Relationships between flood control and cholera in Matlab, Bangladesh

Margaret A. Carrel

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

Chapel Hill
2007

Approved by:

Michael Emch

Melinda Meade

Aaron Moody

Abstract

Margaret Carrel: Relationships between flood control and cholera in Matlab, Bangladesh
(Under the direction of Michael Emch)

Implementation of flood control strategies has been empirically associated with rises in disease rates in the developing world. This research examines the impact of flood protection measures on cholera incidence among a rural Bangladeshi population. Using longitudinal health and demographic data collected over 21 years, analysis of clustering patterns and statistical relationships between cholera incidence and environmental factors was conducted for timeframes prior to and following the introduction of flood control in Matlab, Bangladesh. Results indicate that alteration of normal flooding patterns both temporally and spatially shifted cholera occurrence within the study area, and that these shifts demonstrate further differentiation when information on cholera strain is included in the analysis. These outcomes suggest that introducing flood protection to rural Bangladesh will have significant but complex effects on cholera incidence patterns.

Acknowledgements

I am indebted to my friends and family, whose excitement and encouragement keep me moving, and to my colleagues at the University of North Carolina-Chapel Hill and the International Centre for Diarrhoeal Disease Research, Bangladesh who give me the skills and the knowledge to do what I love.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Abbreviations

HDSS	Health & Demographic Surveillance System
ICDDR,B	International Centre for Diarrhoeal Disease Research, Bangladesh
MDIP	Meghna-Dhonagoda Irrigation Project
O1	O1 serogroup *Vibrio cholerae*
O139	O139 serogroup *Vibrio cholerae*
OLS	Ordinary Least Squares regression

Chapter 1

Introduction

Diarrheal diseases are responsible for the deaths of over 2 million people annually, representing 4% of total worldwide mortality, and disproportionately affect the poverty-stricken (World Health Organization, 2006). The agents of diarrheal diseases vary from viruses to bacteria to amoebas. This wide-ranging group of causative agents, all with a single common and deadly symptom, makes spatial distribution of diarrheal disease incidence a fascinating and complex topic of study. Each disease has its own particularities, including variables such as vastly different infective doses, temperature of water preferred, populations affected, and seasonality. Bangladesh is a prime area to study spatial patterns associated with these diseases because it is a country where millions of people live in close proximity not only to other people, but also to open and unsafe water sources. It is also a country that is actively engaged in alteration of its aquatic ecosystems, a process often associated with changed disease ecologies.

Bangladesh is subject to both annual and abnormal flooding. The country sits at the confluence of three major Asian rivers, the Ganges, Brahmnaputra and Meghna. These rivers have a catchment area of over 1.5 million km^2, more than 11 times the size of Bangladesh itself. This large catchment, when combined with the annual monsoon rains, means that in a normal flood year over 20% of the country may be underwater (Thompson, 1995). In bad years, such as in 1988, more than half the country can be flooded. In response to both annual and unexpected devastation due to flooding, the Government of Bangladesh has implemented a series of flood control programs. These programs usually entail construction of embankments along the country's larger rivers, with the hope that such structures will protect the enclosed populace.

During the creation and implementation of successive water management plans, consideration was given by the government to possible impacts on agricultural land availability and production, economic benefit, gender equality and other social factors, and flood control capabilities (Centre for Water Policy and Development , 2001). Much less attention was paid, however, to the possible impacts of embankments on the transmission of water-related diseases, such as cholera. Such an oversight, given the continuing presence of these diseases among Bangladeshis, is striking. This proposal seeks to determine what, if any, impact flood control embankments have had on the patterns of incidence rates of cholera in Matlab, Bangladesh.

Multiple empirical studies have examined the impact of water construction projects on communicable disease rates. With the exception of onchocerciasis, whose vector habitat is fast moving streams, the rates of nearly every disease whose host or vector depend on standing water sources increase when water resource management programs are implemented by governments or NGOs (Hunter, 2003; Ali et al., 2002a; Emch, 1999; Hunter et al., 1982; Keiser et al., 2005; Singh et al., 1999; Sow et al., 2002; Waddy, 1975). This is true in the developing world as well as in the United States. Hughes & Hunter (1970) argue that programs which alter man's environment result in the formation of new 'ecological contracts,' contracts that typically have hidden or unanticipated costs. As Sow et al. (2002) contend, "feasibility studies mainly emphasize the economic benefits rather than the environmental and health hazards of water-resources developments." The impacts of water management strategies, which often require but do not receive long-term upkeep and continued investment, must be examined longitudinally, as their effects will be felt for decades.

In Matlab there is a unique opportunity to conduct just such a longitudinal analysis. Since 1966 the resident population has been under demographic and health surveillance conducted by the International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B). In Matlab in 1989 a Government of Bangladesh project was completed under the auspices of the Flood Action Plan, the Meghna-Dhonagoda Irrigation Project (MDIP), which introduced flood protection to approximately half the study area. The ICDDR,B data spans the pre- and post-MDIP time periods and both the protected and unprotected populations, allowing us to examine how the completion of the MDIP may have impacted the cholera experience of Matlab's residents. The overall research questions are 1) Does intra- and inter-area spatio-temporal variation in cholera incidence exist between the flood protected and non-protected areas of Matlab? 2) Can this variation be attributed to the implementation of flood protection? To address these questions, two separate, but interrelated, analyses were conducted. The structure of this thesis is as follows: an overall introductory chapter, a second chapter outlining the first analysis, a brief bridge chapter, a fourth chapter describing the second analysis and lastly an overall conclusion chapter.

The first analysis examined over a 21-year timeframe (1983-2003) whether cholera incidence in Matlab clustered in both space and time and whether this clustering behavior had changed with the introduction of flood protection. In addition to flood protection status, other variables that could possibly modify the cholera incidence relationship, such as seasonality of incidence and cholera type, were examined via cluster analysis.

The second analysis examined how the strength of two environmental variables, flood protection status and river proximity, in explaining cholera outcomes had changed since the introduction of flood protection to Matlab. The analysis stratified the interactions according to season and cholera strain. Changes in the correlative relationships between flood protection status and river proximity on cholera incidence were explored at the individual *bari* level and then at the neighborhood level, to examine whether the impact of the MDIP construction varied according to scale.

Background

Since the 1950s, flood control and water management have been central issues to first the East Pakistani government and later the Bangladeshi government. Starting with the 20-year Water Master Plan in 1964, and continuing through today, heavy emphasis was placed on the construction of embankments around the country (Rogers, Lydon, Seckler & Pitman, 1994). Initially, water management schemes were top-down oriented. There was little or no public input on decisions that had a very tangible effect on the daily lives of Bangladeshis. Water plans since 1999, however, have involved greater public participation and worked to address issues such as gender equity, social justice and environmental awareness (Centre for Water Policy and Development, 2001). Yet, there is still little consideration given to the potential implications for disease.

Among the thousands of kilometers of embankments that have been constructed is the Meghna-Dhonagoda Irrigation Project (MDIP), built in 1987-8 and located in Matlab. The

MDIP, completed between two particularly sizeable flood years for Bangladesh, consists of a 60km ring embankment, irrigation and drainage canals, culverts, bridges and two pumping stations (Emch, 2000). Matlab is a rural region, located ~50km southeast of Dhaka where the Ganges and Meghna rivers join to form the Lower Meghna. Running from north to south through Matlab is the Dhonagoda River (Figure 1). The MDIP embankment along the Dhonagoda divides Matlab into two parts, one which experiences the seasonal flooding of the Dhonagoda and one which is typically protected. The protected area makes up about 40% of Matlab's 184km^2 (Ali et al. 2002a).

The population of Matlab is over 200,000 persons, with a population density of ~1000 persons per square kilometer (Ali et al. 2002a). The people of Matlab live within a structure of *baris*, or patrilineal household groupings. Since 1966 the population of Matlab has been under demographic and health surveillance by the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). Each resident of Matlab is assigned an identification code, as is their household, and twice a month each household is visited by a trained community health worker who records demographic information such as births, deaths, marriages and illnesses (Emch, 1999). Health information about the population of Matlab is gathered both at ICDDR,B's hospital and treatment centers and by the community health workers who visit each *bari*. Surveillance data shows that diarrheal and other infectious diseases are highly endemic to the population.

Cholera

The bacterium *Vibrio cholerae (V. cholerae)* is the agent responsible for cholera and is a natural part of the aquatic environment of Bangladesh (Faruque et al. 2005; Jensen et al.

2006). There are multiple types of *V. cholerae* that cause epidemic cholera. Until 1992, all cholera cases in Matlab were caused by the O1 serogroup and its two biotypes, Classical and El Tor. Classical *V. cholerae* was responsible for the majority of cholera between 1966 and 1972, El Tor predominated from 1973-1982, and then Classical and El Tor cocirculated until 1988 when Classical disappeared. In 1993 a new serogroup, O139 was identified. O139 and El Tor have cocirculated since that time (Longini et al., 2002).

Exposure to the *V. cholerae* bacteria, of any type, often does not result in infection and symptoms, as the infective dose is approximately 1 million bacteria. Exposure can either be primary, through contact with *V. cholerae* in aquatic reservoirs, or secondary, through fecal-oral contamination from an infected individual. When infection does occur, treatment with antibiotics and oral rehydration therapy are effective 99% of the time. For individuals who do not receive treatment the death rate can reach 60% (Huq et al., 2005).

Cholera in Bangladesh exhibits high seasonality, with large outbreaks in September through December, at the end of monsoon season. A smaller outbreak occurs in April, just prior to the monsoons (Faruque et al., 2005; Longini et al., 2002). To help explain this seasonality, cholera studies have both examined the genetics of the bacterium and focused on environmental determinants such as water temperature, rainfall amount, and plankton levels. Recent studies also suggest that the presence or absence of bacteriophages in water may impact numbers of *V. cholerae*, and thus the infectivity of water supplies (Faruque et al., 2005; Jensen et al., 2006).

Several studies have been undertaken to determine risk factors associated with cholera in Matlab. Ali et al. (2002a) found an association between cholera and proximity to surface water bodies. Emch (1999) describes six factors that are statistically significant in cholera transmission: multiple households using common latrines, living within a flood-controlled area, high-density use of tubewells, population within a *bari*, areal size of *bari* and local neighborhood population density. Myaux et al. (1997) found a correlation between cases of "cholera-like" diarrhea and parental education status, population density and use of sanitary latrines. Ali et al. (2002b) also found that educational status and living within a flood-controlled area were risk factors for cholera.

Theoretical Framework

The field that supplies the theoretical and conceptual frameworks for this study is medical geography. Spatial and ecological analysis of disease incidence is foundational to the field, and practitioners draw on information and knowledge from multiple fields, including epidemiology, sociology, hydrology, and biology. Medical geographers such as Jacques May (1958) and John Hunter (1974) argued that a disease does not exist independently of an environment and a host, and that a comprehensive understanding of illness must address these factors in addition to the characteristics of the illness itself. Cholera, or rather the adverse effects of infection by *V. cholerae*, exists in a *person* at a *place*. Two fields of literature in particular will prove especially useful in the proposed study. The first views human health as the result of a series of intricate interactions, represented graphically as the triangle of human ecology. The second represents a history of empirical studies that examine how alterations to environments, especially aquatic ecosystems, change or sustain disease interaction and systems.

The Triangle of Human Ecology

Dubos (1987) describes the "process of living" as an interaction between humans' internal and external environments. Melinda Meade (1977) proffers an alternative to this dualistic view with the more comprehensive and versatile triangle of human ecology (Figure 2). The triangle's three vertices are culture, environment and population, and the interplay between the three provides a foundation for integrating and analyzing factors that contribute to disease ecologies.

Culture in this framework is used to mean not only observable aspects of behavior, such as dietary preferences and house type, but also perceptions of reality and understanding, for instance the perceived risk of cholera in impounded water. The environment point of the triangle refers not to an exhaustive study of every aspect of the surrounding world, but rather a systemic understanding of a daily habitat. Population within this framework is a different category than for anthropologists and sociologists. In Meade's triangle it denotes the characteristics of people as a biological organism (age, sex and genetics) which directly impact their interaction with culture and environment, rather than people as social beings. The flexibility and utility of Meade's triangle is that it is not static, a temporal aspect may be applied to the interactions the triangle describes. This framework can thus be utilized to examine a changing human disease ecology through time.

The holistic view proffered by Meade's triangle is essential in medical geography. While the nature of infective agents do impact the way the disease is experienced, of greater import is the way an individual lives their life, their choices and habits and daily environmental interactions. To truly understand a disease pattern, you must study people and aspects of their lives (Dubos, 1965).

Environmental Alteration & Disease Systems

In *Mirage of Health*, Rene Dubos writes that any modification of nature's balances, large or small, comes with consequences because of the complexity of interrelationships in the natural world. He cautions against hasty or ill-considered changes to an ecosystem that could have ramifications for the health of the surrounding populace. Charles Hughes and John Hunter (1970) state in "Disease and 'Development' in Africa" that programs that alter man's environment, be it population movement or dam construction, result in the formation of a new "ecological contract," one which typically has hidden costs.

One of the most powerful ways that humans have taken control of their environment and altered it to suit their needs is through development and control of water resources. Multiple studies have examined the impacts of water management systems on disease in Africa and Asia (Ali et al., 2002; Emch, 1999; Hunter et al., 1982; Keiser et al., 2005; Singh et al., 1999; Sow et al., 2002; Waddy, 1975). Hunter, Rey & Scott (1982) advise that the best way to determine the health impact of changes to water systems is to directly compare disease data pre- and post-development using both quantitative and qualitative methods. If this is not possible, an alternative method is to evaluate disease within the bounds of the developed area against a similar region outside the bounds. They go on to argue that increases in parasitic and infectious diseases due to ecological interference in water systems are empirically predictable.

Human imprint on the earth and its systems is ever-increasing, with short-term anthropocentric organization replacing long-term, finely tuned and delicate systems (Farvar, 1973). As Audy (1961) puts it, with a dangerous combination of new methods and an ignorance of the possible consequences, new niches for disease can be created with increasing frequency. Although the way humans experience disease is being dramatically altered by current environmental change, such negative impacts are not inevitable. The costs to human health must be calculated in the planning of future human/environment interactions and negative impacts mitigated.

Studying the impact of the MDIP on cholera in Matlab can provide policymakers with information on how introducing flood protection measures in poor, rural and flood-prone areas can change waterborne diarrheal disease incidence. Analyzing cluster patterns of cholera prior to and after the introduction of flood protection is one method for determining whether this environmental modification has significant impacts on disease occurrence.

Figures

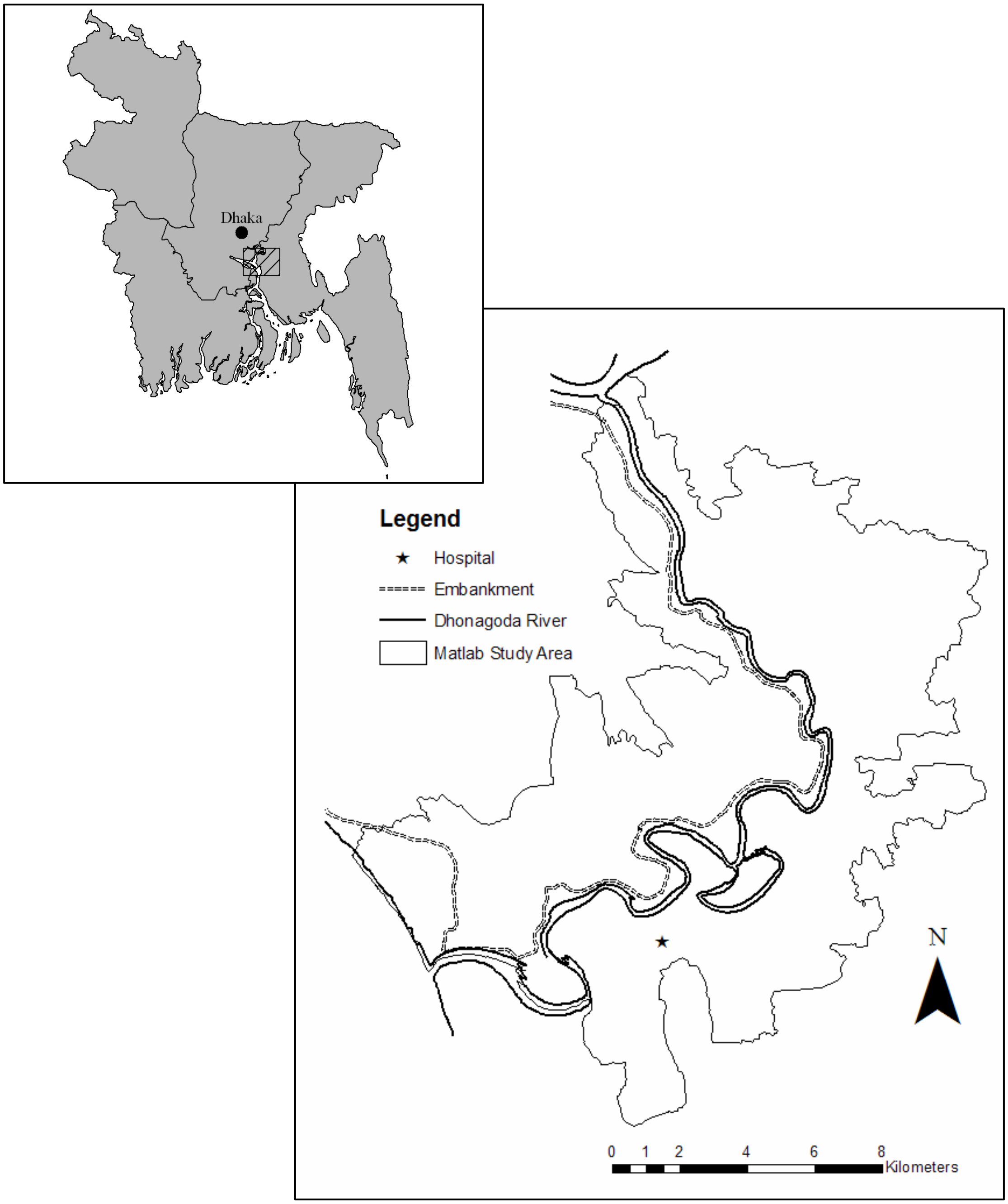

Figure 1.1: Location and Main Physical Features of Matlab, Bangladesh

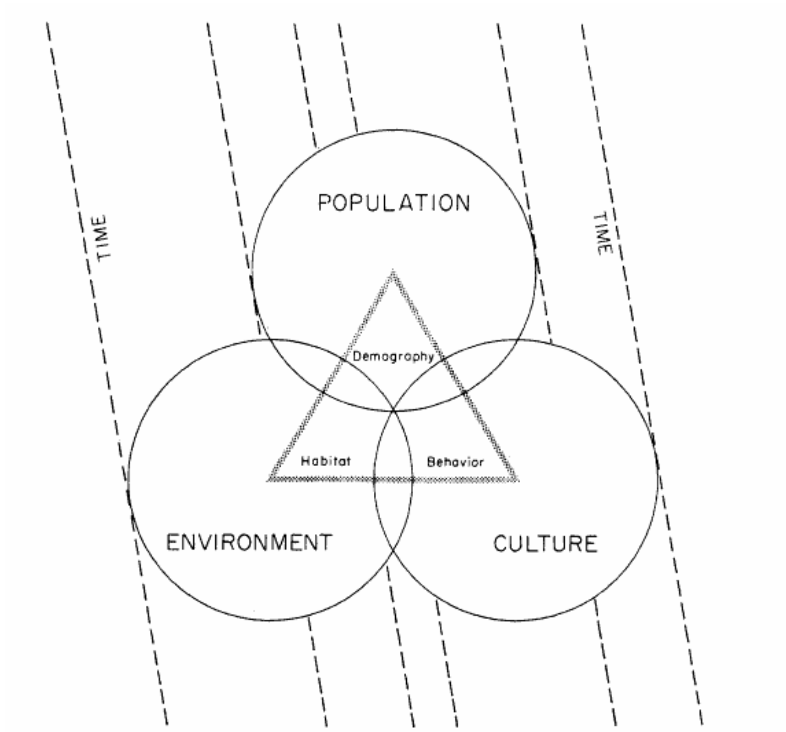

Figure 1.2: Triangle of Human Ecology (Meade, 1977)

References

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 (1982), 55*(6), 1015-1024.

