Library A-0000211

ttachment 1.
FACE SHEET) ETHICAL REVIEW COMMITTEE, ICDDR,B. Date April 9 1995
2/7/95

rincipal Investigator Michael Ezra
Trainee Investigator (if any) _____
pplication No. 95-018
Supporting Agency (if Non-ICDDR,B) _____
itle of Study Risk Factors for Diarrheal Disease in Matlab, Bangladesh: A Medical Geographic Approach
Project status:
(X) New Study
( ) Continuation with change
( ) No change (do not fill out rest of form)

ircle the appropriate answer to each of the following (If Not Applicable write NA):

1. Source of Population:
   (a) Ill subjects — Yes (No)
   (b) Non-Ill subjects — Yes (No)
   (c) Minors or persons under guardianship — Yes (No)
2. Does the study involve:
   (a) Physical risks to the subjects — Yes (No)
   (b) Social Risks — Yes (No)
   (c) Psychological risks to subjects — Yes (No)
   (d) Discomfort to subjects — Yes (No)
   (e) Invasion of privacy — Yes (No)
   (f) Disclosure of information damaging to subject or others — Yes (No)
3. Does the study involve:
   (a) Use of records, (hospital, medical, death, birth or other) — (Yes) No
   (b) Use of fetal tissue or abortus — Yes (No)
   (c) Use of organs or body fluids — Yes (No)
4. Are subjects clearly informed about:
   (a) Nature and purposes of study — (Yes) No
   (b) Procedures to be followed including alternatives used — Yes No NA
   (c) Physical risks — Yes No NA
   (d) Sensitive questions — Yes No NA
   (e) Benefits to be derived — Yes No NA
   (f) Right to refuse to participate or to withdraw from study — (Yes) No
   (g) Confidential handling of data — (Yes) No
   (h) Compensation &/or treatment where there are risks or privacy is involved in any particular procedure — Yes No NA
5. Will signed consent form be required:
   (a) From subjects — (Yes) No
   (b) From parent or guardian (If subjects are minors) — Yes No NA
6. Will precautions be taken to protect anonymity of subjects — (Yes) No
7. Check documents being submitted herewith to Committee:
   - X — Umbrella proposal - Initially submit an overview (all other requirements will be submitted with individual studies).
   - X — Protocol (Required)
   - X — Abstract Summary (Required)
   - X — Statement given or read to subjects on nature of study, risks, types of questions to be asked, and right to refuse to participate or withdraw (Required) (See MSU IRC (Human Subjects) application)
   - ___ — Informed consent form for subjects
   - ___ — Informed consent form for parent or guardian
   - ✓ — Procedure for maintaining confidentiality
   - ✓ — Questionnaire or interview schedule *

* If the final instrument is not completed prior to review, the following information should be included in the abstract summary:
1. A description of the areas to be covered in the questionnaire or interview which could be considered either sensitive or which would constitute an invasion of privacy.
2. Examples of the type of specific questions to be asked in the sensitive areas.
3. An indication as to when the questionnaire will be presented to the Ethic. for review.

e agree to obtain approval of the Ethical Review Committee for any changes involving the rights and welfare of subjects before making such change.

Michael Ezra
Principal Investigator

Trainee

1. PRINCIPAL INVESTIGATORS	Michael Emch Jacques Myaux
2. OTHER INVESTIGATORS	Md. Yunus
CONSULTANTS	
3. TITLE OF PROJECT	"Risk Factors for Diarrheal Disease in Matlab, Bangladesh: A Medical Geographic Approach"
4. EXPECTED STARTING DATE	Oct, 1995
5. EXPECTED DATE OF COMPLETION	Nov, 1996
6. TOTAL BUDGET REQUESTED	US$ 5030
7. FUNDING SOURCES	United States Department of Education Foreign Language and Area Studies Fellowship American Institute of Bangladesh Studies Association of American Geographers

This protocol has been approved by the Community Health Division

Head of the Division
Dr. K.M.A. Aziz

Date June 27, 1995

DISSERTATION PROPOSAL

Risk Factors for Severe Diarrheal Disease in Matlab, Bangladesh:

A Medical Geographic Approach

Submitted by
Michael E. Emch
Department of Geography
Michigan State University
East Lansing, MI 48824

June 1995

Abstract

Risk Factors For Severe Diarrheal Disease in Matlab, Bangladesh:

A Medical Geographic Approach

Michael Emch

Diarrheal diseases cause one-third of the 15 million annual deaths in children under five years old in the developing world. Because of resource constraints in developing countries like Bangladesh it is necessary to identify risk factors so preventative health programs can focus on specific interventions. Assessing risk for diarrheal disease requires knowledge of the complex and dynamic interaction of biological, socio-economic, behavioral, and environmental factors over time and space.

The objective of this study is to advance such knowledge in the context of rural Bangladesh. Specifically, the study will identify the variables related to diarrheal disease risk and analyze the relationship between these variables over time and space. Colwell et al. (1985) recently found evidence of an aquatic reservoir for cholera which dramatically changes longstanding conceptions of the ecology of cholera. I hypothesize that two types of watery diarrhea, cholera and non-cholera, have markedly different spatio-temporal patterns and somewhat different risk factors because the two etiological groups have different reservoirs.

The study will be conducted at the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), where I was part of a team that created an extended-household geographic information system (GIS) database in 1993.

Table of Contents

Table

Figures

1. Introduction

Diarrheal diseases cause one-third of the 15 million annual deaths in children under five years old in the developing world (Snyder & Merson, 1982). Because of resource constraints in developing countries like Bangladesh it is necessary to identify risk factors so preventative health programs can focus on specific interventions. Assessing risk for diarrheal disease requires knowledge of the complex and dynamic interaction of biological, socio-economic, behavioral, and environmental factors over time and space. The objective of this study is to advance such knowledge in the context of rural Bangladesh. Specifically, the study will identify the variables related to diarrheal disease risk and analyze the relationship between these variables over time and space.

Humans were the only known reservoir of Vibrio cholerae until the mid-1980s when theories of the ecology of cholera were substantially revised. During this time, Colwell et al. (1985) found that vibrios can live freely in an aquatic environment even under conditions of nutrient deprivation if the environment is not sodium-free. Before Colwell et al. discovered this free-living, non-culturable state, it was maintained that cholera was only transmitted by ingestion of fecally contaminated food or water. It is now understood that transmission can occur through water without fecal contamination. These findings dramatically change longstanding conceptions of the ecology of cholera.

This study will differentiate between two types of diarrhea, cholera and non-cholera. Cholera is defined as watery diarrhea caused by the bacterium Vibrio cholerae. Non-

cholera watery diarrhea is defined as watery diarrhea caused by microorganisms other than Vibrio cholerae. Ideally, this study would distinguish between all of the non-cholera diarrheal agents, however, the microbiological tests associated with obtaining this information would be exorbitantly expensive.[1] Given contraints of time and money, non-cholera watery diarrhea is a logical and useful grouping because none of the organisms in this group have an environmental reservoir while the organisms in the cholera group presumably do.

This study has two distinct parts. The first part of the study will analyze the differences between the spatio-temporal patterns of cholera and non-cholera watery diarrhea. I believe that they will be different because an environmental reservoir has been identified for cholera and one has not been identified for non-cholera watery diarrhea. The spatio-temporal patterns of these two disease types have not been thoroughly described and they have not been differentiated. Differentiating between the spatio-temporal patterns of the diseases will provide supporting evidence concerning the existence and importance of the cholera reservoir as well as give us some insight into spatio-temporal forecasting of both diseases. Proving or disproving the existence and importance of this environmental reservoir is impossible without extensive and expensive environmental microbiological testing, however, this study is a descriptive first step to understanding this reservoir.

---

[1] This would require a prospective study (possibly community-based) which would take several years.

The second part of this study will differentiate between risk factors of cholera and non-cholera diarrhea. I believe that different environmental reservoirs, and thus different spatio-temporal patterns, will be accompanied by different risk factors.[1] There will be differences in the risk factors because the ecology of the diseases[2] exist within a dynamic spatio-temporal framework. That is, when the spatio-temporal patterns are different then exposure to the diseases will be different. If all important variables that cause[3] a disease were spatio-temporally homogeneous then these spatio-temporal patterns would not affect the risk factors; however, in the real world these variables are not spatio-temporally homogeneous. Identifying specific risk factors will allow health programs to focus on interventions that have the greatest impact. Risk factors of the two disease types will be calculated so that the relative importance of risk (for several independent variables) for the two disease categories can be compared.

---

[2] The medical geographic theory of disease ecology is discussed in section II of this proposal.

[3] The cause of a disease is not a simple concept. The doctrine of specific etiology cannot provide a complete account of the causation of disease. Microbiological evidence of a disease is an essential part of understanding a disease but is only the first step to explaining the disease process. Some call the specific etiological agent the direct cause and factors affecting the outcome of disease indirect determinants. We can thus refer to a causal pathway in which more distant indirect determinants lead to the direct determinants of disease. It is within the realm of the biological sciences to describe properties of the direct determinants but studying the indirect determinants requires an interdisciplinary effort (Dubos, 1965). Statistical association does not always mean that a variable is in the causal pathway, there could be a spurious association. The only way one can jump from association to the causal pathway is through logic; that is, the association must make theoretical sense.

This study is within the medical geographic theoretical approach called disease ecology[2] and will use a methodological approach called ecological association analysis which uses quantitative methods to model spatial and temporal disease variation. This approach facilitates understanding disease causation in a spatio-temporal framework; no such study has previously been conducted on watery diarrhea. Understanding the complexities of risk for watery diarrhea is important for ameliorating this significant health problem in Bangladesh, as well as in other developing countries throughout the world.

This research will provide essential information about the disease ecology of severe watery diarrheal disease. Specifically, it will (1) offer corroborating evidence concerning the existence and importance of an environmental reservoir for cholera by differentiating between spatio-temporal patterns of cholera and non-cholera watery diarrhea; (2) identify and compare risk factors for cholera and non-cholera watery diarrhea; and (3) extend the use of geographic information system (GIS) as a tool in disease modeling.

The study will be conducted at the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), where I was part of a team that created an extended-household GIS database in 1993. Medical geography studies have historically used aggregate data sources because individual data are seldom available and extremely expensive to collect. The ICDDR,B was chosen as a research institution because of its unique watery diarrhea data collection system which makes a micro-level study possible.

