, e.g. due to marriage into a different family. The actors in the network are thus individuals with some kinship-based relation that may lead to relocation from one physical residence to another.

The network was created and used under the assumption that when an individual moves, he or she maintains contact with the previous *bari* of residence. As a result, there is mutual interaction between the old and new *bari*, forming a non-directional social connection. One limitation in using this type of network is that it will not encompass all social interactions in the lives of the Matlab population, only those that mainly include family members. Individuals will therefore also come into contact with friends and acquaintances, interactions that we were unable to document. However, kinship-based relations are appropriate and useful in this particular study population, given that these types of networks are an integral component of social interaction in rural and lower socio-economic settings (Guest and Chamratrithirong, 1992; Hollinger and Haller, 1990). Traditional customs such as shared meals and familial visits further encourage social and physical interaction between kin.

As stated above, the network was based on individual-level migrations that create a connection between *bari*s, which are the "nodes," or units of analysis in the network. An individual migration from *bari* x to *bari* y creates a linkage between *bari* x and *bari* y, also known as a dyad. A list that contains all known connections between pairs in a network is called an edgelist. An edgelist can then be represented in a number of forms, such as a nodelist, a visual graph, or a matrix. A social adjacency matrix was used here, containing values of either 1 or 0, in which 1 represented the presence of a single, non-directional social connection between two *bari*s and 0 was equivalent to no social connection. Multiple social networks and consequent matrices were created for each year of data, which did not consider any connections created in the previous year, due to uncertainty regarding how long social linkages actually last. The result was twenty-one individual networks for each year.

The spatial component for this analysis included creating a neighborhood at which level to measure neighborhood clustering, as well as creating environmental variables to control for in the

analysis. Firstly, for each *bari*, all other *baris* located within a 1000-meter distance buffer were identified using ArcGIS software. A distance-band spatial matrix of all *baris* was created, where 1 represented a common neighborhood between two *bari*s and 0 represented no common neighborhood. The neighborhood buffer was then used to compare spatial clustering at various scales. Secondly, for each *bari* the distances to the nearest road, river, pond, and tubewell were measured and variables created representing this distance, in kilometers. In addition, a variable for the depth of the nearest tubewell was included. Not all *bari*s in the population dataset were represented in the GIS; therefore the social network contained only those *bari*s for which both spatial and demographic data were available. The total number of *bari*s used in both the social and spatial analysis was 8,873.
The dependent variable was the rate of cholera in the population in a *bari* for each year, compiled from all individual recorded cases within the disease database. There were 8,765 cases of cholera in Matlab during the twenty-one year study period. Individual-level cases were assigned to a *bari* for each year using the unique RID of the individual diagnosed with cholera at a treatment center. The total number of cases that year was then divided by the total population of the *bari* to produce a cholera rate. Each year thus had an *n* x1 vector of *bari*-level dependent cholera values.

The two 8,873 x 8,873 matrices, one for the social network (W1) and one representing the shared spatial 1000-meter neighborhood (W2), were row-standardized into weights matrices. Social affiliates and spatial neighbors were thus granted equal "weight" in terms of their influence on a certain *bari*. The matrices were then each multiplied by the *n* x 1 vector of cholera rates per *bari*, generating a lag operator which represents the average rate of cholera in neighboring *bari*s, or those either socially-affiliated (social lag) or spatially connected (spatial lag).

The lagged variables were used in a global Moran's *I* model to identify social and spatial clustering. Used as a measure of spatial autocorrelation, the Moran's *I* is also used in the social sciences to identify clustering within non-spatial networks, representing elements such as language or cultural variables (Dow, 2007). Running the model for the social network and the spatial neighborhood separately, however, does not take the other into account. Therefore any social

clustering may actually be the result of spatial clustering, i.e. individuals who are socially connected are more likely to live close to one another and therefore be affected by the same environmental risk factors, rather than transmitting the disease through personal contact. A model that estimates social effects while controlling for both known independent variables and unknown underlying spatial effects is thus desirable here.