Ali, M., Emch, M., von Seidlein, L., Yunus, M., Sack, D. A., & Rao, M. et al. (2005). Herd immunity conferred by killed oral cholera vaccines in Bangladesh: A reanalysis. *Lancet, 366*(9479), 44-49.

Ali, M., Emch, M., Yunus, M., & Sack, R. B. (2001). Are the environmental niches of vibrio cholerae O139 different from those of *Vibrio cholerae* O1 El tor? *International Journal of Infectious Diseases : IJID : Official Publication of the International Society for Infectious Diseases, 5*(4), 214-219.

Audy, J. R. (1961). Man-made maladies. In S. m. Farber, & R. H. L. Wilson (Eds.), *The air we breathe: A study of man and his environment* (pp. 100-112). Springfield, Illinois: Charles C. Thomas.

Centre for Water Policy and Development (CWPD). (2001). *Water resource management in Bangladesh: A policy review*. Leeds, UK: School of Geography, University of Leeds.

Dubos, R. J. (1987). *Mirage of health : Utopias, progress and biological change*. New Brunswick: Rutgers University Press.

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

Emch, M. (2000). Relationships between flood control, kala-azar, and diarrheal disease in Bangladesh. *Environment and Planning A, 32*, 1051-1063.

Emch, M. (1999). Diarrheal disease risk in Matlab, Bangladesh. *Social Science & Medicine (1982), 49*(4), 519-530.

Faruque, S. M., Naser, I. B., Islam, M. J., Faruque, A. S., Ghosh, A. N., & Nair, G. B. et al. (2005). Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proceedings of the National Academy of Sciences of the United States of America, 102*(5), 1702-1707.

Farvar, M. T., Milton, J. P., & Washington University . Center for the Biology of Natural Systems. (1972). *The careless technology: Ecology and international development*. Garden City, N.Y.: Natural History Press.

Hughes, C. C., & Hunter, J. M. (1970). Disease and "development" in Africa. *Social Science & Medicine (1982), 3*(4), 443-493.

Hunter, J. M. (1974). The geography of health and disease. In J. M. Hunter (Ed.), *The challenge of medical geography: Studies in geography no. 6*. Chapel Hill: Department of Geography, University of North Carolina.

Hunter, J. M. (2003). Inherited burden of disease: Agricultural dams and the persistence of bloody urine (*Schistosomiasis hematobium*) in the upper East region of Ghana, 1959-1997. *Social Science & Medicine (1982), 56*(2), 219-234.

Hunter, J. M., Rey, L., & Scott, D. (1982). Man-made lakes and man-made diseases. towards a policy resolution. *Social Science & Medicine (1982), 16*(11), 1127-1145.

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., & Ali, A. et al. (2005). Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Applied and Environmental Microbiology, 71*(8), 4645-4654.

Jensen, M. A., Faruque, S. M., Mekalanos, J. J., & Levin, B. R. (2006). Modeling the role of bacteriophage in the control of cholera outbreaks. *Proceedings of the National Academy of Sciences of the United States of America, 103*(12), 4652-4657.

Keiser, J., De Castro, M. C., Maltese, M. F., Bos, R., Tanner, M., & Singer, B. H. et al. (2005). Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *The American Journal of Tropical Medicine and Hygiene, 72*(4), 392-406.

Longini, I. M.,Jr, Yunus, M., Zaman, K., Siddique, A. K., Sack, R. B., & Nizam, A. (2002). Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *The Journal of Infectious Diseases, 186*(2), 246-251.

May, J. M. (1958). *The ecology of human disease: Studies in medical geography*. New York: MD Publications.

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

Myaux, J., Ali, M., Felsenstein, A., Chakraborty, J., & de Francisco, A. (1997). Spatial distribution of watery diarrhea in children: Identification of "risk areas" in a rural community in Bangladesh. *Health & Place, 3*(3), 181-186.

Rogers, P., Lydon, P., Seckler, D., & Pitman, G. T. K. (1994). *Water and development in Bangladesh: A retrospective on the food action plan*. Arlington, Virginia: Irrigation Support Project for Asia and the Near East.

Singh, N., Mehra, R. K., & Sharma, V. P. (1999). Malaria and the Narmada-River development in India: A case study of the Bargi Dam. *Annals of Tropical Medicine and Parasitology, 93*(5), 477-488.

Sow, S., de Vlas, S. J., Engels, D., & Gryseels, B. (2002). Water-related disease patterns before and after the construction of the Diama Dam in northern Senegal. *Annals of Tropical Medicine and Parasitology, 96*(6), 575-586.

Thompson, P. M., & Sultana, P. (1996). Distributional and social impacts of flood control in Bangladesh. *The Geographical Journal, 162*(1), 1-13.

Waddy, B. B. (1975). Research into the health problems of manmade lakes, with special reference to Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 69*(1), 39-50.

World Health Organization. (2006). *Water-related diseases.* Retrieved April 16, 2006 from http://www.who.int/water_sanitation_health/diseases/diarrhea/en/.

Chapter 2

Spatio-temporal clustering of cholera:
The impact of flood control in Matlab, Bangladesh, 1983-2003

Abstract: The provision of flood protection to a vulnerable population could have significant impacts on the incidence of cholera, a waterborne disease. Using longitudinal health and population data gathered over 21 years, cluster analysis was conducted to determine if substantive changes have occurred in the spatial distribution of cholera incidence since the construction of flood protection structures in a rural area of Bangladesh. Results indicate that both temporal and spatial shifts in cholera incidence have occurred, but that these shifts are not universal, differing according to cholera type.

Keywords: cholera, Bangladesh, cluster, flood protection, spatial scan statistic

Introduction & Background

Alteration of natural environments for development purposes often have unanticipated consequences for disease ecologies as traditional human-environmental interactions are changed. Studies of modified aquatic systems in Africa and Asia, involving the damming or impoundment of water sources, have almost universally shown a magnification of disease incidence (Hughes & Hunter, 1970; Hunter, 2003; Hunter, Rey & Scott, 1982; Keiser et al., 2005; Singh, Mehra & Sharma, 1999; Sow, de Vlas, Engels & Gryseels, 2002; Waddy, 1975). This study sought to determine whether the introduction of flood protection to a rural region of Bangladesh in the late 1980s resulted in a similar change in cholera incidence through an exploration of spatio-temporal clustering patterns.

Matlab, Bangladesh is a rural region located approximately 50km southeast of the capital city, Dhaka (Figure 1). Approximately 200,000 people live in Matlab, with a population density of nearly 1000 people per square kilometer. The majority of Matlab's residents are Muslim and engaged in agricultural production, primarily of rice. Since 1966, Matlab has been the site of demographic and health surveillance administered by the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). According to ICDDR,B hospital records, cholera and other diarrheal diseases were endemic to the area in the 1960s and continue to persist today.

Running from north to south through Matlab is the Dhonagoda River. The Dhonagoda floods annually during the monsoon season, resulting in fields, roads and households under several feet of water. In the 1980s, Matlab was targeted by the Government of Bangladesh as a site for flood mitigation efforts. Completed in 1989, the Meghna-Dhonagoda Irrigation Project (MDIP) resulted in flood protection for approximately half of the Matlab study area (Emch, 2000). The MDIP consisted of construction of a large earthen embankment along the northern edge of the Dhonagoda, as well as the installation of culverts, bridges and pumping stations (Ansary, Fulton, Bhuiya & Chowdury, 1997). The population inside the flood protected area still receives massive amounts of rainfall during the monsoon, but is not subject to the Dhonagoda overflowing its banks as water from upstream pours into the area. For Matlab residents living outside the flood protected area, seasonal patterns continue much as they did before, with perhaps slightly increased flood levels as water that previously spread across the area north of the river is now forced south by the embankment.

To explore this division in how Matlab's residents experience the rainy season and its possible impact on their cholera experience, we looked at occurrence of hot-spot clusters in both space and time. We endeavored to answer two questions: 1) Does incidence of cholera cluster in both space and time in Matlab? 2) Have cluster patterns changed since flood protection was introduced? We hypothesized that not only did cholera, a highly infectious and transmissible disease, cluster in Matlab, but that there would be significant changes in both the timing and location of clusters. These changes in cluster patterns would act as indicators that human-environmental interactions which lead to cholera infection were changed by the introduction of flood protection. Previous studies conducted over shorter time period have shown residence in the flood protected area of Matlab to be a risk factor in cholera incidence (Ali, Emch, Donnay, Yunus & Sack, 2002b; Emch, 1999; Emch, 2000), and we were interested in exploring whether this relationship would be expressed in our longitudinal cluster analysis.

Cholera in Bangladesh exhibits high seasonality, with incidence peaking in September through December. There is typically a smaller outbreak in April or May, just prior to the monsoon (Faruque et al., 2005; Longini et al., 2002). Although the precise cause triggering the outbreaks during the rainy season are not understood, they have been correlated with increases in water levels, copepod counts, plankton levels and bacteriophage populations (Colwell et al., 1992; Faruque, 2005; Huq, 2005). The dry season outbreaks, in comparison, occur when surface water sources are at their lowest point. At this point in the year, residents of Matlab are using depleted water sources harboring potentially dense populations of *Vibrio cholerae* bacteria (hereafter *V. cholerae*).

There are multiple types of *V. cholerae* bacteria that cause epidemic cholera, and well-documented and significant shifts have taken place in the types responsible for the majority of cholera cases in Matlab. Two cholera serogroups exist, O1 and O139, and within the O1 serogroup there are two biotypes: Classical and El Tor (Longini et al., 2002). Classical and El Tor cholera co-circulated in Matlab at the beginning of the study period until 1988 when Classical disappeared. El Tor then predominated until the highly virulent O139 emerged in 1993 (Ali, Emch, Donnay, Yunus & Sack, 2002a; Longini et al., 2002).

We anticipated finding no differences across pre-MDIP Matlab in spatial patterns of cholera clusters, thinking instead that the reported clusters would be most defined by seasonal variation rather than spatial. In comparing the post-flood protection clusters to the pre-flood protection clusters we thought that distinct spatial patterns would be seen, particularly in the rainy season. The rainy season would be the time of most pronounced spatial differences in clustering if flood protection did impact cholera incidence since it is only in the rainy season that the MDIP embankment confers any flood protection. In addition to comparing pre-MDIP and post-MDIP clustering for indications of protection or risk afforded by flood protected status, we chose to explore cholera clustering by biotype and serogroup, to see if flood protection had different mediating impacts based upon cholera type. It was anticipated that all types of cholera would exhibit similar clustering patterns to those established in the 1983-1989 and 1990-2003 analyses.

Data

We drew on health, demographic and geographic datasets to evaluate our research questions. Health and population data was gathered under the auspices of ICDDR,B's Health & Demographic Surveillance System (HDSS), in operation since 1966. Every resident of Matlab is assigned a unique identification number in the HDSS that connects them to a village, a *bari* and a household. A *bari* is a patrilineally connected grouping of households. Twice monthly, each *bari* is visited by a trained ICDDR,B community health worker and information on demographic events such as births, deaths and migrations is recorded (Myaux, Ali, Felsenstein, Chakraborty & de Francisco, 1997).

In addition to recording demographic information and providing basic health and nutrition information, the ICDDR,B community health worker also inquires about illness in the *bari* and makes referrals for hospitalization at either the main ICDDR,B hospital in Matlab town or one of the three subcenters located in the area. Treatment at these facilities is both free and specialized in the treatment of diarrheal diseases, and residents are also provided free transportation to the treatment centers if they cannot make the trip on their own (Ali, Emch, Donnay, Yunus & Sack, 2002a; Emch & Ali, 2003). For this reason, we make the assumption that all cholera cases are reported to and treated by ICDDR,B health facilities.

The identification numbers of Matlab's residents are recorded by the hospital and connected to the demographic information recorded for their households and *baris* by database managers at ICDDR,B's headquarters in Dhaka.

We created a database of all 9580 laboratory-defined cholera cases observed in Matlab from January 1, 1983 to December 31, 2003. This study period was chosen to give a seven year baseline of cholera distribution before the MDIP construction was completed (1983-1989), followed by 14 years (1990-2003) where cholera patterns could differentiate between the flood protected and unprotected area. From the HDSS, we were also able to gather information about annual mid-year populations for each household, aggregated to the *bari* level. The *bari* is the unit of analysis in this study because it is the smallest scale population unit that can make use of a Geographic Information System (GIS).

A GIS database that is accurate within 10m was created for Matlab by digitizing photographs and satellite images (Figure 2) (Emch, 1999; Ali, Emch, Ashley & Streatfield, 2001). Each *bari* within Matlab has a unique identification number as assigned by the HDSS. This *bari* ID is the common unit that allows for integration of cholera cases, background population counts and geographic location within the study area. A total of 7490 *baris* had both background population data and their geographic locations mapped in the GIS. A master database of monthly cholera counts and mid-year population for each of the 7490 *baris* was created and queried for subsequent cluster analysis.

Methods

This study is interested in local clustering of cholera incidence, the scale at which a spatial scan statistic has been found to be good at detecting clusters (Song & Kuldorff, 2003).

SaTScan™ is freely available software that implements a space-time scan statistic for point data (Kulldorff, 1997; Kulldorff & Nagarwalla, 1995). It has been frequently used in cluster analysis of both infectious and chronic diseases (Gosselin, Lebel, Rivest & Douville-Fradet, 2005; Polack et al., 2005; Fang, Kulldorff & Gregorio, 2004), and was previously used by Emch & Ali (2003) to analyze cholera incidence over a 3-year timeframe. SaTScan detects, at the local level, events that are excessive and then tests whether those excesses could have occurred randomly. It does so by moving a cylindrical window over the study area, centering on one data point (in this case each *bari*) after another. At each point the radius of the scanning window is varied according to user-defined spatial parameters. At the same time, the height of the cylinder is varied according to user-defined temporal parameters. The result of this process is a collection of overlapping cylinders that each represent a possible space-time cluster. For each cylinder the null hypothesis, that disease risk is the same inside as outside the cylinder, is tested against the alternative hypothesis, that risk is elevated within the cylinder. A p-value is calculated based on Monte Carlo simulations, in this case 999, as is a likelihood-ratio.

SaTScan offers a number of advantages for the analysis of cholera in Matlab. Firstly, when utilizing the Poisson process option, inhomogeneity of background population is not a confounding factor. Under the Poisson null hypothesis, the expected number of cases at each point is assumed to be determined by the proportion of the total population connected to that point. In this way, clusters that are detected are not an artifact of higher populations in some *baris*. Secondly, SaTScan allows for the introduction of temporal variation in background population. This is important when considering cholera over a 21-year timeframe, as new *baris* are established or old *baris* are destroyed by flooding. Thirdly, SaTScan does not make *a priori* assumptions about cluster location, size or duration, but rather considers all potential options. As a result, clusters indicated by SaTScan are not biased by either a *bari*'s flood status or whether cholera was diagnosed in the rainy or dry season. And fourthly, SaTScan not only indicates that the null hypothesis has been rejected, but at what location in the study area it was rejected. The interface between SaTScan and commercially available GIS software (ArcGIS 9.1) is such that detected clusters can easily be mapped and assigned either flood protected or unprotected status, based on the *bari* defined as the cluster centroid.

Our initial SaTScan analysis utilized all cholera cases over the entire 21-year timeframe and looked for purely temporal and purely spatial clusters in addition to the spatio-temporal clusters explored in all subsequent analyses. The space and time limitations were set to 50% of the study population and 50% of the study period. This allowed the program to scan for clusters of both large and small size and duration. The purpose of this analysis was to gain a sense of how cholera clustered in Matlab for the entire 21-year study period, whether the most statistically significant clusters occurred before or after the MDIP divided the study area into two. Results reported by SaTScan do not indicate typical cholera patterns in Matlab, instead they represent cholera events that are unusually high or low given background rates. Exploring clusters over the entire 21-year timeframe informed future analyses by highlighting certain years and certain areas that had uncharacteristic cholera incidence.