II. Theoretical Framework

In order to advance philosophical and theoretical implications of this study it is important to situate the study within the field of geography and within the sub-field of medical geography. In this section, I begin with an overview of various philosophies of geography and identify the philosophies which guide this study. This is followed by a more focused discussion of the sub-field of medical geography and disease ecology which more narrowly characterize this study. Next, I describe the methodological approach of ecological association analysis which utilizes the theory of disease ecology. Finally, I discuss past disease ecology and ecological association analysis studies which have informed this study.

Philosophies of Geography

The field of geography is complex and multi-faceted. Different geographers have divided philosophies of geography in different ways. Pattison (1963) created four divisions including:

1) Spatial - concerned with distance, form, direction, position, and spatial relationships.
2) Area studies - concerned with characteristics of places and their differentiation.
3) Man-land - concerned with effects of land (nature) on man and man on land.
4) Earth science or physical - concerned only with the physical environment; hard science approach.

Taaffe (1974) as well as Morrill (1970) divided geographic thought into three categories:

1) Spatial - concerned with spatial relationships.
2) Regional - concerned with characteristics of places and their differentiation.
3) Man-land - concerned with effects of land (nature) on man and man on land.

During the first half of this century the regional/area studies approach was the dominant

tradition in geography. Hartshorne (1939, 1959) argued that geographers should primarily be involved in region construction of the earth's surface; he called this areal differentiation. Because Hartshorne's objective was primarily concerned with description, his type of geography is characterized as an idiographic science.

Hartshorne's areal differentiation approach has come into question by many geographers. Schaefer (1953) suggested that the descriptive nature of Hartshorne's area studies school did not do justice to explaining distributions of phenomena on the earth's surface. In contrast, Schaefer believed that spatial relationships were most important and that causal interactions should be determined through quantitative studies. Harvey (1969) also advocated the use of quantitative methods in geography, as well as the creation of a nomothetic discipline. The school of spatial organization, which also includes the work of Garrison, Berry, Bunge, and Haggett, became dominant in the 1950s and 1960s during the quantitative revolution of geography.

The man-land (now called human-environment) tradition has evolved substantially during this century. Early proponents of this tradition highlighted their beliefs about the effects of the physical environment on humans. This sometimes extreme and discriminatory approach has been called environmental determinism (Semple, 1911; Brigham, 1915; Davis, 1915; Huntington, 1924). In 1923 Barrows introduced a less extreme geographic human-environment tradition at the University of Chicago called human ecology. The reference to ecology is derived from the biological sciences and refers to the existence of a human ecosystem. According to Haggett (1977) Barrows' human ecology used a social science

perspective to study relationships of human society and its physical environment. Since the early days of the human-environment tradition several other ecological traditions have emerged. Cultural ecology, firmly established by Sauer (1923), focuses on cultural manifestations of the human landscape. Political ecology, although not a unified approach which can be easily defined, is a holistic approach to understanding human-environment relations (Blaikie, 1994). Campbell and Olson (1991) developed a political ecology model for studies of human-environment relations called the Kite. This heuristic model purports that one must understand political, economic, environmental, socio-cultural variables at different spatial scales to fully understand the relationship between society and the environment. As I will argue later in this discussion, the sub-discipline of medical geography has developed a parallel theory to some of these human-environment theories in the sub-disciplinary tradition of disease ecology.

As Pattison (1963) and Taaffe (1974) as well as most writers on the philosophy of geographic thought have mentioned in their articles, the aforementioned traditions are not mutually exclusive. This study is situated within two of the traditions of geography, the human-environment and the spatial traditions. The study is spatial because it is interested in the distribution of a phenomenon (disease) in space and time. And it is ecological in that it is based on the theory of disease ecology, a holistic approach to understanding disease in the context of human-environment interaction.

Medical Geography/Disease Ecology

Analyzing risk of contracting watery diarrheal disease in Bangladesh requires a

conceptual framework that addresses the complexities of biological, socio-economic, behavioral, and environmental factors over time and space. A medical geographic theoretical approach that addresses these issues is disease ecology, which maintains that disease is a dynamic complex of variables that coincide in time and space (May, 1958, 1977; Mayer, 1982, 1984; Mayer & Meade, 1994; Meade, 1977; Meade et al., 1988; Learmonth, 1988; Paul, 1985; Pyle, 1977, 1979). Hunter (1974) argued that we must not have a pathogencentric view of disease by focusing only on the disease agent. On the contrary, he suggested that our studies of disease "must cojointly involve pathogen, host, and environment" (p. 1). He defined environment broadly as "its diverse physical, biological, social, cultural, and economic components" (p. 3). Hunter defined geography as a discipline that bridges the social and environmental sciences and said that "its integration and coherence derive from systems-related analysis of man-environmental interactions through time and over space" (p. 3).

My outlook of medical geography parallels Hunter's in that I believe that the medical geographic approach should be holistic; that is, I believe that I should study the integration of many different types of variables which result in the coincidence of agent and host in time and space. While the types of variables to be investigated have been classified in many different ways, I find Mayer's (1986) classification system the most useful. He differentiated between biological, socio-economic, behavioral, and environmental variables.

Biological variables are those that describe biological characteristics of the host. Sometimes geographers are only interested in these biological variables because they must control for them to isolate the effects of variables in which they are most interested. Different patterns of socio-economic, behavioral, and environmental variables result in different rates at which agent and host come into contact with one another, and therefore different spatio-temporal patterns of disease. Virtually every disease exhibits spatial and temporal variation and medical geographers attempt to explain this variation. Behavioral variables are those that describe individual or group behaviors and may be related to culture or individual decision making. Environmental variables are those of the physical environment. Socio-economic variables are variables that indirectly affect the coincidence of agent and host such as wealth or class.[4]

The theory of disease ecology is related to both the human-environment and spatial traditions of geography. This theory is a bit more specialized than the previously mentioned human-environment theories in geography because the variable of study in disease ecology is always disease. However, human-environment[5] interaction can essentially be viewed as the cause of disease. The spatial tradition of disease ecology is

---

[4] Variables may be thought of as directly or indirectly spatial. The directly spatial variables are those that are related to the physical world and do not change locations unless the physical environment changes. Indirectly spatial variables are those that are associated with human activity. Behavioral and socio-economic variables may be thought of as indirectly spatial (Personal communication with Bruce Pigozzi, 1995).

[5] Environment as defined broadly by Hunter (1974).

evident in that most definitions and studies are interested in the spatial distribution of disease.

Ecological Association Analysis

A methodological approach which utilizes the theory of disease ecology, called ecological association analysis, holds that quantitative studies which associate environmental, physical, and cultural variables can explain the spatial and temporal variation of disease occurrence (McGlashan, 1967; Mayer, 1986). The fundamental question asked using this approach is, what factors are associated with the spatio-temporal variation of disease? Mayer states that "the term ecological association implies the existence of specific links between the environment and both individuals and groups" and that "in the context of medical geography, the focus is on those relationships which are consequential in disease pathogenesis" (p. 66). He also states that "in ecological analysis, the emphasis is therefore on the complex set of interactions between people and their environment" (p. 66).

The method of ecological association is an approach that must be used within a logical framework and is best used within a theoretical approach such as disease ecology. Mayer addresses one of the main challenges of ecological association analysis.

> One of the most vexing problems in ecological analysis is that of moving from statistical association to causal relationships. It is one thing to identify cultural, environmental, or social factors which are associated statistically with disease occurrence. This may be

> accomplished in the absence of a theoretical framework, or a logical association, between the disease and the environment. Correlation between the disease, and a host of related independent variables, may be so spurious as to defy the formation of meaningful causal hypotheses. For example, there is a very strong correlation between multiple sclerosis prevalence and annual per capita steel consumption, at the national level. The relationship may be tenuously meaningful, in that steel consumption may be a surrogate for concepts such as economic development, or the correlation may be meaningless, since it may be coincidental that multiple sclerosis and steel consumption show the same pattern of variation. (p. 66)

When medical geographers use ecological association they must distinguish between causally meaningful relationships and spurious correlations. Several researchers (e.g., King, 1979; Mayer, 1982) have also identified the ecological fallacy as a serious problem in many aggregate-level ecological association studies. The ecological fallacy states that conclusions made at the aggregate level (i.e. county, state, national levels) are not always true at the individual level. For example, if an association is found between cancer and smoking when grouped by county, one cannot be sure that the association exists at the individual level. That is not to say that finding an association is not important information, but that further investigations must be conducted. Mayer (1982) suggested that to alleviate the problem of the ecological fallacy, individual-level case-control[6] studies should be conducted. To date, very few medical geography studies have been done at the individual level because these data are expensive to collect and thus are seldom available.

[6] Discussed in section VII of this proposal.

Related Disease Ecology and Ecological Association Analysis Studies

Past disease ecology and ecological association analysis studies have ranged from speculative studies to multi-variate explanatory studies. Jacques May (1958) wrote several voluminous descriptive studies of the ecology of many infectious diseases including brucellosis, poliomyelitis, tuberculosis, and leprosy. These studies are recognized as the formal beginning of the disease ecology tradition but have been criticized as being atheoretical and overly idiographic. Burkitt (1962) described the existence of a "lymphoma belt" straddling the equator where a childhood cancer occurred (later named Burkitt's Lymphoma). He found that this cancer only occurred in specific ecological niches. Roundy (1976) identified associations between disease and altitude in Ethiopia. Hunter (1992) described hyperendemic lymphatic filariasis areas in north east Ghana that were associated with rice irrigation projects. Kloos (1985) found that schistosomiasis in the Awash valley in Ethiopia was associated with migrant labor. Hunter (1982) brought attention to associations between irrigation projects and infectious disease throughout the tropical world and called for health policy considerations when development projects are implemented.

Much of the disease ecology literature has been devoted to infectious diseases especially in the developing world. Recently, however, the approach has been used for chronic and environmental diseases in the developed world. Hunter (1976, 1977) described

a seasonal cycle for childhood lead poisoning and identified geographic concentrations of the disease in older residential areas along traffic arteries in the United States. Meade (1980) studied cardiovascular mortality in the southeastern United States, and Glick (1979, 1980) analysed the spatial characteristics of cancer mortality in Pennsylvania using a GIS.

Since the inception of the disease ecology and ecological association approaches, greater attention has been paid to temporal patterns, spatial scale, and statistical methods. Temporal changes in biological, socio-economic, behavioral, and environmental variables affect how agents and hosts come into contact with one another; thus studies should always have a temporal dimension. Disease associations at one spatial scale may not be present at other spatial scales; therefore, multi-scale studies should be conducted whenever possible. The use of GIS makes spatial analysis methods more efficient. Also, the interface between GIS and statistical methods has recently begun to be explored. I think it is important to note that GIS and statistics are useful tools that can be employed to conduct disease ecology and ecological association studies. The technology of GIS is only useful within some theoretical and methodological framework. It provides researchers with an accurate spatial modelling environment in which to conduct research. Scholten and de Lepper (1991) have reviewed the benefits of the diverse applications of GIS in health. A few of the studies in the above literature review have utilized a GIS to analyze their data and many have used statistics.

