A STOCHASTIC ANALYSIS OF THE DIFFUSION OF CHOLERA IN INDIA IN 1971

By

SUJATA WARRIER MITRA

Bachelor of Arts

University of Calcutta

Calcutta, India

1976

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
May, 1980

A STOCHASTIC ANALYSIS OF THE DIFFUSION

OF CHOLERA IN INDIA IN 1971

Thesis Approved:

Robert E. Harris
Thesis Adviser

Stephen W. Tweedie

K.D. Harries

Norman N. Durham
Dean of the Graduate College

PREFACE

This study examined the diffusion patterns of cholera within the country of India. The primary objective was to analyze the diffusion patterns as related to migration by the use of the Markov Chain technique and predict the future distributions of deaths from cholera. The objective was to list the adequacy and the applicability of the technique as a tool for disease research.

This investigation owes its existence to the geography graduate program at O.S.U. The flexibility and academic freedom of the program has allowed me to pursue my own interests. Sincere appreciation is extended to all faculty members of the department and their wives for making my stay here a memorable educational experience.

I want to thank my committee chairman, Dr. Robert E. Norris, for his guidance and assistance throughout the study. Special gratitude is also extended to Dr. Stephen W. Tweedie and Dr. Keith D. Harries for their continued support and advice throughout this study. A note of thanks is also extended to Gary Lauver and Meg Kletke in the preparation of the programming problems. Without all of them, this thesis would never have been possible.

Thanks are also due to all my friends who encouraged me all along and to Linda for typing my thesis in so short a time.

Finally, I would like to thank Steve and his family for helping me learn so much of the American culture and my parents and sister for their understanding and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

AP	Andhra Pradesh
AS	Assam
BI	Bihar
GJ	Gujarat
HY	Haryana
HP	Himachal Pradesh
JK	Jammu and Kashmir
KE	Kerala
MP	Madhya Pradesh
MH	Maharashtra
MN	Manipur
ME	Meghalaya
MY	Mysore (Karnataka)
NG	Nagaland
OR	Orissa
PJ	Punjab
RJ	Rajasthan
TN	Tamil Nadu
TP	Tripura
VP	Uttar Pradesh
WB	West Bengal

CHAPTER I

INTRODUCTION

Man exists in a state of dynamic equilibrium with his total environment. Forming an integral part of the ecological system, he strives to maintain the equilibrium in a constant effort for survival. The phenomena known as life is generated by chains of nucleoprotein--capable of infinite adaptive modifications (50, pp. 8-12). Operating on the principle of survival, man interacts with other nucleoprotein that includes organisms of all sizes, shapes and features thereby maintaining the equilibrium or homeostatic relationship with the environment (23, p. 1). Changes in the nature of either the host or the agent can cause an imbalance which manifests itself in the form of disease. Disease cannot arise without the convergence in space and time of an external stimulus (agent) and the factors conditioning response ot that stimulus (host).

The appearance and character of diseases, as a global phenomena are subject to historical development and varying geographical and demographic characteristics of the population. The overall picture of disease varies temporally and spatially (19, p. 1). The disease ecology is further complicated as agents range from morphologically simple living entities like viruses to complex animals.

Micro-organisms concerned with their own biological survival in human tissue contribute to more than two-thirds of all human illness.

Burton and Smith (10) explain this:

> For the parasite to affect the host fatally is not usually the most successful result, as the infecting agent may perish with the host. Long continued association of parasite with a host population results in the evolution of a biological balance, satisfactory to the parasite and generally tolerable to the host (p. 112).

The situation permitting, the parasites then multiply which then leads to transfer elsewhere. The mechanism used for transmission by different infectious diseases are varied. In come cases, the transmission may be direct from one host to another whereas in other cases, an intermediate vector is involved in the transmission (50, p. 19). Mechanical vectors that carry disease externally are not essential to the host cycle, while the reservoir is free of biological vectors (28, p. 47). In contrast to vectored diseases are the non-vectored diseases where a disease is transmitted from one to another without an intermediate host. In such cases, reservoirs may be either living hosts or nonliving factors in the environment. These reservoirs are later identified as endemic areas where common environmental risk factors are to be found (50, p. 19).

Non-vectored disease, such as influenza, the common cold, cholera, etc. are transmitted from one human being to another (14, p. 81). Three modes of transmissions have been recognized in the case of non-vectored diseases. The first is simple expulsion of the agents from one person where upon they pass through the air and into the respiratory tract of the victim. Transmission also occurs through infected fecal matter that has contaminated water or any other materials which is later handled by the potential victim. The final mode of transmission occurs through direct bodily contact.

Cholera, a non-vectored disease, is a specific gastro-intestinal

infection that some observers found long before Koch discovered Vibrio-Cholerae in 1883. Macnamara, who first discussed cholera in his book of Asiatic cholera in 1876, pointed out that the cause of the disease consisted of a pestilential virus following the modes of transmission outlined above.

The endemic areas of cholera are comparatively limited. The two most important ones are in China and India. The latter is located in the southern part of Bengal with Calcutta as its center, and spreads widely from this focus. Studies of contagious diseases frequently omit analytical studies of the spatial component of the process of epidemiology. Most of these studies have examined man-environment relations, interactions and spatial movement. However, these studies have not examined the importance of internal migration and its relation to the spread (or diffusion) of diseases. Such diffusions have resulted in the creation of havoc in our economic, social and national life. Spatial patterns including the direction of the movement of any disease have to be examined in relation with the mobility of people within the borders of a nation. Such studies are required more in the case of contagious disease, as they are spread by close contact between the victim and the person harboring the disease. The nature of contagious diseases determines their spread and tracing and relating the movement of the disease to migration will help in controlling the spread of the disease.

The increasing complexity of problems associated with major contagious diseases in underdeveloped countries demand new solutions and new ideas for both individual and collective welfare. The association of cholera and contact between people in the spread of the disease makes

makes clear the need to examine spatial aspects of this contagious disease. The occurrence of any infectious disease is a given series of interaction between the host, the infectious agent and the environment. Where the population is fairly mobile the likelihood of interaction increases and so does the spread of disease.

The study of the diffusion of cholera in India affords an ideal way of using a spatial-environmental analysis, it would help in furthering the understanding of the disease. Also, it would further the understanding of its relations with other phenomena especially that of internal mobility, the associative occurrences in space and time which might serve as a tool in the eradication of the disease.

Cholera: The Disease

Cholera assiatica or Asiatic cholera (from the Greek word vibrio cholerae alluding to evacuation) is caused by the microbe spirillium vibrio cholerae. It is an acute infectious disease that is characterized by profuse, uncontrollable diarrhea and vomiting. This leads to the symptomatic excessive loss of fluids and chlorides from the tissues and ultimately result in muscular cramps and collapse. The causal organism also known as vibrio comma is described by Pullen (49) as a

> motile, comma-shaped organism, 1.5 microns by 0.2 to 0.4 microns, with a single polar flagellem; it stains easily with weak carbol-fuschia, is negative, and grows easily on ordinary bacteriological medium at 37°C (p. 948).

Cholera is invariably contracted in the immediate vicinity of cholera patients through direct contact. Other modes of transmission are the partaking of foodstuffs and using eating utensils that have been contaminated by the vibrio while kept in the vicinity of the patient.

Transmission also occurs through water which is the most common medium of infection. Roaches and houseflies which contaminate food are also contributing factors (49, p. 951).

Cholera passes through the following stages in its development. The first sequence is the incubation period which is 24-72 hours. During this stage there are no visible outer signs or symptoms of the disease. Hence, the vibrio having entered through the mouth, passes the first barrier--the stomach. During the second stage, after the incubation period, the vibrio moves on and finds a favorable medium in the small intestine where it starts to multiply in the lumen of the bowel. In this second stage the toxic changes that occur in the body leads to a fall in blood pressure and a stagnation of blood especially in the kidneys. The internal changes in the second stage manifest itself in profuse, watery diarrhea and vomiting and sets the stage for dehydration, severe muscular cramps, shock and azotemia. The symptoms mentioned above may last from one to more than five days and are known as the stage of evacuation.

The stage of evacuation either terminates naturally in death or the victim may recover if a treatment of saline drips is utilized. If this final stage is of unusually long duration--over seven days--the recovery is usually associated with severe toxemia. On the other hand, if the stage is short, the recovery is uneventful (49, p. 952).

Factors Influencing the Contagion of Cholera

Evidence in ancient Indian literature indicates the existence of a syndrome showing clinical symptoms that resemble cholera. But, cholera, in its pandemic form in India is a relatively new disease dating

back only as far as 1817. McNamara in his work History of Asiatic Cholera, indicates the presence of this disease in India. Researchers working on the evidence found have since indicated certain factors that influence the contagion of cholera.

The factors are important in understanding the epidemiology of cholera are that: 1) man is the only true reservoir of infection as no other animal is truly susceptible to infection and as the vibrio does not survive for long as a saprophyte; 2) in the true sense, there is no such thing as a cholera carrier as the infection is maintained by a rapid passage from person to person; 3) that there is no solid immunity after an attack of cholera. Hence, it is possible for a person to have an attack of cholera a few years after a previous attack. Therefore, the number of people susceptible to the disease are never exhausted (49, p. 950).

The factors controlling the epidemiology, evidently can be very easily maintained under:

> sanitary conditions that allow more or less free transference of infection from person to person, where there is a sufficient density of population to ensure an inexhaustible supply of susceptible persons, where the climatic conditions are suitable, that is where there is a moderately high temperature and a high humidity throughout the year (49, p. 950).

In endemic areas, the disease is never totally absent but outbreak might occur simultaneously at places distant from one another and under suitable climatic conditions then spreads in a wave-like form to areas generally free of cholera (46, p. 53). The spread of the disease relates closely to major lines of communication particularly when they are disorganized, as in the case of seasonal or periodic labor migrations and those related to religious fairs and pilgrimages (46, p. 53).

The major problem in dealing with cholera is that the infection may be symptomless or vary in intensity from mild diarrhea to the severe and rapidly fatal cholera. The close resemblance of the early stages of cholera to that of mild diarrhea results in the infection passing unnoticed and being treated for diarrhea. Moreover, from the point of epidemiology and prevention, many of these patients pass the vibrio in their stools. And mainly from this point, it becomes necessary to make early bacteriological diagnosis even if the case is a sporadic one.

In order to curb the deaths that result from the disease, it is essential to trace the movement of the disease as related to migration so that not only is the endemic area more clearly defined but the susceptible areas are also known. The knowledge gained thereby will help in affecting better control measures.

Scope of the Investigation

As stated earlier, Asiatic cholera is caused by the microbe spirillium vibrio cholerae descended from those vibrio living in water particularly those found in tanks or surface water which are abundant in India, especially in Bengal in the Ganges-Brahmaputra delta. Over the millenia people drinking this water must have been infected with these thereby acquiring them as saprophytes. With the steady increase in population, there was more rapid passage of the organism from one human being to another. The process of pathogenic selection in this area was hastened by the position of Calcutta and its numerous millions and resulting lack of sanitation. Bengal, with Calcutta as its focus, constitutes the main endemic focus in India. Furthermore, the key position

of Calcutta as a world trade route focus, completed the picture and launched cholera on its career as a pandemic disease (12 , p. 178).

Cholera, when compared to other contagious diseases, is one of the most easily preventable of diseases which has been banished in all communities where a reasonable standard of living has been maintained. There being a definite relationship between the weather, especially the monsoon, overcrowding and unsanitary living conditions, the propagation of the disease once introduced is decisively influenced by various factors. The factors include the pH content in the tank water, onset of the monsoon, level of personal hygiene, aggregation of people at fairs and pilgrimages and most important of all, the movement of people. But even though research carried on by the World Health Organization has been able to control the disease in other parts of the world, it has had problems in controlling the disease in the endemic areas and preventing it from spreading to other localities.

Purpose

It has been recognized that infectious diseases, are to a large extent spread by migrants moving from areas of outbreaks to uninfected localities. Thes outbreaks are further enhanced during the months of pilgrimages in cultures where they are an important aspect. Thus, in keeping with the facts that have been presented earlier, the specific purpose of this study is to establish the nature of the relationship between distance and time and the effects of a general migration pattern in the diffusion of cholera. The latter purpose is achieved by isolating certain features of reality and constructing a model that represents an idealistic situation.

In correspondence to the purpose of this study the following hypothesis is proposed:

The simulated patterns of the diffusion of cholera will be found to correspond with the observed patterns as the disease is closely related to the general mobility of people within the nation.

Study Area and Limitations

The location of India between 8° and 37°N latitudes and 61° and 97°E longitudes, along with its isolation from the greater land mass of Asia, its high concentration of population, and the physical aspects of the land affords the ideal setting for disease, especially cholera, that thrives in such environments.

The homeland for the disease is India with the endemic seat of infection in the Bengal basin. The possibilities of simulating diffusion patterns of cholera are numerous. But the main focus, in the study, is the year 1971.

The one major limitation to this study is that the lower Ganges-Brahmaputra delta cannot be studied as one whole unit. The major reason for this is that the western part of that area is today Bangladesh. Ecologically, Bengal and Bangladesh are two different units because Bengal has a large urban population concentrated in the city of Calcutta in contrast to Bangladesh with its rural populations. But where the cholera vibrio is concerned, the two sides should be considered as one unit in a study attempting to explain the diffusion of the disease. Political differences between the two countries pose an inherent difficulty in treating the region as one unit. Therefore, only one part of the whole region namely that of Bengal can be included in this study.

Major pilgrimage spots and their role in the spread of diseases cannot be treated separately other then when included in the general migration and the resulting diffusion patterns. Also the effect of the pH content in the water and its effect on cholera are not part of the study again due to data problems.

Methodology, Data, and Time Period

The emphasis of the inquiry, namely an analysis of the migration flows as related to the diffusion patterns of the disease will be followed by prediction of future movements. The major problems that remain are those of research methodology, source of data and the period of time.

Data for this study were obtained from the Indian Statistical Bureau for the year 1971. Years prior to 1971 could not be used because of difficulties in obtaining as well as inconsistency and inaccurracy of the available data. Moreover, in the absence of reliable migration data for the year of study, the gravity model was used to simulate a general migration pattern. The exact procedure involved is discussed in Chapter IV. In tabular form, the data were published by the Office of the Registrar General of India (Vital Statistics Division) and obtained from the Government of India by the author. The acquired data contained death rates from cholera on a monthly basis for the year 1971 with the units of observation being the 21 states in India. Union Territories in India--of which there are nine--were not part of the study as the data were not included.

The major focus of this study is the diffusion patterns of cholera as related to migration. Accordingly, the five time categories used

were: January and February being t_0; March and April being t_1; May and June being t_2; July and August being t_3; September and October being t_4; and November and December being t_5. Keeping in mind the onset of the monsoons and the months of pilgrimage, time periods t_2 and t_3 were chosen to coincide with the factors.

Methodologies and techniques related to the study of diffusion especially that of disease are numeroud and varied. However, the concern of this study was to employ a technique that is both original as well as simple to execute. Within these limits, the range of choices are drastically reduced.

The purpose being to simulate a diffusion pattern and then to test whether it corresponds to the actual patterns, three different techniques were employed. The first technique used was the Gravity Model to approximate migration flows between the state of India. In the absence of actual migration data the resulting matrix of movement then was used in the second technique--the Markov Chain Model--to simulate diffusion patterns by the prediction of death rates during the different time periods as related to migration. The predicted death rates were then compared to the actual death rates by computing residuals to determine the amount of error in prediction.

Based on the principle of interaction, the gravity model was used to generate migration flows between the states of India. The obtained matrix of movement was the basis on which future predictions are made with the use of the Markov Chain Model. The overriding feature of the Markov Chain Model is in its simplicity in describing dynamic movement processes.