Doreian (1982), and Anselin (1988) describe a combined linear spatial effects-spatial disturbance model, estimated using maximum-likelihood methods. Initially, there is the spatial effects model:

$$y = \rho_1 W_1 y + X\beta + e$$

where ρ_1 is the spatial effects parameter; W_1 is the spatial weights matrix; X is a matrix of independent variable observations; β is a vector of parameters to be estimated; and e is a randomly distributed error term. However, there is the possibility that e is spatially autoregressive, or that the off-diagonal elements of the covariance matrix also exhibit spatial dependence. In this case, the error term is represented as

$$e = \rho_2 W_2 e + \nu$$

where ρ_2 is a spatial parameter for the disturbance term, and ν is an n x 1 vector of a randomly distributed error term. The model that integrates both spatial effects, or spatial lag, and spatial error appears as

$$y = \rho_1 W_1 y + X\beta + \rho_2 W_2 e + \nu$$

and, if estimated using typical OLS procedures, would be inefficient due to the autoregressive nature, or correlation of the Wy term and the error. Furthermore, the standard errors produced would be biased. Therefore, maximum-likelihood estimation methods are preferable for measuring the effects of interest (Doreian, 1982).

The spatial effects-spatial disturbances model was applied in this analysis as described above, where y was the rate of cholera in a *bari* of interest. The primary difference is that W_1 became a social weights matrix, and ρ_1 a social effects parameter. The remaining elements remained as described above, or the spatial weights matrix with a spatial disturbance term, and a random error component. The combined model was appropriate here, as the social effect, or the primary outcome of interest, was assessed in terms of both existence and strength, while the spatial disturbance was used to correct the bias potentially created by autocorrelation of cholera-related variables in space. Using the social and spatial weights matrices, the above model was run for each year using MATLAB 7.7.0 and the LeSage Econometrics Toolbox, available online. The environmental control variables were also included, related to the local *bari* environment and potential relation to cholera outcome.

Results

At the 1000-meter neighborhood level, significant spatial clustering occurred for every year with the exception of 1990 (Table 1). This indicates that *bari*s located within the same neighborhood tended to have similar rates of cholera during the majority of the years examined. In contrast, running the Moran's I for the social network across the twenty-one years yielded significant clustering only in 1989, 1993, and 2000 (Table 2). This indicates that *bari*s socially connected in those years had comparable rates, but only during certain periods within the dataset.

The data was then analyzed using the combined social effects/spatial error model, the results of which are presented in Table 3 and Figure 2. Certain environmental control variables, as described above, were also found to be significant in certain years and are noted if so. The spatial error, presented first, was significant for every year at the $p<0.05$ level. The parameter represents the extent to which the clustering of cholera rates not explained by measured independent variables--nor the social effect--can be accounted for by the clustering of the error term. In other words, unaccounted-for variables related to similarity within the local environment were significant in all years.

When this underlying spatial error was controlled for, the social effects parameter was significant at the $p<0.05$ level for five out of twenty-one years, specifically for 1983, 1985, 1989, 1993, and 2000. The values represent the extent to which cholera rates clustered in the network; the lower coefficients are a result of the small number of overall cholera cases given the population size. The values of the social effect do not exhibit consistent increases or decreases over time; in fact, when viewed graphically in Figure 2, they show an overall decreasing, wave-like pattern, mirrored somewhat by the spatial error. Of the environmental control variables, results showed varying levels of significance for distances to geographic features or tubewell depth during certain years. These are also listed in Table 3.

Year	**z-score and probability: clustering of cholera rates**	
1983	8.61	p<0.0001
1984	8.53	p<0.0001
1985	13.14	p<0.0001
1986	17.29	p<0.0001
1987	5.97	p<0.0001
1988	8.02	p<0.0001
1989	4.01	p<0.0001
1990	1.58	p=0.1146
1991	4.37	p<0.0001
1992	6.20	p<0.0001
1993	18.15	p<0.0001
1994	9.63	p<0.0001
1995	2.66	p=0.0079
1996	6.06	p<0.0001
1997	11.28	p<0.0001
1998	4.16	p<0.0001
1999	8.82	p<0.0001
2000	6.61	p<0.0001
2001	3.61	p=0.0003
2002	3.76	p=0.0002
2003	8.42	p<0.0001

Table 4.1: Significance of spatial clustering of cholera rates from Moran's I model, 1000-meter neighborhoods.