Next, cholera cases and background *bari* populations were divided into two time periods, 1983-1989 and 1990-2003, which were analyzed separately for significant spatio-temporal clusters. The first dataset was defined as a baseline of how cholera initially clustered in Matlab, in the years before there was any differentiation in human-environment interactions due to flood protection. The second dataset was used to determine if there was observable differentiation in size and timing of clusters in flood protected and unprotected areas. Rather than using the SaTScan defaults of 50% space and 50% time as the upper bounds for detected clusters, we chose to artificially lower the time limit to one month and the space limit to first 5% and then 10% of the study area. This artificial spatio-temporal scaling allowed us to explore micro-scale clustering of cholera within Matlab. By setting the time limit to one month, we were able to detect clusters that could be categorized as either occurring during the rainy season (June through November) or the dry season (December through May).

The low spatial bounds provided two benefits, firstly, they meant that the majority of clusters were wholly contained within the protected or unprotected areas and secondly, cholera in Matlab has been shown to be affected most by those neighbors who live within a small distance of a resident's *bari*, rather than by residents several kilometers away. The results returned by SaTScan with these small bounds were therefore more representative of small pockets of cholera that could be seen as neighborhood-level interactions rather than a process occurring at the level of the entire study area.

Finally, the 9580 total cholera cases from 1983-2003 were grouped according to laboratory-defined serogroups and biotypes. Cluster analysis was performed according to four divisions: Classical, El Tor, O1 (both Classical and El Tor) and O139. Each was analyzed over the entire 21-year time period, with one month temporal bounds and 5% spatial bounds.

We did not conduct separate cluster analyses for the flood protected and unprotected areas. Doing so would have set another artificial bound on the analysis, given that the significance of clusters would be reported in terms of only half of the study area's population and case counts. Additionally, comparability of the two areas is already possible given SaTScan's identification of specific cluster centroids and the ability in ArcGIS to determine if said centroid is in the flood protected or unprotected area. SaTScan results were imported into geographic information system software (ArcGIS 9.1) and joined to the Matlab *bari* layer to create maps of cluster centers and radii.

Results

Results from the 1983-2003 analysis showed five significant clusters, three high and two low (Table 1). We explored both high and low clusters because areas with lower than expected rates are of as much interest as those places with higher rates. The first low cluster, Cluster #1, was purely temporal, and took place between 1999 and 2003 over the entire study area. Purely temporal clusters are exempt from the 50% spatial bound, and the inclusion of all of Matlab's *baris* in this cluster is an indication that cholera prevalence in Matlab has, irregardless of spatial location, fallen considerably in the past several years. This is likely the result of overall improvement in socioeconomic status of rural Bangladeshis. The other low cluster (Cluster #2) was purely spatial and centered in the unprotected northeast section of Matlab, and lasted the entire 21-year time period.

Of the three high clusters reported by SaTScan, one was purely spatial and the other two were spatio-temporal. All three were located in the southern and western portion of Matlab (Figure 3). Cluster #3 is a purely spatial cluster that took place between 1983 and 2003, and serves as an indication that consistently higher-than-expected prevalence of cholera existed in this southern portion of the study area irregardless of temporal considerations. The two spatio-temporal clusters detected in the overall 1983-2003 analysis are similar in terms of size, statistical significance and relative risk. The first (Cluster #4) took place between 1983 and 1987, before MDIP construction, and is located south of the Dhonagoda River. The other (Cluster #5) occurred between 1992 and 1995, after MDIP construction, and was centered in the flood protected region. Thus, given every case of cholera over the entire study period, the location of unusually high incidence of the illness shifted from south of the river to north after MDIP construction.

<u>1983-89 Clusters vs. 1990-2003 Clusters</u>

For all subsequent analyses, only those clusters that include 5 or more *baris* and with significant (<.05) p-values are reported in order to present only those results that can be considered both spatially and statistically meaningful. Though SaTScan was looking for both high and low spatio-temporal clusters, only high clusters were detected. The fact that no low clusters were reported in the 5% and 10% analyses though some were detected in the initial 21-year analysis suggests that artificially lowering the spatio-temporal bounds has an effect on the detection of lower-than-expected trends in cholera incidence.

Fifteen significant clusters were detected in the 1983-1989, 5% space, one month time analysis. These clusters were then classified by their location in either the future-flood protected or future-unprotected areas of Matlab in order to examine whether any spatial variation in clusters existed prior to 1989. Of the fifteen clusters, 8 were centered in the future flood protected area and the remaining 7 in the future unprotected area (Figure 4). Increasing the spatial bounding to 10% but holding the temporal bound at one month resulted in the detection of six clusters, 2 with centers in the future-protected area and 4 centered in the future-unprotected area. The results of both analyses suggest that unexpectedly high incidence of cholera was evenly distributed over the study area and that no noticeable differences in the cholera experience of the to-be-divided populations is observable. The spatial pattern observed in the 5% clusters from 1983-1989 appears to have little to do with whether a cluster centroid is located north or south of the river and more to do with its proximity to the Dhonagoda River.

Seventeen clusters between 1990 and 2003 were identically detected in both the 5% and 10% spatially bound analyses, out of 27 total reported in the 5% and 21 total in the 10%. This high level of overlap suggests that artificially limiting the spatial bounds of analysis had little impact on the detection of significant clusters during the 1990-2003 period and that the results are robust. Unexpectedly, the results of the 1990-2003 cluster analysis also show little distinct spatial variation. Mapping the 27 clusters detected in the 5% analysis shows that ten were centered in the flood protected area, seventeen were in the unprotected area (Figure 5). For both the flood protected and unprotected clusters, proximity to the Dhonagoda River still seems somewhat important, although the clusters appear more dispersed across the study area than in the 1983-1989 period. The post-flood protection clusters are mainly distributed across the southern portion of Matlab, with few clusters found in the northeast section of the study area.

Although no spatial differences may be observed between the 1983-1989 clusters and the 1990-2003 clusters, as is shown in the previous figures, a strong temporal shift took place in cholera clustering after construction of the MDIP. Prior to the introduction of flood protection in Matlab, clusters were primarily detected during the dry season. Of the 15 clusters reported in the 5% analysis (Figure 6), only two occur in the six months of the rainy season, here defined as June-November. These two 1983-1989 clusters take place in October and November, at the very end of the rainy season when water levels are the highest. The remaining thirteen clusters take place either in December and January, the very beginning of the dry season when flood waters are still receding, or in May, when Matlab is at its driest or only beginning to experience rainfall. The 10% analysis returned similar temporal results: only one of the 6 clusters reported in the 10% analysis occurred in the rainy season, and the dry season clusters again take place at the beginning of that period.

After the MDIP was completed, clusters occurred mostly in the rainy season. Only seven out of twenty-seven clusters detected at the 5% level in 1990-2003 took place in the dry season (Figure 7). The 10% analysis returned similar results, only five out of twenty-one total significant clusters occurred in the dry season. For the 1990-2003 period, the majority of clusters were detected in the rainy months of June-November. And although location in the study area logically had little apparent impact on cluster timing in pre-MDIP Matlab, in the 1990-2003 analysis there are distinct differences observed in the timing of flood protected clusters as opposed to unprotected clusters. Of the twenty-seven 1990-2003 reported clusters, ten were centered in the flood protected portion of Matlab, seventeen in the unprotected section. The timing of these ten flood protected clusters as compared to the other seventeen supports the hypothesis that clustering in the flood protected area occurs later than in the unprotected area. Flood protected clusters occur solely in May, at the very end of the dry season when water resources are at most depleted, and in September through November, the final months of the rainy season when water resources are most abundant. The unprotected clusters, in contrast, occur fairly evenly throughout the months, excepting February and March. Clustering for the entire Matlab area has thus shifted from occurring mainly in the dry season prior to MDIP construction to occurring mainly in the rainy season, and the clusters are reported earlier in the rainy season for the unprotected area than for the flood protected area.

At the yearly time scale, the results of both the 5% and 10% analyses for the 1983-1989 time period show no significant clusters occurring after 1986. Indeed, the majority, 8 out of 15 in the 5% analysis and 4 out of 6 in the 10% analysis, take place in 1983. Additionally, both the 5% and 10% results indicate clusters occurring in December of 1985 and January of 1986, suggesting that these two months were a time of high cholera prevalence for certain sections of Matlab. After 1986, though, no clusters are reported. This gap in reported cluster years continues in the 1990-2003 analyses (Figure 8). For both the 5% and 10% 1990-2003 analyses, no clusters are reported until 1992. For three years on either side of the 1989 embankment construction, then, cholera prevalence in both the flood protected and unprotected areas was static, no statistically significant clusters of high or low rates were detected. In addition to this lapse in cluster occurrence, the clusters detected in the 1983-1989 period are evenly distributed between the future-protected and future-unprotected areas of Matlab, suggesting that any differences observed in the reported 1990-2003 clusters can be correlated with the introduction of flood protection under the MDIP.

At the annual scale, there appear to be cluster timing differences between the flood protected and unprotected areas. The lapse in detected clusters ends sharply in the early 1990s, during the time that the highly virulent O139 arose in Bangladesh. In the initial years of the post-O139 introduction the majority of clusters are detected in flood protected areas of Matlab. In 1997 this shifts, however, and many more clusters are detected in the unprotected areas. This suggests that the flood protected areas of Matlab were initially more impacted by the arrival of O139, that the alteration of the normal flood/drought cycle by the construction of the MDIP accelerated the impact of O139. Further analysis was undertaken at the biotype level to determine if this was an accurate assessment of the underlying patterns being expressed in the 1990-2003 clusters.

Biotype/Serogroup Analysis

In the biotype-level cluster analysis we explored the possibility of spatial and temporal differences in clustering occurring at the bacterial level. Cluster analysis was completed on four datasets over the entire 1983-2003 period, with cases aggregated in the following categories: Classical, El Tor, O1 (Classical & El Tor together), O139. As was previously discussed, Classical cholera was endemic in Matlab until 1988 when it disappeared, leaving El Tor to circulate on its own until O139 arrived in 1993. Differences in the spatio-temporal clustering exhibited within the O1 serogroup (Classical & El Tor), and between the serogroups (O1 vs O139) reveal interactions between the types, the season and flood protection status.

The Classical analysis returned 14 significant clusters, nine of which occurred in 1983, at the very beginning of the study period. This suggests that Classical incidence was at its peak, with uncharacteristically high rates given the entire 21-year period, in the early 1980s, before slowly being replaced by El Tor. Of the fourteen Classical clusters, only one took place in the rainy season (previously defined as June-November). The El Tor cluster analysis, in contrast, returned significantly more clusters in the rainy season (21) than the dry season (8).

The O1 serogroup analysis (Classical & El Tor aggregated) detected 23 significant clusters. Of these, only 2 took place before 1988 and could possibly represent Classical clusters. This means the remaining 21 clusters must be El Tor clusters, and that El Tor cholera rates exhibit greater variation in space and time than Classical. This is logical given that El Tor was present in Matlab over the entire 21-years analyzed and thus has more opportunity to exhibit clustering. No O1 clusters were reported between 1986 and 1994, consistent with the previously described findings that indicated a lack of significant cholera clustering on either side of the 1989 MDIP completion. The last O1 cluster reported occurs in 2000, three years before the end of the study period.

The O139 cluster results can be viewed as highly robust, 5 of the 15 reported clusters take place in early 1993. This is the year that O139 is known to have arrived in Matlab and been at its most virulent. No O139 clusters are detected after 2000, the same endpoint as the O1 clusters. A comparison at the month timescale (Figure 10) shows that O1 clusters mirror typical seasonality of cholera in Bangladesh, with spikes at the end of the dry season right before the monsoons arrive and then again in the middle of the rainy season when water levels are at their peak. O139, in contrast, spikes in the middle of the dry season and shows a slow gradual climb in clusters at the tail end of the rainy season. Overall, the O1 clusters occurred more in the rainy season than the dry, while the O139 clusters occurred almost evenly between the two seasons.

Differences in season of occurrence between O1 and O139 are then further enhanced when flood protection status is considered. Within the flood protected area, six of the seven O1 clusters took place in the rainy season (Figure 11). Of the five flood protected clusters detected in the O139 analysis, two took place in the rainy season and three in the dry season

(Figure 12). Overall spatial differences can also be seen between the O1 and O139 clusters. The majority of O1 clusters fall outside the protection of the MDIP, while the O139 clusters are evenly divided between the flood protected and unprotected areas.

Discussion & Conclusions

In analyzing longitudinally how cholera in Matlab has been affected by the introduction of flood protection structures by the government of Bangladesh in the late 1980s, we found that the introduction of flood protection to Matlab radically altered patterns of cholera. Cluster patterns observed at the 21-year timescale changed when pre-MDIP and post-MDIP cholera were considered separately, and stratifying incidence according to cholera strain showed that the arrival of a new type of cholera in Matlab superceded the effect of flood protection status on cholera clustering. While our cluster analysis does not determine whether flood protection is a causal factor in cholera transmission, the findings do suggest that the construction of the MDIP had a significant impact on regional incidence of cholera.

A shift from clustered cholera primarily south of the Dhonagoda before the introduction of flood protection to clustered cholera north of the river after flood protection was suggested by the results of the 21-year overall analysis. This pattern was not repeated, however, when the data was stratified by time period. An even spatial distribution across Matlab was observed in the 1983-1989 cluster patterns. This was as we expected, as prior to the introduction of flood protection there was little, if any, differentiation in environment and water-use behaviors between the areas north and south of the Dhonagoda River. Surprisingly, however, there was not a strong spatial variation observed in the clusters reported for 1990-2003. Approximately two thirds of the clusters had their centers located south of the river in the unprotected area, the remaining were located within the MDIP embankment. This suggests that the introduction of flood protection to Matlab did not have major impacts on the spatial distribution of cholera clusters. The majority of clusters, regardless of time period or flood protection status, were located in the central and southern sections of Matlab. Only one cluster in the 1983-1989 analysis and four clusters in the 1990-2003 analysis were located in the northern area of Matlab. Similarly, in the 21-year overall analysis the only low, purely spatial cluster detected was located in the northern part of Matlab. These findings are consistent with previous studies that identified southern Matlab as the foci of cholera incidence (Emch & Ali, 2003). Reasons for this pattern are uncertain, though it would be easy to assume the patterns exist because the northeast area is up-river, and thus cholera flushes to the south and west. This view does not stand up to scrutiny, however, given that the boundaries of Matlab are artificial political creations, and that the northeast portion has as much downstream exposure to rivers and streams located outside of Matlab. Other explanations, such as lower population density or different socioeconomic status in the northeast, are similarly inadequate. Higher cholera levels in the southern portion of Matlab are often attributed to the proximity of the confluence of the Meghna and Dhonagoda Rivers. The reasoning is that water levels are more subject to fluctuation and thus cholera transmission is both easier and more frequent.

Rather than distinct spatial shifts we had expected to see between the pre- and post-flood protection periods, the construction of the MDIP had a large effect on the temporality of cholera clustering. Classifying clusters according to season suggested that a shift from dry season to rainy season cluster has taken place since the introduction of flood protection. This shift did not occur equally between the flood protected and unprotected areas, however, as cholera clusters were detected earlier in the rainy season outside the MDIP embankment than within. While the overall shift from dry to rainy clustering was an unexpected result, the delayed clustering within the flood protected area is logical. Without the rising waters of the Dhonagoda, cholera incidence within the MDIP is more heavily influenced by monsoon rainfall. Being protected from swift river flooding delays cholera clustering inside the flood protected area. The transition from dry season clustering to rainy season clustering has implications for both Matlab's residents and for ICDDR,B health workers. Local knowledge and awareness of cholera is based on long-term patterns of incidence, and shifts in these patterns can leave residents open to unexpected and unanticipated infections in times when they are not as guarded in their behaviors. Our findings also suggest that ICDDR,B can expect to have greater demand for treatment and prevention in the rainy season than they have previously.

In addition to changes in temporal patterns of cholera clusters since flood protection's introduction in 1989, differences in cholera patterns exist when the Matlab data is stratified by bacterial strain. Cluster analysis by strain type revealed striking differences in how Classical, El Tor and O139 cholera is experienced in space and time by Matlab's residents. While the O1 cholera clusters follow traditional seasonal patterns and show distinct splits between the flood protected and unprotected areas, O139 clusters are fairly evenly distributed in both space and time. This is likely a reflection of the relatively new arrival of O139 in Matlab, and its greater virulence. Perhaps O139's infectivity is less mediated by density within water sources that are either flooded during the monsoon or shrinking during the dry season, so that it clusters equally across the flood protected and unprotected areas of Matlab irregardless of season. O1 cholera, in contrast, has long-established patterns that fluctuate in concert with the seasons, so it is greater affected by the impacts of the introduction of flood protection on water supplies in Matlab.

Our study used seven years of cholera data prior to the construction of MDIP flood protection structures to determine baseline cholera clustering patterns. While we make the assumption that the years between 1983 and 1989 are representative of how cholera incidence clustered in Matlab before the population was divided in half, this could be incorrect. It is possible that cyclical patterns of cholera clustering exist for Matlab that we are unable to detect within the data available. We therefore make the above statements about temporal changes and differing responses of O1 and O139 with this caution.

Our study did not control for age, sometimes an important confounding factor in cholera studies given high incidence among young children and their mothers. We decided not to control for age based upon the belief that age structures are similar across the Matlab study area. In addition, the study did not account for the steadily decreasing rates of cholera within Matlab over the past two decades due to generally increased socioeconomic status of rural Bangladeshis. Generally decreasing rates are likely across Matlab, but if there are differentials between the flood protected and unprotected areas that would be expressed in cluster detection.

Restrictions within user-imposed limits on the SaTScan software could affect the validity of the results. The space-time cluster detection within SaTScan uses a circular window, not an ideal shape for detection of clusters that occur along a linear route, such as a river (Aamodt, Samuelsen, & Skrondal, 2006). The 5% and 10% spatial bounds and 1 month temporal bound decrease somewhat the strength of the results, but the general overlap between results at both 5% and 10% and high comparability between results and the literature suggest that this isn't an insurmountable limitation. As well, the detection of patterns described in previous literature, such as higher recorded incidence in April/May at the end of the dry season and again during the rainy season, as well as the spiking arrival of O139 in early 1993, suggest that SaTScan is a valid tool for analysis.

The construction of the MDIP flood protection structures seems to have taken place at a time of stable cholera rates in Matlab, as indicated by a lack of clusters reported in the years immediately prior to and directly after MDIP construction in 1989 in both the time-specific analyses and the cholera strain analyses. Additionally, cholera incidence seems smoother in space and time in the later years of the study period, as evidenced by the low cluster detected after 1999 in the 21-year analysis and the absence of any clusters reported after 2000 in the 1990-2003, O1 or O139 analyses.