The resulting predicted death rates were first correlated with the

actual deaths to obtain a measure of how well the predictions are close to the actual deaths. Secondly, residuals were computed to determine the states where the predictions were not close to the observed patterns and to find the reasons as to why they were not closely correlated to the actual deaths. A more detailed account of the research methodology appears in Chapter V.

Overview

The expectations and the objectives of this research would be useful to health planners and researchers involved in the eradication program for cholera. It would be useful for the government of India in planning better health facilities. In a country of pilgrimages, where the disease is spread through human contact, the findings would not only help the governments but also private organizations in planning health facilities to control the movement of the disease.

The remaining chapters are organized as follows. Chapter II contains a review of the literature pertinent to the study and the application of stochastic models. Chapter III relates to the historical background of cholera and its trends in the past. Chapter IV is concerned with the methodology employed in the study and the results of the analysis are presented in Chapter V. The study is concluded in Chapter VI which is a general summary of all the chapters and implicationa of the study.

CHAPTER II

REVIEW OF SELECTED LITERATURE

Introduction

In the continuing search for a better understanding of the major causes of human illness, the number of explanations that have been presented represent a vast literature from a variety of fields--sociology, anthropology, economics, environmental sciences and geography. The fundamental approaches to the study of diseases have been 1) the environmental approach which traces the subtle aspects of the natural environment and its effects on human health and 2) the traditional public health and organized medical approaches which are concerned with the characteristics of both natural and man-made factors that cause diseases.

Early Disease Literature

The history of western medical thought began at the period of religious superstition when major diseases were linked to displeasure of the gods at human beings. The beginning of our understanding of health problems in relation to geographic variations can be traced back to the Greeks. References also can be found in the works of Arab writers during the Dark Ages in Europe. The information found in the literature has enhanced our knowledge of disease specific environments. The lapse of time from the Dark Ages until the early years of the fifteenth

century saw no contribution of any significance. Middle fifteenth century thought admitted the existence of entities of "germs" that caused "disease," but the association of these entities with other stimuli was not propounded.

Geographical influences on diseases are to be found in various accounts written in the eighteenth and nineteenth centuries. The literature of this period reveals that most of the work dealt with outbreak of various diseases within cities or some exotic ailment in distant lands. Amidst such works, the major effort was made by August Hirsch in his three-volume work Handbook of Historical and Geographical Pathology, which later came to be known as the classical work of that era. However, with the breakthrough of microbiology and the identification of specific disease carrying agents, studies relating to geographical environment came to a halt (50, p. 7).

With the development of microbiology and as the knowledge of the agents casuing diseases grew so did the different studies. These studies varied from analysis of any one single disease to a superabundance of works that were classified as medical geography with most parts of the world being included within the geographic area that was covered. Major works during this period include those of McKinley (38) and Simmons (59), where the emphasis was on the environment in relation to the disease.

With the development of the environmental approach, the scope of understanding diseases in relation to the environment increased too. With the increased importance of industries, disease research began to reveal connections with industry related problems. Emphasis began to shift to social and cultural differences and these were taken as

determinants of spatial variability of disease and health in highly urbanized and industrial nations (50, p. 4). One of the prime examples of a work with the emphasis on poverty as a determinant of health conditions was the work by Rowntree (56) in York, England. His emphasis was on the "unsanitary" living conditions of low class industrial workers which was a potential breeding ground for disease.

From the early years of the twentieth century until the period of the World Wars and the years following them, the emphasis on understanding disease in relation to the environment continued to expand and understandably so. The effects of the two wars and the exposure of various opposing forces and troops to various disease producing agents saw the slow expansion for a greater understanding of the environmental approach. The Germans were the first to take advantage of the opportunity and this resulted in The World Atlas of Epidemic Diseases by Rodenwalt and Jusatz (54). This work was brought to completion before World War II with the help of the U.S. Navy. Running parallel to the works of the Germans were were those research works being conducted by the Japanese and the Belgians with differing degrees of uniqueness but Jusatz in Germany remained the pioneer in medical geography. Along with this, Momiyama (44) showed how climate and modernization influenced mortality and Verhasselt (66) of Brussels conducted studies of cancer.

More recent have been some of these by the Soviet scientists who have been concerned with biogeographic aspects of medical geography. In France, the foundation of medical geography was laid down by Max Sorre (62) who is best known today for his concepts in medical ecology.

Disease Ecology

Research involved in the environmental aaproach has focused on the interaction of man with his total environment. They have tended to deal with the various aspects of disease in relation to natural conditions pertaining to climate, terrain and latitude. Most of these avenues had been explored by the researchers that have been involved in understanding the relation of disease to the environment. Among the contributions to this concept was that of Jacques M. May who came to the United States for the production of an atlas of disease. In addition to the production of the atlas, May further explained disease ecology and introduced the concepts of two, three and four factor host environment relationships, i.e. the interaction at various steps among causative agents, intermediate hosts, reservoirs and man (28, pp. 46-80).

Elaborating on the ecological concept, May (42) has explained the importance of environmental stimuli and the response of human tissues to external stimuli. Stimuli, he further suggested, can be of three kinds: organic, inorganic and socio-cultural. The inorganic stimuli include natural environmental aspects of humidity, wind, luminosity, heat and minerals found both in the soil and water. The direct influences of climate, May explained, are best felt in the tropical areas where debilitation is due to exposure to heat and humidity.

In considering organic stimuli, May has maintained that they can be better understood within the context of inorganic stimuli because of their close association. He has suggested that propagation of diseases are the result of certain kinds of ecological niches found within certain climatic environments. This suggestion explains how certain diseases

cannot exist outside their living cells and hence the resulting biological cycle can be correlated with seasonal changes. Although May did stress the importance of cultural and social patterns as stimuli, he did not elaborate on it as his efforts have been in the area of natural environmental explanations. Using this approach, May has tried to explain the geographical ecology of a wide variety of diseases which include cholera, plague, dysentery and other human ailments (42).

Since the time of May's contribution, various researchers have expanded on the disease ecological approach. Foremost among them has been the work of Gregory Knight who explained that this approach worked well in the analysis of vector borne diseases which can be explained as the interaction of the agents and the hosts. This could be studied in detail because of the earlier work of Le Riche and Milner (35) who discovered that certain diseases like malaria were transmitted by insects. Among the more recent works is that of Vornov (67) who stated that:

> Human ecology thus differs from that of other living beings--both plants and animals--in terms of man's unusually powerful impact upon the environment, which in turn has had an impact on man himself; also in terms of the means that man uses to protect himself against environmental hazards because of the ambiguity of man (pp. 230-37).

Derived from May's original contributions, this statement by Voronov can be best understood in the context that it is essential to isolate the dominant causes of diseases in relation to the "landscape" and by viewing the environment as both a stimulator as well as an indirect cause of disease (50, p. 82).

In spite of all these contributions there seems to be a barrier between medical geographers and bio-medical researchers. This becomes evident in the recent five volume work on international health problems (68) where the emphasis is laid on the problem of disease as a hazard

to the general population. Hence, in this work, the role of climate in the spread of any disease has been explained in very general terms. Further, levels of economic development, GNP and infant mortality are used as world health indicators (50, p. 83). This gap has to be bridged in order that bio-medical researchers interested in disease distributions can work together with medical geographers and understand the relation of disease to environmental complexes.

Disease ecology studies by medical geographers are numerous; the emphasis on environmental factors range from the importance of terrain and altitude to geochemical explanations, the influence of climate and socioeconomic conditions. Some of the best known studies in disease ecology have been: Burkitt's Lymphoma (9) and its relation to terrain and altitude in Africa; Scheil and Wepfer's (58) study of goiter in the United States; Fonaroff's (13) study of malaria in relation to climate. Literature consisting of examples of diseases in relation to the geochemical environment are substantial with Robert Mitchell's (43) work in 1972 being one of the first to trace this relationship. This was followed by various studies conducted in Canada, London and Glasgow.

Within the realm of studies concerned with the ecology of disease, one that had been considered to be fairly unique is the geographical analysis of patterns of Rocky Mountain spotted fever distributions in South Carolina by Pyle (52). The geography of this disease is unique in many ways because it has been known to be endemic in one area during a certain period of time and for some reason becomes endemic in another area. Although the ecological features are similar, the ticks which are the vectors in this case are different. One aspect that this study has not proven is whether the changes in the distribution of the disease is a case of relocation diffusion or not.

Diffusion and Disease Literature

The spread or diffusion of any disease from a central source which is often the endemic seat is the result of various mechanisms. Studies of disease diffusion are especially appropriate in the case of infectious diseases. Within the realm of geographical literature studies of disease diffusion have taken a variety of approaches.

The two major approaches in the study of disease diffusion have been: 1) the pathogenic viewpoint which is concerned with the influence of the physical environment wherein studies are closely related with the disease ecological approach; 2) the "causal" approach where "causal inferences" can be drawn from an examination of environmentally related influences or human influences. This has resulted in the fragmentation of literature that is related to disease diffusion as researchers can adopt a wide variety of approaches summarized in Table I.

Following the early contributions of Hirsch (20) in attempting to analyze the spread of diseases from a pathogenetic point of view in his Handbook of Geographical and Historical Pathology, Helmut Jusatz (54) and his colleagues analyzed the diffusion of cerebrospinal meningitis. In this study the causes of the spread of the disease over time and space have been found to exist in the physical environment. Jusatz (54) followed up on the already developed comprehensive review of the occurrence of the disease by Hirsch and analyzed the pathology of the meningococcus causing the problem. The emphasis on this study was placed on the effects of the environmental conditions and it was found that the climate affected the movement of the disease from east to west along transportation lines.

Among recent contributions to studies of disease diffusion is that

TABLE I

RELATED APPROACHES TO DISEASE DIFFUSION

Approach	Historical and General Notes	Geographic Epidemiologies	Public Health Planning
Type of Disease	Vectored and non-vectored infections.	Vectored and non-vectored infections.	Infectious (same); Mostly cause of death reporting.
Geographical scale	International; Multi-regional or mixed; National; aspatial.	Scales; general mapping.	General: larger discrete units e.g. states and counties; aspatial.
Methods of Analysis	Descriptive accounts: "Flows."	Small samples; Selective comparisons; Pathologies/etiology.	Socio-economic; "Seeding;" Mapping tubular comparisons.
Models	Narrative: replication.	Mostly non-parametric.	Descriptions of change regression.
Outcomes/Results	ex post facto provision of general information.	Identification of diffusion mechanisms for small samples.	Indication of change over time; Costs; Health care provision; Policy indicator.

Source: Gerald F. Pyle, Applied Medical Geography. New York: John Wiley and Sons, 1979.

of Hunter (24). He studied the diffusion of ouchocerciasis spread by parasitised black flies and the resulting modification of human behavior in northern Ghana (Sakoti). He found that in the spread of the disease there were cycles of advance and retreat and as a result human settlement near the Red Volta river advanced and retreated with the disease. The conclusions were that during the uprising period, human settlements moved away from the river because of the resulting river blindness and moved closer to the rivers during the retreat period. What was surveyed in this unique study was an example of how human behavior is modified by the cycles of diffusion in a vectoral disease especially in areas where primitive living conditions prevail.

In the study of disease diffusion, there have been historical reconstructions of the diseases. In such case, there is a heavy reliance on archival research (50, p. 127). In this context, one of the most widely known studies in the diffusion of disease is that of the Black Death (Bubonic Plague) in Europe during the middle of the fourteenth century. The pathogen in this case produces the infection in rats which is then transferred to a flea and thence on to humans. Howe (21) in his study, believed that the disease spread into Europe from Asia, existed as a disease of the poor and diffusion was a result of unsanitary living conditions. Improvement in living conditions helped the control of the diffusion.

Cholera and Diffusion

In understanding the differences of any infectious disease it often helps to reconstruct historically the particular epidemic under study because it helps in eradicating the disease. One such disease that

serves as the prime example of the usefulness of historical analysis is that of cholera, as man is the prime reservoir with the vibrios being spread from one person to another. Rodenwalt (54) reconstructed the diffusion of cholera on a world-wide scale from its endemic source in the Ganges-Brahmaputra delta westward across India and Africa in 1863 and 1864 and thence into Europe during 1865 to 1867. Initially the diffusion was along water-borne transportation routes.

Pyle (51) later reconstructed the diffusion of cholera in the United States in 1866. In this case, the disease spread both along water as well as land transportation routes. A more recent study by Kwofie (30) was a trend surface analysis of both the temporal and the spatial diffusion of cholera in West Africa in the 1970s. Using a ten-year period, he examined the three phases in the movement of the disease: the primary saturation stage; the saturation and lastly the waning stage. During the primary saturation stage, the disease spreads from west to east whereas in the saturation stage this circulation pattern changed resulting in a recirculation to the west coast as well as to the interior. In the waning phase, the general trend continued from both east to west and vice versa centralizing in the Sahelian zone.

Continuing this work in greater depth, Stock (64) identified four types of diffusion--coastal, riverine, urban heirarchical and localized radial contact in the movement of cholera in Africa. He identified diffusion routes for these types of diffusion and in doing so linked the general principals of diffusion with disease etiology.

The general principles of diffusion proposed by Hagerstrand have been used in practically all the studies of disease diffusion. The fine meshing of Hagerstrand's general principles with the epidemiology

of infectious diseases are demonstrated in most of the studies of disease diffusion. Examples of this approach are Hunter's (24) study of the diffusion of ouchocerciasis in northern Ghana; in the diffusion of influenza in England and Wales in 1971 by Hunter and Young (25) and Pyle's (50) study of the diffusion of the same diseases in the U.S. during 1975 to 1976 and 1977.

In general, most of the disease studies have traced the diffusion of various diseases historically relating the diffusion to both the physical environment as well as to human behavior. A unique study by Prothero (48) traced the effects of migration in the spread of the disease. This detailed study can be cited as an example of the relation of human activity to the spread of the disease. This continuous maintenance of the reservoirs by the migrants make it very difficult for any control measures to be set up as Prothero has concluded. An interesting feature in relating migration to the movement of disease is the example of migration for religious purposes such as the annual pilgrimage to Mecca as well as in countries where pilgraimages are part of the cultural makeup. Not only are migrants responsible for the spreading of disease but also for bringing disease into areas where they have not been known to exist. Lenczner and Owen (34) discussed this relationship in their study of tropical and parasitic diseases in the Toronto area.

Most of the disease literature has been found to deal with epidemiology and the pathology of diseases. Few publications since the classic Handbook of Geographical amd Historical Pathology (20) have made any attempts to put forth a global prospective on the varying patterns of disease. One of the more recent ones is the publication

titled A World Geography of Human Diseases, edited by Howe (21) which tries to bridge the gap left after Hirsch. The effort of the editor to present it as a twentieth century version of the Handbook generally succeeds. The major problem with the book, however, is that although it is interdisciplinary and therefore incoprporates the works of biochemists, statisticians, geographers, medical researchers, and from various other fields. Hence, it lacks uniformity in style, view and attitude even with careful editing. Useful as a reference book it has shortcomings for educational purposes.

In contrast to Howe's book is Applied Medical Geography by Pyle (50) which deals with the broad spectrum of disease epidemiology, diffusion and health problems. This work deals with the development of medical geography, techniques in the geographic analysis of diseases, and the applications of geography both in the study of diseases as well as for health care problems. Both Pyle's (50) book and Learmonth's (32) latest book on Patterns of Disease and Hunger are spirited expositions of medical geography. The latter examines man-environment interactions in the study of health problems. Both books can be used for introducing ideas as well as general textbooks.