Year	z-score and probability: clustering of cholera rates	
1983	0.27	p=0.397
1984	1.55	p=0.061
1985	1.51	p=0.066
1986	0.04	p=0.485
1987	0.19	p=0.423
1988	0.02	p=0.493
1989	8.49	p<0.0001
1990	0.02	p=0.493
1991	-0.26	p=0.604
1992	-0.43	p=0.666
1993	14.61	p<0.0001
1994	-0.04	p=0.518
1995	0.06	p=0.475
1996	-0.11	p=0.455
1997	0.56	p=0.289
1998	-0.05	p=0.482
1999	-0.04	p=0.484
2000	5.31	p<0.0001
2001	0.13	p=0.450
2002	-0.42	p=0.338
2003	-0.01	p=0.495

Table 4.2: Significance of social clustering of cholera rates from Moran's I model.

Year	Spatial Error	Significance	Social Effect	Signifiance	Other Variables
1983	0.47	***	0.01	***	Pond***, Tubewell*
1984	0.48	***	0.04		Road*
1985	0.58	***	0.05	***	Road*, Pond*
1986	0.61	***	-0.01		Road***
1987	0.54	***	0.00		
1988	0.53	***	0.00		
1989	0.46	***	0.28	***	
1990	0.34	***	0.00		
1991	0.42	***	-0.01		
1992	0.52	***	-0.01		Pond*, Tubewell*
1993	0.56	***	0.15	***	

1994	0.53	***	-0.01		
1995	0.34	***	-0.01		
1996	0.45	***	0.00		Tubewell depth***
1997	0.54	***	0.03		
1998	0.42	***	-0.01		River*, tubewell depth***
1999	0.57	***	-0.01		
2000	0.54	***	0.14	***	
2001	0.43	***	0.00		Road***
2002	0.45	***	-0.01		Tube*
2003	0.51	***	-0.01		

Table 4.3: Spatial error and significance, social effect and significance, and additional environmental variables that affect cholera rate.

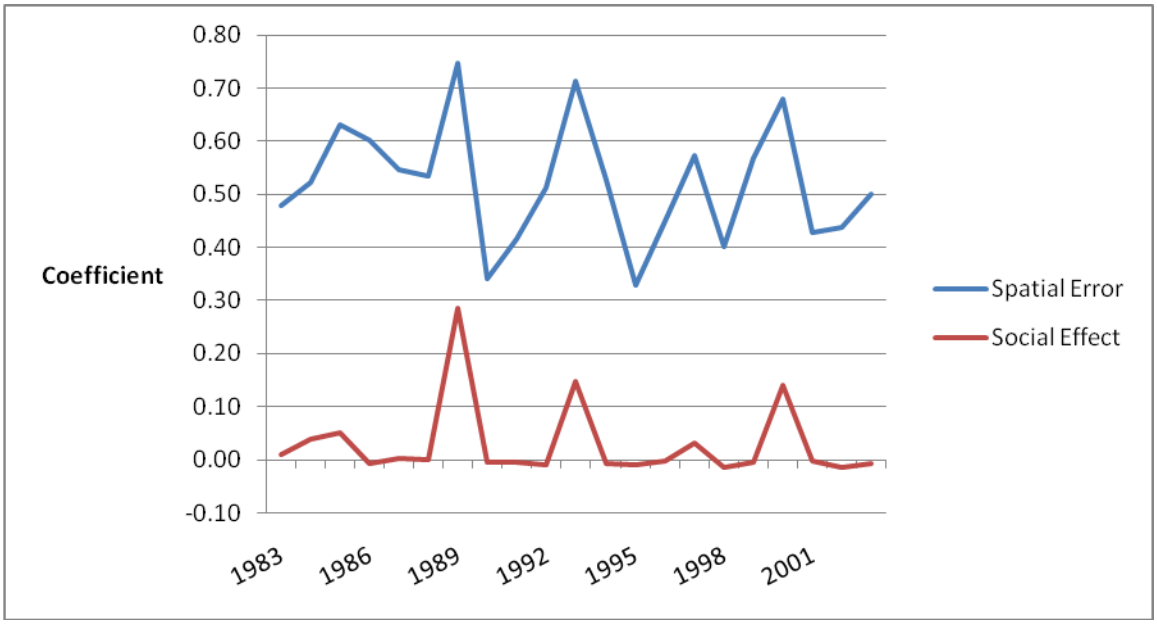

Figure 4.2: Spatial error and social effect of cholera clustering by year. For significant social effect values, see Table 1.