III. Review of Related Diarrheal Disease Studies

Diarrheal disease can be caused by many etiological agents. For practical purposes diarrhea can be classified into two manifestational categories, dysentery and watery diarrhea (Benenson, 1990; ICDDR,B, 1993). This study will focus on watery diarrhea, thus the agents that cause dysentery will not be considered[7]. Two studies were conducted in Matlab, Bangladesh to examine the relative importance of various enteropathogens in diarrheal disease (Table 1) (Baqui et al., 1992; Black et al., 1980). The non-cholera, non-dysentery organisms are all watery diarrheal agents. Cholera is caused by the colonization of the small intestine with Vibrio cholerae O1. During the Spring of 1993, Vibrio cholerae O1 was replaced by another strain, Vibrio cholerae O139. The clinical manifestations of the two agents are identical, however, little is known about the newer strain (Siddique et al., 1994).

In rural Bangladesh, cholera transmission can be divided into primary and secondary types (Colwell & Spira, 1992; Craig, 1988). Primary cases are infected by surface water sources. An example of this is when a person drinks untreated pond water or eats undercooked shellfish and is directly infected with the bacteria. Secondary cases are infected by primary cases through fecal-oral transmission, usually by contaminated drinking water. An example of this is when a family member is infected by a sick member of their family when the sick person puts his/her hands in the family's drinking water pot; cholera

[7] The dysenteric agents that are present in Matlab include Shigella, Campylobacter jejuni, Entamoeba histolytica, enteropathogenic Escherichia Coli (EPEC), and enteroadhesive Escherichia Coli (EAEC).

requires a large innoculum so this is thought to be somewhat rare. Another example of secondary transmission is when a mother is infected by the feces of her baby. Primary transmission is controlled by factors such as temperature, salinity, nutrient concentrations, the number of available attachment sites (plankton), seafood consumption, and contact with water (Colwell & Spira, 1992)..

Table 1: Percent of Diarrheal Episodes Associated with Etiological Agents in Matlab

Pathogen	Black et al. 1981 Community-based	Black et al. 1981 Hospital-based	Baqui et al. 1990 Community-based	Baqui et al. 1991 Hospital-based
Vibrio cholerae 01	0.3	13	0.4	39
Vibrio cholerae non 01	1.1	7	2.9	3
Shigella	12.8	5.5	8.6	11
ETEC	26.9	29	12.2	14
Campylobacter	-	-	17.6	11
Salmonella	-	<1	1	1
EAEC	-	-	34.3	-
EPEC	-	-	13.5	-
Aeromonas	-	-	2	-
Pleismonas	-	-	.1	-
Rotavirus	3.8	24	4.3	-
Entaemoeba histolytica	.2	.4	.4	2
Giardia lamblia	.5	2	2.2	2
Cryptosporidium	-	-	1.9	-
No pathogen	49.5	-	42.1	-

(- not tested) (Derived from Baqui et al., 1994)

In rural Bangladesh cholera transmission is seasonal, with a peak after the monsoon

ranging from September to December (Baqui et al., 1994). There is another seasonal epidemic in Matlab with a peak in April or May. Colwell and Spira (1992) suggested that the post-monsoon season is associated with a heavy bloom of zooplankton, maximum recreational water contact, and maximum available crustacea in the marketplace. They postulated that the presence of a permanent environmental reservoir of Vibrio cholerae in the brackish ponds and canals of rural Bangladesh allows for the marked seasonal variation in cholera transmission.

Many studies have identified risk factors for cholera in rural Bangladesh and they can be roughly divided into the four types of ecological association variable classes, biological, behavioral, environmental and socio-economic. However, some of the risk factors are a combination of two or more of these variable classes. Although not exhaustive, these studies represent the most important findings related to cholera risk and their principal investigators come from many different disciplinary backgrounds.

Three studies have identified biological risk factors. Glass et al. (1985) found that individuals with type O blood are predisposed to cholera. Breast-feeding protects infants against cholera but this might be related to contamination of water during bottle-feeding, a behavioral variable (Glass & Black, 1992). Glass and Black (1992) reported that children aged 2 to 15 are at greatest risk of contracting cholera. Age is a biological variable but the reason for this finding is certainly very complex involving many types of variables

including behavioral.

Several behavioral variables have also been identified. In Bangladesh, women of child-bearing age have high cholera incidence rates presumably because of increased person-to-person contact (Glass & Black, 1992). This is an example of a behavioral variable that is intertwined with culture. Glass et al. (1982) found that villages with daily bazaars have higher cholera rates. This is an example of an aggregate behavioral measure.

Environmental variables have been identified in four studies. Sommer and Woodward (1972) found that people who lived close to tubewells had a much lower incidence rate than those who lived further away because they had access to clean water. Khan et al. (1981) found that cholera attack rates were higher for families with access to canal water as opposed to river water or tank water. In Matlab, Bangladesh, cholera is more common in villages that are not adjacent to the main river (Glass et al., 1982). Hughes et al. (1982) found that rural Bangladeshi families who used contaminated surface water for cooking and bathing were more likely to get cholera than those who did not.

Glass et al. (1982) found that predominantly Hindu villages have higher cholera rates. It is unclear whether this is because of socio-economic or cultural reasons or both. Other studies have been conducted that identify risk factors for specific non-cholera agents or non-specified diarrhea. Becker et al. (1986) found that children in poorer households had a higher proportion of days with undifferentiated diarrhea and rotavirus than more

affluent households. Chen et al. (1981) and Bairagi (1987) found that undernutrition is not a predictor of diarrheal incidence.

There are several significant gaps in the literature on watery diarrheal disease. No studies have used a disease ecology approach and few studies have identified indirect socio-economic determinants of diarrheal disease. While many studies have used simple non-parametric statistical methods, few have identified the multi-variate relationships between the different types of variables. Also, no studies have differentiated between cholera and non-cholera diarrheal disease risk. Craig (1988) looked at spatio-temporal clustering of cholera, but there have been no studies looking at spatio-temporal patterns and their associations with other variables.

Several studies support a hypothesis that the temporal and spatial patterns of cholera and non-cholera watery diarrhea will be different from one another. Little seasonal variation is found in non-cholera watery diarrhea while a significant seasonality occurs in cholera (Baqui et al., 1994). Black et al. (1981) found that incidence of rotavirus is relatively constant except for a small peak in December, and that enterotoxogenic Escherichia coli occurs more frequently in the hot months.[8] An environmental reservoir is not known to exist for non-cholera organisms; humans and animals are the only reservoirs for these organisms (Benenson, 1990; Warren & Mahmood, 1993).

[8] Rotavirus and Escherichia coli are non-cholera watery diarrheal agents.

The dichotomy between cholera and non-cholera diarrhea was chosen because it has been reported that cholera has an environmental reservoir and there is no evidence of one for any of the organisms in the non-cholera diarrhea group. Accurate descriptions of the micro-scale spatio-temporal patterns of the two disease categories have never been completed; the existing GIS database will now make this task manageable. Although this study will not accurately identify the location of the cholera reservoir it can offer corroborating evidence of its existence and importance by describing the spatio-temporal patterns of the disease. If the cholera reservoir is important then the spatio-temporal patterns of cholera and non-cholera watery diarrhea will be very different. The pattern of severe cholera watery diarrhea should generally correspond to the environmental reservoir, especially at the beginning of the season.

Since it is hypothesized that these two disease categories have different spatio-temporal patterns it is thought that their risk factors will be different. The risk factors will be different because the ecology of diseases exist within a dynamic spatio-temporal framework and if the spatio-temporal patterns are different then exposure to the diseases will be different. If funding or time were not a consideration the study would distinguish between all of the non-cholera diarrheal agents, however, the microbiological tests associated with obtaining this information would be extremely expensive and time-consuming.

IV. The Research Setting

Bangladesh suffers markedly from endemic diarrheal disease. The people of Bangladesh suffer not only directly when they contract the disease, but also indirectly by suffering economic hardship because of lost productivity and medical expenses. The research site for the ICDDR,B and for this project (called Matlab because the Centre's hospital is located in Matlab Town) is in south-central Bangladesh approximately 50 km south-east of Dhaka. It is adjacent to where the Padma River meets the Meghna River forming the Lower Meghna River. Figure 1 shows the study location within Bangladesh, the location of two minor rivers in the study area (Gumti and Dhonagoda), and the location of one major river (the Lower Meghna, which empties into the bay of Bengal approximately 100 km south of the study area). Figure 2 shows the location of Matlab relative to major rivers and cities (the river running next to Matlab Town is the Dhonagoda).

The ICDDR,B has operated this field research area since 1963 and has a present study population of approximately 200,000.[9] There are 142 villages in the study area, 128 of which are predominantly Muslim and 14 of which are predominantly Hindu. The study area is almost entirely rural and most people's occupations are in agriculture or fisheries. Increasing population in the past 100 years in combination with the tenure system have led

[9] The study population includes people from villages for which the ICDDR,B maintains data collection procedures. The study area and therefore the number of people in the study population has changed several times since 1963.