Our results indicate that the introduction of flood protection in Matlab has changed patterns of cholera incidence previously experienced in the area. Unusual spikes in cholera incidence, as indicated by the presence of significant clusters, now occur more frequently in the rainy season, perhaps at a time when it is unexpected given historic patterns and residents are less on guard against infection. The introduction of flood protection has not affected all types of cholera incidence equally, however, and further analysis of how the O1 and O139 serogroups have responded to the MDIP construction could suggest different targeting strategies for the two types.

Figures & Tables

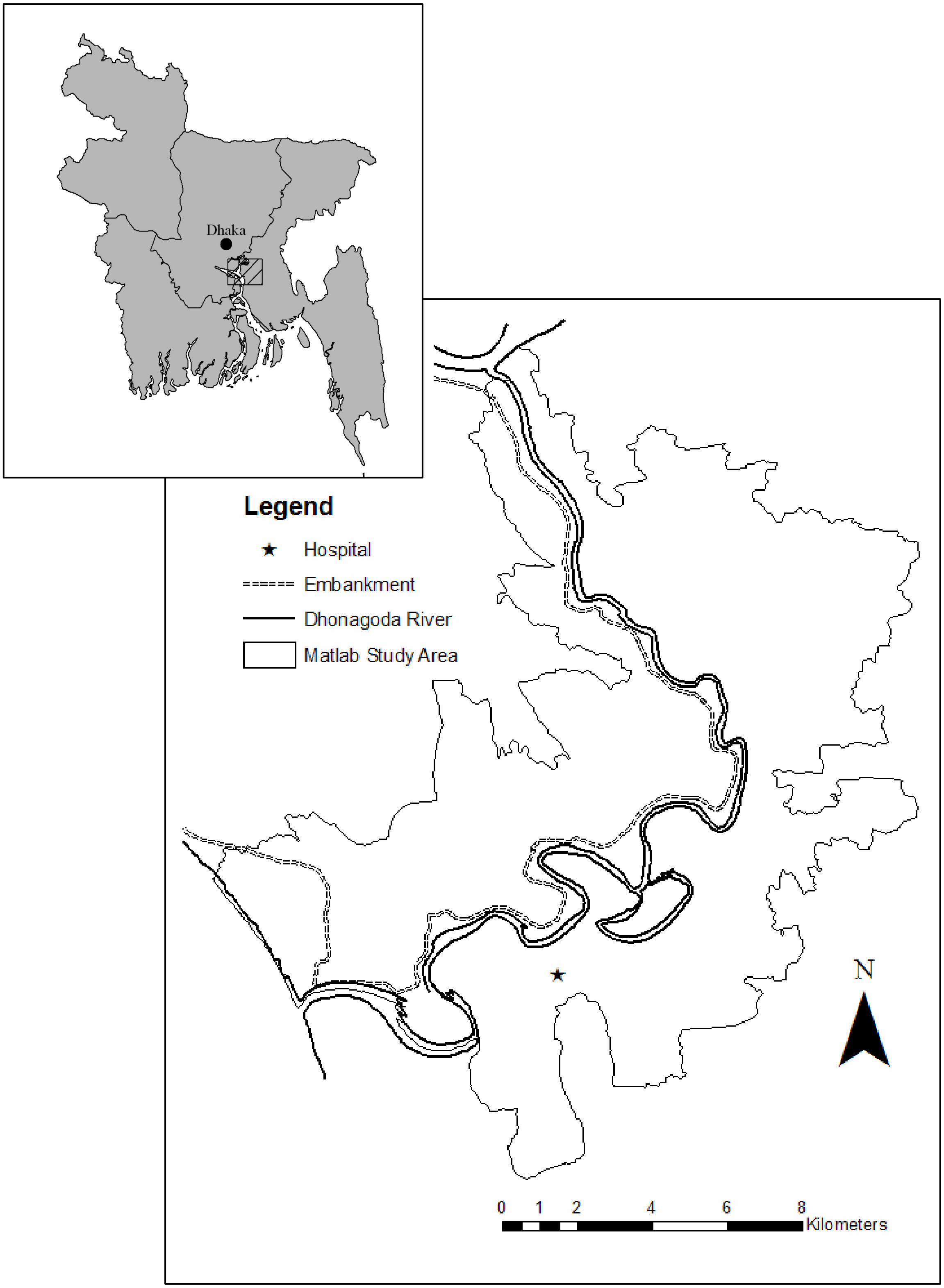

Figure 2.1: Location and Main Physical Features of Matlab, Bangladesh

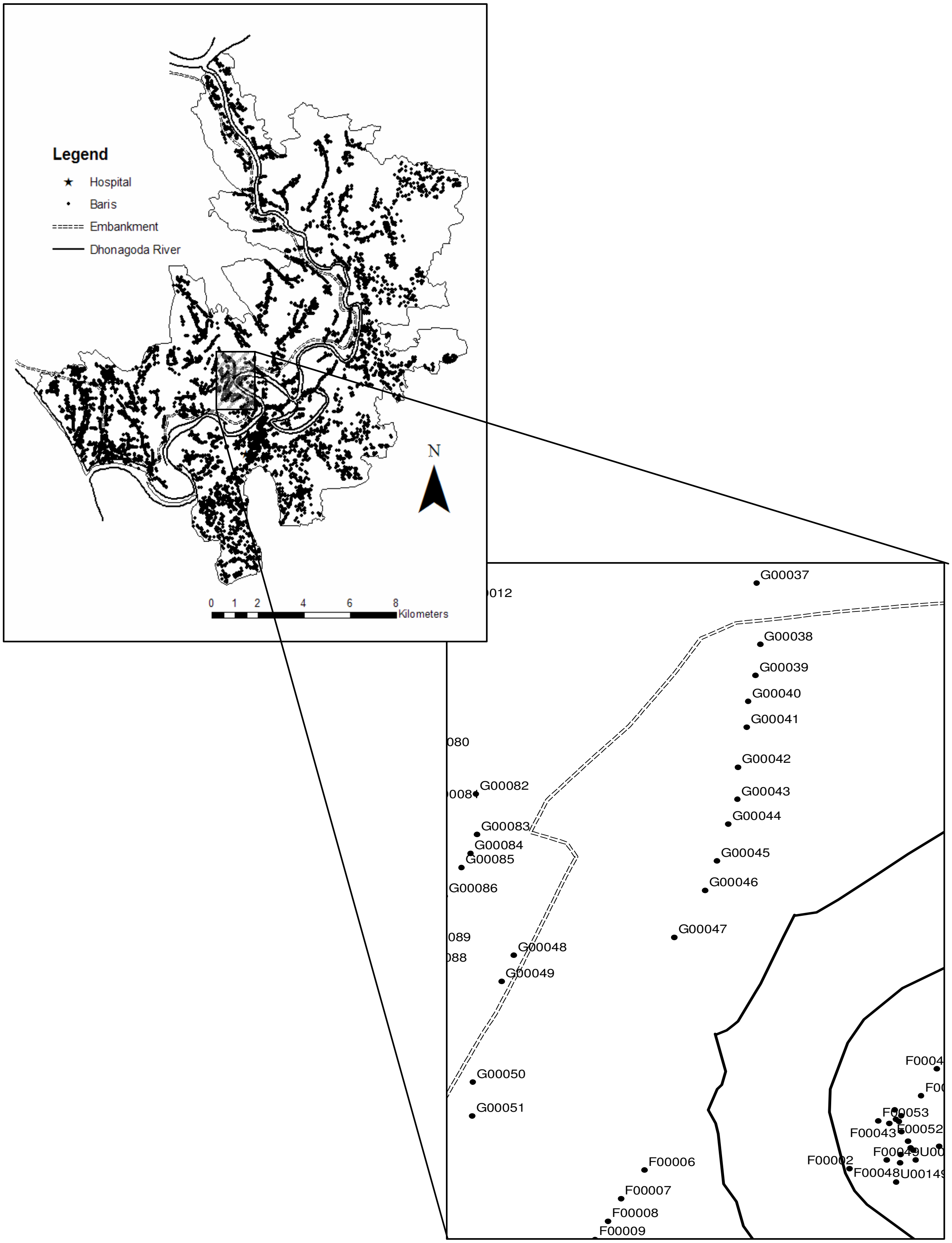

Figure 2.2: Matlab GIS with *Bari* Locations Identified

Cluster	Type	High or Low	Radius (km)	Start Date	End Date	Log Likelihood Ratio	P-Value	Relative Risk
1	Temporal	Low	All	1/1/1999	12/31/2003	856.999	0.001	0.27
2	Spatial	Low	6.86	1/1/1983	12/31/2003	2175.934	0.001	0.10
3	Spatial	High	6.02	1/1/1983	12/31/2003	1482.870	0.001	3.43
4	Space/Time	High	6.50	1/1/1983	12/31/1987	522.021	0.001	2.38
5	Space/Time	High	6.39	1/1/1992	12/31/1995	830.196	0.001	3.00