Discussion

When comparing cholera rates across *bari*s at the 1,000-meter neighborhood level, the results of this analysis suggest that *bari*s within this distance of one another are at similar levels of risk,

regardless of year. However, social clustering only appeared during three out of the twenty-one years. From this, one could infer that during those periods, there was some process related to the social network operating to facilitate cholera transmission. When considering the consistent spatial clustering of cholera over time, however, it can be argued that this social clustering—measured aspatially using the Moran's I—is in fact a product of socially-connected *bari*s also being located close to one another in space. Therefore, underlying spatial and environmental factors may be driving the similarity in cholera rates, as opposed to social network effects.

Using the social effects-spatial disturbances model is a potential way to account for the spatial autocorrelation of omitted predictor variables, or the autocorrelation of the error term, given the network used and the definition of the spatial weights matrix. When applied, the significance of the three years of social clustering identified in the Moran's I remained, and two additional years were revealed as exhibiting a significant social effect. This can be used to infer that within those five years, processes somehow related to kinship-based social networks affected cholera transmission, or produced similar cholera rates in *bari*s. This may be the result of actual physical transmission of the pathogen, via cultural vectors and direct person-to-person transmission. It may also be due to similar behaviors across *bari*s that are related to one another, and these practices either increasing or decreasing collective risk. Examples would include household sanitation, water storage, and so forth.

Significant spatial error parameters estimated by the model for all years, upon first interpretation, suggest the importance of presently unidentified spatial components in producing common cholera rates among socially connected *bari*s. These components may include known risk factors identified in previous literature, such as population density or proximity to a failing latrine. However, there is an additional important potential variable related to the spatial error that is non-environmental: the social relationships and interactions between individuals living in spatially proximate *bari*s not related using the kinship network. In other words, the spatial error parameter may "capture" those social interactions not included by the network definitions used here. Individuals will interact to some degree with their neighbors, if perhaps not even more frequently than with kin.

Whereas separating out these non-kinship networks is difficult without advanced survey methods, it is perhaps possible to use the available spatial information to produce a measure of predicting social interaction. As opposed to Euclidian distance buffers, which may encompass environmental features that significantly limit social interaction, road networks or other facilitators of population mobility may be used as potential predictors of social interaction.

Certain additional considerations apply when interpreting these results. Firstly, the weights matrices used may have affected the findings significantly, as parameter estimates are based on specification of either matrix *W* (Leenders, 2002). The social matrix uses only a binary variable for either absence or presence or a social relationship. All relations that exist are given an equal value of 1 prior to standardization. Based on different theories of social influence, shared behavior, and interaction as related to social networks, a social weights matrix can be constructed in a variety of ways. Here a rather simplistic approach was used; alternatives would include weighing by number of connections (i.e. migrations) between two *bari*s or by number of steps connecting the two *bari*s (e.g., *bari*s not directly related but sharing a connection with another *bari* would be given a value, such as 2). Furthermore, spatial distance, or spatial adjacency, can be represented in a variety of ways beyond the one-kilometer distance band. Varying distances may produce different results, as would using a weights matrix using absolute spatial distance between *bari*s. Alternatively, Thiessen polygons based on *bari* point locations could be used to form a joint-boundary defined spatial weights matrix.

Another consideration is temporal in nature; that is, using years as units may not capture effects related to cholera transmission at a seasonal or monthly scale. Matlab normally features two seasonal cholera epidemics a year, one pre- and one post-monsoon. Certain theories argue that different forms of transmission drive the two. Using a finer time scale may reveal different patterns in the data. Additionally, the prediction of the dependent variable does not consider network ties and cholera cases from previous years, which may be significant. A longitudinal analysis would be appropriate and could be developed for many different time frames.