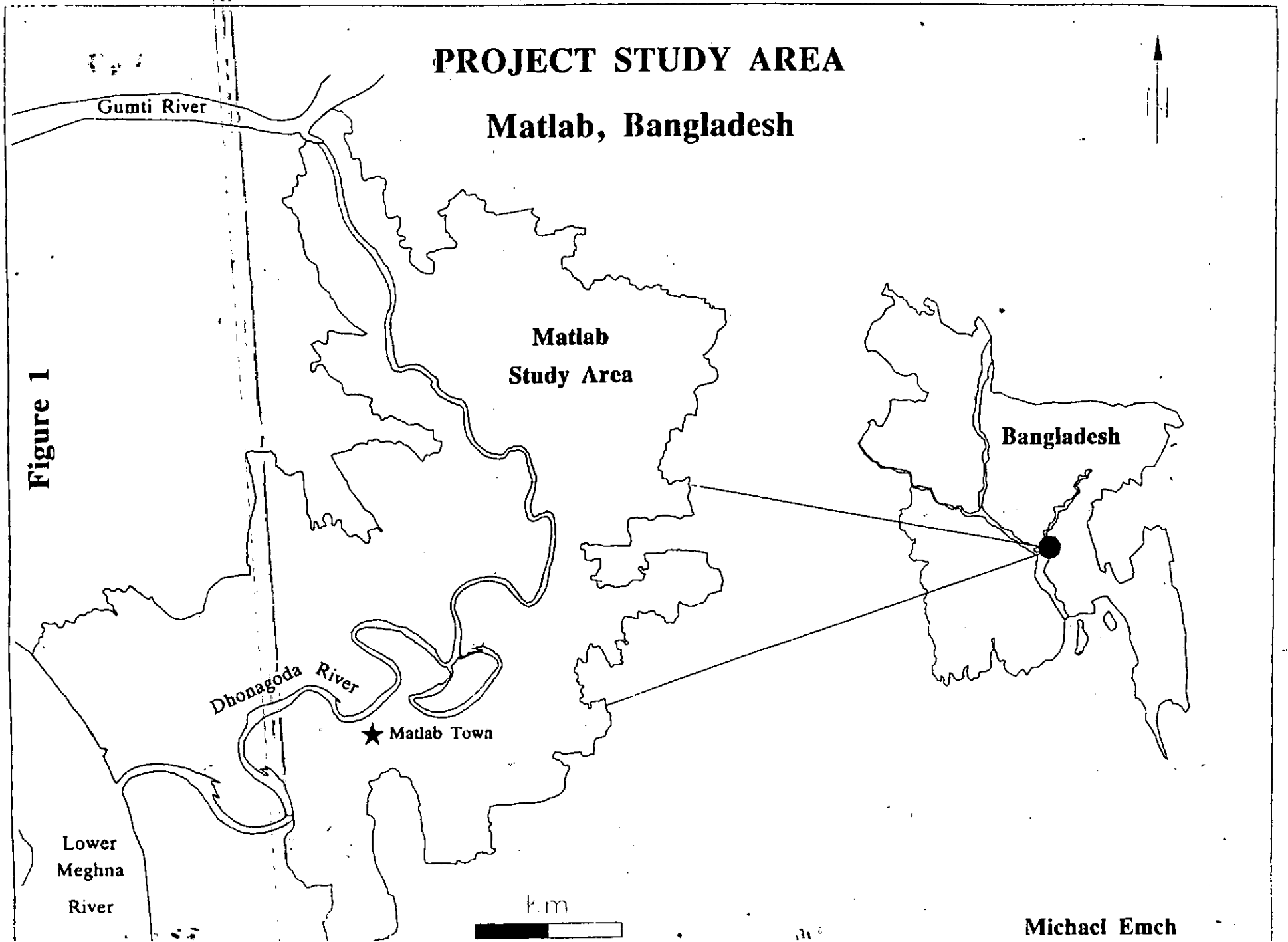
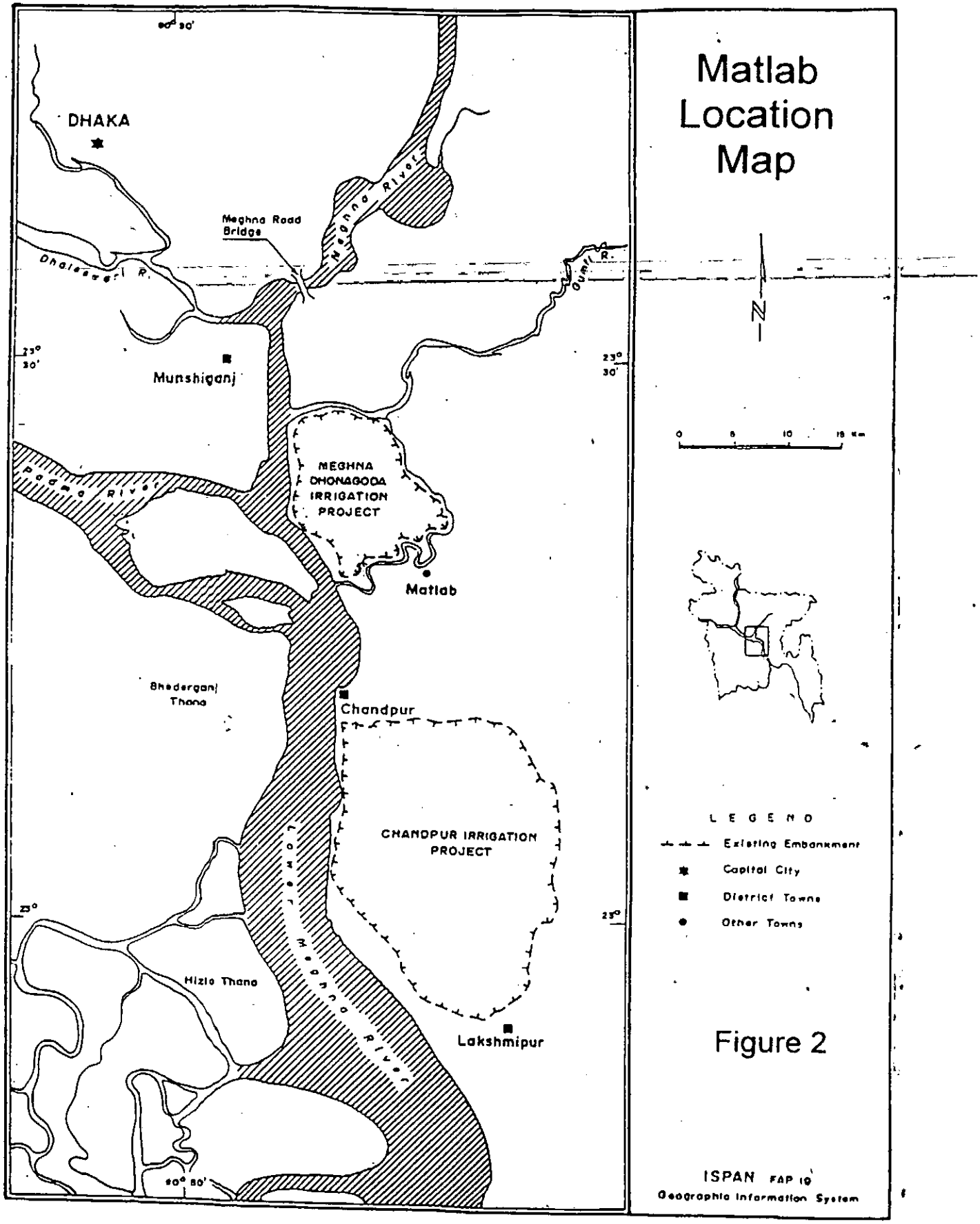

PROJECT STUDY AREA
Matlab, Bangladesh
Gumti River
Matlab
Study Area
Bangladesh
Dhonagoda River
Matlab Town
Lower
Meghna
River
km
Michael Emch

Figure 1


Matlab
Location
Map
DHAKA
Meghna Road
Bridge
Meghna River
Dhaleswari R.
Gumti R.
Munshiganj
MEGHNA
DHONAGODA
IRRIGATION
PROJECT
Padma River
Matlab
Bhedergani
Thana
Chandpur
CHANDPUR IRRIGATION
PROJECT
Hizla Thana
Lakshmipur
Tetulia River
N
0 5 10 15 Km
LEGEND
Existing Embankment
Capital City
District Towns
Other Towns
Figure 2
ISPAN FAP 19
Geographic Information System

to a major problem of landlessness in Matlab. Because of these shortages the land is used almost exclusively for agricultural purposes and trees only exist around raised household areas. The monsoon climate of the study area is characterized by high temperatures, heavy rainfall, and marked seasonal variation (Rashid, 1991).[10]

The Matlab study area has a major environmental division. During the mid-1980s, as part of the Bangladesh Flood Action Plan, a flood-control embankment was built (called the Meghna-Dhonagoda Irrigation Project by the Government of Bangladesh). The embankment is a compartmentalization scheme that regulates the amount of water that enters the embanked area. During the monsoon season flooding is regulated inside the embankment. There are three growing seasons for rice inside the embankment and only two outside the embankment where flooding is unregulated. Figure 2 shows the location of the embankment relative to several rivers and Matlab Town. The Matlab study area is in the highly fertile Meghna Flood Plain and because of its proximity to the confluence of the Padma and Meghna Rivers the southern part of the study area was traditionally subject to

---

[10] The following average monthly rainfall and temperature data were collected at a weather station near Matlab.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rain	.66	2.74	4.98	19.30	28.60	52.20	52.73	41.61	29.13	23.75	4.88	.69
MaxT	26.1	29.5	33.6	33.8	33.6	32.2	31.4	31.9	31.2	31.8	29.9	26.9
MinT	10.7	14.8	17.9	22.2	23.2	23.8	25.1	24.7	24.2	22.9	18.2	11.8

(Data are from a weather station in Comilla. Rainfall data in centimeters, collected from 1947-77. MaxT and MinT are the maximum and minimum monthly temperature averages in degrees Celsius, collected from 1971-81).

massive erosion. During the 1988 flood several square kilometers of land were lost into the Meghna River. The embankment was built in an attempt to contain the floods to decrease erosion and increase the number of rice growing seasons from two to three.