Table 2.1 1983-2003 Cluster Results

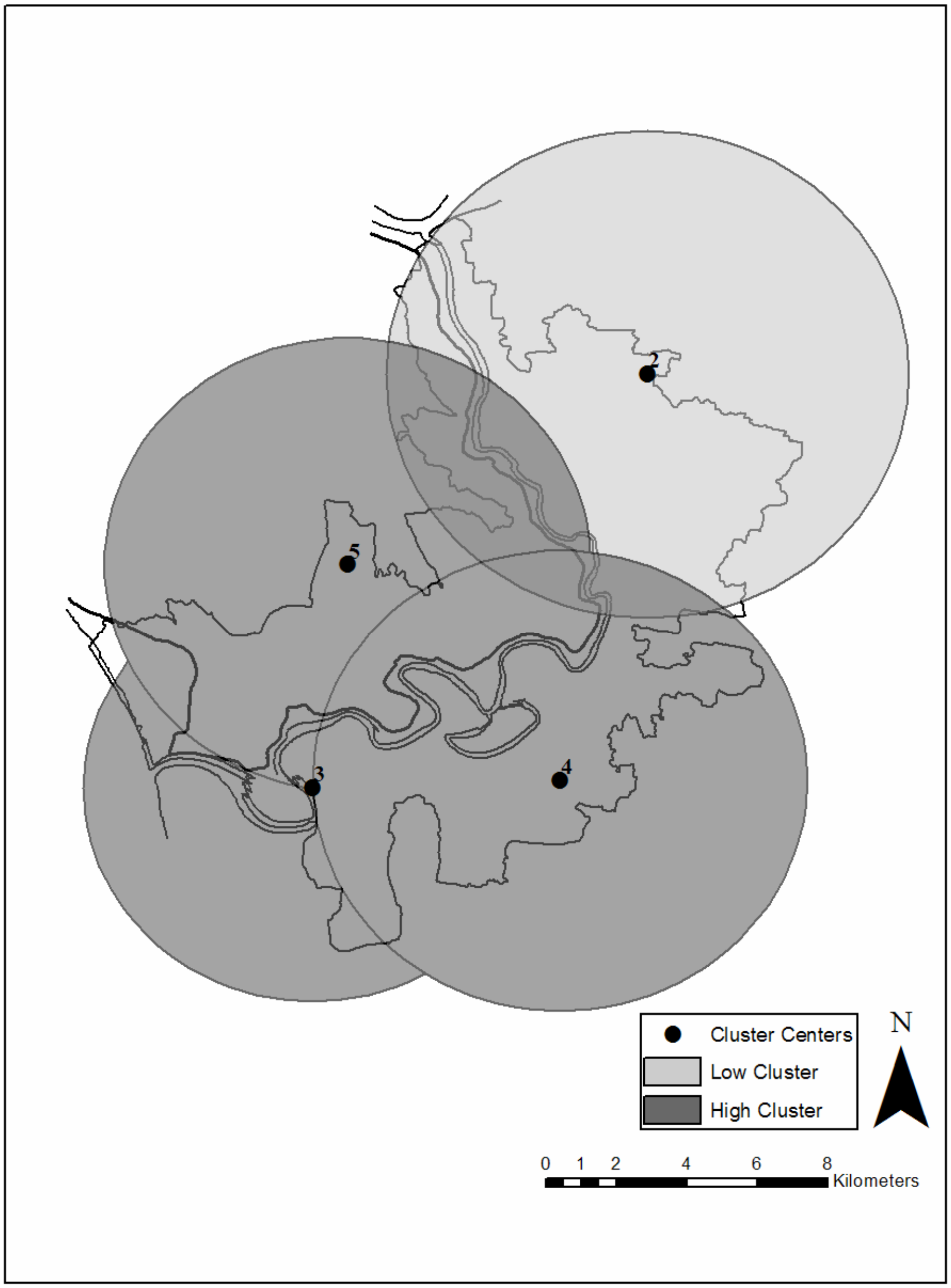

Figure 2.3: 1983-2003 Clusters

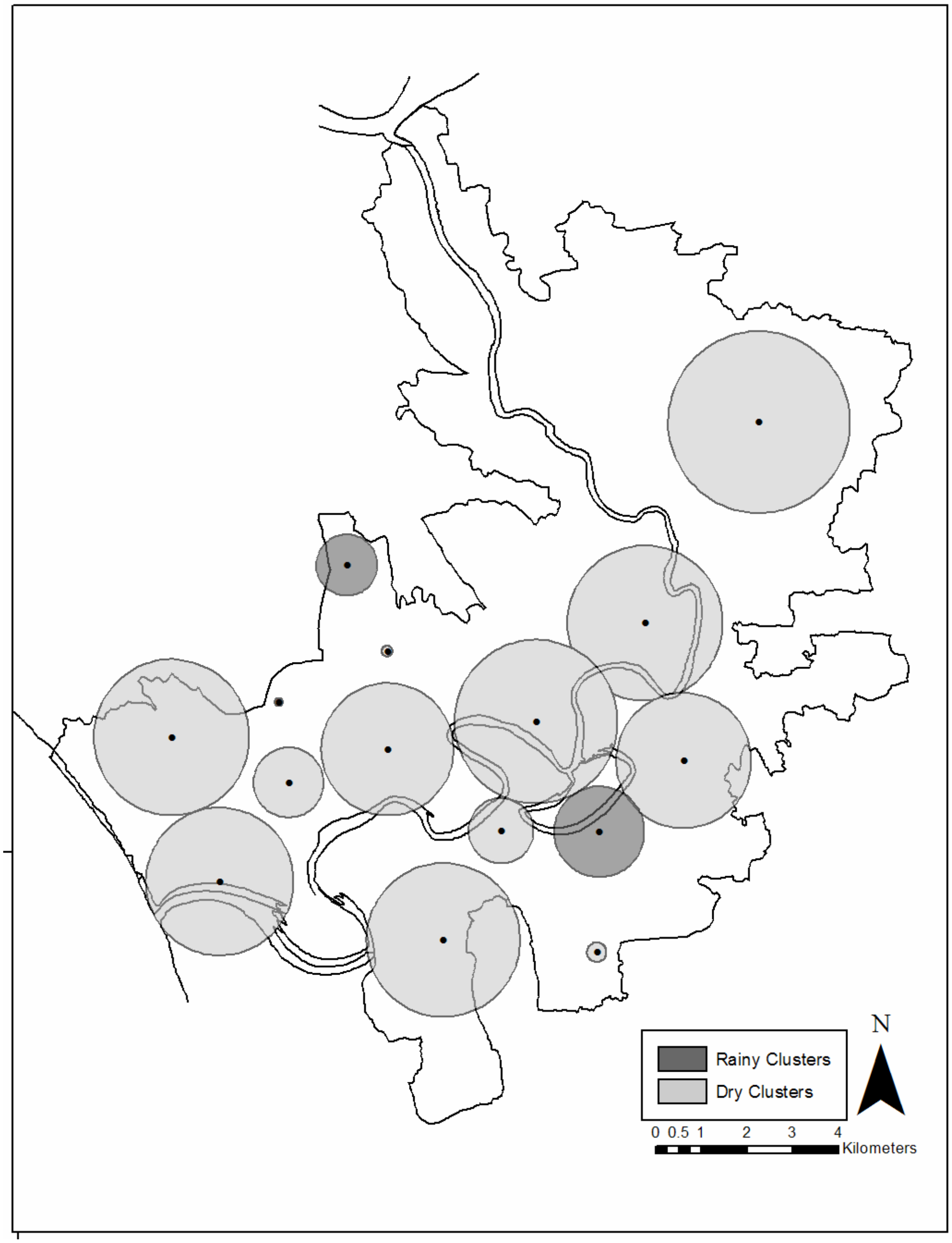

Figure 2.4: 1983-1989 Clusters at 5% Spatial Bound

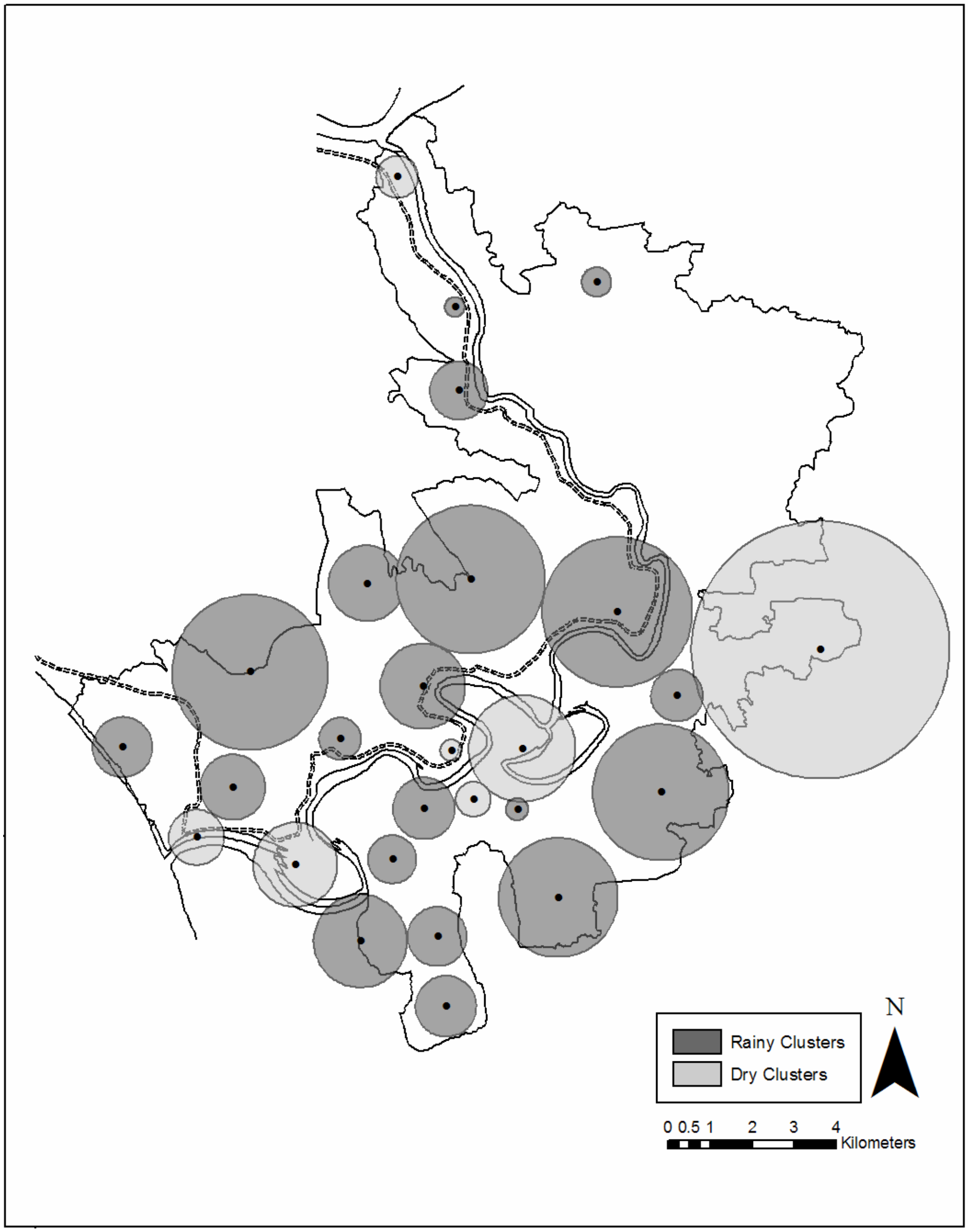

Figure 2.5: 1990-2003 Clusters at 5% Spatial Bound

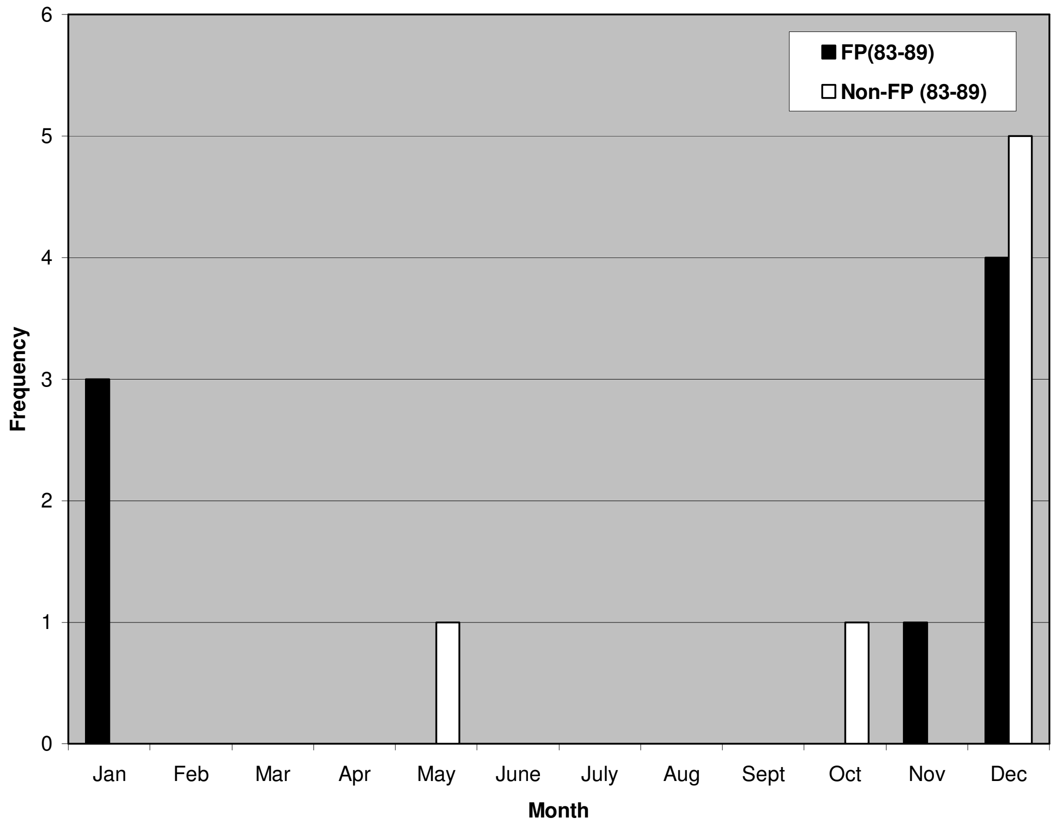

Figure 2.6: 1983-1989 Clusters by Month, 5% Spatial Bound

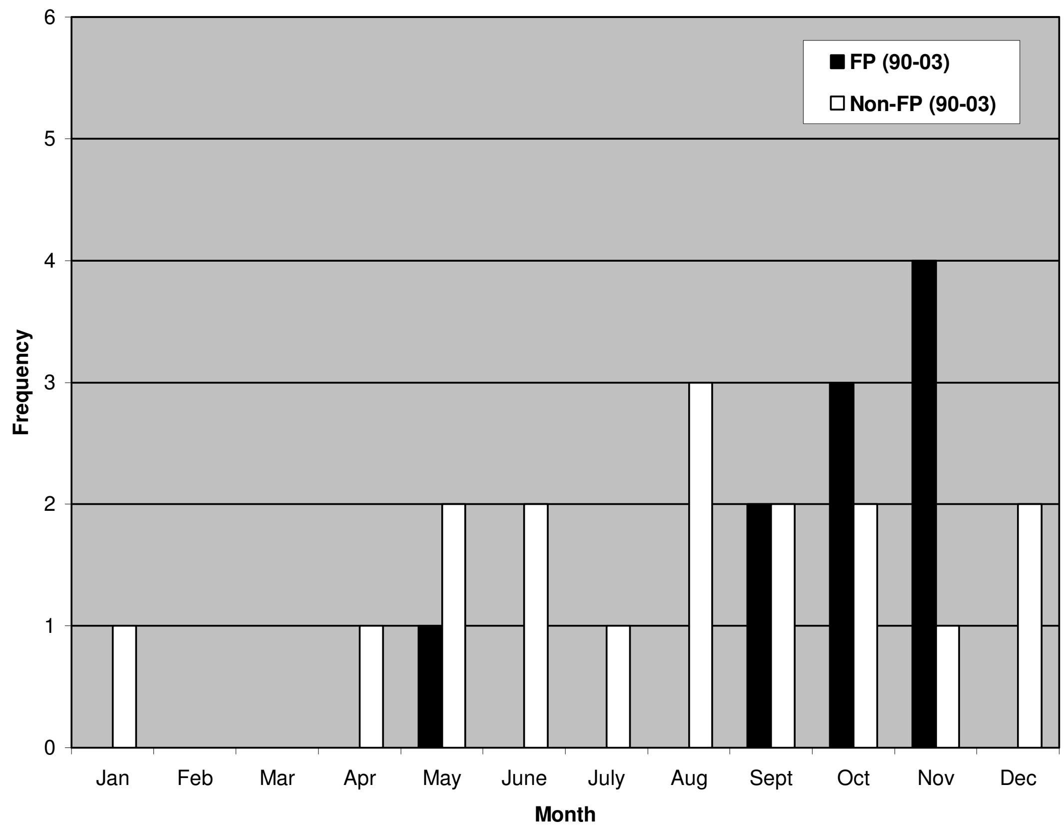

Figure 2.7: 1990-2003 Clusters by Month, 5% Spatial Bound

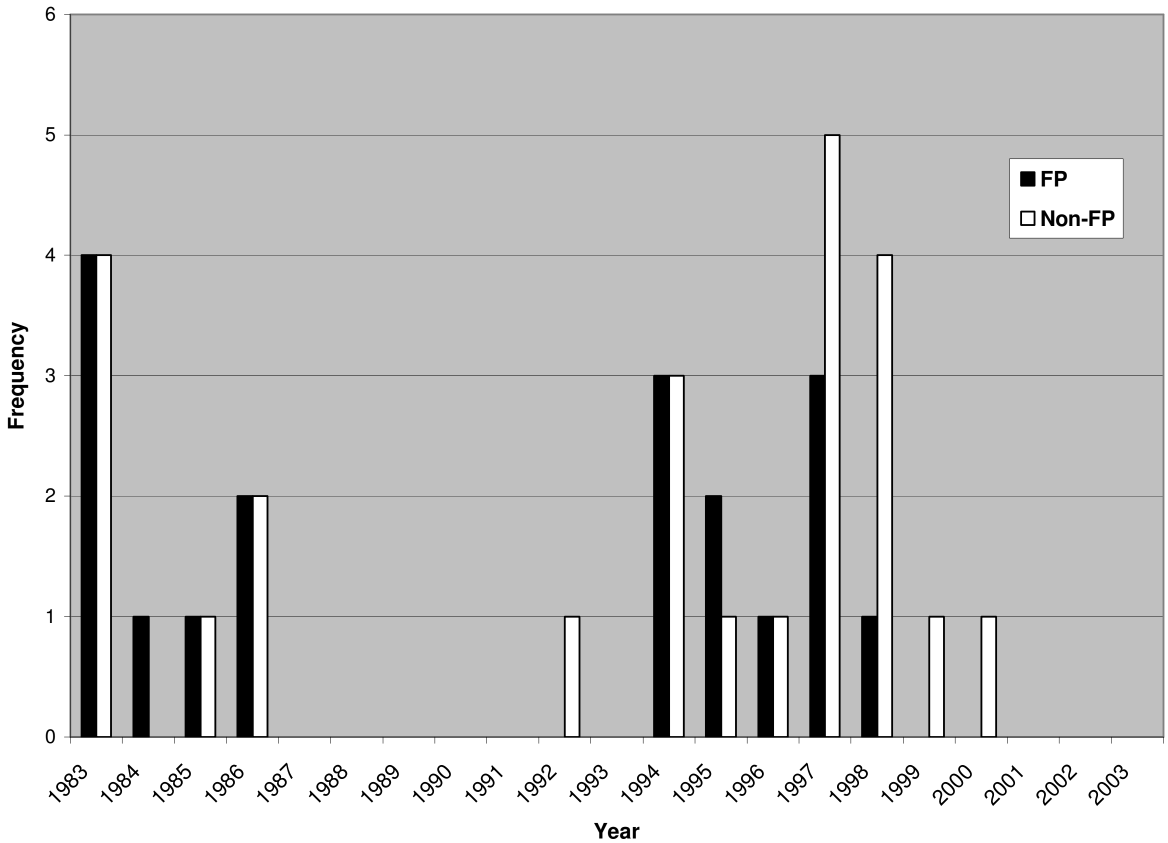

Figure 2.8: 1983-1989 and 1990-2003 Clusters by Year, 5% Spatial Bound

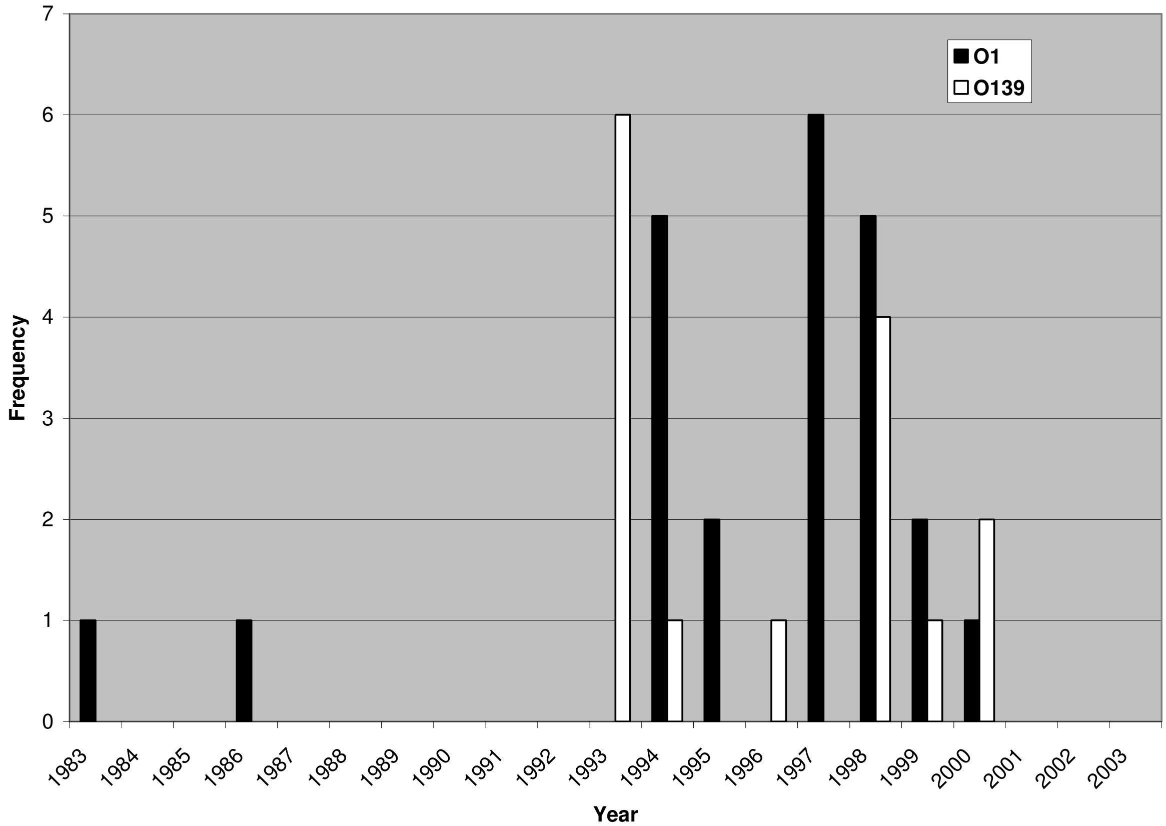

Figure 2.9: O1 and O139 Clusters by Year, 1983-2003

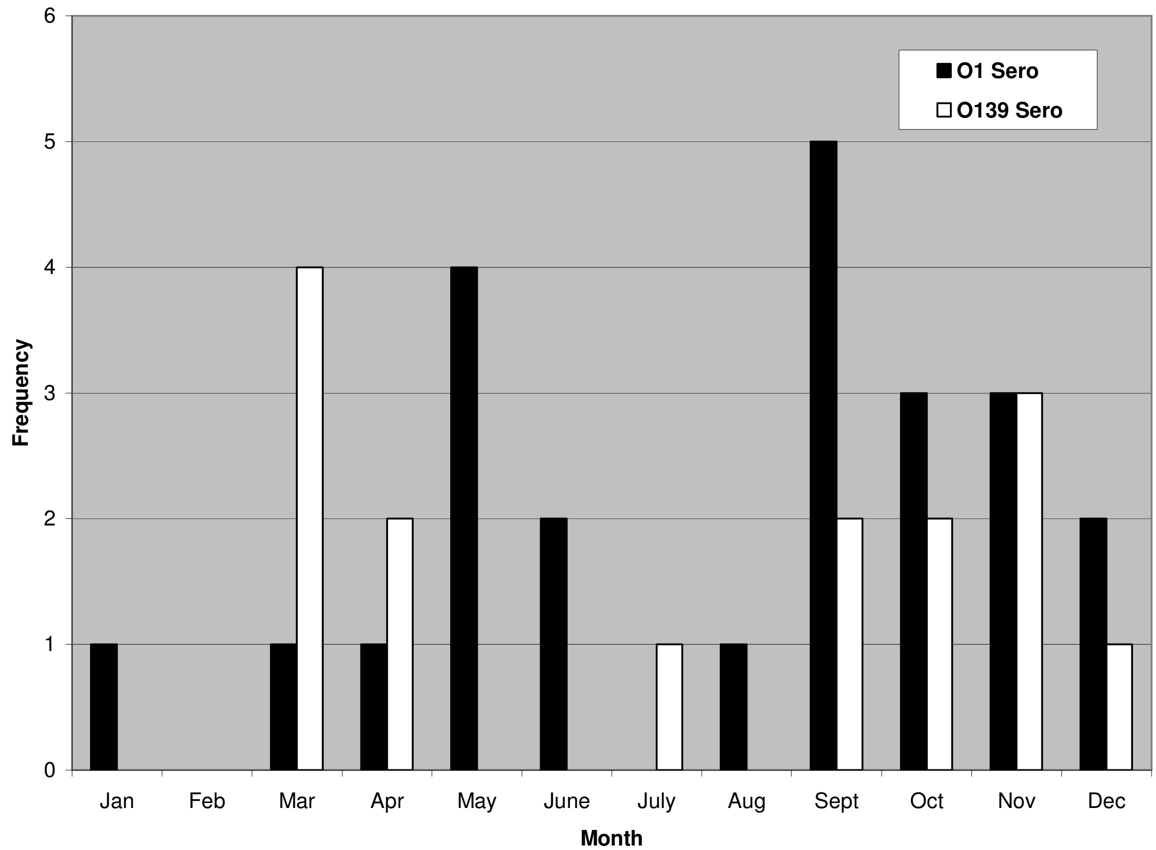

Figure 2.10: O1 and O139 Clusters by Month, 1983-2003

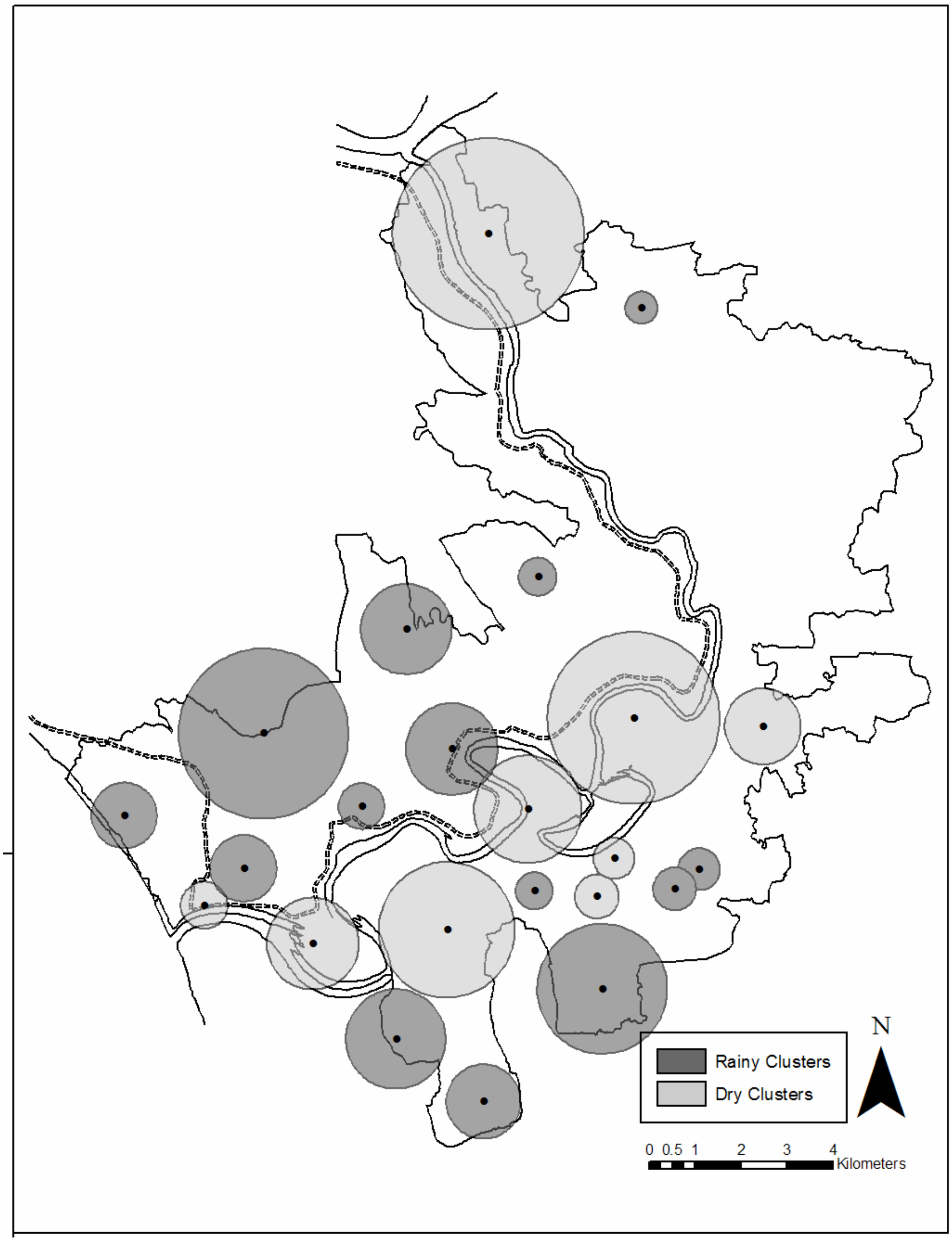

Figure 2.11: O1 Clusters, 1983-2003

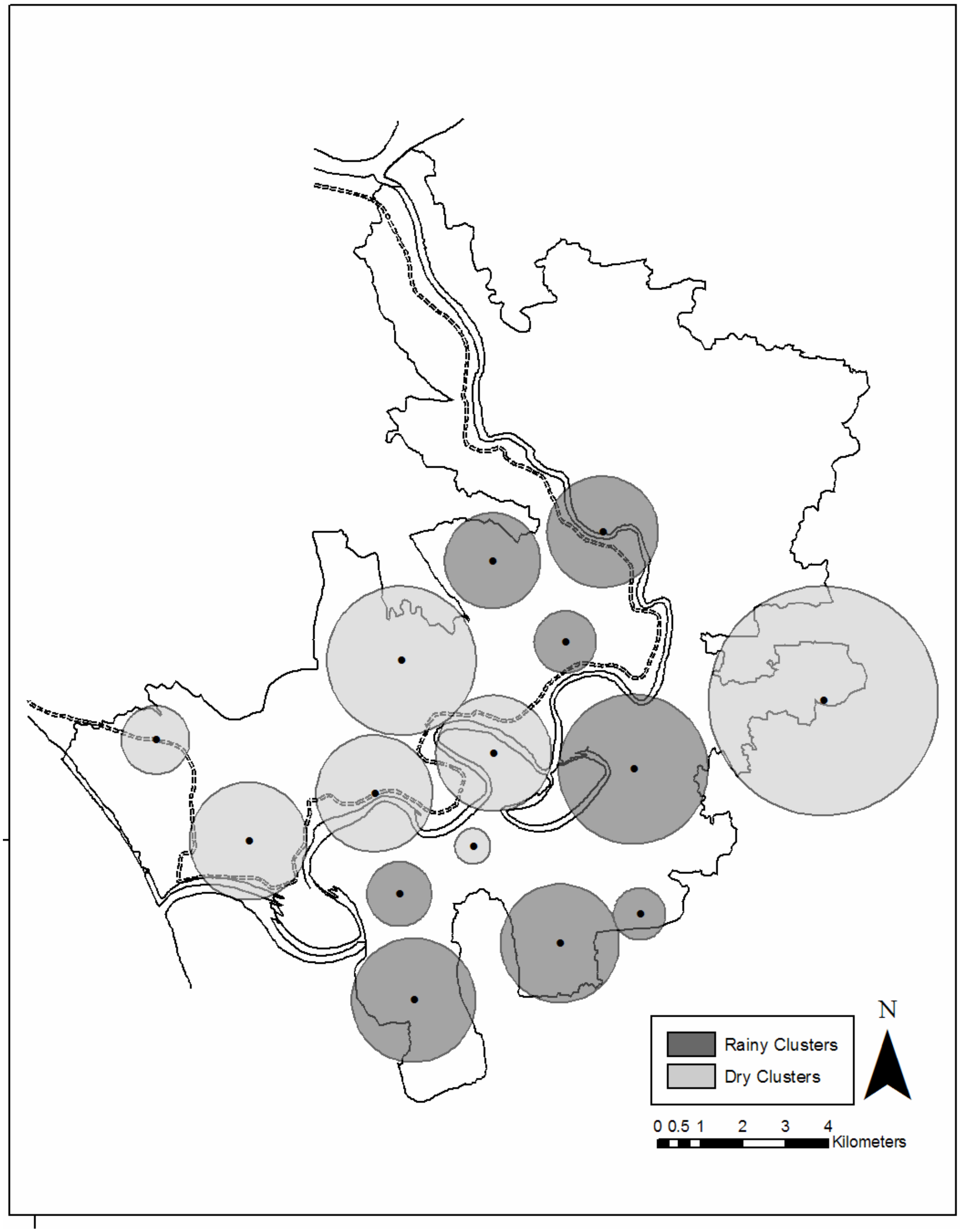

Figure 2.12: O139 Clusters, 1983-2003

References

Aamodt, G., Samuelsen, S.O. & Skrondal, A. 2006, "A simulation study of three methods for detecting disease clusters", *International journal of health geographics,* vol. 5, pp. 15.

Ali, M., Emch, M., Ashley, C. & Streatfield, P.K. 2001, "Implementation of a medical geographic information system: concepts and uses", *Journal of health, population, and nutrition,* vol. 19, no. 2, pp. 100-110.

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,* vol. 8, no. 3, pp. 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 (1982),* vol. 55, no. 6, pp. 1015-1024.

Ali, M., Emch, M., von Seidlein, L., Yunus, M., Sack, D.A., Rao, M., Holmgren, J. & Clemens, J.D. 2005, "Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis", *Lancet,* vol. 366, no. 9479, pp. 44-49.

Ali, M., Emch, M., Yunus, M. & Sack, R.B. 2001, "Are the environmental niches of Vibrio cholerae O139 different from those of Vibrio cholerae O1 El Tor?", *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases,* vol. 5, no. 4, pp. 214-219.

Ansary, S., Fulton, L., Bhuiya, A. & Chowdury, M. 1997, *An Impact Evaluation of the Meghna-Dhonagoda Embankment*, International Centre for Diarrhoeal Disease Reasearch, Bangladesh (ICDDR,B) & the Bangladesh Rural Advancement Committee (BRAC), Dhaka, Bangladesh.

Bennish, M.L., Harris, J.R., Wojtyniak, B.J. & Struelens, M. 1990, "Death in shigellosis: incidence and risk factors in hospitalized patients", *The Journal of infectious diseases,* vol. 161, no. 3, pp. 500-506.

Colwell, R.R., Huq, A., Islam, M.S., Aziz, K.M., Yunus, M., Khan, N.H., Mahmud, A., Sack, R.B., Nair, G.B., Chakraborty, J., Sack, D.A. & Russek-Cohen, E. 2003, "Reduction of cholera in Bangladeshi villages by simple filtration", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, no. 3, pp. 1051-1055.

Emch, M., Ali, M. & Yunus, M. *Risk areas and neighborhood-level risk factors for Shigella dysenteriae 1 and Shigella flexneri: Implications for vaccine development*.

Emch, M. 2000, "Relationships between flood control, kala-azar, and diarrheal disease in Bangladesh", *Environment and Planning A,* vol. 32, pp. 1051-1063.

Emch, M. & Ali, M. 2003, "Spatial Cluster Analysis for Etiological Research and Identification of Socio-environmental Risk Factors" in *Geographic information systems and health applications*, eds. O.A. Khan & R. Skinner, Idea Group Pub., Hershey, PA, pp. 172-187.

Emch, M. 1999, "Diarrheal disease risk in Matlab, Bangladesh", *Social science & medicine (1982),* vol. 49, no. 4, pp. 519-530.

Fang, Z., Kulldorff, M. & Gregorio, D.I. 2004, "Brain cancer mortality in the United States, 1986 to 1995: a geographic analysis", *Neuro-oncology,* vol. 6, no. 3, pp. 179-187.

Faruque, S.M., Naser, I.B., Islam, M.J., Faruque, A.S., Ghosh, A.N., Nair, G.B., Sack, D.A. & Mekalanos, J.J. 2005, "Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 5, pp. 1702-1707.

Gosselin, P., Lebel, G., Rivest, S. & Douville-Fradet, M. 2005, "The Integrated System for Public Health Monitoring of West Nile Virus (ISPHM-WNV): a real-time GIS for surveillance and decision-making", *International journal of health geographics,* vol. 4, pp. 21.

Hughes, C.C. & Hunter, J.M. 1970, "Disease and "development" in Africa", *Social science & medicine (1982),* vol. 3, no. 4, pp. 443-493.

Hunter, J.M. 2003, "Inherited burden of disease: agricultural dams and the persistence of bloody urine (Schistosomiasis hematobium) in the Upper East Region of Ghana, 1959-1997", *Social science & medicine (1982),* vol. 56, no. 2, pp. 219-234.

Hunter, J.M., Rey, L. & Scott, D. 1982, "Man-made lakes and man-made diseases. Towards a policy resolution", *Social science & medicine (1982),* vol. 16, no. 11, pp. 1127-1145.

Huq, A., Sack, R.B., Nizam, A., Longini, I.M., Nair, G.B., Ali, A., Morris, Jr., J. Glenn, Huda Khan, M.N., Siddique, A.K., Yunus, M., Albert, M.J., Sack, D.A. & Colwell, R.R. 2005, "Critical Factors Influencing the Occurrence of *Vibrio cholerae* in the Environment of Bangladesh", *Applied and Environmental Microbiology,* vol. 71, no. 8, pp. 4645-4654.

Keiser, J., De Castro, M.C., Maltese, M.F., Bos, R., Tanner, M., Singer, B.H. & Utzinger, J. 2005, "Effect of irrigation and large dams on the burden of malaria on a global and regional scale", *The American Journal of Tropical Medicine and Hygiene,* vol. 72, no. 4, pp. 392-406.

Kulldorff, M. & Information Management Systems 2006, *SaTScan™7.0: Software for the spatial and space-time scan statistics*.

Kulldorff, M. 1997, "A spatial scan statistic", *Communications in Statistics: Theory and Methods,* vol. 26, no. 6, pp. 1481-1496.

Kulldorff, M. & Nagarwalla, N. 1995, "Spatial disease clusters: detection and inference", *Statistics in medicine,* vol. 14, no. 8, pp. 799-810.

Longini, I.M.,Jr, Yunus, M., Zaman, K., Siddique, A.K., Sack, R.B. & Nizam, A. 2002, "Epidemic and endemic cholera trends over a 33-year period in Bangladesh", *The Journal of infectious diseases,* vol. 186, no. 2, pp. 246-251.

Myaux, J., Ali, M., Felsenstein, A., Chakraborty, J. & de Francisco, A. 1997, "Spatial distribution of watery diarrhoea in children: identification of "risk areas" in a rural community in Bangladesh", *Health & place,* vol. 3, no. 3, pp. 181-186.