Beyond the data and model parameters, it is also critical to look at larger-scale processes that may have driven cholera levels and clustering in the data, such as those related to politics and environment, or the emergence of disease strains and country-wide epidemics. All of these processes may partly be driving results. For example, major floods throughout Bangladesh in 1987-88 and 1998, potentially increasing frequency of human contact with the pathogen through contaminated surface waters and other vehicles of transmission. The years 1992 and 1993 also saw the emergence of a new cholera strain which caused a large-scale epidemic in Bangladesh as well as neighboring countries; another epidemic of the same strain occurred in 2002. Factors such as these need to be considered when making inferences based on these results, as they may explain sudden increases in clustering and rates overall.

Conclusion

This research demonstrates that, using social networks and accounting for spatial autocorrelation, social effects can be isolated as a cause of similar cholera rates between *baris*, but that spatial and neighborhood-level factors are perhaps of greater importance due to their persistent effect over the years. Development of targeted intervention methods will be most effective having greater insight as to which form is most prevalent; therefore understanding the extent to which secondary transmission occurs as a result of social networks is important for public health campaigns and education.

References

Acosta, J.C., Galindo, C.M., Kimario, J., Senkoro, K., Urassa, H., Casals, C., Corachán, M., Eseko, N., Tanner, M., Mshinda, H., Lwilla, F., Vila, J., and Alonso, P.L. 2001. Cholera outbreak in southern Tanzania: risk factors and patterns of transmission. *Emerging Infecrious Diseases (Supplement 3):* 583–587.

Ali, M. Emch, L. von Seidlein, M. Yunus, D.A. Sack, M. Rao, J. Holmgren and J.D. Clemens. 2005. Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis, *Lancet*, 366 (9479) (2005), pp. 44–49.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic.

Aral et al. 1999. Sexual Mixing Patterns in the Spread of Gonococcal and Chlamydial Infections. *American Journal of Public Health* 89, 6, 825-832.

Auerbach, D.M., W.W. Darrow, H.W. Jaffe, and J.W. Curran. 1984. Cluster of Cases of the Acquired Immune Deficiency Syndrome: Patients Linked by Sexual Contact. *The American Journal of Medicine*, 76(3):487-92.

Bates, S.J., Trostle, J., Cevallos, W.T., Hubbard, A., and Eisenberg, J.N.S. 2007. Relating Diarrheal Disease to Social Networks and the Geographic Configuration of Communities in Rural Ecuador. *American Journal of Epidemiology*, 166(9): 1088-1095.

Bell, D.C., Atkinson, J.S., and Carlson. J.W. 1999. Centrality measures for disease transmission networks. *Social Networks*, 21(1): 1-21.

Birmingham, M.E., Lee, L.A., Ndayimirije, N., Nkurikiye, S., Hersh, B.S., and Wells, J.G. 1997. Epidemic cholera in Burundi: patterns of transmission in the Great Rift Valley Lake region. *The Lancet*, 349: 981–85.

Black, R.E., Merson, M.H., Huq, I., Alim, A.R.M.A., Yunus, M.D., 1981. Incidence and severity of rotavirus and Escherichia coli diarrhoea in rural Bangladesh. *The Lancet*, 1:141-143.

Cardenas, V., Saad, C., Varona, M., and Linero, M. 1993. Waterborne cholera in Riohacha, Colombia, 1992. *Bulletin of the Pan-American Health Organization*, 27(4): 313-30.

Colwell, R.R. 1996. Global Climate and Infectious Disease: The Cholera Paradigm. *Science*, 274(5295): 2025-2031.

Colwell, RR, Kaper, J, and Joseph, SW. 1977. Vibrio cholerae, Vibrio parahaemolyticus, and other vibrios: occurrence and distribution in Chesapeake Bay. *Science*. 198:394.

Colwell, RR, Tamplin, ML, Brayton, PR, Gauzens, AL, Tall, BD, Herrington, D, Levine, MM, Hall, S, Huq, A, and Sack, DA. Environmental aspects of Vibrio cholerae in transmission of cholera. In: *Advances in research on cholera and related diarrhoea*, 7th ed Tokyo, KTK Scientific Publishers. 1990. pp. 327–343.

Cook, V.J., Sun, S.J., Tapia, J., Muth, S.Q., Arguello, D.F., Lewis, B.L., Rothenberg, R.B., and McElroy, P.D. 2007. Transmission network analysis in Tuberculosis contact investigations. *Journal of Infectious Diseases* 196: 1517-27.