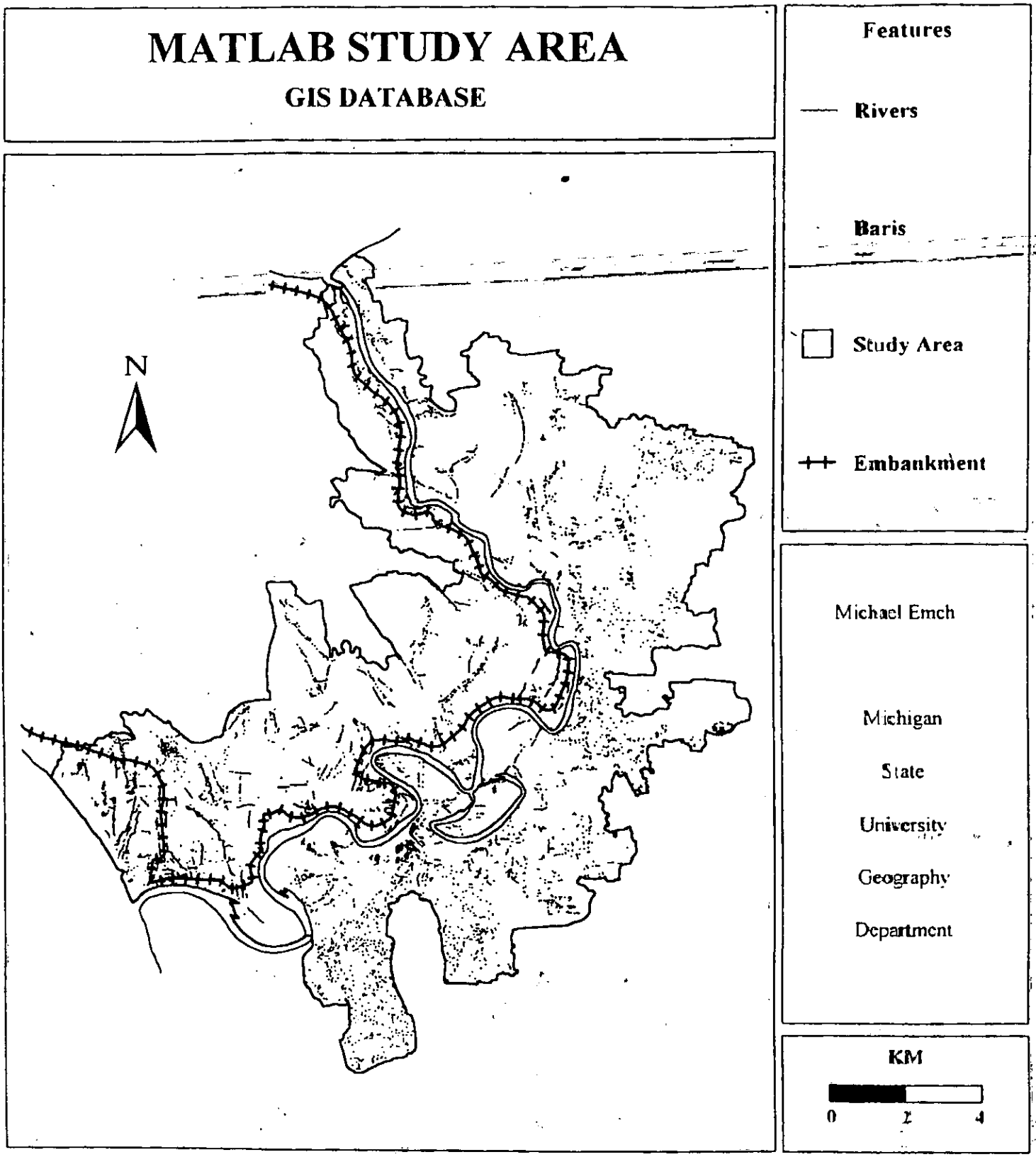
Rice dominates agriculture in the Meghna Floodplain. Rice crops are mainly local varieties including *aman*, *aus*, and *boro* but high-yielding varieties are increasingly used inside the embankment. Other crops include, potatoes, jute (although its production has declined in the past 15 years), mustard, onions, garlic, and chilies. Sugarcane and various vegetables and fruits are grown in small amounts. The other main occupation in Matlab is river fishing. Much of the fishing in Matlab is done from river banks with small nets for subsistence. People also fish from small boats using nets and they sell their catch in the small Matlab market or at the major fishing center of Chandpur ten km south of the study area (Rashid, 1991). The only other notable economic venture is livestock rearing. Almost every household I have been to in Matlab has a couple of underfed, uncaged chickens wandering about and occasionally has a cow or goat. I have never seen any larger scale livestock operations in the area but I do not have data to verify this.

While at ICDDR,B in 1993 I was part of a team that created a vector GIS database of this field research area. Features in digital format include extended household units called *baris*, rivers, roads, schools, religious structures, village boundaries, the flood-

control embankment, and health facilities. Figure 3 shows three features in the GIS database including the flood-regulating embankment, the Dhonogoda River, and *baris*. The area northwest of the embankment is the flood-regulated area and the area southeast of the embankment is the unregulated area. *Baris* are patrilineally-related clusters of households that are raised above the surrounding land area that is used for agriculture. The *baris* are all identified by an ICDDR,B demographic surveillance system (DSS) census number within the structure of the GIS database. This allows me to link attribute data, including disease incidence data, to specific locations.

The Matlab field research center is a diarrhea treatment center (DTC) which has in- and out-patient services, a laboratory for the identification of pathogens, and research facilities. The Matlab DTC treats about 7,000 to 8,000 diarrhea cases per year and to date more than 230,000 diarrhea patients have been treated by the Centre. There are motorized boats which function as a free ambulance service for diarrhea patients so access to the hospital is remarkably good. Also, all DTC services are free. The research center maintains a community-based data collection system. One-hundred and twenty community health workers (CHWs) visit each household every two weeks to collect demographic, morbidity, and other data. The DSS conducts periodic censuses (most recently in 1993) and uses CHWs to update demographic data (births, deaths, and migrations). The DTC laboratory consists of microbiology, clinical pathology, and bio-chemistry units which

Figure 3


MATLAB STUDY AREA
GIS DATABASE
Features
Rivers
Baris
Study Area
Embankment
N
Michael Emch
Michigan
State
University
Geography
Department
KM
0
2
4

provide diagnostic services to the hospital and for field research activities.

V. Specific Research Questions

This research project will measure whether there are differences between spatio-temporal patterns and risk factors of cholera and non-cholera watery diarrhea in Matlab. It will address the following questions: (1) What are the spatio-temporal patterns of these two disease categories?, (2) What are the similarities and differences between the spatio-temporal patterns of the two diseases? (3) What are the biological, socio-economic, behavioral, and environmental variables (risk factors) involved in the disease ecology of cholera?, (4) What are the risk factors involved in the disease ecology of non-cholera watery diarrhea?, and (5) To what degree are the risk factors for these two diseases similar and different?

VI. Data Sources

I will employ a number of data collection methods in this study. These include: (1) collecting diarrheal disease data (dependent variables) from Diarrhea Treatment Center records; (2) collecting data for variables I hypothesize to be related to diarrheal disease by administering a questionnaire to diarrheal disease cases and controls; (3) obtaining demographic data from ICDDR,B Demographic Surveillance System records and community health worker record books; and (4) creating data for spatially constructed variables using the aforementioned GIS database. These four data sources are each

discussed in further detail below.

(1) Diarrheal disease data will be collected for people from the Matlab treatment area who were hospitalized at the DTC with watery diarrhea from 1992 to 1994. The cases will be assigned to one of three diarrheal disease categories which will be used as dependent variables in the analysis stage of the research: (A) Vibrio cholerae O1, (B) Vibrio cholerae O139, and (C) non-cholera watery diarrhea. For each patient admitted to the Matlab diarrhea treatment center from the study area, a stool sample is collected and routinely tested for Vibrio cholerae and Shigella, a dysenteric agent. If the agent is Vibrio cholerae then the strain (O1 or O139) is determined. In this study, laboratory records of the patients will be used to assign one of the three above agent categories. Hospital records specify whether or not there was blood in each patient's stool. Patients who tested positive for Shigella or who had blood in their stool will be excluded because this study is not concerned with dysentery. The cases that did not have dysentery or cholera will be assigned to the non-cholera watery diarrhea category. There are approximately 4000 patients admitted annually to the Diarrhea Treatment Center from the Matlab study area. Approximately 70 percent of these patients have watery diarrhea. Thus, in the three year study period there are approximately 8400 cases of watery diarrhea. All of these cases will be determined for mapping purposes but independent variable data will be collected for a random sample of these cases.

One community-based control will be chosen for each case. After the cases are defined, a list of potential controls will be compiled from DSS records. A person is eligible to be a control if he/she lives in the Matlab surveillance area, was not admitted to the DTC during the study period, and did not die of a diarrheal disease during the study period. If a potential control died during the study period, the cause of death will be determined from DSS records and if one of the causes was a diarrheal disease then this individual will be excluded from the control group. The number of cases and controls will be determined by budgetary considerations. While I hope to collect data for 500-1000 cases and 500-1000 controls, I will collect a minimum of 186 cases and 186 controls. (See Appendix 1 for minimum sample size calculation).

Information will be collected on the following independent variables hypothesized to be related to watery diarrhea. This information will be collected by administering questionnaires, obtaining secondary data from DSS records and community health worker record books, and creating data for spatially constructed variables using the GIS database. These data will be collected for both cases and controls. Interval-ratio level data will be collected whenever possible, however, some of these variables are inherently nominal or ordinal. (See Appendix 2 for individual discussion of variables).

1) People per dwelling area (overcrowding measure). Group-level Behavioral Variable and/or Socioeconomic Variable, collected from CHW books.
2) *Bari*-specific population density (overcrowding measure). Group-level Behavioral Variable and/or Socioeconomic Variable, constructed using GIS and DSS data.

3) Tubewells/person/ area around *baris*. Environmental Variable, constructed using GIS and CHW special study data.
4) Distance from river. Environmental Variable, constructed using GIS.
5) Source of water used for cooking, drinking, and bathing. Behavioral Variable, collected from questionnaire.
6) Flood control (embanked or not embanked area). Environmental Variable, constructed using GIS.
7) Breast feeding status for children under five. Biological or Behavioral Variable, collected from CHW books.
8) Measles vaccine status for children under five. Biological Variable, collected from CHW books.
9) Housing construction material type. Socio-economic Variable, collected from questionnaire.
10) Land and livestock ownership. Socio-economic Variable, collected from questionnaire.
11) Cash income and food stocks. Socio-economic Variable, collected from questionnaire.
12) Maternal and paternal education. Socio-economic Variables, collected from questionnaire.
13) Paternal occupation. Socio-economic Variable, collected from questionnaire.
14) Arm circumference for children under five. Biological Variable, collected from CHW books.
15) Tibial edema (used for kwashiorkor detection) for children under five. Biological Variable, collected from CHW books.
16) Latrine type. Behavioral Variable, collected from CHW special study.
17) Latrine use. Behavioral Variable, collected from questionnaire.
18) Knowledge of etiology and transmission of diarrhea. Behavioral Variable, collected from questionnaire.
19) Consumption of shellfish. Behavioral Variable, collected from questionnaire.
20) Gender. Biological or Behavioral Variable, collected from DSS records.
21) Age. Biological or Behavioral Variable, collected from DSS records.