Polack, S.R., Solomon, A.W., Alexander, N.D., Massae, P.A., Safari, S., Shao, J.F., Foster, A. & Mabey, D.C. 2005, "The household distribution of trachoma in a Tanzanian village: an application of GIS to the study of trachoma", *Transactions of the Royal Society of Tropical Medicine and Hygiene,* vol. 99, no. 3, pp. 218-225.

Singh, N., Mehra, R.K. & Sharma, V.P. 1999, "Malaria and the Narmada-river development in India: a case study of the Bargi dam", *Annals of Tropical Medicine and Parasitology,* vol. 93, no. 5, pp. 477-488.

Song, C. & Kulldorff, M. 2003, "Power evaluation of disease clustering tests", *Int.J.Health.Geogr.,* vol. 2, no. 1, pp. 9.

Sow, S., de Vlas, S.J., Engels, D. & Gryseels, B. 2002, "Water-related disease patterns before and after the construction of the Diama dam in northern Senegal", *Annals of Tropical Medicine and Parasitology,* vol. 96, no. 6, pp. 575-586.

Waddy, B.B. 1975, "Research into the health problems of manmade lakes, with special reference to Africa", *Transactions of the Royal Society of Tropical Medicine and Hygiene,* vol. 69, no. 1, pp. 39-50.

Chapter 3

Bridge

The results of the cluster analysis indicate that cholera did cluster spatio-temporally in Matlab both before and after completion of the MDIP, but that patterns of clustering had shifted after the introduction of flood protection. These results were used to inform further exploration of longitudinal cholera incidence via spatial regression analyses at the neighborhood level. The outcome of the SaTScan cluster analysis suggested that seasonality of cholera incidence was an important variable to take into account when considering the impact of neighborhood on cholera outcomes, since there had been a significant shift in seasonality of cluster occurrence. The cluster results also indicated that a *bari's* distance to the Dhonagoda River was an important variable to consider in addition to the flood protection status of a *bari*. The size of the clusters, even with the artificial spatial and temporal bounds imposed, suggested that examining small neighborhood sizes would be most appropriate, as the majority of clusters had radii of less than 2000 meters. We thus chose to create spatial lags at 500m, 1000m and 2000m in order to test whether the spatio-temporal patterns observed pre- and post-MDIP in the cluster analysis would be seen in the results of the regression analysis.

Chapter 4

The impact of flood protection on spatial autocorrelation of cholera relationships in Matlab, Bangladesh.

Abstract:

We examined how the spatial autocorrelation of relationships between two environmental variables, river proximity and flood protection status, and cholera incidence outcomes, stratified by cholera strain and season, varied at three neighborhood scales. Ordinary Least Squares (OLS) and spatially-lagged regression analysis was performed on a cholera dataset drawn from 1983-2003 records from Matlab, Bangladesh. Stronger increases in model fit observed in the 1990-2003 data than 1983-1989 data under spatial lags suggests that the division of the study area into two sections, flood protected and unprotected, increased the spatial autocorrelation of all tested relationships. Seasonally stratified results indicated that dry season cholera is more spatially autocorrelated than rainy season cholera, regardless of time period. Stratification by cholera strain showed evidence that older and more established O1 cholera exhibits higher spatial autocorrelation in relationships with environmental variables than does newer O139 cholera. Significance tests also suggest that certain relationships retain their statistical significance once spatial autocorrelation of relationships is accounted for while others do not, suggesting that introducing flood protection has affected not only spatial patterns of relationships but strengths of association as well.

Keywords: spatial autocorrelation, cholera, Bangladesh, flood protection

Introduction

Cholera, all but eradicated in the developed world, remains a significant cause of morbidity and mortality in the developing world. Although present at low levels year-round in places such as Bangladesh, large outbreaks occur in seasonal cycles. The exact mechanism(s) that lead to these consistent peaks and valleys in cholera incidence are not known, but several strong relationships have been found between the arrival of monsoon rains and copepod, algae, and bacteriophage levels (Jensen et al., 2006; Faruque et al., 2005; Huq et al., 2005; Colwell et al., 2003; Islam et al., 1993). Social/behavioral and environmental factors have also been associated with cholera incidence, such as handwashing, latrine-type, educational status, etc (Ali et al., 2002; Emch, 1999; Myaux et al., 1997). By determining variables that are necessary precursors to epidemic cholera, as well as those that are associated with increased opportunity for infection, public health and educational interventions can be better targeted.

Increasingly used in public health, epidemiology and medical geography studies is the theory of neighborhood effects on health outcomes. Neighborhoods contain both compositional, i.e. demographic, and contextual, i.e. environmental, factors that interact to influence the health outcomes of local populations. Neighborhood studies operate under the premise that some variables which modify health outcomes take place only on micro-scales, interacting with other variables at higher or lower spatial scales (Ali et al., 2005). By studying these contextual and compositional factors as they relate to one another across scales, a localized health outcome can be better understood or predicted. The environmental and social variables that have been associated with cholera incidence are not static in space and time, but likely have different influences at different times and distances from households. In other words, maybe neighborhood influences on cholera are more important to consider in certain times or certain places. For instance, perhaps the proximity of a household to a main river influences cholera outcomes only up to a kilometer, and then that influence drops off, or the location of a neighborhood within a flood protected area influences cholera outcomes only in the rainy season.