Stochastic Models and Social Science Research

The disease literature has revealed the use of various geographic concepts and techniques in the study of disease diffusion and epidemiology. The use of techniques have been limited to mapping, regression and correlation, factor analysis, trend surface analysis, the correlation of various rates, etc. The use of stochastic models though widely

prevalent in other fields of social science research has been used limitedly in the realm of medical geography, the severe limitations being in the inherent data requirements. Recognition of the use of stochastic models especially that of Markov Chain technique in geographic, economic and social research has come about only during the last couple of decades. Developed by the Russian mathematician, A. A. Markov early in the twentieth century, this theory stands out as a prime example of a central development in abstract mathematics which is applicable to practical problems.

The principle use of Markov Chain technique has been in economics and that to mostly in the measurement of industrial concentration. The Simon and Bonini paper (60) is a theoretical discussion on the use of stochastic models in studying industry concentration through time as opposed to the more traditional models.

A significant problem to which Markov Chains have been applied is that of occupational and social mobility of the population. Prais (48) studied intergenerational mobility of the British population among social classes as indicated by social groupings. The ultimate occupation structure was predicted in the construction of a transition matrix indicating the probability of a given man's son going into various occupational groups based on the father's occupation. In economics, researchers have studied various industry concentrations, size and distribution of firms through time, income distribution, and regional growth of gross income using the Markov Chain technique as the technique. Other than in economics, the theory has been applied to study occupational structure through marriage licenses and to survey potential voters. In addition to the application mentioned here, there have been

countless applications to problems in the "hard sciences." Interest here is primarily related to social science problems specifically those related to disease and migration.

Traver and Gurley (64) were the first to study internal migration using a Markov Chain model as the basis. Their study focused on the movement of whites and non-whites between the nine census divisions in the U.S. using 1960 census data to estimate long range population distribution. Following this Rogers (55) also noted the simplicity of using this model in studying migration as related to a large number of socio-economic, political and psychological factors. More recently Salkin et al, provided additional support for the use of this model in migration studies adding to this study a simple method to account for natural increase in population growth to the derived estimates from the model.

The strength and weakness of this model in movement research is critically examined by Brown (7). He points out in his paper that the strengths of the model lie in 1) that it focuses on the dynamic aspect of movement in comparison to other transition models, 2) the assumptions are built into the model in contrast to the Monte Carlo simulation model where the final results depend upon the assumption, 3) several descriptive measures are provided and 4) the parameters are easily estimated. On the other hand, Brown cautions that the model suffers from the inability to account for natural increase and the assumption of stationary transition probabilities. However, he suggests that when this model is used as a descriptive tool, the weaknesses should not be considered as a major flaw.

It should be mentioned here that a lot of criticism has been leveled against the Markov Chain technique, since the theory was used in

migration studies. The one major criticism relates to an assumption that the population under study is relatively homogeneous with the inclination to migrate i.e. all members of a population are potential migrants to some degree. Blumen et al. (5) and Goodman (15) were the first to challenge this assumption of population homogeneity and suggested that the population may be divided into movers and stayers which led later to further stratification by age, sex, occupational structure and duration of residence. This stratification led to the development of the Semi-Markov Chain where the mover-stayer approach is extended. Later researchers developed this model further to allow for more stratification. The various research works using different factors of stratification mentioned above have revealed that this Semi-Markovian model should be used with caution as the forecasting capability of the model is entirely dependent on the availability of suitable forecasts of any phenomena (45, p. 133).

The review of literature pertaining to the use of stochastic models has revealed that these models have been widely used in social science research but they are not widely prevalent in studies of disease diffusion. Thus, the model has proved to be useful for research in various fields.

This review of literature has not dealt with the growth and development of the Gravity Model, which has been used to set up the Markov Chain Model. Its use has been widely recognized especially in the works of Isard (26). A fairly thorough discussion is made both in Smith (61) and in Abler et al, (1) on the use of the Gravity Model and its application to imaginary data. Abler et al. (1) discuss to some length on fitting migration data to this model as well as simulating migration flows

between two centers using population at each center and intervening distance. Various other contributions to the use of the Gravity Model are too exhaustive to mention.

Summary of Literature Review

This review of literature is by no means an exhaustive work; it only seeks to sketch the development of medical geography, the different approaches used in studying the various aspects of disease and to review briefly the use of stochastic models in social science research.

Medical thought began with religious superstition. Relating health problems to geographic variations began with the Greeks and development in the relationship of disease to environment was slow. The lapse of time from the Dark Ages until the early fifteenth century saw no significant contribution. From the fifteenth to the eighteenth centuries, most of the works dealt in the outbreak of various diseases in distant lands. Amidst such work, Hirsch (20) wrote his classical Handbook of Geographical and Historical Pathology. However, with the advancement in microbiology, the emphasis shifted to studying the disease causing agents. Work relating to the environment was carried on by McKinley (38), Simmons (59), Rowntree (56), Rodenwalt and Jusatz (54).

The disease ecological approach was introduced by May whose significant contribution was to explain the importance of external environmental stimuli in relation to disease. Although his work was carried on by Burton and Smith (10), McGlashan (36) and Knight (28) there has been a definite communications barrier between biomedical researchers and medical geographers. Recent works by Howe (21), Learmonth (32) and Pyle (50) have tried to bridge the gap.

Within the realm of disease diffusion studies, the spread of various diseases have been traced historically, their causality being traced back to the environment as well as human behavior. Pioneering the work has been Hirsch (20) followed by Jusatz (54), Hunter (23) and Pyle (51). Studies of the diffusion of cholera have traced various diffusion patterns with most of these studies being conducted in Africa by Kwofi (30) and Stock (64).

Prothero (47) stands out by relating the diffusion of malaria to migrant behavior. The work is one of the earliest recorded during the recent past that attempts to study the effects of migration in the spread of the disease. Most of these works have dealt with the epidemiology and the pathology of diseases using simple techniques such as correlation and regression, mapping, trend surface analysis, etc.

Stochastic models have not been used in medical geography. Most of the works where these models have been applied have been in economics, politics and migration studies. The growth and development of the Gravity Model have not been traced in this review as works in the field are too exhaustive to mention. In contrast, the application of the Markov Chain Analysis is still at an infantile stage, although its importance has been recognized as a more plausible approach to other transition models.

The choice, therefore, in the final analysis is not a clearcut one of the best technique but one to be made between the practical and the ideal. In contrast to other transition models offering increased precision but harder to operationalize, the Markov Chain Model affords easy operation at the price of precision.

CHAPTER III

THE HISTORY OF CHOLERA IN INDIA

Introduction

The origin of cholera has drawn more speculation and argument than most other diseases. It has a fascinating history provided by evidence from ancient writers. Later writings by physicians have contributed to furthering our knowledge about the disease.

The Early Years; Ancient Times Until 1800

The evidence found in early Indian literature leads to the deduction that Asiatic cholera--a specific infection caused by Vibrio Cholerae--existed in India in epidemic form has been evaluated by various writers. Those who favor the early existence of the disease point out the description of a particular syndrome with clinical features identical to those of true cholera in ancient Indian literature particularly in the writings of Susruta (46, p. 11). However, McNamara (39) has argued against this point in his History of Asiatic Cholera and suggests that a careful study of these early writings reveal that cholera in its epidemic or Asiatic form never existed then. On the other hand, Sticker (63) although expressing some misgivings regarding the authenticity of the text did feel that the description in these early texts does refer to a description of true cholera.

Even if the above mentioned reference does prove to be unreliable, a second category of evidence reveals the existence of religious rites that were invoked to ward off the ravages of the disease. The people of Lower Bengal had for a long time worshipped the goddess of cholera and evidence supporting this is presented by Pollitzer (46) from McNamara's History of Asiatic Cholera suggesting that the goddess appears:

> according to tradition, that, at an early period, the date of which cannot be ascertained, a female while wandering about in the woods met with a large stone, the symbol of the goddess cholera. The worship of the deity through this stone was, according to the prevailing ideas of the Hindoos [sic], the only means of preservation from the influence of this terrible disease. The fame of the goddess spread and people flocked from all parts of the country to come and pray at her shrine in Calcutta (p. 12).

The malady must have raged at times with violence or it would not have been necessary to propitiate the deity on account of it. Statements such as this do strongly suggest that cholera existed in India since time immemorial. Irrefutable proof of its presence during historical times is furnished by European observers who after the arrival of Vasco da Gama in 1498 on the Malabar coast had been given an opportunity to become acquainted with the disease. Gasper Correa was the first of the Portugese to give detailed descriptions of the ravages of the disease in his Lendas da India (Legends of India). This work was later carried on by the other Dutch, French and British observers (46, p. 13).

The early records of cholera to be made by Europeans refer exclusively to the west coast of India because the British did not gain a foothold on the Cormandel coast and Bengal until a century after the Portugese had landed in Goa. Earliest accounts of the occurrence of cholera on the Coromandel coast is to be found in the writings of an

English physician Dr. Paisely (1774) and recorded in Curtis' work Diseases of India. Even then the work of Paisely was brought to light only 33 years later because the British classified the disease as being a spasmodic affectation instead of recognizing it as an affectation sui generis. This could be regarded as one of the major reasons why no descriptions of the diseases are to be found in the writings of early British physicians even during the latter half of the eighteenth century and at the beginning of the nineteenth century. Further, the Hospital Boards of Calcutta and Madras were established only in the year 1786, hence no reports on the incidence of cholera among the European and native soldiers were ever recorded. After the establishment of the Hospital Boards, the records available for 1786 and 1790 reveal violent outbreaks of the disease in many different parts of the country but the available records are scanty.

Although the evidence is fragmentary and incomplete, it does show that cholera was present in India during ancient times and continued to exist and manifest itself periodically in widespread conflagrations. Sticker (63) points out that even during these early times it is easy to perceive the role of the pilgrimagees, movement of people and military operations that were responsible in the propagation of the disease. The known history of these early times however do not furnish any clues to the epidemiological importance of Bengal as the cradle (46, p. 16). Observations made in the area since 1817 fill in the gap regarding the knowledge of cholera in Bengal.

The First Through Sixth Pandemics - 1817 to 1899

In the history of cholera, 1817 marks the beginning of a new epoch

when a series of pandemics, after having gained impetus in India, spread to different parts of the world. The spread paid no heed to distance, natural obstacles or any attempts to ward off the disease by cordons and other quarantine measures. Since then a disease that has been limited to India for such a long period became a serious concern to the rest of the world.

The years prior to 1817 were marked by abnormal meteorological conditions which probably led to the onset of the pandemic. According to Sticker, the year 1816 was unusually hot and sticky whereas 1815 was marked by abnormally heavy rainfall with still more in 1817 which was followed by heavy floods and harvest failures. Subsequently, cholera began to show an unusual violence.

The disease continued to be active in the following years of 1818 to 1820. By 1820, it became localized in the Bengal basin but "the epidemic which had arisen in 1817, well nigh covering India within the three succeeding years had not subsided" (46, p. 19). But, in the meantime, the disease had spread widely beyond the confines of the subcontinent by water-borne transportation routes to as far east as Indonesia and as far west as Astrakhan in Russia by 1823 to 1824.

The infection at Astrakhan in Russia during 1824 persisted long after the disease had been controlled elsewhere and a recrudescence of the infection is believed by some researchers to be the origin of the second pandemic in 1829. However, this cannot explain the presence of the disease at Orenburg (Chkalov) in 1829 a year earlier than when it manifested in Astrakhan. Chinese researchers have also recorded that the infection was again borne from India to China; reaching Peking, sweeping through Mongolia and eventually traveling to Moscow (46, p. 21).

Although this surmise might explain the presence of cholera at Orenburg, it does not explain or account for the second inroad of infection west of the Caspian Sea.

On the basis of historic accounts available, it can be surmised without doubt that the origin of the second pandemic can be traced back to the Bengal basin. It followed pretty much the same route as the first pandemic moving westwards along the Ganges and Jamuna rivers, through Punjab. By 1829, the disease was rampant in Afghanistan penetrating into Persia and carried into the southeast corner of European Russia. Here, a new epidemic broke out at the end of 1829 from where the disease started to spread northwestward.

The years 1829 to 1851 saw the progress of cholera through Europe, Western Russia and even North Africa with outbreaks occurring in different countries in these areas. By 1849 it had reached America by sea route directly from Europe and came to be known as "America's greatest scourge." From America the infection spread into Mexico and as far south as Panama.

The third pandemic in 1852 is harder to trace as its commencement was a combined result of recrudescence as well as a temporary entrenchment of local infection. This time the disease took only serious tolls in northern Europe while in the southern sections it ran rampant. In the endemic center in India it showed signs of increased activity. By the end of 1859, Europe seemed to become free temporarily of the infection, whereas in India it remained entrenched.

The fourth pandemic lasting ten years from 1863 to 1873 stood in marked contrast to the previous pandemics because the disease did not penetrate into Europe over ancient trade routes through Persia and the

Caspian sea ports. The diffusion of the disease was by new traffic routes that had opened up over Arabia into Egypt, Constantinople, southern France and Italy.

The disease, as noted earlier, had its origin in the endemic center of Bengal and reached Mecca by pilgrims who arrived there by ship from India. An important aspect of this is that an extraordinarily large number of pilgrims had assembled in 1865 and the disease found good outlets for continuing its movements, The outbreaks of the disease that followed was marked by extreme violence with about one third of the assembled 90,000 pilgrims falling victim to the disease at Mecca. It also encouraged the spread of the disease by returning pilgrims thereby penetrating practically all the countries in the world. The information available indicates that areas in India, southeast and east Asia suffered from cholera throughout the pandemic years; the most serious outbreaks being between 1877 to 1879 when 158,204 cases and 89,207 deaths were recorded (46, p. 38).

Notwithstanding the wide areas of the fifth pandemic held sway from 1881 to 1896, it caused less havoc than its predecessors. This is significant in the hostory of the disease as Koch discovered Vibrio Cholerae in 1883. In India, it remained serious with violent outbreaks and was again carried into Mecca and then on to Europe. But in Europe it remained confined to France, Italy and Spain with most of the countries taking careful measures. In Russia, Germany and France it did become widespread by 1892 but never had the same intensity as the earlier pandemics.

The exacerbation of the cholera situation in India in 1899 led to violent outbreaks in Calcutta and Bombay in the year 1900. Like

previous pandemics, it showed a tendency to spread eastwards through maritime routes. In Europe. the area within which the disease prevailed was far more limited than in previous pandemics. The visitation was restricted to only sporadic cases from 1924 but was fairly widespread in the years 1915 through 1920 because of the First World War. In the East Asian countries and the Indian sub-continents the situation was serious in 1908 and 1909. These years were preceded by a period lasting from 1905 to 1908 when the disease ran rampant in India (46, p. 45). This is shown by the cholera deaths for the period in Table II.

TABLE II

CHOLERA DEATHS IN INDIA
(1904 TO 1909)

Years	Cholera Deaths	Years	Cholera Deaths
1904	189,855	1907	400,024
1905	439,439	1908	579,814
1906	682,649	1909	227,842

Source: R. Pollitzer, Cholera. Geneva: World Health Organization, 1959.

Following these years, the cholera mortality exceeded half a million annually in 1918 (556, 533 deaths) and in 1919 (565,166 deaths) (46, p. 45).