Craig, M. 1988. Time-space clustering of vibrio cholerae 01 in Matlab, Bangladesh, 1970-1982. *Social Science and Medicine*, 26 (1): 5-13.

D'Souza, S., 1981. A population laboratory for studying disease processes and mortality: the demographic surveillance system in Matlab, Bangladesh. Special Publication No. 13, ICDDR,B, Dhaka, Bangladesh.

Doreian, P. 1982. Maximum likelihood methods for linear models: Spatial effect and spatial disturbance terms. *Sociological Methods Research*, 10(3): 243-269.

Dow, M.M. 2007. Galton's problem as multiple network autocorrelation effects: cultural trait transmission and ecological constraint. *Cross-Cultural Research*, 41(4): 336-363.

Emch, M. 1999. Diarrheal disease risk in Matlab, Bangladesh. *Social Science & Medicine*, 49(4): 519-530.

Emch, M., and Ali, M. 2001. Spatial and temporal patterns of diarrheal disease in Matlab, Bangladesh. *Environment and Planning*, 33: 339-350.

Fitzpatrick, LK et al. 2001. A Preventable Outbreak of Tuberculosis Investigated through an Intricate Social Network. *Clinical Infectious Diseases*, Vol. 33, no. 11, 1801-1806.

Franco, A.A., Fix, A.D., Prada, A., Paredes, E., Palomino, J.C., Wright, A.C., Johnson, J.A., McCarter, R., Guerra, H., Morris, J.G. Jr. 1997. Cholera in Lima, Peru, correlates with prior isolation of Vibrio cholerae from the environment. *American Journal of Epidemiology*, 146(12): 1067-1075.

Fuduka, J.M, Augusta, Y., L. Chaparro, M. Campos, and E. Chea. 1995. Clinical characteristics and risk factors for *Vibfio cholerae* infection in children. *The Journal of Pediatrics*, 126(6): 882-886.

Ghani, A.C., Swinton, J., Garnett, G.P. 1997. The Role of Sexual Partnership Networks in the Epidemiology of Gonorrhea. *Sexually Transmitted Diseases* 24, 1, 45-56.

Guest, P. and Chamratrithirong, A. 1992. The social context of fertility decline in Thailand, in: Goldscheider C, ed., Fertility Transitions, Family Structure, and Population Policy, Boulder, CO, USA: Westview Press.

Gunn, R.A., Kimball, A.M., Mathew, P.P., Dutta, S.R., and Rifaat, A.H.M. 1981. Cholera in Bahrain: epidemiological characteristics of an outbreak. *Bulletin of the World Health Organization* (WHO), 59(1): 61-66.

Halloran et al. 2002. Containing Bioterrorist Smallpox. *Science* 298, 5597, 1428 – 1432.

Hanneman, R.A. 2001. *Introduction to social network methods*. Riverside, California: Department of Sociology, University of California at Riverside.

Hollinger, F., and Haller, M. 1990. Kinship and social networks in modern societies: a cross-cultural comparison among seven nations. *European Sociological Review*, 6: 103-124.

Islam, M.S., Drasar, B.S., and Bradley, D.J. 1990. Long-term persistence of toxigenic *Vibrio cholerae* 01 in the mucilaginous sheath of a blue-green alga,*Anabaena variabilis*. *Journal of Tropical Medicine and Hygiene*, 93 (2): 133–139.

Islam, M.S., Drasar, B.S., and Sack, R.B. 1993. The aquatic environment as a reservoir of *Vibrio cholerae*: a review. *Journal of Diarrhoeal Disease Research*, 11 (4): 197–206.

Jaffe, H.W., et al. 1983. National case-control study of Kaposi's sarcoma and Pneumocystis carinii pneumonia in homosexual men: part 1, epidemiologic results. *Annals of Internal Medicine*, 99(2):145-51.

Kinney et al. 2003. Roles of Religious Involvement and Social Support in the Risk of Colon Cancer among Blacks and Whites. *American Journal of Epidemiology*. 158, 11, 1097-1107.

Klovdahl, A.S. 2001. Networks and tuberculosis: an undetected community outbreak involving public places. *Social Science & Medicine*, 52(5): 681-694.