Variation in time and space of social and environmental factors aside, neighborhood associations are useful when studying cholera given the nature of the agent and disease symptoms. Cholera is a highly infectious bacterial disease whose agent thrives in shared water supplies, and an infection of one person within a neighborhood can influence infection of surrounding people. The definition of precisely what geographic (or other) area a neighborhood comprises remains open to debate. Suggested methods of definition include individuals' activity space, social networks or areas around households at set Euclidean distances. Cholera in Bangladesh has been studied using Euclidean distance neighborhoods, with between ½ and 2 kilometers found to be appropriate levels of analyses (Ali et al., 2007 *in press*; Emch et al., 2007; Emch et al., 2006).

In the rural area of Matlab, Bangladesh, two environmental variables, residence within a flood protected sector and distance to the main river, have been found to influence the incidence of cholera (Ali et al., 2002a; Ali et al., 2002b; Emch, 1999). Our study further examined the impact of these two variables on cholera incidence. We were interested in determining whether the relationships existed between flood protection and cholera and river distance to household and cholera, whether these relationships were spatially autocorrelated, at what neighborhood scales the relationships were most spatially autocorrelated, and whether levels of spatial autocorrelation had changed since the introduction of flood protection. Cholera outcomes were then stratified by season of incidence and by cholera strain to explore whether occurrence of cholera exhibits differing levels of spatial autocorrelation according to season and if taking cholera type into account changes measures of spatial dependence. We hypothesized that spatial autocorrelation of cholera relationships would increase after the introduction of flood protection, particularly in the rainy season when neighborhoods are more likely to share a flood or non-flood experience, but that stratifying by cholera strain would not show significant differences in levels of spatial autocorrelation. By exploring measures of spatial autocorrelation, we were able to understand whether the changing explanatory power of the two environmental variables post-flood protection was due to in-*bari* variation or to differences at larger scales. In addition to exploring changing levels of spatial autocorrelation, the statistical significance of all relationships was calculated in order to assess which associations between cholera incidence and flood protection and river proximity were important once spatial autocorrelation was accounted for.

Study Area and Data Sources

Our study utilized data collected in Matlab, Bangladesh. Matlab is a small, rural area located approximately 50km southeast of the capital city of Dhaka (Figure 1). Population density is high, with roughly 200,000 people living in Matlab's 184km^2 (Ali et al., 2002a). Households in Matlab are organized in a system of patrilineally-related households called *baris*. The average household size is 5.6 individuals (Colwell et al., 2003) and the average number of households in a *bari* is six (Ali et al., 2002a). The majority of residents are engaged in agricultural or day labor, with 70% of households landless or functionally landless in terms of farmland, a shift from previous decades when the majority of households engaged in self-supporting agricultural production (ICDDR,B, 2007b). Between 1974 and 2005 ownership of four socioeconomic indicators (blanket, hurricane lamp, radio and a television) rose dramatically, as did the use of tin roofs and walls in household construction (ICDDR, B, 2007a). Despite such increases in socioeconomic status and 90% of households using tubewell water in 2005, diarrheal diseases, including cholera and shigellosis, persist in Matlab (ICDDR,B, 2005b).

The causative agent of cholera, *Vibrio cholerae* bacteria, are naturally occurring in the estuaries and waterways of the Indian subcontinent (Faruque et al., 2005; Jensen et al., 2006). Epidemic *V. cholerae* are classified according serogroup and biotype. Two serogroups exist, O1 and O139; the O1 serogroup is subdivided into two biotypes, Classical and El Tor. Classical and El Tor cholera cocirculated in Matlab until 1988 when Classical disappeared. El Tor was responsible for all cholera infections in Matlab until 1993, when O139 cholera first appeared and began to cocirculate with El Tor (Longini et al., 2002). Regardless of which strain of *V. cholerae* is predominant, outbreaks of cholera epidemics occur in regular seasonal cycles in Matlab. A small peak in incidence takes place in late spring to early summer at the end of the dry season, a larger peak occurs in late fall and early winter, at the end of the rainy season (Faruque et al., 2005; Huq et al., 2005).

Running from north to south through the Matlab study area is the Dhonagoda River. The Dhonagoda experiences flooding annually during monsoon rains. In the 1980s the Government of Bangladesh started work on the Meghna-Dhonagoda Irrigation Project (MDIP), intended to protect residents from the devastating impact of annual monsoon flooding and in an effort to increase agricultural production: the MDIP allows triple-cropping and the planting of high yield varieties of rice (Vaughn, 1997). Completed in 1989, the main feature of the MDIP is a 60km earthen embankment that encircles approximately 17,000 hectares (ICDDR,B, 1992). The embankment roughly follows the northern bank of the Dhonagoda and protects the northern portion of Matlab from seasonal river flooding.

Since 1966, the Matlab study area has been under demographic and health surveillance by the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). ICDDR,B's Health and Demographic Surveillance System (HDSS) consists of active demographic and passive health surveillance. At birth or migration into Matlab each individual is issued an identification number that links them to their household, *bari* and village. ICDDR,B-trained community health workers (CHWs) visit each *bari* twice a month and record vital events such as births, deaths, marriages and migrations, linking each event to a resident's file based on their identification number (Emch, 1999). The CHWs also inquire about individuals' health within the *bari*, and make recommendations for treatment at either a hospital in Matlab or one of four health subcenters, all run by ICDDR,B. Transportation to and treatment at an ICDDR,B facility is free, so all cases of cholera within Matlab can be assumed to have been recorded (Ali et al., 2002a). A patient undergoing treatment for cholera has their stool laboratory-tested, and a serogroup and biotype are attached to their record along with date of treatment (Ali et al., 2002b; Longini et al., 2002). All cholera cases can thus be assigned to a month of incidence and location of residence.

We created a database of all cholera cases in Matlab between January 1, 1983 and December 31, 2003. This twenty-one year time frame spans the seven years prior to the completion of the MDIP and the fourteen years following. For 7490 *baris* we had data on 9580 total cholera cases between 1983-1989 and 1990-2003, as well as cholera cases subdivided according to serogroup and biotype. Cases were also grouped according to season, with cases taking place between June and November defined as rainy season cases and those taking place between December and May defined as dry season cases. From the demographic surveillance we were able to calculate mid-year *bari* populations, which were then summed for the years of 1983-1989 and 1990-2003. We split the data into these two time periods in order to compare cholera pre- and post-flood protection. Cholera counts (total, type and season) for each *bari* were then divided by total *bari* populations over the two time periods and multiplied by 1000 to gain a cholera rate per 1000 person years for 1983-1989 and 1990-2003.

Cholera cases and the population living in each *bari* were then linked to a specific geographic location in Matlab's Geographic Information System (GIS) in order to calculate two environmental variables. Matlab's GIS is accurate within 10m and includes *bari*, river, MDIP embankment and health center features (Ali et al., 2001; Emch, 1999). Using the GIS, each *bari* was assigned a flood protection status in the database, either 1 for protected or 0 for unprotected. *Baris* in the 1983-1989 dataset were assigned the 1/0 flood protection status as an indicator of their future condition, allowing us to explore whether there were inherent differences between the flood protected and unprotected areas of Matlab prior to their division by the MDIP. Using the ArcGIS software a Euclidean distance to the Dhonagoda River was calculated for each *bari*.

Methods

Our objective was to examine whether correlative interactions, stratified by season and strain, between flood protection status and river proximity and cholera incidence were spatially dependent. Statistical analyses were performed in GeoDa™0.9.5-I, which is intended for exploratory spatial data analysis (ESDA), including calculation of local autocorrelation statistics. A GIS layer of Matlab's *baris* with their attendant cholera incidence and environmental measures was imported into GeoDa. Using this layer, we created three neighborhood scales, corresponding to spatial weights at Euclidean distances of 500m, 1000m and 2000m. Spatial weights represent possible spatial interaction among data points (Anselin et al., 1996). Threshold distances rather than nearest neighbors were implemented because of the spatial heterogeneity of *bari* locations in Matlab and because geographic distance is an important modifying variable in an environmentally responsive agent such as *V. cholerae* (Emch et al., 2006). Distances up to 2000m were chosen because local neighborhood effects were of interest rather than the influence of cholera incidence at greater distances.

We at first conducted ordinary least square (OLS) regressions for all the models. Measures in the OLS output indicate when spatial dependence is present and that a spatial lag model would be a better fit for the data, including Moran's I z-value and Lagrange Multiplier (LM) tests for lag and error (Anselin, 2004). It was anticipated that the majority of models created would demonstrate signs of spatial dependence and indicate that a spatial lag would improve fit, given the highly infectious nature of cholera and its tendency to spatio-temporally cluster. We then created spatial lags at 500m, 1000m and 2000m and re-ran the models at all three scales. The spatial lag models, based on a maximum likelihood method, account for spatial dependence by including an autoregressive term of the dependent variable (Anselin et al., 2006). All spatial dependence effects are captured by the lagged term.

Three types of approaches are commonly used to compare models: maximizing fit (adjusted R^2), null hypothesis tests (likelihood ratios) and model selection criteria (Akaike information criterion (AIC) & Schwarz criterion (SC)). Adjusted R^2 is an indication of fit, likelihood ratios and AIC indicate fit and complexity, and SC indicates fitness, complexity and sample size (Johnson & Omland, 2004). Although R^2 is not an appropriate statistic to rely upon in a spatial lag model, the log-likelihood ratio (LLR), AIC and SC are. An increase in the log-likelihood value and a decrease in both AIC and SC between the spatial lag model and the OLS suggests that there is an improved fit by using the lagged model. Levels of spatial autocorrelation present in each relationship were assessed by comparing LLR, AIC and SC values in the OLS and spatial lag models. Greater gains in LLR and falls in AIC and SC were taken as an indication that higher levels of spatial autocorrelation were present in certain relationships than in others. The statistical significance of the relationships at each spatial lag was determined by the probability measure associated with the z-score of the independent variable. Our purpose in examining this measure was to ascertain which relationships remained significant once spatial autocorrelation was accounted for.

Our initial analysis used the 1983-1989 and 1990-2003 cholera rates per 1,000 person-years-lived as dependent variables to explore whether place of residence and proximity to the river had similar levels of spatial autocorrelation before and after the introduction of flood protection. Also modeled as explanatory variables for overall cholera outcomes were the rainy and dry season cholera rates for each timeframe, to test whether occurrence of cholera in one season over the other was a stronger predictor of spatial autocorrelation in overall cholera incidence rates.

We then used the 1983-1989 and 1990-2003 rainy and dry season cholera rates as dependent variables to explore whether the spatial autocorrelation associated with relationships between cholera incidence and flood protection and river distance differed by season. We also tested whether rainy or dry season cholera in 1990-2003 had highly spatially autocorrelated relationships with 1983-1989 rainy or dry season cholera, exploring the possibility that previous cholera history was of stronger influence than either flood protections status or river distance. The last of our analyses focused on exploring differences in levels of spatial autocorrelation based on cholera strain. Rates of Classical, El Tor, O1 and O139 cholera divided by time period were used as dependent variables, while flood protection status, river distance and seasonal rates were explored as effect modifiers.

Results

In the overall cholera rate analyses the 500m and 1000m lag models were better fits for the data while the 2000m lag was a worse fit, as determined by rises in the LLR and decreases in both the AIC and SC. Greater improvements in fit can be taken as indicators of higher levels of spatial autocorrelation present in the tested relationships, while a decrease in model fitness at a spatial lag indicates a lack of spatial autocorrelation to be controlled for. There were greater gains in model fitness for all tested relationships at the 1000m spatial lag than the 500m spatial lag, suggesting that spatial autocorrelation of cholera is strongest at 1000m in Matlab. For every relationship modeled, the 1990-2003 data showed greater improvements in fit than the 1983-1989 data when a spatial lag was implemented (Table 1), indicating that cholera relationships post-MDIP construction have greater levels of spatial autocorrelation. Each of the four1990-2003 relationships retained statistical significance between the OLS and 1000m spatial lag model, while the 1983-1989 relationships between cholera and flood protection and cholera and river distance were insignificant at the 1000m spatial lag.

Results from the seasonal cholera analyses were similar to those of the overall analyses in that the 500m and 1000m spatially lagged models better fit the cholera relationships tested, with the 1000m model showing the greatest improvement in fit over the OLS regression. In contrast to the large gains in model fitness seen with a 1000m spatial lag, the 2000m spatially lagged seasonal models were a worse fit for the data as compared to the OLS. At both the 500m and 1000m lags, the greatest spatial autocorrelation was seen in 1990-2003 dry season cholera models, regardless of independent variable tested (Table 2), followed closely by the 1990-2003 rainy season cholera. Much smaller levels of spatial autocorrelation were shown for the 1983-1989 dry and rainy season rate relationships, with the 1983-1989 rainy season models seeing the smallest enhancement by implementing a spatial lag.

Although the 1990-2003 models evidenced greater spatial autocorrelation than did the 1983-1989 models, the pre-flood protection cholera relationships were more statistically significant after a spatial lag was implemented. The only 1990-2003 relationship that remained statistically significant under the 1000m spatial lag model was that between the rainy season rate and flood protection status. In contrast, half of the 1983-1989 seasonal cholera relationships were significant: two which explored the prediction power of one season's cholera rate on another, the other which modeled the dry season rate and river distance relationship.

Stratifying cholera by strain returned the same results as stratifying by season: better model fits at 500m and 1000m, but worse fits at 2000m. Once again, the 1990-2003 cholera rates saw the greatest amounts of spatial autocorrelation controlled for by the spatial lag models (Table 3). Within the 1990-2003 models, greatest improvements in fit by implementing the 1000m spatial lag were shown for overall O1 cholera relationships (Classical & El Tor combined), then for El Tor alone and lastly O139 cholera, indicating that higher levels of spatial autocorrelation exist among O1 cholera relationships than O139.

The statistical significance of the strain stratified models, once spatial autocorrelation was accounted for, varied greatly between the pre-MDIP and post-MDIP periods. The majority of 1990-2003 models were statistically significant even when spatially lagged. The only statistically insignificant model in 1990-2003 was that of O139 cholera's interaction with river distance. In contrast, the majority of 1983-1989 models were statistically insignificant once spatially lagged. The only significant models in 1983-1989 were those representing the relationship between seasonal rates and total O1 cholera. In fact, all of the models which used the seasonal rates of cholera as independent variables were highly statistically significant ($p<.05$). However, temporal differences in levels of spatial autocorrelation exist between the pre-MDIP and post-MDIP periods. In the 1983-1989 period, O1 cholera was most spatially autocorrelated in its relationship to the rainy season. Between 1990-2003, however, both the O1 and O139 cholera relationships with the dry season were most spatially autocorrelated.

Discussion & Conclusions

For all three sets of data, the overall cholera rates, the cholera rates by season and the cholera rates by strain, implementing a 500m spatial lag model resulted in a better fit of the data. Increasing the spatial lag to 1000m resulted in an even greater increase in fit for all three datasets. Increasing the spatial lag to 2000m, however, reversed this trend: the classic OLS model better fit the data than the 2000m spatial lag model. This suggests that within

Matlab spatial autocorrelation of cholera relationships increases between .5 and one kilometer, but then falls off at two kilometers. The 1983-1989 and 1990-2003 cholera datasets responded to the spatial lags in the same way, indicating that this distance effect was not impacted by the introduction of flood protection in 1989, but that spatial autocorrelation of cholera relationships in Matlab has remained highly localized, strongest at less than 2 kilometers, since 1983.

Across all three categories of analysis (overall cholera rates, seasonal rates, strain rates) the greatest improvements in fit by implementing spatially lagged models were seen in 1990-2003 outcome variables. Although there were improvements in fit for the 1983-1989 outcome variables, they were much smaller than for the time period after implementation of the MDIP flood protection. This suggests that, regardless of stratification, relationships between environmental variables and cholera incidence became more spatially autocorrelated across Matlab after MDIP construction. Spatial autocorrelation is an indication that events being examined are not occurring independently, but rather are being affected by other neighboring events (in this case most strongly at 1km). The consistency with which the 1990-2003 data showed signs of greater spatial autocorrelation means that introducing flood protection to Matlab in 1989 increased the influence that one *bari*'s cholera relationships has on other *bari*'s cholera relationships. This is likely due to increased importance of shared neighborhood experiences once the MDIP is completed and the study area is divided into two radically different sections.

In the overall cholera analysis, flood protection status and river distance were more spatially autocorrelated with cholera outcomes in both time periods than were rainy or dry season rates. Although seasonal cholera was less influential on a neighborhood scale than were the two environmental variables, all the models which used temporal independent variables were statistically significant, indicating that both dry and rainy season cholera are important in determining overall cholera. In contrast, the environmental variables of flood protection and river distance remained statistically significant variables only in the spatial lag models post-1990, not in the 1983-1989 period. Thus, even as spatial autocorrelation has increased post-MDIP, so has the influence of these two environmental variables in predicting cholera outcomes. Increases in both spatial autocorrelation and the importance of these environmental factors has implications for public health targeting when cholera outbreaks occur in Matlab. Not only should *baris* within 1 kilometer of a choleric household be monitored and warned, those *baris* that share flood protection status or proximity to the Dhonagoda River should be made particularly aware.

When stratified by season, cholera relationships in both time periods were more spatially autocorrelated in the dry season than the rainy, regardless of independent variable. The seasonality of cholera in Matlab means that there are traditionally fewer cases occurring in the dry season than in the rainy season, when incidence peaks as flood waters rise. The higher levels of spatial autocorrelation seen in dry season cholera models are thus likely due to the fact that those cases which do occur in dry season are likely to initiate infection in other households which are in close proximity rather than those that are far away: there is no flood water to distribute the bacteria across great distance. The importance of water sources to spread cholera is demonstrated by the fact that, prior to 1990, it was a *bari*'s proximity to the Dhonagoda that was most statistically significant in determining its cholera experience in the dry season.

Although spatial autocorrelation of cholera is strongest in the dry season, the only statistically significant relationship in the seasonally stratified 1990-2003 models was between rainy season cholera and flood protection status, even though it had the smallest level of spatial autocorrelation. The relationship between rainy season cholera and flood protection status in 1983-1989 also exhibited low spatial autocorrelation, but the relationship was not statistically significant. This means that changes in cholera relationships seen after 1990 can be attributed to MDIP completion rather than to basic differences in cholera patterns between the areas north and south of the Dhonagoda River. Introducing flood protection in 1990 thus created a new, highly significant variable in determining *bari* cholera outcomes.