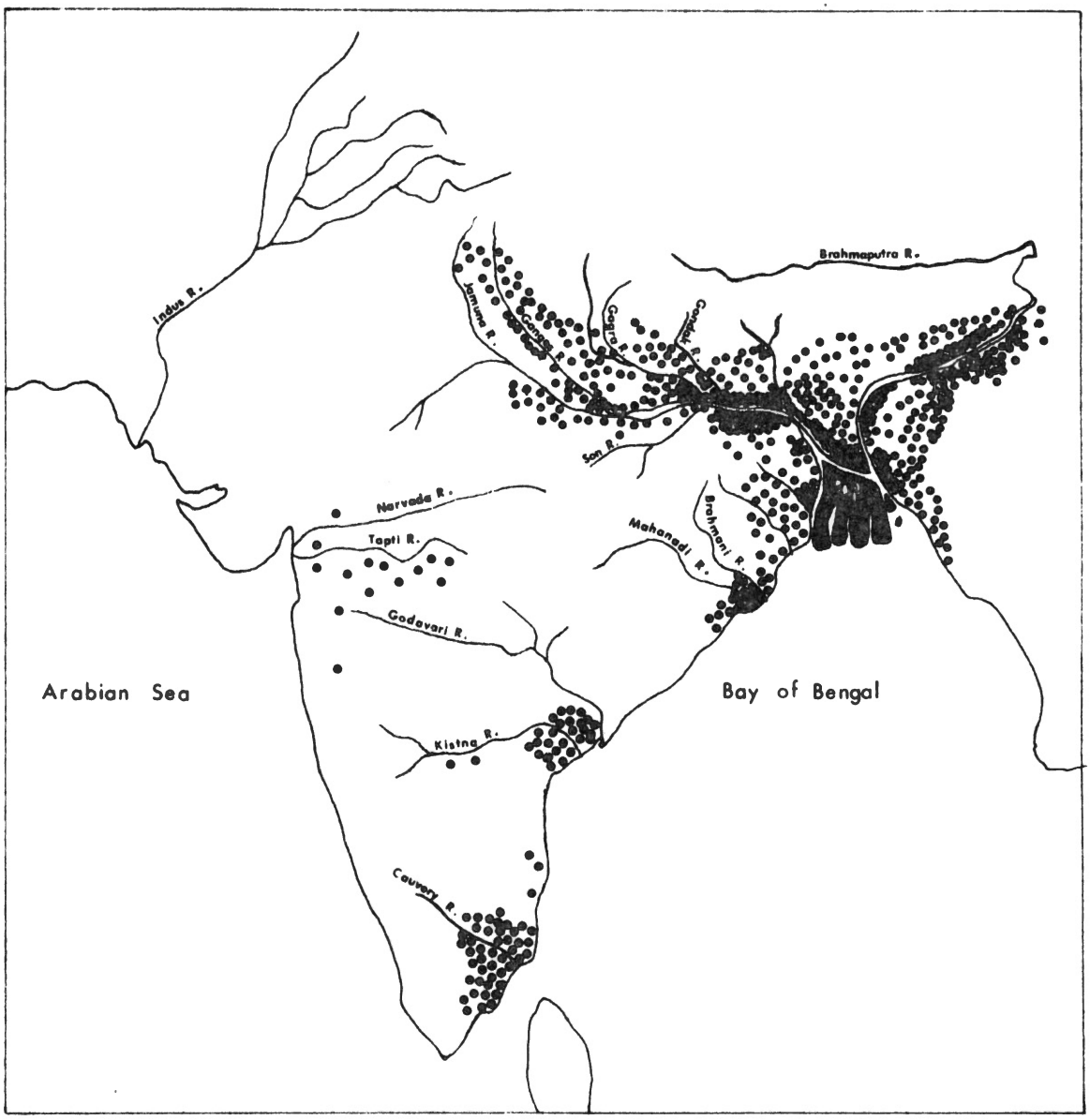
The Endemic Focus

The history of these pandemics reveal that each major epidemic spread seems to have originated in the endemic home of the disease in India. The disease has been entrenched in this area long before 1817, probably since time immemorial.

On this basis of recent studies that have been undertaken, it is possible to demarcate areas which harbor the endemic foci in contrast to areas where epidemics occur only occasionally. Areas which remain free from infection for considerably long periods of time or where the epidemic occur only occasionally can be identified in the central states of India which include Madhya Pradesh, Maharashtra and Punjab. On the other hand the coastal states of Bengal and Orissa along with the states of Bihar and Assam can be demarcated as areas where the disease continues to be present from year to year assuming epidemic proportions periodically. The demarcation of these areas are based on data collected over the years 1901 to 1945 (45, p. 52).

Common factors can be identified for the largest endemic foci in the deltaic region of the Ganges and Brahmaputra delta in both West Bengal and East Bengal (Bangladesh today) extending into Assam and westwards into Bihar and Uttar Pradesh. The minor epidemic of considerably less importance can also be identified in the Cauvery and Krishna deltas (Figure 1).

These factors are that they are situated in close relation to surface water systems in densely populated areas near the coast. The altitude hardly exceeds 50 feet above sea leval. It is believed today, that with some improvement in the level of sanitation, the extent of the foci in the Bengal basin has shrunk in size. Presumably, if an



Source: R. Pollitzer, Cholera. Geneva: World Health Organization, 1959, p. 53.

Figure 1. Cholera Endemicity Level In India, 1901-45

infection held sway covering an extensive area for such a long time period, this focus must have been extensive especially during the major pandemics. Further, the presence of the disease in the Bengal basin for such a long period has led recent researchers like Bryden (8) to believe that this foci in the Bengal basin has constituted the reservoir of infection and the starting point of all epidemics. It is futile, however, to trace the origin of any individual outbreaks to any single locality within the area. According to Pollitzer (46):

> one must presume that the epidemic outbursts in these regions lead to the production of a great volume of infection in different localities, sometimes simultaneously at places distant from one another and that under favourable climatic conditions the disease then spreads in a wave-like form to areas generally free from cholera (p. 53).

Influence of Seasonal Factors in the Spread of Cholera

Incidence rates of cholera collected over the years have revealed a marked seasonal variation in individual parts of India together with striking variation from area to area.

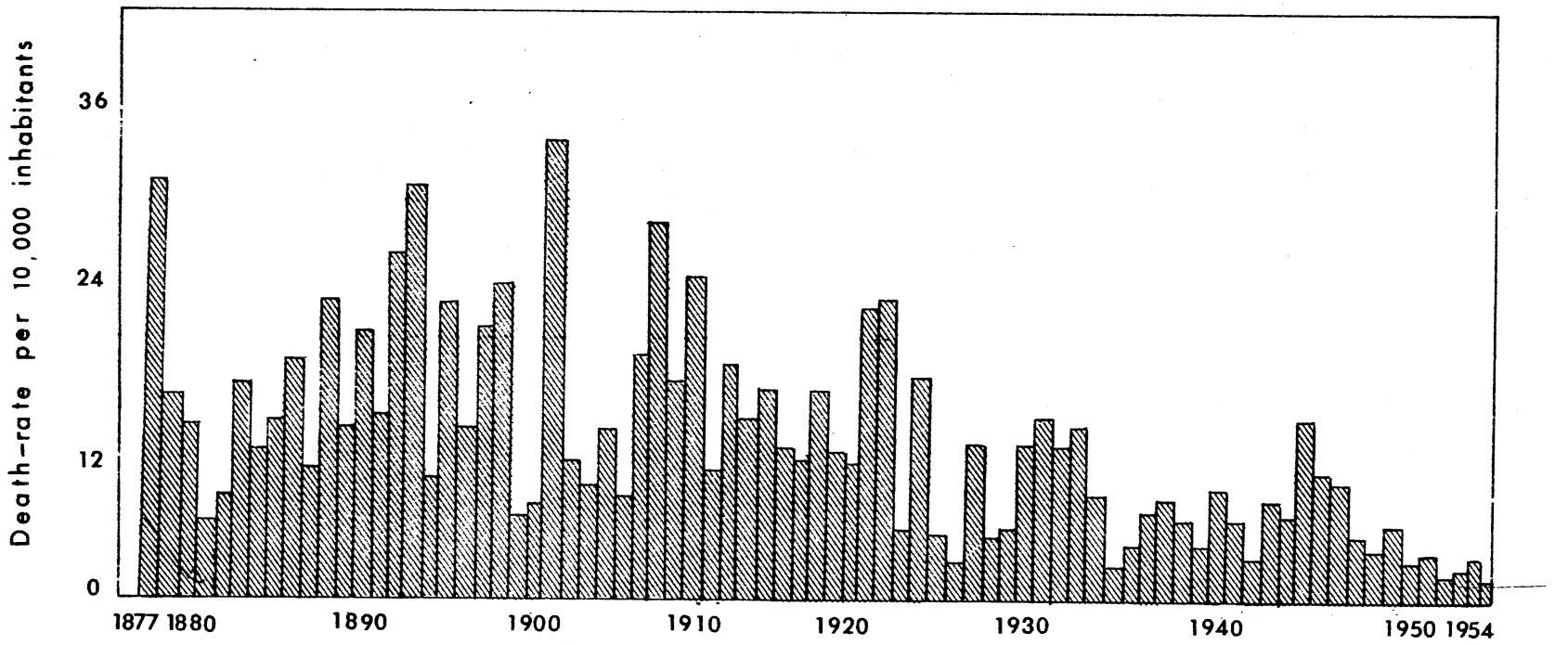
In Bengal, the disease shows considerable variation in the individual areas within the state. Another feature of the disease in the endemic foci is the tendency of the disease to peak after the monsoons during the months of September to November and late in March to May when the mean monthly deaths from cholera are taken into consideration for a year. The data from which the information was drawn is for the year 1925 to 1946, when both West and East Bengal were taken into consideration. Today, with this area being partitioned, it is harder to collect data for the endemic foci as a whole.

Another important feature is that as one moves westwards, away from Bengal, the disease shows a tendency to peak later especially in the northwest states of India after a period of subsidence during the winter to begin a new period of renewal. This was one of the essential features in the spread of the first pandemic in 1817. The history of the spread of the disease is usually characterized by the rise of a seasonal wave in the endemic home, followed by a recrudescence in the upper northern states moving along into Afghanistan (46, p. 56).

Incidence of Cholera in India

Cholera, in India, has shown a marked variance with a maximum of 805,698 deaths in 1900 to a minimum of 60,000 deaths in the more recent years up to 1953. In the sub-continent the last major epidemic in 1943 claimed 459,930 lives. This unusual event, in the then British India, was attributable to severe famine conditions in Bengal and other shortcomings caused by World War II.

Cholera death rates from 1877 to 1954 (Figure 2) show a marked decrease in India. During this span of years, there were three epidemics. The first lasted for a period of four years from 1875 to 1879 followed by the second from 1927 to 1931 which never attained a high peak. The last major epidemic reached its peak in 1943 and lasted from 1941 until 1943. Figure 2 does reveal a marked decline in the death rates for the country especially if figures from 1900 are charted. However, the highest peak attained during the 1943 epidemic which lasted five years throws some doubt as to whether the cholera epidemics have been brought under control through improved sanitation. The establishment of public health services in 1923 and its expansion in the following years have



Source: R. Pollitzer, Cholera. Geneva: World Health Organization, 1959, p. 79.

Figure 2. Annual Cholera Death Rate in India, 1877-1954

helped control the disease to some extent. Further, Figure 2 is based on the mortality rates for the whole country, hence it does reveal the heterogenity of the individual provinces, the length and severity of each epidemic and the periodicity of the incidence (46, p. 79).

However, since 1948, cholera has maintained relatively high levels in different states with a marked increase in 1952 in the states of Bengal and Bihar spreading to neighboring states in 1953. It is clear from the foregoing account that cholera continues to exist in the endemic center in Bengal and the adjoining provinces fluctuating when climatic conditions are favorable.

Summary

Along with the bubonic plague, malaria and syphillis, cholera has played an important part in the shaping of history. In spreading to almost every corner of the world, neither physical distance, natural barriers or man-made obstructions have revented the disease from running rampant. The increased improved methods for treatment and control brought about the great decrease in the incidence rates for bubonic plague. On this basis, there can be no doubt that the increasing use of these facilities would bring about a decrease in the incidence rates for cholera also.

Unfortunately, in the case of cholera, especially in India, the disease has shown no signs of a satisfactory decrease. Compared to the early years, there has been some decrease but control measures have to be used with more efficiency. Unlike the plague which can today be considered res gesta, the many unsolved problems of cholera continue to call for urgent attention.

It is to help provide some insight into the movement of the disease in relation to the migration of people that this thesis was undertaken. The problems has never been approached from a purely geographical point of view. Hopefully, the method of analysis used in this study will help to alleviate the problem. The procedure and the results of the analysis are presented in the following chapters.

CHAPTER IV

METHODOLOGY

Introduction

The real world and real situations are invariably complex. In order to simplify the complex world it becomes essential to isolate certain features of reality and devlop a more or less ideal situation by constructing a model. The idealized system is thus symbolically represented in the model which also contains certain features of reality.

Any real world situation is never replicated but merely represented in a model. A model can help us to visulize hidden aspects and patterns in the real world. It stands for the structural and functional attributes of any situation helping to make implicit relationships explicit (17, p. 33).

Models are known to be of three types on the basis of the purpose they serve--descriptive, prescriptive, and predictive. The latter of the three models is best suited to the study of this research problem as predictive models provide a basis for deductions that can later be compared with observations in the real world. Thus, in order to idealize the problem of cholera in relation to the movement of people and then compare the predicted results with the observed. The following steps were taken:

1) the conceptualization or idealization of the problem which was

stated in Chapter I.

2) Symbolization or the construction of a model that represents the idealized problem. In the case of this research, the problem was abstracted by calculating a migration flow matrix by employing the Gravity Model.

3) The performance of logical operatuons on the model to find solutions to the idealized problem. In this study, the resulting matrix, which represented the movement of people, was then transferred to the Markov Chain Model to predict diffusion patterns for cholera.

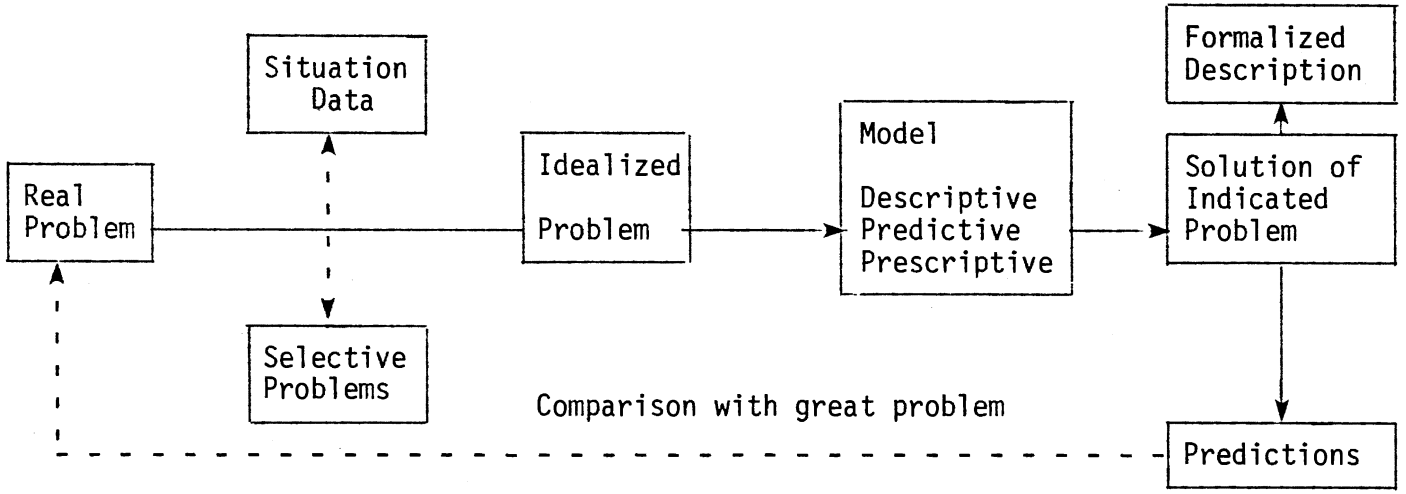
4) Evaluation or testing the ability of the model to represent the real situation by comparing the model based predictions with the real-world observations.