(2) Data from questionnaires will be collected from a random sample of cases and their controls. The questionnaire will be administered by four ICDDR,B trained enumerators. Upon arrival to Bangladesh the questionnaire will be translated into Bengali and back to English by another person to test the accuracy of the translation. A pre-test will then be conducted. Variable measurement will be refined and variables may be added or subtracted based on the pre-test results. I will add a variable if it is thought to be

responsible for some of the spatio-temporal variation of either cholera or non-cholera watery diarrhea. A complete list of these variables cannot be compiled until a pre-test is completed (preliminary questionnaire variables are identified in the aforementioned variable list and in Appendix 2). After the questionnaire is completed it will be administered to all of the chosen cases and controls. If the individual who is the case or control is present, they will be asked the questions directly. If this person is a minor, has diminished mental capacity, or died during the study period then the head of household will be asked the questions. Because these variables are mainly socio-economic variables (i.e., they are household level variables) asking the head of household should accurately measure the variable. If a case has migrated out of the study area then there may be a problem of bias. I will attempt to collect whatever data I can from family members. If the family has migrated out of the area then I will try to determine why they left from neighbors so that I can comment on any apparent patterns to the migration process that might be related to bias.

(3) The Matlab study area has a population of approximately 200,000. The ICDDR,B DSS has a computerized database of everyone in the area. The age and gender of each of the cases and controls will be determined. Also, the cause of death of each potential control will be determined so that controls can be chosen. Variable number two will be created using the GIS with *bari* population data that will be taken from DSS

records.

(4) Several variables (numbers 2, 3, 4, and 6) will be spatially constructed using the GIS database and other attribute data. These variables involve spatial calculations using features in the GIS database as well as attribute data derived from other sources. (See Appendix 2 for detailed descriptions of the creation of each of the variables).

The data collection stage of this research will take 12 months to complete. The questionnaire pre-test and GIS database update will be conducted from October through December 1995. During this time the diarrhea cases will be determined from laboratory and hospital records at the Diarrhea Treatment Center. Also, a list of controls will be chosen. From January through September 1996 the questionnaires will be administered and secondary data will be collected from the community health worker records and DSS records.

VII. Analytical Methods

Analytical methods used in this study will include: (1) disease mapping; (2) a *bari*-level logistic regression analysis; (3) a case-control study; and (4) an individual-level logistic regression analysis. These methods are discussed in further detail below.

(1) I will map the *bari* locations of the three watery diarrhea groups (O1, O139, and non-cholera) in two week periods so that spatio-temporal patterns can be identified. I will determine whether there are any differences between the spatio-temporal patterns of the two

strains of cholera and non-cholera watery diarrhea. While the aquatic reservoir for cholera can not be mapped in this study, knowing the differences in the spatio-temporal patterns of cholera and non-cholera diarrhea may give insight into the general location of the reservoir. The spatio-temporal patterns will also be analyzed by building spatial and temporal variables into logistic regression models (see parts 2 and 4 of this discussion of analytical methods).

(2) I will create separate logistic regression models of cases/non-cases for each of the three categories of watery diarrhea using the variables created in the GIS (2, 3, 4, and 6) as the independent variables.[11] This *bari*-level analysis will be conducted using all hospitalized diarrhea cases. *Baris* in the study area that experienced at least one case of watery diarrhea or cholera during the study period are cases and all *baris* in the study area that no individual contracted disease are non-cases. This can be done for all the *baris* in the study area because variables 2, 3, 4, and 6 do not require questionnaires or manual data searches. They will be created using a computerized database, thus, collecting data for a larger number of observations will not be a more difficult task. I will also incorporate seasons into the models as dummy variables to add a temporal component. The *bari* will be

[11] After the data are collected I will test to see if my independent variables are multi-collinear. Since multi-collinearity is a violation of multiple logistic regression, if I find multi-collinearity in my independent variables I will either conduct a principle components analysis and use a grouped component in place of multi-collinear independent variables or I will drop independent variables from the model.

the unit of measurement for this analytical method. I will test to see whether the variables are related to incidence of the three types of watery diarrhea. I will also analyze the residuals of the logistic regression models to see if there is any spatial pattern. This will indicate whether the fit of the model is better in some areas versus others. If there is a spatial pattern to the residuals then I will investigate the reason for the pattern. I will determine whether the residuals are spatially autocorrelated so that I may identify whether there are apparent spatial patterns in the unspecified variables.

(3) I will conduct a case-control study to identify which of the aforementioned variables are risk factors for cholera and non-cholera watery diarrhea as well as to compare the risk factors for the three types of diarrhea. In case-control studies, comparisons are made between a group of persons that have a disease and a group that does not. Those individuals with the disease are cases and those without the disease are controls. Exposures or risk factors of interest can be compared between cases and controls. The proportion of cases possessing a risk factor of interest can be compared to the corresponding proportion in the control group. Statistical comparisons of the frequencies of individuals with and without risk factors provide information about what variables put an individual at highest risk for a disease. Individuals will be the unit of study and all of the independent variables will be used in this part of the study. Cases are the hospitalized diarrhea cases (the two types of cholera and non-cholera diarrhea) and the controls are from the community

(discussed in section VI of this proposal). I will calculate risk for each of the independent variables by calculating odds ratios (a non-parametric statistical comparison of ratios) and their significance values. A comparison of risk factors of cholera and non-cholera watery diarrhea will be conducted by using non-cholera diarrhea as the control group. This will determine the differences in relative importance of different risk factors for the three types of watery diarrhea.

(4) I will conduct a second logistic regression analysis at the individual level using the cases and controls as the binary variable. This analysis is different than the logistic regression performed in section two of this analytical methods section because this analysis is done at the individual level. The cases will be given a value of 1 and the controls a value of 0. All of the independent variables will be used to create this model.[12] As with the *bari*-level analysis of only the GIS variables, I will incorporate seasons into the model as dummy variables to add a temporal component. I will test to see whether the variables are related to incidence of the three types of watery diarrhea. As with the *bari*-level analysis, I will analyze the residuals of the logistic regression models and determine whether the residuals are spatially autocorrelated.

Mapping the incidence of watery diarrhea every two weeks will accurately describe spatial and temporal patterns of the two diarrheal disease categories. It is hoped that these

[12] See footnote 11.

maps will offer corroborating evidence of the existence and importance of an environmental reservoir for cholera. If cholera has an important environmental reservoir then there should be seasonality and spatial association with the aquatic reservoir. If there is not an environmental reservoir for non-cholera diarrhea there will be less temporal clustering of variables and no association with the aquatic reservoir. As mentioned earlier, the aquatic reservoir cannot be located without detailed environmental microbiological studies, however, describing the spatio-temporal patterns can provide insight into the existence and importance of this reservoir. The case-control study and logistic regression analyses identify and compare risk factors of cholera and non-cholera diarrhea.

Potential Biases

There are potential biases in the study area because of the existence of Community Operated Treatment Centres (COTC). Access to the COTCs may vary geographically throughout the study area; thus there is a possibility that people in the areas around the COTCs will report to them rather than to the Diarrhea Treatment Centre (DTC). After the data are collected I will use the GIS database to analyze whether or not incidence rates are related to proximity to the COTCs. Their locations are included in the GIS database. If they are related then I have no choice but to limit the study area to the areas that are not near the COTCs. This is roughly half of the Matlab study area. Since I plan to collect data for many more individuals than needed statistically, this will not be a problem to do after the data have been collected.

Another potential bias in this study is that there may be differences in clinical presentation between cholera and non-cholera watery diarrhea at the DTC and COTCs. This potential bias can also be investigated using the GIS after the data are collected. Using the GIS I will determine whether both types of diarrheal incidence are related to proximity to the DTC and if these relationships are similar to one another. If the relationships between the two types of diarrhea and proximity to the DTC are significantly different from one another then I must limit the study area to the area not within the COTC catchment areas. An alternative way to deal with these problems of potential bias is to make the areas not near the COTCs the project study area from the beginning. I do not want to do this because I think the exploration of different use patterns will be interesting data in and of themselves. Also, since I will be collecting more data than statistically required, I will not have the problem of too few observations.

VIII. Research Affiliation and Authorization

This research will be conducted under the auspices of both Michigan State University (MSU) and the ICDDR,B. The proposal has already been accepted by the MSU Human Subjects Committee and has been submitted to the ICDDR,B Research Review Committee. While at the ICDDR,B I will be working within the Community Health Division (CHD) where I worked in 1993 and the summer of 1994. This research will be funded by the following grants: the Association of American Geographers (AAG) dissertation grant, the Foreign Language and Area Studies Grant (FLAS) from the United States Department of Education, and the American Institute of Bangladesh Studies (AIBS)

dissertation grant. In the process of obtaining AIBS funding this proposal received research clearance from the Government of Bangladesh.

IX. Summary of Research Schedule

I will begin my dissertation project in October 1995 and complete the study within one year. I will pretest and finalize survey questions by December 1995. During this time (October to December 1995), I will also update the GIS database which was created during 1993. From January to September 1996, I will employ four community health workers to administer the survey questionnaires. I will then enter the survey data digitally and begin to conduct preliminary analyses. Upon completion of the field research, I will return to MSU to complete the dissertation. After defending the dissertation, the research results will be given to the ICDDR,B and Dhaka University libraries.

Appendix 1: Minimum Sample Size Calculation

The minimum sample size will be determined by budgetary considerations. I will maximize the number of cases and controls within given time and money constraints. However, the following discussion shows calculations for a minimum sample size. Hsieh (1989) explained sample size determination for logistic regression analysis based on methods described in Whitemore (1981). Required input data using this method for simple logistic regression include:

(1) the probability (P) of events at the mean value of the covariate (.01-.5).
(2) the odds ratio (**r**) of disease corresponding to an increase of one standard deviation from the mean value of the covariate.

Hsieh (1989) provided tables of sample size (S) for different probability values based on event proportion and odds ratios. This figure must be multiplied by the number of controls that will be used for each case (x). With multiple logistic regression a multiple correlation coefficient (ρ) must be estimated and the simple logistic regression sample size is then divided by ($1-\rho^2$). Thus, the formula to determine a multiple logistic regression sample size is:

$$\frac{Sx}{1-\rho^2}$$

where Sx is based on P and **r**

First one must start with the covariate that is thought to explain the most variation of the dependent variable and estimate the odds ratio (**r**). Then the simple logistic regression sample size (S) should be looked up in the table. And lastly, the simple logistic regression sample size must be divided by ($1-\rho^2$).

Since this study is a case-control study the probability ($\mathcal{P}$) of events is .5. I estimate the odds ratio (**r**) for one of my covariates (latrine type) to be 2. And I will accept a multiple correlation coefficient (ρ) of .6. For a 95 per cent confidence level the sample size is 119. I will double this number (238) because I will sample one control for every case. Thus, the multiple regression sample size is:

$$\frac{238}{1-.6^2} = \frac{238}{.64} = 371.875$$

This is the sample size for both cases and controls. Thus, I will divide this number by two to get the sample size for cases and controls. The number is 185.9375 which is rounded up to 186 cases and 186 controls.