Our analysis of cholera stratified by strain showed that O1 cholera is more spatially autocorrelated in its relationships to environmental variables and seasonal patterns than is O139 cholera. O1 cholera relationships are also more often statistically significant when spatially lagged in the 1990-2003 period than the 1982-1989 period. These higher levels of spatial autocorrelation and statistical significance among O1 cholera post-MDIP completion indicate that the older, more "traditional" types of cholera seen in Matlab reacted strongly to the introduction of flood protection. Lower levels of spatial autocorrelation exhibited by O139 cholera are perhaps a factor of its relatively recent arrival in Matlab and typically greater virulence, as its incidence and spread is less dependent on scale effects. Lower overall spatial autocorrelation aside, however, O139 incidence is still mediated by flood protection status as demonstrated by the strong significance ($p<.05$) of the model. These results, in conjunction with those of the overall and the seasonally stratified analyses, indicate that introducing flood protection in 1989 had large impacts both on spatial autocorrelation of cholera relationships and on the statistical significance of those relationships.

The models in this study typically had low R^2 values, indicating that the independent variables used did not have strong explanatory power for the outcome variables tested. This is logical, given that cholera results from a complex interaction of several environmental, social and behavioral variables and that no one variable completely explains cholera incidence. We do not suggest that the relationships uncovered between cholera incidence and neighborhood sizes are the whole story, but do indicate that cholera is subject to variation in spatial autocorrelation by both season and strain type.

The disappearance of Classical cholera in 1988 and the arrival of O139 cholera in 1993 add a level of ambiguity into our comparisons of scale dependent spatial autocorrelation according to serogroup and biotype. Additionally, the artificial demarcation of neighborhoods as 500m, 1000m, or 2000m boundaries around each *bari* fails to take into account the reality of environmental or social interactions actually taking place at each location. However, such a level of detail could only be understood by doing a detailed socio-environmental network analysis.

The influences of household flood protection status and river proximity on cholera incidence in Matlab, Bangladesh are spatially autocorrelated at the local level. The influence of neighboring *baris'* experiences on individual *bari* cholera outcomes has increased since the MDIP was built, suggesting that response to cholera outbreaks in rural Bangladesh should focus not only on individual affected *baris* but on closely neighboring *baris*. Our findings also suggest that considering these variables in conjunction with seasonal and cholera strain factors is important for understanding how cholera patterns have changed longitudinally in Matlab.

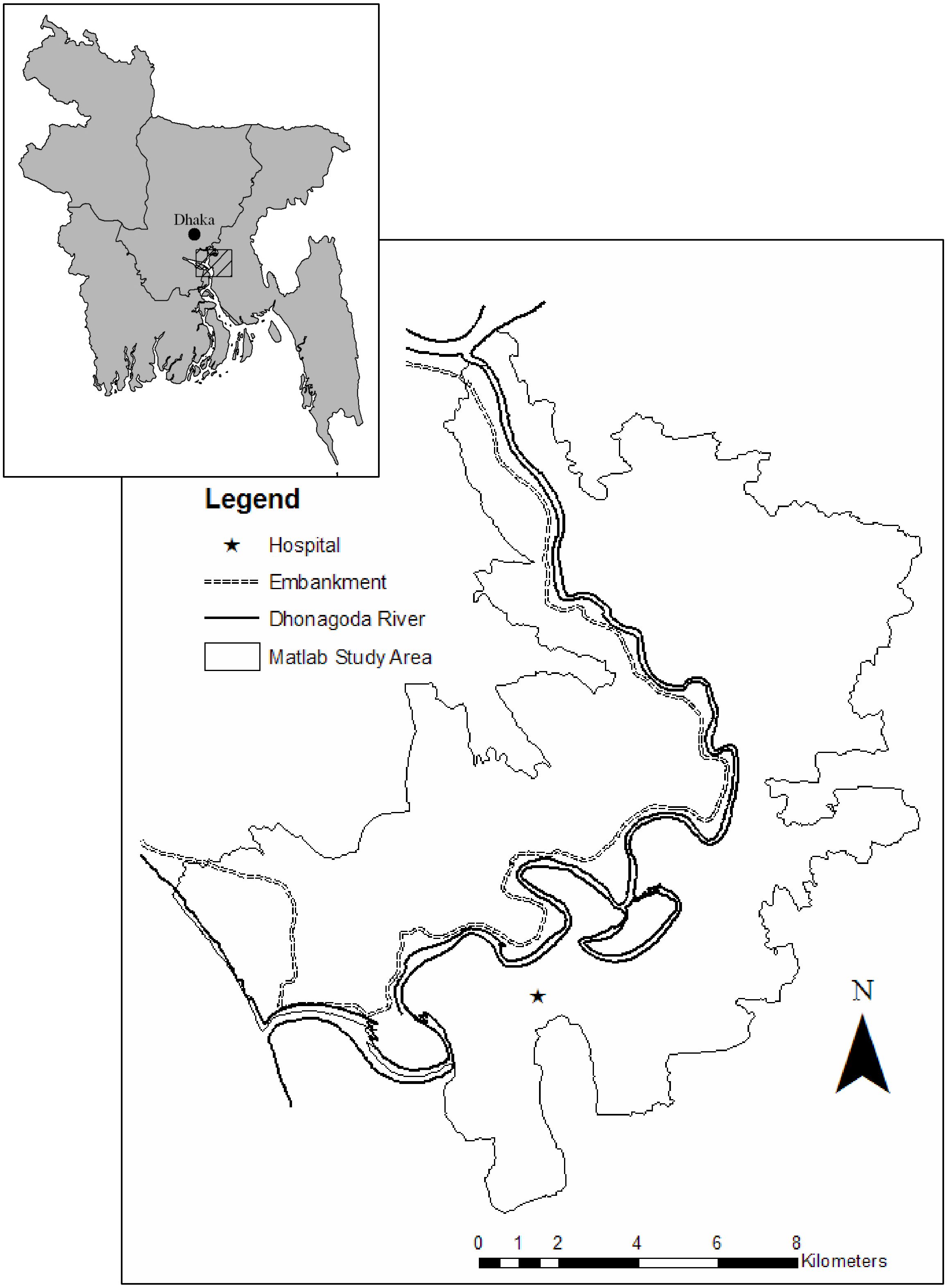

Figure 4.1: Location and Main Physical Features of Matlab, Bangladesh

Dependent Variable	Independent Variable	Sig.
1990-2003 cholera rate	river distance	*
1990-2003 cholera rate	flood protection status	**
1990-2003 cholera rate	1990-2003 dry season rate	**
1990-2003 cholera rate	1990-2003 rainy season rate	**
1983-1989 cholera rate	flood protection status	NS
1983-1989 cholera rate	river distance	NS
1983-1989 cholera rate	1983-1989 rainy season rate	**
1983-1989 cholera rate	1983-1989 dry season rate	**

Table 4.1: OLS and 1000m Spatial Lag Model Results, Overall Rates
Arranged in descending order from greatest spatial autocorrelation (greatest rise in log-likelihood ratio (LLR) and greatest drop in Akaike Information Criterion (AIC) & Schwarz Criterion (SC)) to least.
**** $p<.05$**
*** $p<.10$**

Dependent Variable	Independent Variables	Sig.
1990-2003 dry season cholera rate	flood protection status	NS
1990-2003 dry season cholera rate	1983-1989 dry season cholera rate	NS
1990-2003 dry season cholera rate	river distance	NS
1990-2003 rainy season cholera rate	river distance	NS
1990-2003 rainy season cholera rate	1983-1989 rainy season cholera rate	NS
1990-2003 rainy season cholera rate	flood protection status	**
1983-1989 dry season cholera rate	flood protection status	NS
1983-1989 dry season cholera rate	1983-1989 rainy season cholera rate	*
1983-1989 dry season cholera rate	river distance	*
1983-1989 rainy season cholera rate	river distance	NS
1983-1989 rainy season cholera rate	flood protection status	NS
1983-1989 rainy season cholera rate	1983-1989 dry season cholera rate	**

Table 4.2: OLS and 1000m Spatial Lag Model Results, Seasonal Rates Arranged in descending order from greatest spatial autocorrelation (greatest rise in log-likelihood ratio (LLR) and greatest drop in Akaike Information Criterion (AIC) & Schwarz Criterion (SC)) to least.
**** p<.05**
*** p<.10**

Dependent Variable	Independent Variable	Sig.
1990-2003 O1 cholera rate	river distance	**
1990-2003 O1 cholera rate	flood protection status	*
1990-2003 El Tor cholera rate	river distance	**
1990-2003 O1 cholera rate	1990-2003 dry season rate	**
1990-2003 El Tor cholera rate	flood protection status	**
1990-2003 O139 cholera rate	river distance	NS
1990-2003 O1 cholera rate	1990-2003 rainy season rate	**
1990-2003 O139 cholera rate	flood protection status	**
1983-1989 Classical cholera rate	flood protection status	NS
1983-1989 Classical cholera rate	river distance	NS
1983-1989 O1 cholera rate	flood protection status	NS
1983-1989 O1 cholera rate	river distance	NS
1990-2003 O139 cholera rate	1990-2003 dry season rate	**
1983-1989 O1 cholera rate	1983-1989 rainy season rate	**
1983-1989 El Tor cholera rate	river distance	NS
1983-1989 El Tor cholera rate	flood protection status	NS
1983-1989 O1 cholera rate	1983-1989 dry season rate	**
1990-2003 O139 cholera rate	1990-2003 rainy season rate	**

Table 4.3: OLS and 1000m Spatial Lag Model Results, Bacteria Strain Rate Arranged in descending order from greatest spatial autocorrelation (greatest rise in log-likelihood ratio (LLR) and greatest drop in Akaike Information Criterion (AIC) & Schwarz Criterion (SC)) to least.
**** p<.05**
*** p<.10**

References

Ali, M., Emch, M., Yunus, M., Sack, D., Lopez, A.L., Holmgren, J. & Clemens, J. 2007. Vaccination of adult women against cholera protects infants and young children in rural Bangladesh. The Pediatric Infectious Disease Journal, *in press*.

Ali, M., Emch, M., Ashley, C. & Streatfield, P.K. 2001. Implementation of a medical geographic information system: concepts and uses. Journal of Health, Population, and Nutrition, 19 (2), pp. 100-110.

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), pp. 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), pp. 1015-1024.

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

Anselin, L., Bera, A.K., Florax, R. & Yoon, M.J. 1996. Simple diagnostic tests for spatial dependence. Regional Science & Urban Economics, 26 (1), pp. 77-104.

Anselin, L., Syabri, I. & Kho, Y. 2006. GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis, 38, pp. 5-22.

Colwell, R.R., Huq, A., Islam, M.S., Aziz, K.M., Yunus, M., Khan, N.H., Mahmud, A., Sack, R.B., Nair, G.B., Chakraborty, J., Sack, D.A. & Russek-Cohen, E. 2003. Reduction of cholera in Bangladeshi villages by simple filtration. Proceedings of the National Academy of Sciences of the United States of America, 100 (3), pp. 1051-1055.

Emch, M. 2000. Relationships between flood control, kala-azar, and diarrheal disease in Bangladesh. Environment and Planning A, 32, pp. 1051-1063.

Emch, M. 1999. Diarrheal disease risk in Matlab, Bangladesh. Social science & medicine (1982), 49 (4), pp. 519-530.

Emch, M., Ali, M., Acosta, C., Yunus, M., Sack, D.A. & Clemens, J.D. 2007. Efficacy calculation in randomized trials: global or local measures?. Health & place, 13 (1), pp. 238-248.

Emch, M., Ali, M., Park, J.K., Yunus, M., Sack, D.A. & Clemens, J.D. 2006. Relationship between neighbourhood-level killed oral cholera vaccine coverage and protective efficacy: evidence for herd immunity. International journal of epidemiology, vol. 35 (4), pp. 1044-1050.

Faruque, S.M., Naser, I.B., Islam, M.J., Faruque, A.S., Ghosh, A.N., Nair, G.B., Sack, D.A. & Mekalanos, J.J. 2005. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proceedings of the National Academy of Sciences of the United States of America, 102 (5), pp. 1702-1707.

Huq, A., Sack, R.B., Nizam, A., Longini, I.M., Nair, G.B., Ali, A., Morris, Jr., J. Glenn, Huda Khan, M.N., Siddique, A.K., Yunus, M., Albert, M.J., Sack, D.A. & Colwell, R.R. 2005. Critical Factors Influencing the Occurrence of Vibrio cholerae in the Environment of Bangladesh. Applied and Environmental Microbiology, 71 (8), pp. 4645-4654.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) 2007a. 2005 Socio-economic Census, ICDDR,B Health & Demographic Surveillance Unit, Public Health Sciences Division, Dhaka, Bangladesh.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) 2007b. Registration of Health and Demographic Events 2005, ICDDR,B Health & Demographic Surveillance Unit, Public Health Sciences Division, Dhaka, Bangladesh.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) 1992. Demographic, Health and Nutritional Impact of the Meghna-Dhonagoda Embankment, ICDDR,B, Dhaka, Bangladesh.

Islam, M.S., Drasar, B.S. & Sack, R.B. 1993. The aquatic environment as a reservoir of Vibrio cholerae: a review. Journal of diarrhoeal diseases research, 11 (4), pp. 197-206.

Jensen, M.A., Faruque, S.M., Mekalanos, J.J. & Levin, B.R. 2006. Modeling the role of bacteriophage in the control of cholera outbreaks. Proceedings of the National Academy of Sciences of the United States of America, 103 (12), pp. 4652-4657.

Johnson, J.B. & Omland, K.S. 2004. Model selection in ecology and evolution. Trends in ecology & evolution (Personal edition), 19 (2), pp. 101-108.

Longini, I.M.,Jr, Yunus, M., Zaman, K., Siddique, A.K., Sack, R.B. & Nizam, A. 2002. Epidemic and endemic cholera trends over a 33-year period in Bangladesh. The Journal of Infectious Diseases, 186 (2), pp. 246-251.

Myaux, J., Ali, M., Felsenstein, A., Chakraborty, J. & de Francisco, A. 1997. Spatial distribution of watery diarrhoea in children: identification of "risk areas" in a rural community in Bangladesh. Health & Place, 3 (3), pp. 181-186.

Vaughn, C.S. 1997. An Impact of Environmental Changes and People's Perceptions of the Meghna-Dhonagoda Embankment, International Centre for Diarrhoeal Disease Reasearch, Bangladesh (ICDDR,B) & the Bangladesh Rural Advancement Committee (BRAC), Dhaka, Bangladesh.

Chapter 5

Conclusion

The results of the spatially-lagged regression analyses confirmed the earlier cluster results indicating that changes in cholera incidence have taken place in Matlab since the introduction of flood protection to a portion of the study area. The two analyses presented, cluster and regression, together form a part of the picture explaining the changes that have occurred in cholera incidence in Matlab since the introduction of flood protection measures. The results of both studies suggest that the strongest changes in incidence are not strictly spatial, such that cholera incidence went up inside or outside the flood protected area, but rather are more intricate, involving changes in seasonality of occurrence and variance by cholera type.

Prior to the introduction of flood protection, the occurrence of cholera clusters in Matlab appeared to be mediated by proximity to the Dhonagoda River. In the 1990-2003 analysis, river distance seemed to have a lesser role in variation, although clusters still appeared evenly distributed across the flood protected and unprotected portions of Matlab. Despite a lack of observable spatial differentiation between pre-MDIP clusters and post-MDIP clusters, significant temporal variation is present. Our findings suggest that the seasonal timing of clusters shifted post-MDIP construction, from occurring predominantly in the dry season to taking place in the rainy season. Additionally, rainy season clusters occur earlier in the unprotected portion of Matlab than within the flood protected area.

Important seasonal variation also appeared in the spatial regression analysis. When cholera was analyzed according to season of incidence the greatest levels of spatial autocorrelation were seen in dry season cholera, regardless of time period or independent variable. This suggests that cholera incidence in Matlab is highly influenced by neighboring cholera incidence in the dry season, and that introducing flood protection in 1989 did little to affect this. Completing the MDIP did, however, affect levels of spatial autocorrelation seen in rainy season cholera incidence, and introducing flood protection to half the study area greatly increased the influence of surrounding cholera on individual *bari* cholera across Matlab.

Division of cholera incidence according to bacterial strain demonstrated further differentiation between pre-MDIP Matlab and the post-MDIP divided Matlab. Classical cholera clusters took place predominantly in the dry season, El Tor clusters occurred in the rainy season and O139 clusters were split evenly between the two seasons. When the flood protection status of each type-cluster was considered, flood protection status did not appear to mediate the occurrence of O139 clusters. In contrast, the O1 clusters were sited mainly outside of the protection of the MDIP. Differentiation in results according to bacterial strain was also apparent in the spatial regression analysis. O1 cholera relationships were more spatially autocorrelated than their O139 counterparts, suggesting that O139 cholera is less mediated by surrounding cholera events than is O1 cholera.

The strong differences in the cluster maps and in the results of the regression analyses suggest that there are significant differences between cholera in 1983-1989 and 1990-2003. These differences can be correlated with the completion of the Meghna-Dhonagoda Irrigation Project in 1989, in keeping with previous literature on the changes that water construction programs can have on disease outcomes. Viewing the human-environmental interactions taking place in Matlab that lead to cholera outcomes in terms of the triangle of human ecology, we can postulate about changes in behaviors that might have occurred as a result of MDIP construction. At the population level, increases in nutritional status experienced within the flood protected area as the result of the ability to double or triple crop could alter cholera patterns, as cholera is more likely to occur in malnourished individuals. From the environmental perspective, water sources that were previously flooded by the Dhonagoda in the early months of the rainy season remained at normal levels within the flood protected area. The MDIP embankment could also have altered the social environment, cutting off social networks that previously spanned the river. Behaviors of those individuals living inside the MDIP were also likely affected, as patterns of water usage (i.e. going to the river to wash dishes or laundry) were shifted by the building of the earthen embankment. Our results only hint at the changes to the highly intricate and complex relationships between social, environmental and biological variables that interact to result in cholera infection.