5) From this comparison it is then possible to derive new generalizations or hypotheses whcih can then become the starting point for further investigations (17, p. 39). These steps can be diagrammatically represented as shown in Figure 3.

The major purpose of this chapter is to comment further on Step 3 which entails performing certain logical steps so as to find a solution to the idealized problem.

The Gravity Model and Migration Simulation

The first step, then, was to simulate migration flows between the 21 states of India in order to relate migratory movements of the people in the country to the diffusion of cholera. The Gravity Model was used to generate the migration matrix. The roots of the Gravity Model lie in the "least effort" principles of Zipf and the social physics concept of Stewart. These concepts were derived from the earlier works of



Source: F. Grundy and A. Reinke, "Health Practice Research and Formalized Managerical Methods," Geneva: World Health Organization Public Health Papers No. 51, 1973.

Figure 3. Diagrammatic Representation of the Scientific Approach, Utilizing a Model

Ravenstein who had observed from migration data that a population center attracted migrants from other centers in direct proportion to the population of the sending center and inversely proportional to the distance between the sending center and the receiving center. The Gravity Model which was derived from the concept of Ravenstein states that the interaction between two masses is proportional to the product of the size of the two masses and inversely proportional to the space over which the interaction takes place. That is:

$$I_{ij} = \frac{P_i P_j}{d_{ij}^{\alpha}}$$

where I_{ij} = the interaction or flow between a point i and another point j.

P_i = population of place i.

P_j = population of place j.

d_{ij} = distance between places i and j.

α = exponent of the distance between i and j (in the application of α to this study, the function of the distance between i and j was squared to obtain a more standard form).

Using this formula, interactions between all 21 states of India were computed using the population data for each state in 1971. Linear distances between the center of each state were obtained in inches and approximated in miles.

The application of the model to this study is demonstrated in the following hypothetical example using imaginary data (Tables III, IV and V). The example contains a set of four states, A, B, C, and D. During a given time period, people move between the states. Using the population of each state as the measure of mass at the origin and destination,

TABLE III

HYPOTHETICAL EXAMPLE OF COMPUTING INTERACTION BETWEEN STATES

From State (Population)	To State (Population) B(60)	C(60)		Total Interaction
A(80)		42+ $\frac{82\times60}{42^2}$* $=\frac{4800}{1764}=2.72$**	45 $\frac{80\times40}{45^2}=\frac{3200}{2025}=1.58$	70 $\frac{80\times50}{70^2}=\frac{4000}{4900}=0.81$	5.11‡
B(60)	42 $\frac{60\times80}{42^2}=\frac{4800}{1764}=2.72$		18 $\frac{60\times40}{18^2}=\frac{2400}{324}=7.4$	30 $\frac{60\times50}{30^2}=\frac{3000}{900}=3.33$	13.45
C(40)	45 $\frac{40\times80}{45^2}=\frac{3200}{2025}=1.58$	18 $\frac{40\times60}{18^2}=\frac{2400}{324}=7.4$		35 $\frac{40\times50}{35^2}=\frac{2000}{1225}=1.6$	10.58
D(50)	70 $\frac{50\times80}{70^2}=\frac{4000}{4900}=0.81$	30 $\frac{50\times60}{30^2}=\frac{3000}{900}=3.33$	35 $\frac{50\times40}{35^2}=\frac{2000}{1225}=1.6$		5.7

\+ Distance between states.

‡ Total interaction of state i with all other states in the matrix.

* $(P_i P_j/d_{ij}^2)$ = the interaction factor

** The resulting interaction between state i and state j.

TABLE IV

ESTIMATED PERCENTAGE OF INTERACTION BETWEEN STATES

From State (Population)	To State (Population) A	B	C	D	Total Interaction
A		$\frac{2.72^*}{5.11} \times 100 = 53^{**}$	$\frac{1.58}{100} \times 100 = 30.9$	$\frac{0.81}{5.11} \times 100 = 15.8$	5.11
B	$\frac{2.72}{13.45} \times 100 = 20.2$		$\frac{7.4}{13.45} \times 100 = 55$	$\frac{3.33}{13.45} \times 100 = 24.7$	13.45
C	$\frac{1.58}{10.58} \times 100 = 14.9$	$\frac{7.4}{10.58} \times 100 = 69.9$		$\frac{1.6}{10.58} \times 100 = 15.2$	10.58
D	$\frac{0.81}{5.73} \times 100 = 14.3$	$\frac{3.33}{5.73} \times 100 = 58.11$	$\frac{1.6}{5.73} \times 100 = 27.9$		5.73

* Interaction Factor (Refer to Table III)

** Percentage of interaction derived from interaction between A and B/interaction for state A and then converted to percentage

TABLE V

ESTIMATED MIGRATION FLOWS BETWEEN STATES

From State	To State A	B	C	D	Total Population
A	$80 \times \frac{90}{100} = 72$	$8 \times \frac{53}{100} = 4$	$8 \times \frac{30.9}{100} = 3$	$8 \times \frac{15.8}{100} = 1$	80
B	$6 \times \frac{20.2}{100} = 1$	$60 \times \frac{90}{100} = 54$	$6 \times \frac{55}{100} = 3$	$6 \times \frac{24.7}{100} = 2$	60
C	$4 \times \frac{14.9}{100} = 0$	$4 \times \frac{69.9}{100} = 3$	$40 \times \frac{90}{100} = 36$	$4 \times \frac{15.1}{100} = 1$	40
D	$5 \times \frac{14.3}{100} = 1$	$5 \times \frac{58.1}{100} = 3$	$5 \times \frac{27.9}{100} = 1$	$50 \times \frac{90}{100} = 45$	50

an interaction equal to ($P_i P_j/d_{ij}^2$) was calculated for each cell in the matrix.

The first step, then was to compute the interaction between the states. The interaction for any given state with all other states in the matrix were summed to obtain the total interaction for that given state e.g. the total interaction for state A is 5.11. Similarly, total interaction for the remaining states were computed.

The next step involved the calculation of the percentage of interaction of any given state with each other state j in the matrix based on the total interaction for that particular state with all others. For example, the total interaction for state A with all others--B, C, and D in the matrix is 5.11. The interaction factor of state A with B is 2.72. Hence, 2.72 is 53 percent of the total interaction for state A (2.72/5.11 = 53 percent). Similarly, figures for all other states were computed (Table IV). The computing of percentages accounts for the proportion of A's residents moving to B being different from those moving from B to A. The result is consistent with the inherent tendencies of certain tain states to attract more people in contrast to other backward states.

When considering the propensity of a population to migrate, it is inappropriate to assume that the population is homogenous and all members are potential migrants, because there are movers and stayers. With regard to India, government data sources for the year 1961 indicate that between 10 to 15 percent of the population in any given state move. On this basis it was assumed for this study that 90 percent of the population are stayers and the remaining ten percent are movers. The ten percent movers for any given state i could move from state i to any other given state j. With reference to the hypothetical example in

Tables III and IV, state A has a population of 80 people and if 90 percent of the people are stayers, then 90 percent of 80 is equal to 72 or 72 persons remain in the state. The remaining eight could move to either states B, C or D.

Of the eight movers, the exact number moving from state i to state j depends on the interaction between the two states. This has already been computed (Tables III and IV). With reference to state A in the computed example, the percentage of interaction with state B is 53 percent, to state C 30.9 percent and state D 15.8 percent. This is true if 100 people are moving out. If however, only eight people are moving out of state A, the result would then be four people moving to state B, three to C and one to D all calculated on the basis of the interaction (Table V). A simulated migration flow is thereby obtained. This indicates a general migration flow. The same procedures were then applied to actual data for the 21 states of India using actual population data for the year 1971. The obtained migration flow matrix was then transferred to the Markov Chain model in order to predict the movement patterns of cholera in relation to the migratory patterns of the people in the country.

The Markov Chain Theory

Mathematical models are said to be stochastic when their outcome depends partly on chance, or when the outcome of any one individual event within a sequence of events depends upon some probability p. The Markov process assumes dependence between one set of events and the immediately preceding set of events. In the language of probability theory, "the outcome of the directly preceding trial (and only on it)" (27, p. 171).

It is assumed that there us a finite series of outcomes of an experiment--t_1, t_2, t_3, ... t_n. Any single outcome of the experiment t_j is dependent upon the outcome of the preceding outcome t_{j-1}. In the Markov literature outcomes are termed "states" with each "state" representing changes in the distribution over time, will be referred to as time periods, or stages, in this study. The term states will be used to refer to the areal units, which are the states of India. Each stage is linked with the preceding stage by a set of probabilities termed the transition matrix. If the initial stage of the system is known (or the initial distribution at t_0) and information regarding the transition probabilities is known, then the behavior of the system over time can be computed. The transition probabilities can be arranged in a matrix similar to the example shown in Table V. In general the matrix form is:

$$P = \begin{matrix} P_{11} & P_{12} \cdots P_{ij} \cdots\cdots P_{1N} \\ P_{21} & P_{22} \cdots P_{2j} \cdots\cdots P_{2N} \\ P_{i1} & P_{i2} \cdots P_{ij} \cdots\cdots P_{iN} \\ P_{N1} & P_{N2} \cdots P_{Nj} \cdots\cdots P_{NN} \end{matrix}$$

where P_{ij} denotes the probability of movement from state i to state j during a single time period. The sum of probabilities in any row equals one and thus accounts for the redistribution of the residents of each areal unit from one time period to the next.

The initial stage of the system t_0, may be described as a vector of the form:

$$t_0 = S_1^{(0)}, S_2^{(0)}, S_3^{(0)} \ldots\ldots S_n^{(0)}$$

where $S_i^{(m)}$ represents the number of individuals in state i at time m. The second time period is:

$$t_1 = S_1^{(1)}, S_2^{(2)}, S_3^{(3)}, \ldots\ldots S_n^{(1)}.$$

calculated by multiplying the initial vector by the transition matrix

$$t_1 = t_0 * P$$

Similarly, the next stage can be expressed as:

$$t_2 = t_1 * P \text{ or}$$

$$t_2 = t_0 * P * P \text{ or}$$

$$t_2 = t_0 * P^2$$

and in general

$$t_K = t_{k-1} * P^{(k)}$$

The frequencies in each vector represent the distribution of the individuals among the states at a given time period.

These are the rudiments of the Markov Chain theory that are used in this study. For a complete discussion of other features such as mean stay times, ergodic and transient sets, regular and absorbing Markov Chains etc., the reader is referred to Kemeny and Snell (27) and Lauver (33).

The Application of the Theory to the Diffusion of Cholera

In considering the application of the Markov Chain theory to migration and the spread of cholera, the 21 states of India were used as the areal units. The distribution of cholera in the system at t_0 can be defined as $S_1^{(0)}, S_2^{(0)}, S_3^{(0)}, \ldots\ldots S_{21}^{(0)}$, where $S_1^{(0)}$, denotes the number of deaths from cholera in the last of the 21 states arranged in alphabetic order. The transition probability matrix discussed earlier, based on

the gravity model, was used to simulate migration flows within India (see pp. - and Table). In building the Markov Chain model, the following assumptions are of prime importance:

1) The spread of cholera or the rise in the number of deaths in each state is the result of the movement of people.

2) The spread of the disease in any given time period is regarded as a stochastic event with some probability p of occurrence.

3) The mobility of cholera victims is proportional to the mobility rates of the population.

4) The system is a closed one (i.e. no one can enter the system through birth or migration and no one can leave the system through death or migration).

5) The probability of the disease moving from state i to state j, summed for all j will equal unity; (i.e. $\sum_{i=1}^{21} P_{ij} = 1$).

6) The migration probabilities between any two states do not change over time.

7) The initial distribution of the death rates for cholera are known.

8) The number of infected persons in a given state (the potential carriers) is proportional to the number of deaths from cholera in that state.

The results of the Markov Chain process are purely hypothetical and are constrained by the assumptions outlined above and the regularities of finite mathematics.

The Markov Chain in this study describes a diffusion process which takes place through several time periods. The initial time t_0 represents

the time period prior to that from which the diffusion process is traced.

The initial vector contains 21 elements, one for each of the 21 states of India. Each of the elements in this vector contains the number of deaths from cholera during t_0--defined as January and February in this study. The sum of the elements of the vector is equal to the total number of cholera deaths in the country at t_0.

The second component of the model is a stochastic matrix containing the probability of people moving from state i to state j. This matrix consists of 21 rows each representing the same 21 states and 21 columns again representing the same 21 states with the order being the same for both the rows and the columns as in the vector t_0. Each element in the matrix is the probability of people moving from state i to state j and vice versa. The diagonal of the matrix represents the probability of people staying within the state.

The predicted distribution of deaths at t_1 results from matrix multiplication:

$$\underset{(1x21)}{t_1} = \underset{(1x21)}{t_0} * \underset{(21x21)}{P}$$

and in general the distribution of deaths at any given time period $t_{(K+1)}$ is a function of the distribution at the previous time period:

$$t_{K+1} = t_K * P$$

Note that the model does not predict the total number of deaths at any given time period, but rather predicts the distribution of the disease at each time period.

In order to compare predicted deaths with the actual deaths from cholera it was necessary to multiply the predicted vector t_1 by a

constant to equalize the total number of deaths in the system at each time period. That is, if $A_1 = (d_1, d_2, \ldots\ldots d_{21})$ represents the actual number of deaths in the system then:

$$\sum_{i=21}^{21} d_i = C_1 \sum_{i=21}^{21} S_i^{(1)} = \sum_{i=1}^{21} C_1 S_i^{(1)}$$

A single time interval in the theoretical Markov process has no fixed meaning in reality. One matrix multiplication could represent anything from a minute to a century. Thus, it is necessary to look at a series of predicted vectors and arbitrarily select time intervals, or powers of the transition matrix, that seem to allow the diffusion process to work long enough to cause a significant redistribution.

In this study the incremental power was arbitrarily set at ten. Predicted values were calculated for t_{10}, t_{20}, t_{30}, t_{40},t_{50}, which represent actual time periods t_1 (March-April), t_2 (May-June), t_3 (July-August), t_4 (September-October), and t_5 (November-December). The predicted deaths for each time period were compared with the actual deaths using two techniques. First, Pearson's correlation coefficients between actual and predicted values were calculated. Second, in order to determine the error in prediction as derived from the Markov Chain process, the predicted deaths were subtracted from the actual values.

The residuals were then mapped to ascertain the states where the Markov process either over-estimated or under-estimated the death rates. This helped not only in establishing the ability of the Markov process in predicting the diffusion of the disease by migration but also in helped to reveal factors that might effect its spread.

Note that the residuals, in this study, were not computed by linear regression. Use of regression would calculate a second set of predicted

values based on the Markov predicted values and regression residuals would then be derived from this second set of predicted values. Such residuals would be only secondarily related to the Markov process and are in a sense an approximation of an approximation.

CHAPTER V

RESULTS OF THE ANALYSIS

Introduction

In this chapter the results of the analysis are presented with the focus being on two aspects. First, an attempt is made to describe the movement of cholera as related to the movement of people within the country. Secondly, the Markov Chain process is assessed as a descriptive tool in disease research. The limitations of the data should be kept in mind throughout this chapter (i.e. the input to the model consisted of the interstate migration flows estimated by the Gravity model and the initial distribution of deaths from cholera in the 21 states of India).