Appendix 2: Individual Discussion of Independent Variables

1) People per dwelling area (overcrowding measure). **Group-level Behavioral Variable and/or Socioeconomic Variable, collected from CHW books.** To create this variable the number of persons in a dwelling must be divided by the area of that dwelling. The CHWs document the area of each dwelling in Matlab. When a dwelling is renovated, the new area is recorded by the CHW. The DSS collects data on the population of each household in the study area. I hypothesize that watery diarrhea will be associated with crowding because it will increase the likelihood of human contact and, thus, person-to-person transmission.

2) *Bari*-specific population density (overcrowding measure). **Group-level Behavioral Variable and/or Socioeconomic Variable, constructed using GIS and DSS data.** I will calculate population density for each *bari*, using DSS census data and buffering techniques in the GIS. I have the same hypothesis for this variable as for variable one.

3) Tubewells/person/ area around *baris*. **Environmental Variable; constructed using GIS and CHW special study.** In 1988, the CHWs identified the location of all of the tubewells in the Matlab study area. Presently, these data are being updated. Using these data and the GIS database, I will calculate the number of tubewells within a half kilometer buffer around each *bari*. I hypothesize that the greater the number of tubewells near a *bari*, the less diarrheal disease will occur. This is based on the assumption that the existence of

tubewells generally indicates access to clean water.

4) Distance from river. **Environmental Variable, constructed using GIS.** In a village level study in Matlab from 1968 to 1977, Glass et al. (1982) found that people living in villages adjacent to rivers were less likely to contract cholera. With the GIS database I can test this more precisely by determining the distance of each *bari* from the river. I hypothesize that distance from the river is positively related to the incidence of cholera and not related to non-cholera watery diarrheal incidence. This conforms to Colwell's (1992) assertion that the most significant cholera reservoir is in the brackish canal environments further from the main river.

5) Source of water used for cooking, drinking, and bathing. **Behavioral Variable, collected from questionnaire.** A series of questions will be asked to determine the different sources of water that are used for different purposes by the cases and controls. These questions will reveal the result of how access and personal choice affect how water is used.

6) Flood control (embanked or not embanked area). **Environmental Variable constructed using GIS.** In the mid-1980s, a major flood control program was implemented in part of Matlab. An enclosed embankment (polder) protects part of the study area from flooding, while the area outside the embankment remains unprotected. I will determine whether each *bari* is inside or outside the embankment using the GIS. I hypothesize that watery diarrhea will occur at a higher rate outside the embankment than inside because water sources are

more likely to be contaminated when flooded.

7) Breast feeding status for children under 5. **Biological and/or Behavioral Variable, collected from CHW books.** The CHWs record the breast feeding status of all children under 5 years old. I hypothesize that for children under five, incidence of watery diarrhea will be lower if they are breast-fed because of acquired immunity and/or reduced ingestion of contaminated water sources.

8) Measles vaccine status for children under 5. **Biological Variable, collected from CHW books.** The CHWs record the vaccine status of all children under 5 years old. I hypothesize that watery diarrhea will be associated with children who are not immunized against measles because it was found that the measles vaccine is related to decreased diarrheal incidence (Feachem & Koblinsky, 1983).

9) Housing construction material type. **Socio-economic Variable, collected from questionnaire.** The CHWs record the construction material of each dwelling, including renovations, in the Matlab study area. Assuming that construction material (as well as variables 10, 11, 12, and 13) are indicators of socio-economic status, and that incidence of diarrheal disease is related to socio-economic status, I hypothesize that less expensive materials (and low socio-economic status in variables 10, 11, 12, and 13) will be associated with diarrheal disease. I do not think that the housing materials (or any of the other socio-economic variables) are directly involved in diarrheal disease transmission but that housing

construction material is an indicator of socio-economic status. If socio-economic status is related to diarrheal disease there are certainly complex reasons why. Access to clean water or latrines may be related to socio-economic status. These issues can be explored after the data are collected.

10) Land and livestock ownership. **Socio-economic Variable, collected from questionnaire.** The amount of land and livestock owned by each study household will be determined through a questionnaire designed to gather socio-economic data.

11) Cash income and food stocks. **Socio-economic Variable, collected from questionnaire.** The amount cash income and biannual food stocks for each study household will be determined through the questionnaire.

12) Maternal and paternal education. **Socio-economic Variables, collected from questionnaire.** The CHWs record the maternal and paternal educational levels in the record books.

13) Paternal occupation. **Socio-economic Variable, collected from questionnaire.** The CHWs record the father's occupation in the record books.

14) Arm circumference for children under 5. **Biological Variable, collected from CHW books.** The CHWs record the arm circumference of children every 3 months. This is an indicator of malnutrition. I hypothesize that malnutrition is not related to diarrheal disease based on previous studies identified in the literature review and I would like to verify this.

15) Tibial edema (used for kwashiorkor detection) for children under 5. **Biological Variable, collected from CHW books.** The CHWs record if a child has tibial edema. I hypothesize that kwashiorkor, a form of malnutrition, is not related to diarrheal disease.

16) Latrine type. **Behavioral Variable, collected from CHW special study.** A special latrine data collection project was conducted during 1988 and was updated by CHWs during 1994. Several nominal categories of latrine are recorded (i.e. closed or open pit; permanent or makeshift structure). I hypothesize that individuals who live in households with unsanitary latrine types (open pit & makeshift structure) or no latrine are more likely to get diarrheal disease.

17) Latrine use. **Behavioral Variable, collected from questionnaire.** A series of questions will be asked to determine whether and how often a latrine is used by cases and controls. These questions will reveal the result of how access and personal choice affect how latrines are used.

18) Knowledge of etiology and transmission of diarrhea. **Behavioral Variable, collected from questionnaire.** A series of questions will be asked about people's perceptions of diarrheal etiology and transmission. I hypothesize that individuals who do not understand diarrheal etiology and transmission are more likely to contract the diseases because they are less likely to modify their behaviors to avoid sources of the diseases.

19) Consumption of shellfish. **Behavioral Variable, collected from questionnaire.** Data

about people's shellfish consumption will be collected in the questionnaire. I hypothesize that individuals who eat more shellfish are more likely to contract the diarrheal diseases.

20) Gender. **Biological and/or Behavioral Variable, collected from DSS records.** Gender for all individuals is recorded by the DSS. I will do gender-matched analyses because incidence rates may differ by gender because of different behavioral patterns.

21) Age. **Biological and/or Behavioral Variable, collected from DSS records.** Age for all individuals is recorded by DSS. I will do age-matched analyses because incidence rates differ by age because of immunological differences and behavioral patterns.

Appendix 3: Research Budget

Mr. Michael Emch, P.I.

---

1. Stationary **$500**

Paper, copying, and office supplies.

---

2. Research/Enumeration Support **$ 4530**

Enumeration Costs: Including employment of 4 enumerators (24 person-months) to assist in survey interviews, enumerator transportation, data entry people, and translation of questionnaire.

Enumerator Salaries: 24 person-months * $100/month	$2400
Enumerator Transportation:	$400
Data Entry Salaries: 12 person-months * $140/month	$1680
Questionnaire Translation:	$50

Total $ 5030

Bibliography

Barrows, H.H. **1923** Geography as human ecology. *Annals of the Association of American Geographers* 1-14.

Baqui, A. H.; Black, R.E.; Mitra, A.K; Chowdhury, H.R.; Zaman, K.; Faveau, V.; and Sack, R.B **1994**. Diarrhoel diseases: Matlab experience. (unpublished manuscript)

Baqui, A.H.; Sack, R.B.; Black, R.E.; Haider, K.; Hossain, A.; Alim, A.R.M.A.; Yunus, M.; Chowdhury, H.R.; and Siddique, A.K. **1992**. Enteropathogens associated with acute and persistent diarrhea in Bangladeshi children <5 years of age. *The Journal of Infectious Diseases*. 166: 792-796.

Bairagi, R.; Chowdhury, M.K.; Kim, Y.J.; Curlin, G.T.; and Gray, R.H. **1987**. The association between malnutrition and diarrhoea in rural Bangladesh. *International Journal of Epidemiology*. 16(3): 477-481.

Becker, S.; Black, R.E.; Brown, K.H.; and Nahar, S. **1986**. Relations between socioeconomic status and morbidity, food intake and growth in young children in two villages in Bangladesh. *Ecology of Food and Nutrition*. 18: 251-64.

Benenson, A.S. **1990.** *Control of communicable diseases in man*. Washington, DC.: American Public Health Association.

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

Black, R.E.; Merson, M.H.; Rahman, A.S.M.M.; Yunus, M.; Alim, A.R.M.A.; Huq, I.; Yolken, R.H.; and Curlin, G.T. **1980**. A two-year study of bacterial, viral, and parasitic agents associated with diarrhea in rural Bangladesh. *The Journal of Infectious Diseases*. 142(5): 660-664.

Blaikie, P.M. **1994**. Political ecology in the 1990s: An evolving view of nature and society. CASID Distinguished Speaker Series No. 13, Michigan State University, Center for Advanced Study of International Development, East Lansing, Michigan.

Brigham, A.P. **1915**. Problems of Geographic Influence. *Annals of the Association of American Geographers*. pp 3-25.

Burkitt, D. **1962.** A "Tumour Safari" in east and central Africa. *British Journal of Cancer*. 16(3): 379-386.

Campbell, D.J. and Olson, J.M. **1991**. Framework for environment and development: The kite. Michigan State University, Center for Advanced Study of International Development, East Lansing, Michigan.

Chen, L.C.; Huq, E.; and Huffman, S.L. **1981**. A prospective study of the risk of diarrheal diseases according to the nutritional status of children. *American Journal of Epidemiology*. 114(2): 284-292.

Colwell, R.R.; Brayton, P.R.; Grimes, D.J.; Roszak, D.R.; Huq, S.A.; and Palmer, L.M. **1985**. Viable, but non-culturable Vibrio cholerae and related pathogens in the environment: implications for release of genetically engineered micro-organisms. *Bio/Technology*. 3: 817-820.

Colwell, R.R. and Spira, W.M. **1992**. The ecology of Vibrio cholerae. In D. Barua and W.B. Greenough (Eds.), *Cholera*. (107-128). New York: Plenum Medical Book Company.

Craig, M. **1988**. Time-space clustering of Vibrio cholerae O1 in Matlab, Bangladesh, 1970-1982. *Social Science and Medicine*. 26(1): 5-13.