Koo, D., Aragon, A., Moscoso, V., Gudiel, M., Bietti, L., Carrillo, N., Chojoj, J., Gordillo, B., Cano, F., Cameron, D.N., Wells, J.G., Bean, N.H., and Tauxe. R.V. 1996. Epidemic Cholera in Guatemala, 1993: Transmission of a Newly Introduced Epidemic Strain by Street Vendors. *Epidemiology and Infection*, 116(2): 121-126.

Lasch et al. 1984. Cholera in Gaza in 1981: epidemiological characteristics of an outbreak. *Trans R Soc Trop Med Hyg*, 78(4):554-7.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A.S.G., and Colwell, R.R. 2000. Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proceedings of the National Academy of Sciences* 97(4): 1438–1443.

Miller, J.C., and Hyman, J.M. 2007. Effective vaccination strategies for realistic social networks. *Physica A* 386, 780–785.

Miller, C.J., Feachem, R.G., and Drasar, B.S. 1985. Cholera epidemiology in developed and developing countries: new thoughts on transmission, seasonality, and control. *Lancet*, 8423: 261-3.

Nalin, DR. 1976. Cholera, copepods, and chitinase. *Lancet*, 2 (7992): 958.

Nalin, D.R., Daya, V., Reid, A., Levine, M.M., and Cisneros, L. 1979. Absorption and growth of *Vibrio cholerae* on chitin. *Infection and Immunology*, 25 (2): 76-770.

Quick et al, 1995. Epidemic Cholera in Rural El Salvador: Risk Factors in a Region Covered by a Cholera Prevention Campaign. *Epidemiology and Infection*, 114(2): 249-255.

Ramakrishna, B. S., Kang, G., Rajan, D. P., Mathan, M., Mathan, V. I. 1996. Isolation of Vibrio cholerae O139 from the drinking water supply during an epidemic of cholera. *Tropical Medicine & International Health*, 1(6):854-858.

Ries, A.A., Vugia, D.J., Beingolea, L., Palacios, A., Vasquez, E., Wells, J.G., Baca, N.G., Swerdlow, D.L., and Pollack, M. 1992. Cholera in Piura, Peru: A modern urban epidemic. *Journal of Infectious Diseases,* 166(6): 1429-433.

Sengupta et al. 1994. Epidemiology of *Vibrio cholerae O139 with Special Reference* to Intrafamilial Transmission in Calcutta, *Journal of lnfection,* 31: 45-47.

Shapiro, R.L., Otieno, M.R., Adcock, P.M., Phillips-Howard, P.A., Hawley, W.A., Kumar, L., Waiyaki, P., Nahlen, B.L., and Slutsker, L. 1999. Transmission of epidemic Vibrio cholerae O1 in rural western Kenya associated with drinking water from Lake Victoria: an environmental reservoir for cholera? *American Journal of Tropical Medicine and Hygiene,* 60(2): 271-276.

Siddique, AK, Zaman, K, Baqui, AH, et al. 1992. Cholera epidemics in Bangladesh: 1985–1991. *J Diarrhoeal Dis Re.* 10: 79–86.
Snow, J. 1855. *On the mode of communication of cholera*. 2nd ed. London, England: J. Churchill.

Spira et al. 1980. Microbiological surveillance of intra-neighborhood El Tor cholera transmission in rural Bangladesh. *Bulletin of the World Health Organization* (WHO), 33(5):627-36.

Swerdlow et al. 1992. Waterborne transmission of epidemic cholera in Trujillo, Peru: lessons for a continent at risk. *The Lancet*, 340(8810):28-33.

Swerdlow, D.L., Malenga, G., Begkoyian, G., Nyangulu, D., Toole, M., Waldman, R.J., Puhr, D.N.D., and Tauxe, R.V. 1997. Epidemic cholera among refugees in Malawi, Africa: treatment and transmission. *Epidemiology and Infection*, 118: 207-214.

Tamayo et al. 1965. Studies of Cholera El Tor in the Philippines. *Bulletin of the World Health Organization (WHO)*;33(5):645-649.

Tauxe et al. 1988. Epidemic Cholera in Mali: High Mortality and Multiple Routes of Transmission in a Famine Area. *Epidemiology and Infection*, 100(2): 279-289.