The transition matrix of the Markov Chain captures and embodies the process of movement. All other descriptive measures are drawn directly from this matrix, which in this study, represents the movement of disease between all possible pairs of the 21 states; hence a 21 x 21 matrix. The analysis is then presented under the following headings in order to partition each segment:

1) Mapped patterns of cholera
2) Predictions by the Markov Model
3) Error in prediction by state for all time periods
4) Salient flows as related to migration.

Mapped Patterns of Cholera for 1971

Before discussing the results of the analysis, it may prove to be useful to examine a map of the study area. The location of the 21 states of India are illustrated in Figure 4.

Historically, cholera has shown a markedly varying incidence from year to year with a maximum of 805,698 deaths in 1900 to about 60,000 deaths in 1932 (96, p. 78). Over the years, cholera deaths have shown a long term decrease as seen in Figure 2. Although the figure does reveal that the magnitude of the disease has decreased with the successive years, the death rates for 1971 (the total deaths for the year is 4,226) does leave room for doubt as to whether with all the public health services, the disease has been brought under control.

The death rates for the six different time periods reveal a marked difference between the northern and the southern states (Figure 5). The endemic focus in Bengal shows a marked tendency to peak soon after the monsoon hits the coastal areas in the Bay of Bengal (i.e. after late May/June). Away from the endemic center, the disease peaks at a later period. The same foci shows another peak during the end of the year in November and December. The spread of the disease seems to be free from this endemic center to areas generally free from cholera under favorable climatic conditions.

In provinces where the effects of the monsoons are felt the most, the peak periods are closely related to the seasonal effects of the monsoon. This effect is more marked along the Indo-Gangetic valley in the states of Bengal, Bihar and Uttar Pradesh.

The southern states of Andhra Pradesh, Mysore, Tamil Nadu, Kerala

Figure 4. The 21 States of India

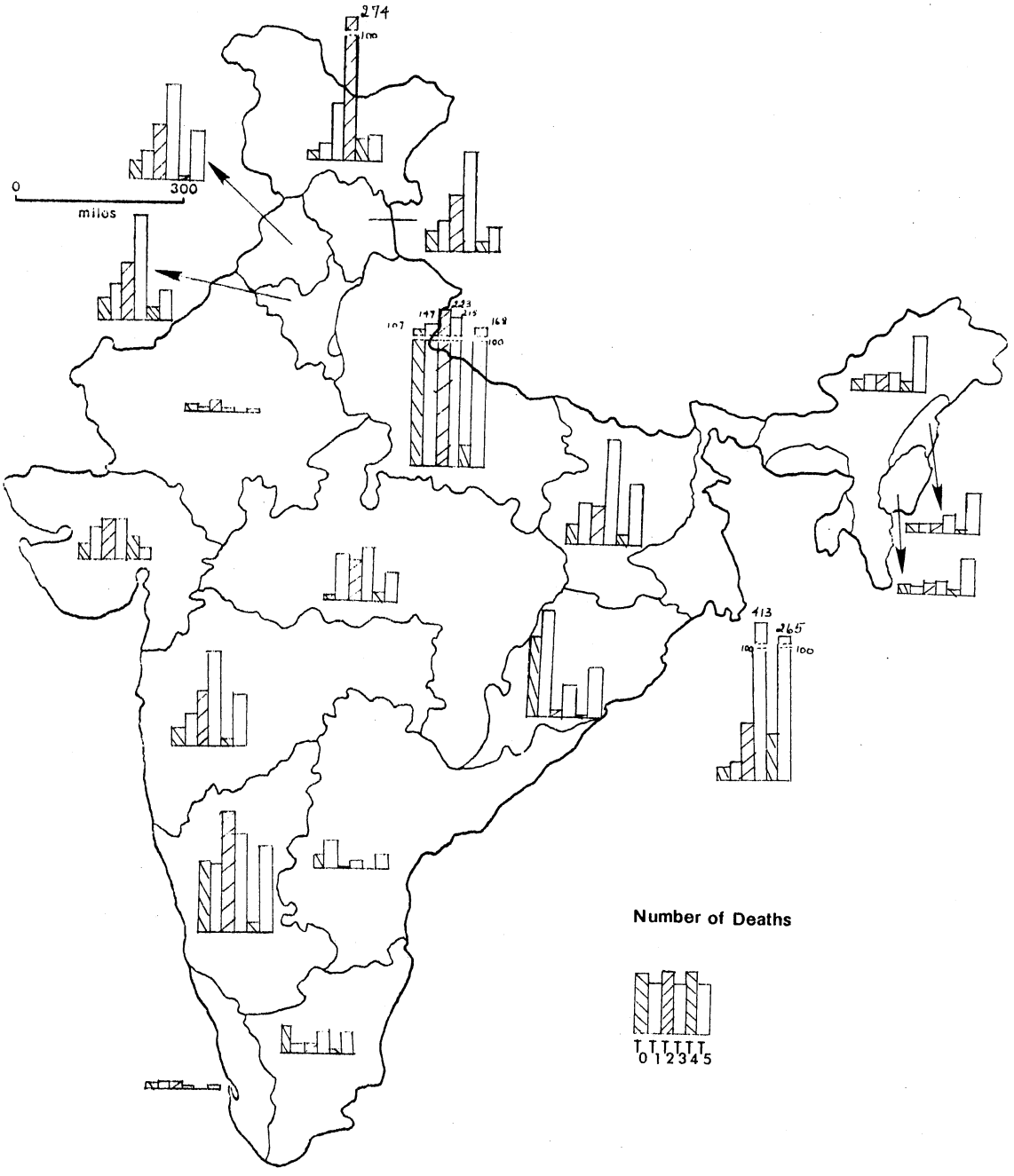


Figure 5. Total Number of Deaths by Cholera for Six Time Periods - By State in 1971

and Maharashtra reveal tendencies to peak at later periods, after the disease has already peaked in the northern states. Presumably, this can be related to the movement of people from the northern states where the disease already exists. The one exception to the general trend in the southern states appears to be Mysore where the peak in May and June probably is related to the outbreak of the monsoon on the west coast in May.

Incidence rates for cholera per 100,000 inhabitants for each state were computed and mapped (Figures 6-11) revealing that for the first two time periods t_0 (January-February) and t_1 (March-April) the incidence rates were low ranging from about 0 to 10 deaths per 100,000. Incidence rates for t_2 (May-June) show a definite increase from 0 to 10 deaths to 11-25 deaths per 100,000. Mysore, in the south, also shows an increase from about 11 to 25 to 26-50 deaths per 100,000.

Time period t_3 (July-August) reveals the patterns to be the same as the previous two time periods for the central states of India. However, Bengal shows a definite increase from 11-28 deaths to 51 to 100 deaths per 100,000. With the advent of the monsoon, all the states in the Gangetic valley--Uttar Pradesh, Bihar, Himachal Pradesh and Haryana show a marked increase as the monsoon moves up the valley from the Bengal basin. This pattern continues on into t_4 (September-October) with the outer states such as Assam, Orissa and Maharashtra showing a slight increase as the infection spreads. During the latter period t_5 (November-December), most of the states show a marked decrease from the incidence rates in the previous time periods although the disease has not totally disappeared.