Davis, W.M. **1915**. The principles of geographic description. *Annals of the Association of American Geographers*. 61-105.

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

Feachem, R.G. and Koblinsky, M.A. **1983**. Interventions for the control of diarrhoeal diseases among young children: measles immunization. *Bulletin of the World Health Organization*. 61(4): 641-652.

Glass, R.I,; Becker, S.; Huq, M.I.; Stoll, B.J.; Khan, M.U.; Merson, M.H.; Lee, J.V.; and Black, R.E. **1982**. Endemic cholera in rural Bangladesh, 1966-1980. *American Journal of Epidemiology*. 116(6): 959-970.

Glass, R.I. and Black, R.E. **1992**. The epidemiology of cholera. In D. Barua and W.B.Greenough (Eds.). *Cholera*. (pp. 129-154). New York: Plenum Medical Book Company. New

Glass, R.I.; Holmgren, J.; Haley, C.E., Khan, M.R.; Svennerholm, A.; Stoll, B.J.; Hossain, K.M.B.; Black, R.E., Yunus, M. and Barua, D. **1985**. Predisposition of cholera of individuals with O blood group. *American Journal of Epidemiology*. 121(6): 791-796.

Glick, B.J. **1980**. The geographical analysis of cancer occurrence: Past progress and future directions. In Meade, M.S. (Ed.), *Conceptual and methodological issues is medical geography*. 170-193. Chapel Hill, NC: University of North Carolina. Studies in Geography No. 15.

Glick, B.J. **1979**. The spatial autocorrelation of cancer mortality. *Social Science and Medicine*. 13D: 253-56.

Haggett, P. **1977**. *Locational analysis in human geography*. London: Edward Arnold.

Hartshorne, R. **1939**. *The nature of geography: A critical survey of current thought in the light of the past*. Lancaster, PA: Association of American Geographers.

Hartshorne, R. **1959**. *Perspective on the nature of geography*. Chicago: Rand McNally.

Harvey, D. **1969**. *Explanation in geography*. London: Edward Arnold.

Hsieh, F.Y. **1989**. Sample size tables for logistic regression. *Statistics in Medicine*. 8: 795-802.

Hughes, J.M.; Boyce, J.M.; Levine, R.J., Khan, M.; Aziz, K.M.A.; Huq, M.I. and Curlin, G.T. **1982**. Epidemiology of eltor cholera in rural Bangladesh: Importance of surface water in transmission. *Bulletin of the World Health Organization*. 60(3): 395-404.

Hunter, J.M.; Rey, L. and Scott, D. **1982**. Man-made lakes and man-made diseases: Towards a policy resolution. *Social Science and Medicine*. 16: 1127-1145.

Hunter, J.M. **1992**. Elephantiasis: A disease of development in north east Ghana. *Social Science and Medicine*. 35(5): 627-649.

Hunter, J.M. **1977**. The summer disease: An integrative model of the seasonality aspects of childhood lead poisoning. *Social Science and Medicine*. 11: 691-703.

Hunter, J.M. **1976**. Aerosol and roadside lead as environmental hazard. *Economic Geography*. 52(2): 147-160.

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

Huntington, E. **1924**. Geography as natural selection. *Annals of the Association of American Geographers*. 1-6.

ICDDR,B. **1993**. Briefing file of ICDDR,B Matlab Health & Research Centre. (unpublished manuscript)

Khan, M.; Mosley, W.H.; Chakraborty, J.; Sarder, A.M.; and Khan, M.R. **1981**. The relationship of cholera to water source and use in rural Bangladesh. *International Journal of Epidemiology*. 10(1): 23-25.

King, P.E. **1979**: Problems of spatial analysis in geographical epidemiology. *Social Science and Medicine*. 13D: 249-52.

Kloos, H. **1985**. Water resources development and schistosomiasis in the Awash Valley, Ethiopia. *Social Science and Medicine*. 20: 609-25.

Learmonth, A. **1988**. *Disease ecology*. New York: Basil Blackwell.

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

May, J.M. **1977.** Medical Geography: Its methods and objectives. *Social Science and Medicine*. 11: 715-730.

Mayer, J.D. and Meade, M.S. **1994**. A Reformed Medical Geography Reconsidered. *The Professional Geographer*. 46: 103-06.

Mayer, J.D. **1986**. Ecological associative analysis. In M. Pacione (Ed.), *Medical geography: Progress and prospect*. (64-81). London: Croom Helm.

Mayer, J.D. **1984**. Medical geography: An emerging discipline. *Journal of the American Medical Association*. 251: 2680-83.

Mayer, J.D. **1982**. Relations between two traditions of medical geography: Health systems planning and geographical epidemiology. *Progress in Human Geography*. 6:216-30.

McGlashan, N.D. **1967**. Geographical evidence on medical hypotheses. *Tropical and Geographical Medicine*. 19: 333-343.

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

Meade, M.S. **1979**. Cardiovascular mortality in the southeastern United States: The coastal plain enigma. *Social Science and Medicine*. 13D: 257-66.

Meade, M.S.; Florin, J.W.; and Gesler, W.M. **1988**. *Medical geography*. New York: The Guilford Press.

Morrill, R. **1970**. The spatial organization of society. Belmont, CA: Wadsworth Publishing.

Pattison, W.D. **1963**. The four traditions of geography. *Journal of Geography*. 211-216.

Paul, B.K. **1985**. Approaches to medical geography: An historical approach. *Social Science and Medicine*. 20:399-409.

Pyle, G.F. **1979**. Elements of disease ecology. In G.F. Pyle (Ed.), Applied medical geography. New York: John Wiley and Sons.

Pyle, G.F. **1977**. International communication and medical geography. *Social Science and Medicine*. 11: 679-682.

Rashid, E.R. **1991**. *Geography of Bangladesh*. Dhaka: University Press Limited.

Roundy, R.W. **1976.** Altitudinal mobility and disease hazards for Ethiopian populations. *Economic Geography*. 52(2): 103-115.

Sauer, C.O. **1925.** *The morphology of landscape*. Berkeley: University of California Publications in Geography.

Schaeffer, F.K. **1953**. Exceptionalism in geography: A methodological examination. *Annals of the Association of American Geographers*. 226-249.

Scholten, H.J. and de Lepper, M.J.C. **1991**. The benefits of the application of geographical information systems in public and environmental health. *World Health Statistical Quarterly*. 44: 160-170.

Semple, E.C. **1911**. *Influences of geographic environment on the basis of Ratzel's system of anthropo-geography*. New York: Henry Holt.

Siddique, A.K.; Zaman, K.; Akram, K.; Mutsuddy, P; and Sack, R.B. **1994**. Emergence of a new epidemic strain of Vibrio cholerae in Bangladesh: An epidemiological study. *Tropical and Geographical Medicine*. 46(3): 147-150.

Snyder, J.D. and Merson, M.H. **1982**. The magnitude of the global problem of acute diarrhoeal disease: A review of active surveillance data. *Bulletin of the World Health Organization*. 60: 605-613.

Sommer, A. and Woodward, W.E. **1972**. The Influence of protected water supplies on the spread of Classical/Inaba and El Tor/Ogawa cholera in rural East Bengal. *The Lancet*. 2: 985-987.

Stanton, B.F. and Clemens, J.D. **1986.** Soiled saris: A vector of disease transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 80: 485-488.

Taaffe, E.J. **1974**. The spatial view in context. *Annals of the Association of American Geographers*. 1-16.

Warren, K.S. and Mahmoud, A.A.F. **1993.** *Tropical and Geographical Medicine*. New York: McGraw-Hill Book Company.

Whittemore, A.S. **1981**. Sample size for logistic regression with small response probability. *Journal of the American Statistical Association*. 76(373): 27-32.

CONSENT FORM

The International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) is conducting a study on watery diarrhoea. We are trying to find out what are the factors that are involved in the occurrence of the disease in your community.

For that purpose we need to ask you several questions related to your environment, customs and your health status. The questionnaire will be addressed only once, and will take about half an hour. We invite you to participate in this study, and refrain from answering one of the question if you feel so.

By answering to these questions, we expect to learn more on the disease and find new ways for its prevention.

The information provided by you will be treated as confidential, and kept by the Principal Investigator.

If you agree to participate, please sign your name below, or give your left thumb impression.

Thank you.

Signature of the Principal Investigator

Date:_____

Signature or left thumb impression of the interviewee

Date:_____

MICHIGAN STATE
UNIVERSITY

April 28, 1995

TO: Michael Emch
316 Pinoak #2
E. Lansing, Mi. 48823

RE: IRB#: 95-213
TITLE: RISK FACTORS FOR DIARRHEAL DISEASE IN MATLAB, BANGLADESH: A MEDICAL GEOGRAPHIC APPROACH
REVISION REQUESTED: N/A
CATEGORY: 2-H
APPROVAL DATE: 04/27/95

The University Committee on Research Involving Human Subjects'(UCRIHS) review of this project is complete. I am pleased to advise that the rights and welfare of the human subjects appear to be adequately protected and methods to obtain informed consent are appropriate. Therefore, the UCRIHS approved this project and any revision listed above.

RENEWAL: UCRIHS approval is valid for one calendar year, beginning with the approval date shown above. Investigators planning to continue a project beyond one year must use the green renewal form (enclosed with the original approval letter or when a project is renewed) to seek updated certification. There is a maximum of four such expedited renewals possible. Investigators wishing to continue a project beyond that time need to submit it again for complete review.

REVISIONS: UCRIHS must review any changes in procedures involving human subjects, prior to initiation of the change. If this is done at the time of renewal, please use the green renewal form. To revise an approved protocol at any other time during the year, send your written request to the UCRIHS Chair, requesting revised approval and referencing the project's IRB # and title. Include in your request a description of the change and any revised instruments, consent forms or advertisements that are applicable.

PROBLEMS/ CHANGES: Should either of the following arise during the course of the work, investigators must notify UCRIHS promptly: (1) problems (unexpected side effects, complaints, etc.) involving human subjects or (2) changes in the research environment or new information indicating greater risk to the human subjects than existed when the protocol was previously reviewed and approved.

If we can be of any future help, please do not hesitate to contact us at (517)355-2180 or FAX (517)336-1171.

Sincerely,

David E. Wright, Ph.D
UCRIHS Chair

DEW:pjm

cc: Edward A. Whitesell

OFFICE OF
RESEARCH AND GRADUATE STUDIES
ersity Committee on Research Involving Human Subjects (UCRIHS)
ichigan State University
Administration Building
East Lansing, Michigan
48824-1046
517/355-2180
FAX: 517/432-1171