Valente, T.W., Watkins, S.C., Jato, M.N., A. Van Der Straten, and L.M. Tsitsol. 1997. Social Network Associations with Contraceptive Use Among Cameroonian Women in Voluntary Associations. *Social Science and Medicine*, 45, no. 5, 677-687.

Wasserman, S. and Faust, K. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Weber et al. 1994. Epidemic Cholera in Ecuador: Multidrug-Resistance and Transmission by Water and Seafood. *Epidemiology and Infection*, 112(1): 1-11.

Wylie, J.L., Shah, L., and Jolly, A. 2007. Incorporating geographic settings into a social network analysis of injection drug use and bloodborne pathogen prevalence. *Health and Place*. 2007 Sep;13(3):617-28.

Chapter 5

Conclusion

This research introduces new methods to understand pathogen transmission, within the theoretical framework of disease ecology. Social network analysis is an innovative tool that has proved useful in examining the diffusion of infectious diseases amongst actors connected by some form of relationship, mainly physical and/or sexual. However, networks can also be used to predict transmission along ties that involve social interaction, such as kinship or friendship. This is appropriate for diseases that spread via more complex means, such as respiratory or water- and food-borne illnesses. Using network methods to examine cholera risk is an innovative approach to a disease that has multiple forms of transmission. In particular, transmission from person-to-person is known to occur in various ways, primarily through contaminated food or water, but has never been predicted using networks. In this study, rates of cholera amongst *baris* connected through kinship ties were compared to identify potential clustering. Similar rates would indicate that these networks may play a role in facilitating transmission, perhaps through behavioral factors such as food-sharing, physical interaction, and common sanitation practices. Understanding the extent to which these behaviors increase cholera risk can then aid intervention efforts.

The two analyses performed in this study considered social networks as routes of transmission, but also identified the role of spatial and neighborhood-level factors, as these are potentially significant for cholera risk as well. The results strongly supported that local environmental factors are important, which was expected. Individuals living near features such as failing latrines and water sources more likely to harbor the cholera pathogen may encounter greater risk. High population density and demographic factors such as socioeconomic status, which likely clusters in space, are also

relevant at the neighborhood level and thus may explain much of the autocorrelation of cholera that was detected here. These variables were not specified in the first analysis, but some were identified in the second based on available data and were significant only during certain years. Therefore, additional unknown factors somehow caused similar rates of cholera across *baris*. Though the idea of environmental risks is stressed here, it is entirely plausible that relationships with non-kin within the neighborhood are also a cause, whether it be through contamination of a communal water source by a carrier, or simply household visits which make transmission from person to person possible. Unfortunately, one can only speculate about the role of these networks in this context, as data on non-familial networks usually requires thorough interview and survey processes. At smaller community levels, this is feasible and thus may be possible in the future for smaller villages within the larger Matlab study area.

The same technique to test for concentration of cholera was applied to a social network created for the purpose of this research, and potentially for future studies as well. Kinship network-based clustering was strongly significant in certain years and remained so despite unknown spatial effects. From these results, one can gather that either additional unknown factors unrelated to space account for some or all of the clustering, or that social networks are indeed an important consideration for cholera prevention efforts. One explanation for the former is that *bari*s related to one another maintain similar customs in terms of hygiene, sanitation, and food preparation. Such common behaviors would place households at risk, but not necessarily due to transmission from non-household members. However, if the latter is indeed true, then members of the common network with cholera may place others at risk through interactions within and outside of the *bari*. This may occur through visits, but would also include social interactions that occur beyond the household.

The actual mechanisms at work are complex and beyond the scope of this research. However, illustrating that cholera clusters in *bari*s connected by social interaction as well as kinship and, most likely, similar behavioral and demographic characteristics, informs understanding of secondary transmission. The difference between primary and secondary transmission is often vague. However,

networks as predictors of risk fall clearly under the latter form, and thus help in developing a more focused knowledge of how other humans specifically affect the risk of developing a disease that is thought to be primarily contracted through the environment.

From a disease ecology perspective, considering environment, population, and behavior are critical to truly understanding a health outcome. The integration of spatial and social effects, as done here, is a new method for examining cholera transmission in a way that has not yet been used in the literature. Though environmental concentration of the disease is clearly a significant factor, the dynamics of social network transmission are worth examining further, for the purposes of understanding what behaviors drive infection.