The comparison of the maps between these different time periods

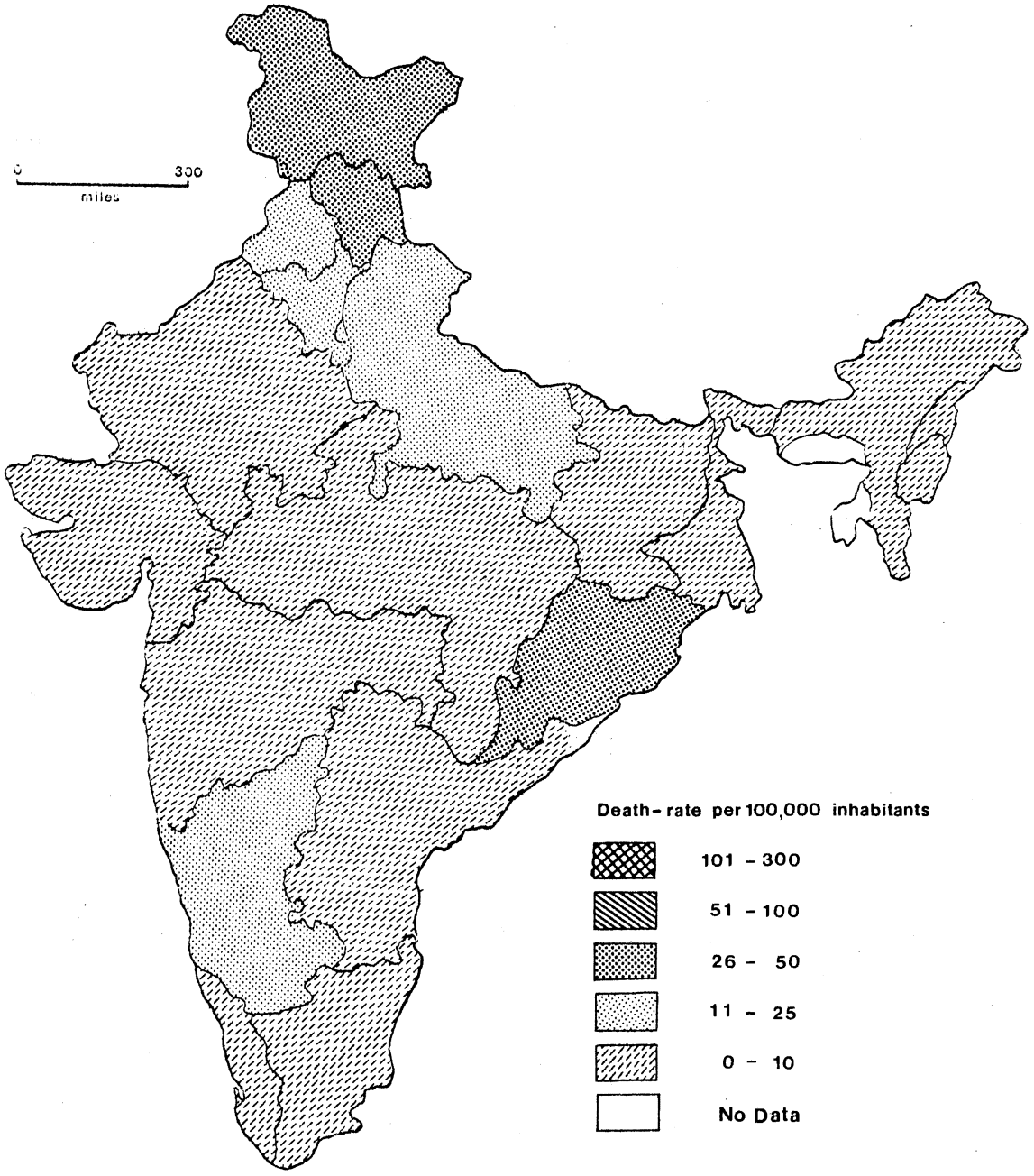


Figure 6. Incidence Rate Per 100,000 Inhabitants for t_0

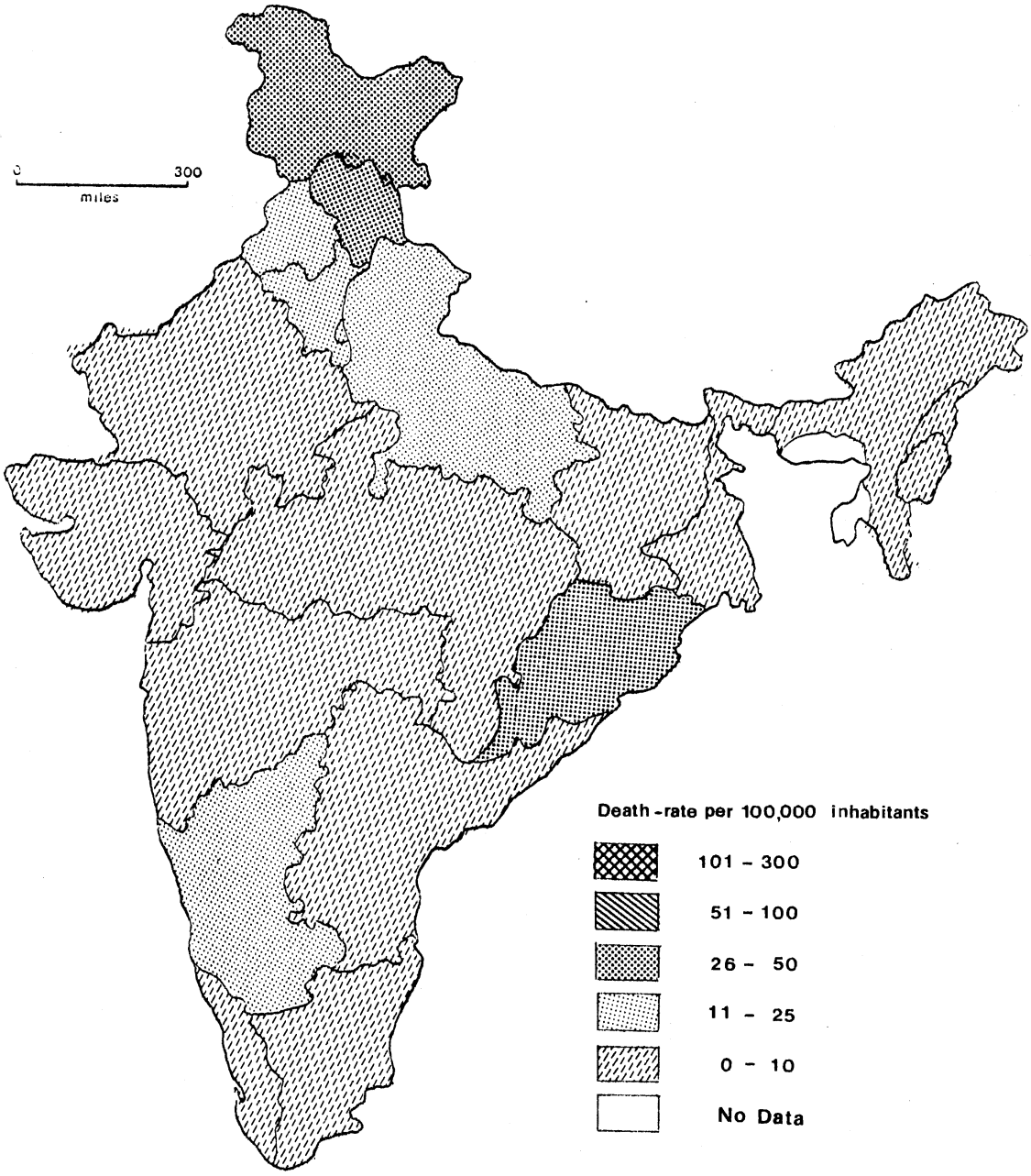


Figure 7. Incidence Rate Per 100,000 Inhabitants for t_1

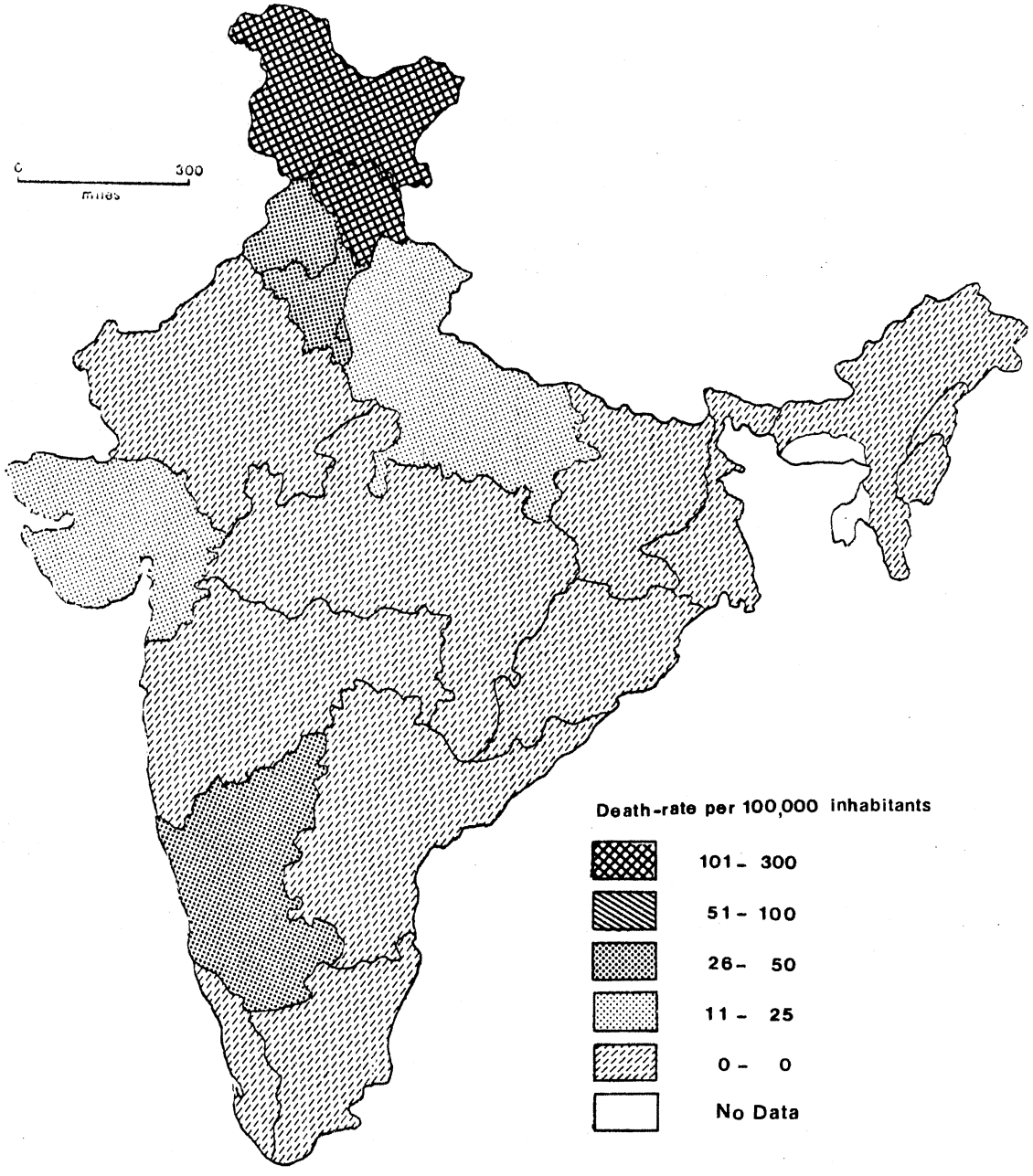


Figure 8. Incidence Rate Per 100,000 Inhabitants for t_2

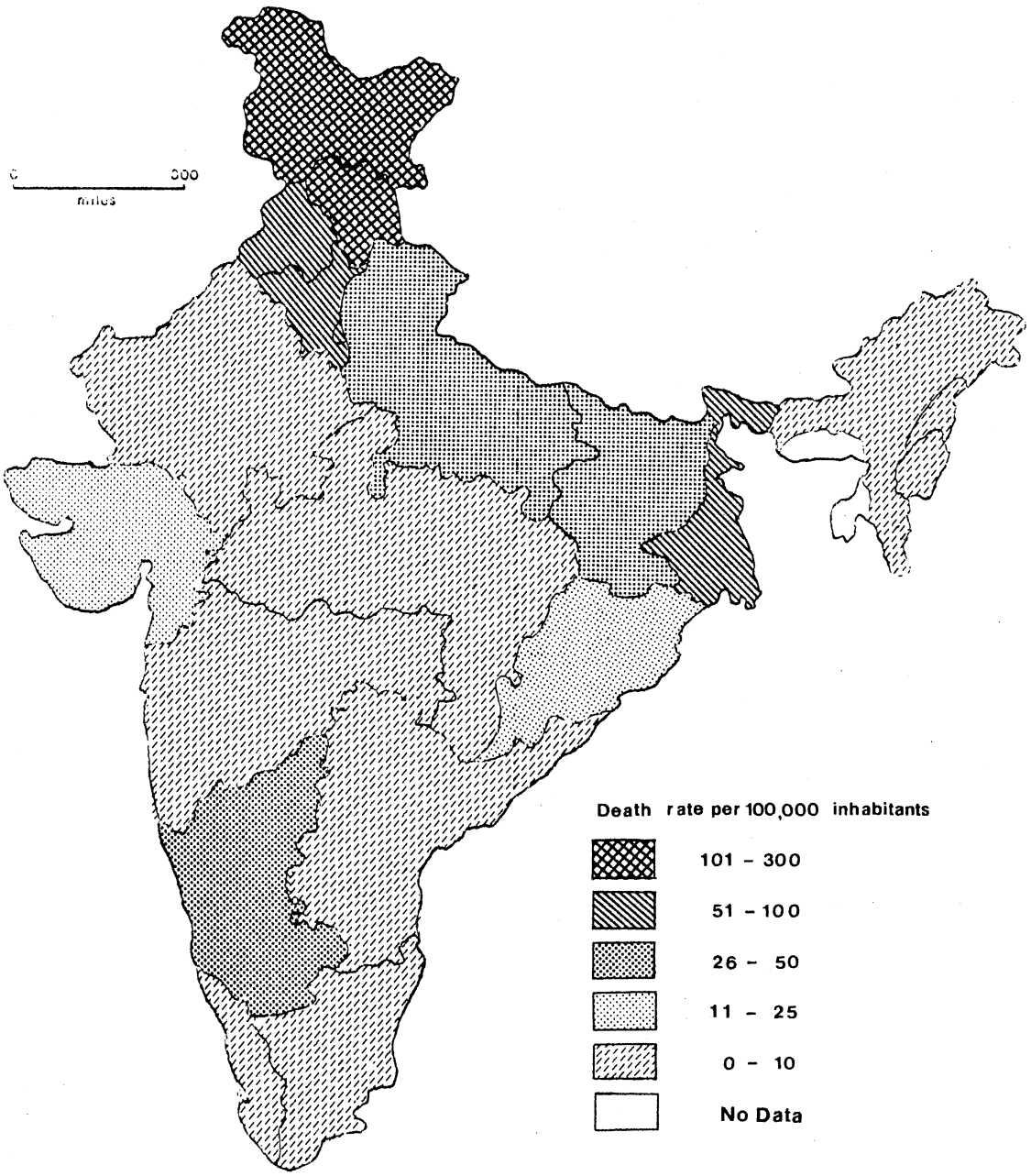


Figure 9. Incidence Rate Per 100,000 Inhabitants for t_3

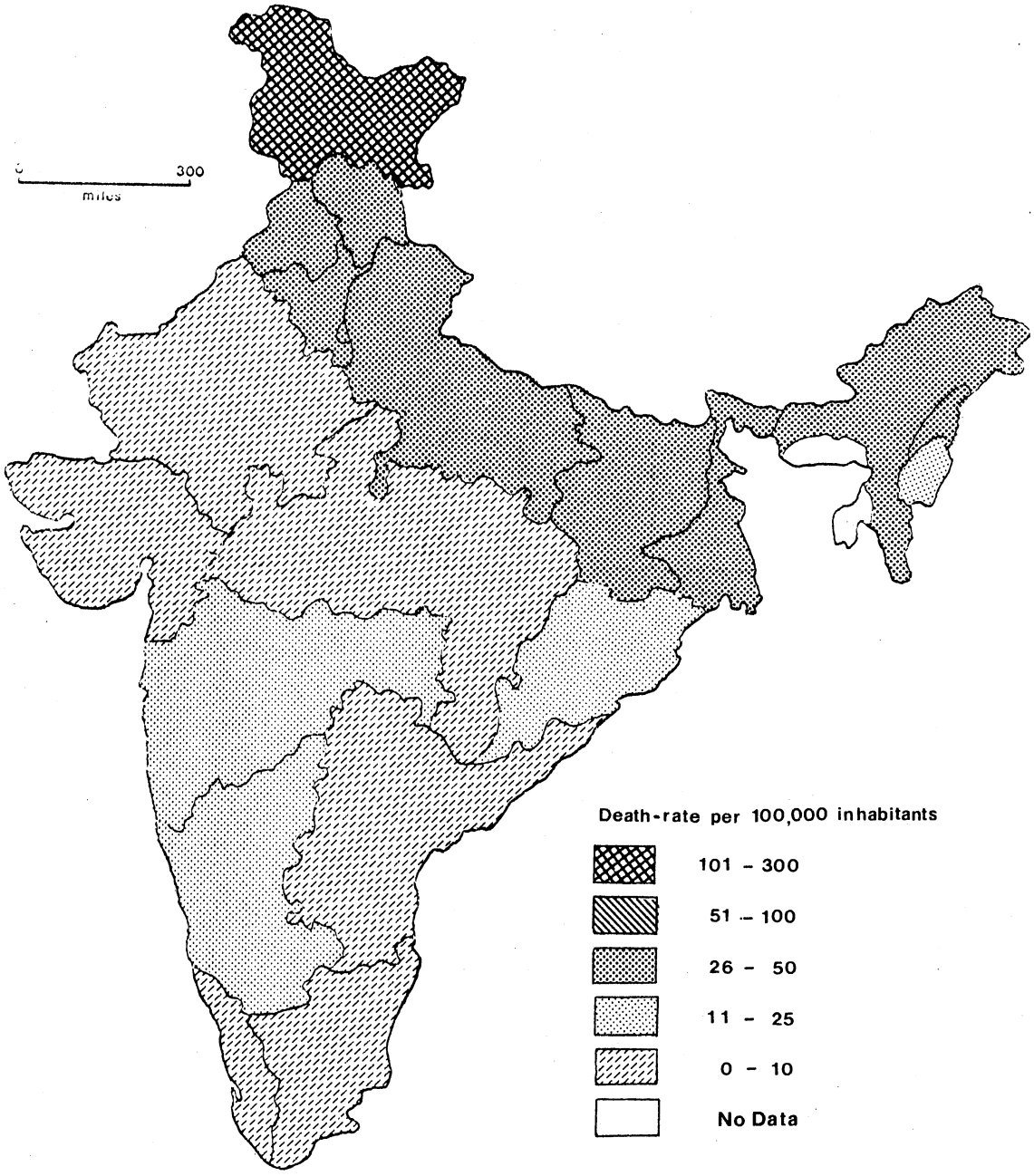


Figure 10. Incidence Rate Per 100,000 Inhabitants for t_4

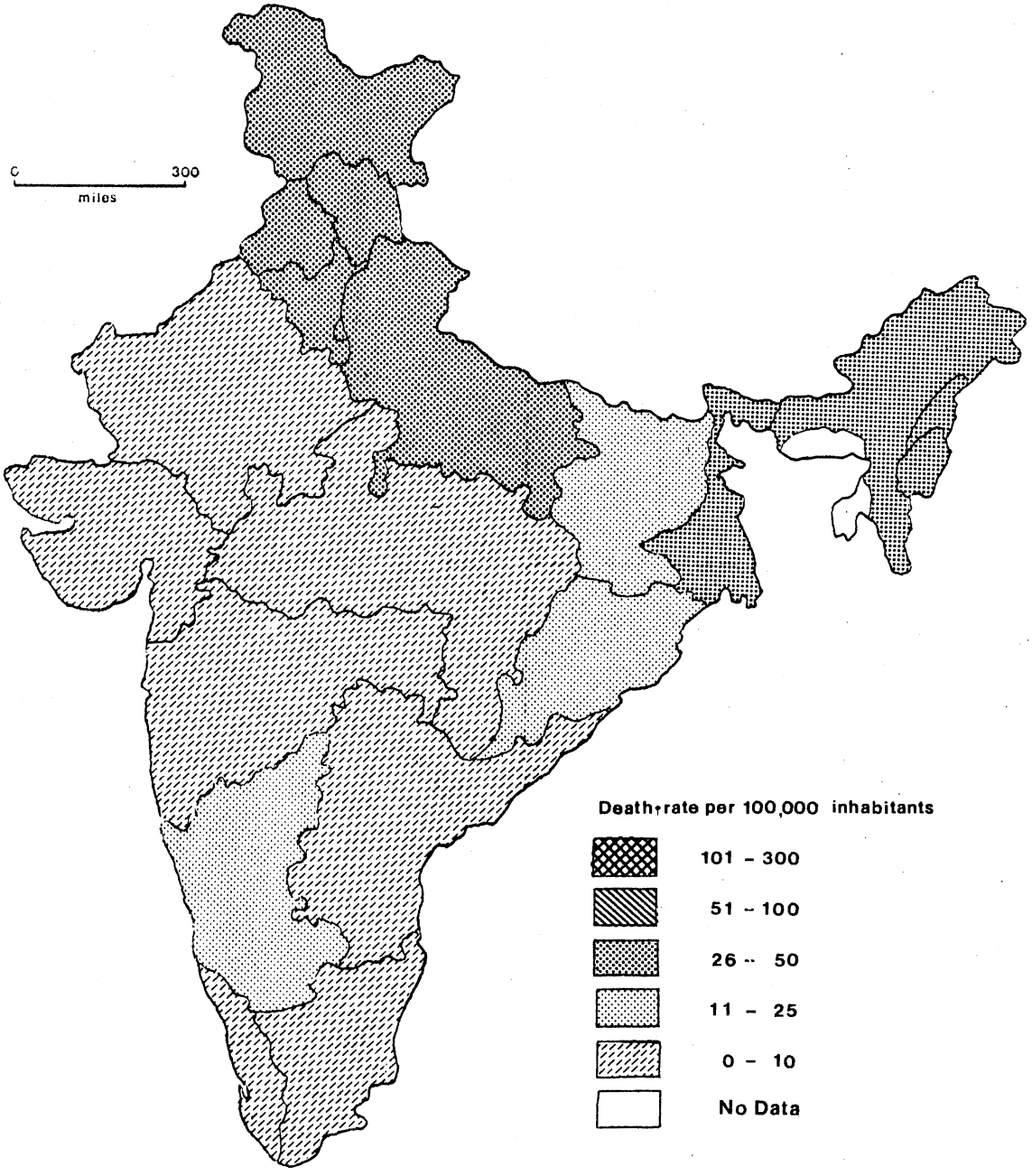


Figure 11. Incidence Rate Per 100,000 Inhabitants for t_5

reveals a definite seasonal trend with the peak periods varying in the northern and the southern states. In the northern states the peak periods coincide with the southwest monsoon with a reversal in the south where the peak period coincides with the northeast monsoon.

In some provinces like Uttar Pradesh and Jammu and Kashmir, the peak incidence can be explained possibly by the holding of large scale pilgrimages (e.g. the peaking of cholera in Jammu and Kashmir during t_2 coincides, presumably, with the Amarnath pilgrimage). Fairs and pilgrimages held at Allahabad and Hardwar in Uttar Pradesh and Punjab correspond with the increase in death rates of cholera in these states.

To a large extent mapping the incidence rates reveals the extent of the problem areas of the disease. In states like Bengal and Uttar Pradesh with large populations 493 or even 500 deaths do not significantly affect the population. But in states like Jammu and Kashmir which are sparsely populated, other than during the pilgrimage months when there is a large influx of people, even 20 to 30 deaths affect the population to a much larger extent. The mapping of the incidence rates helps to demarcate the endemic foci from the outlying states. It becomes clear that Bengal continues to harbor the disease.

Predictions by the Markov Chain Process

The distribution of deaths for t_0 (January-February) formed the base vector from which future predictions were arrived at, using the fundamental transition matrix in Table XV (see Appendix). Table VI presents a comparison of the actual deaths from cholera for the year 1971 compared to the predicted deatha arrived at by the Markov process.

Comparisons between the actual and predicted deaths reveal that

TABLE VI

COMPARISON OF ACTUAL DEATHS (AD) AND MARKOV PREDICTIONS (PD) DEATHS FROM CHOLERA FOR t_1, t_2, t_3, t_4 AND t_5 (IN 1971)

State	AD_1	PD_1	AD_2	PD_2	AD_3	PD_3	AD_4	PD_4	AD_5	PD_5
AP	23	16	2	36	7	89	3	80	17	55
AS	11	14	11	22	14	52	77	48	44	33
BH	31	23	30	38	71	82	79	74	48	50
GJ	21	18	30	31	31	74	18	68	9	96
HY	27	27	45	32	82	63	41	55	21	35
HP	25	23	46	20	80	26	40	19	19	13
JK	12	8	47	18	274	11	146	6	17	4
KE	6	7	6	7	2	26	3	22	2	15
MP	37	7	31	11	44	97	44	93	23	66
MH	36	17	31	36	41	93	56	87	11	57
MA	6	8	9	7	k4	15	18	13	31	6
ME	0	0	0	0	0	0	0	0	0	2
MY	53	71	92	50	79	71	59	48	67	31
NG	5	7	7	4	10	4	19	0	29	0
OR	82	85	9	58	28	78	25	51	40	33
PJ	22	27	42	41	78	89	42	77	40	52
RJ	3	3	8	4	3	7	1	6	1	4
TN	8	28	8	36	14	71	15	61	28	42
TP	0	1	0	1	0	3	0	3	0	4
UP	147	152	223	187	215	379	215	69	168	368
WB	22	31	47	81	413	167	391	158	265	114

the predictions for t_1 are close to the actual deaths for t_1, the correlation being r = .95 (see Table VI). The predictions for t_2 are also close to the actual deaths except in the state of Mysore and Uttar Pradesh where the predicted deaths are higher than the actual deaths. The correlation, in this case, is still relatively high, with r =.88. However, these high correlations between the actual and predicted deaths for t_1 and t_2, drop drastically to r = .53 at t_3 (May-June) then increase to r = .59 and r = .74 to t_4 and t_5 respectively. In the case of t_3, t_4, and t_5 respectively. In the case of t_3, t_4, and t_5 the predicted deaths are either less or grossly higher than the actual deaths for most of the states in the country. The apparent reason for this decline in the Markov process's predictive ability seems to be related to the monsoon factor.

Assuming that the onset of the monsoon does interrupt the diffusion of the disease via migration, and that this is the reason for the poor fit for t_3, the model was calibrated again using the actual deaths for t_3 as the base vector. The same migration matrix and procedures were used to predict the future distribution of cholera given the new initial distribution.

Correlations of the actual and predicted deaths were greatly improved for both t_4 and t_5. The correlation for t_4 increased to .93 and t_5 increased to .88. The state to state predictions are fairly close to to the actual deaths for those two time periods (Table VII). This indicates that the diffusion of cholera for t_4 and t_5 also can be predicted fairly well by the Markov process given the distribution at t_3. Thus, only the change in the distribution of cholera deaths between t_2 and t_3 appears to be weakly related to the hypothesis of diffusion by migration.

TABLE VII

COMPARISONS OF ACTUAL DEATHS AND PREDICTED DEATHS FOR t_4 AND t_5 USING t_3 AS BASE VECTOR

State	AD_4	PD_4	AD_5	PD_5
AP	3	39	17	41
AS	77	38	44	32
BH	79	62	48	44
GJ	18	44	9	37
HY	41	63	21	38
HP	40	39	19	19
JK	146	88	17	24
KE	3	13	2	12
MP	44	69	23	58
MH	56	57	11	47
MA	18	14	31	10
ME	0	1	0	1
MY	59	46	67	28
NG	19	3	29	1
OR	25	46	40	34
PJ	42	113	40	69
RJ	1	6	1	4
TN	15	42	28	35
TP	0	3	0	3
UP	215	215	108	166
WB	391	283	265	162

Error in Prediction - By State

Residuals computed by subtracting the predicted deaths from the actual deaths (refer to Chapter IV, p.57) for all time periods reveal the significance of the predictive capabilities of the Markov process. For t_1, the predictions arrived by the process show an accurate estimation because the residuals for the most part do not exceed $\pm$ 10. The major exception, in this case, are the states of Madhya Pradesh and Maharashtra where there are more deaths than predicted in contrast to Tamil Nadu and Mysore where the predicted deaths are higher than the actual deaths (Table VIII, Figure 12).

During t_2 (Table IX, Figure 13), all the states along the east coast--Tamil Nadu, Andhra Pradesh, Orissa and Bengal show negative residuals; the predicted deaths are higher than the actual deaths. In contrast the central and northwest states have more deaths than predicted. This trend is further exhibited in t_3 (Table X, Figure 14) when the states of Bihar, Tamil Nadu, Andhra Pradesh, Orissa and Maharahtra show the predicted deaths to be higher than the actual deaths. In Bengal, Jammu and Kashmir and Himachal Pradesh there were many more deaths than predicted. The movement of the disease into the north central and northeast states during this time period appears to be a function of the monsoon and not one of migration. In any case, the migration model does not adequately account for the spread or appearance of the disease in this area.

The monsoon, to some extent determines the rise in the number of deaths in some areas where its influence is greatly felt. The southwest monsoon begins in the latter half of May and lasts until around

TABLE VIII

RESIDUALS FOR t_1 - BY STATE

State	Actual Deaths	Predicted Deaths	Residuals
AP	23	16	7
AS	11	14	-3
BI	31	23	8
GJ	21	18	3
HY	27	27	0
HP	25	23	2
JK	12	8	4
KE	6	7	-1
MP	37	7	30
MH	36	17	19
MN	6	8	-2
ME	0	0	0
MY	53	71	-18
NG	5	7	-2
OR	82	85	-3
PJ	22	27	-5
RJ	3	3	0
TN	8	28	-20
TP	0	1	-1
VP	147	152	-5
WB	22	31	-9

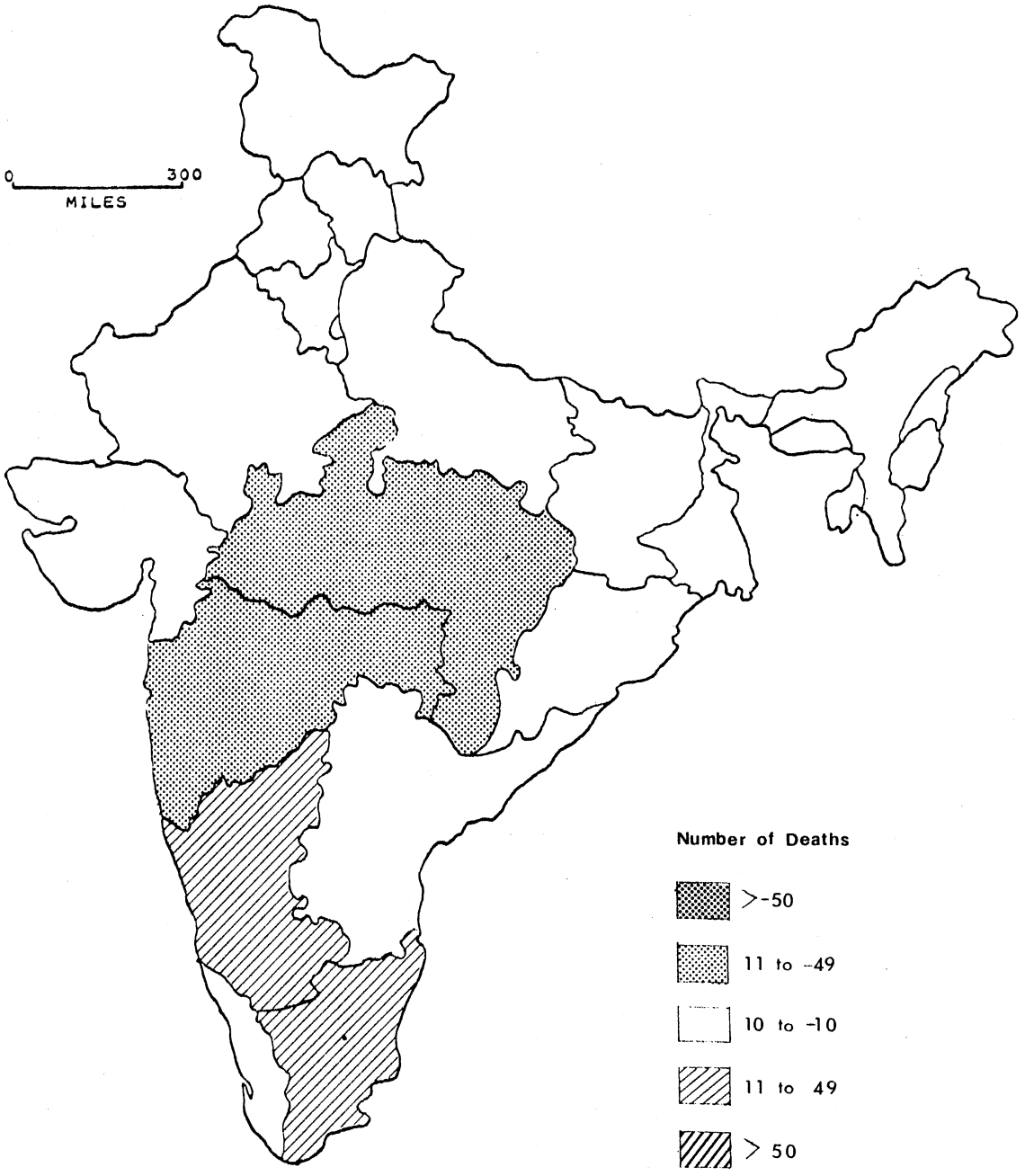


Figure 12. Residuals for t_1 - By State

TABLE IX

RESIDUALS FOR t_2 - BY STATE

State	Actual Deaths	Predicted Deaths	Residuals
AP	2	36	-34
AS	11	22	-11
BI	30	38	-8
GJ	30	31	-1
HY	45	32	13
HP	46	20	26
JK	47	18	29
KE	6	7	-1
MP	31	11	20
MH	31	36	-5
MN	9	7	2
ME	0	0	0
MY	92	50	42
NG	7	4	3
OR	4	58	-49
PJ	42	41	1
RJ	8	4	4
TN	8	36	-28
TP	0	1	-1
VP	223	187	36
WB	47	81	-34

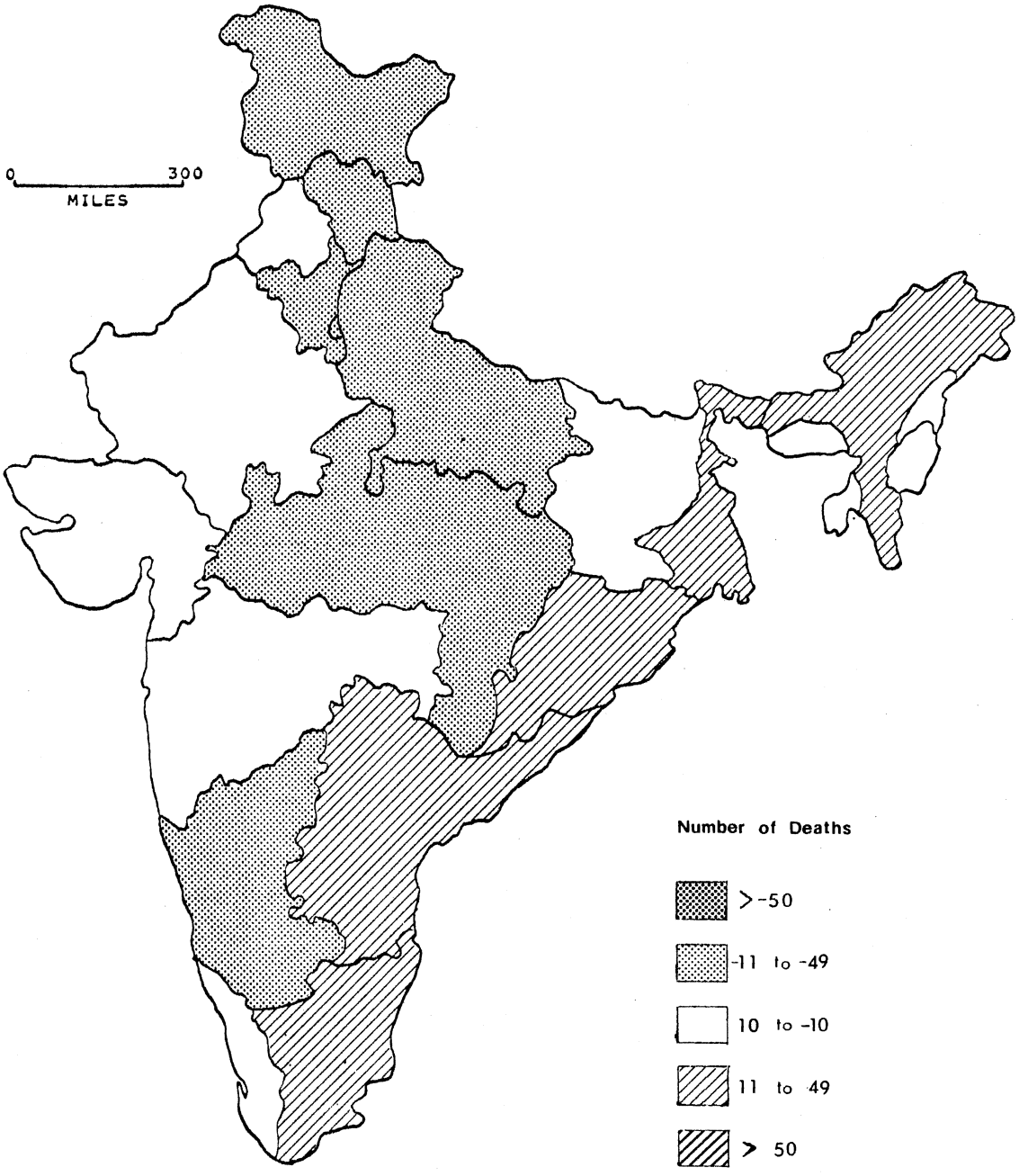


Figure 13. Residuals for t_2 - By State

TABLE X

RESIDUALS FOR t_3 - BY STATE

State	Actual Deaths	Predicted Deaths	Residuals
AP	7	89	-82
AS	14	52	-38
BI	71	82	-11
GJ	31	74	-43
HY	82	63	17
HP	80	26	54
JK	274	11	263
KE	2	26	-24
MP	48	97	-49
MH	41	93	-52
MN	14	15	-1
ME	0	0	0
MY	79	71	8
NG	10	4	6
OR	28	78	-50
PJ	78	89	-11
RJ	3	7	-4
TN	14	71	-57
TP	0	3	-3
VP	215	379	-164
WB	413	167	246

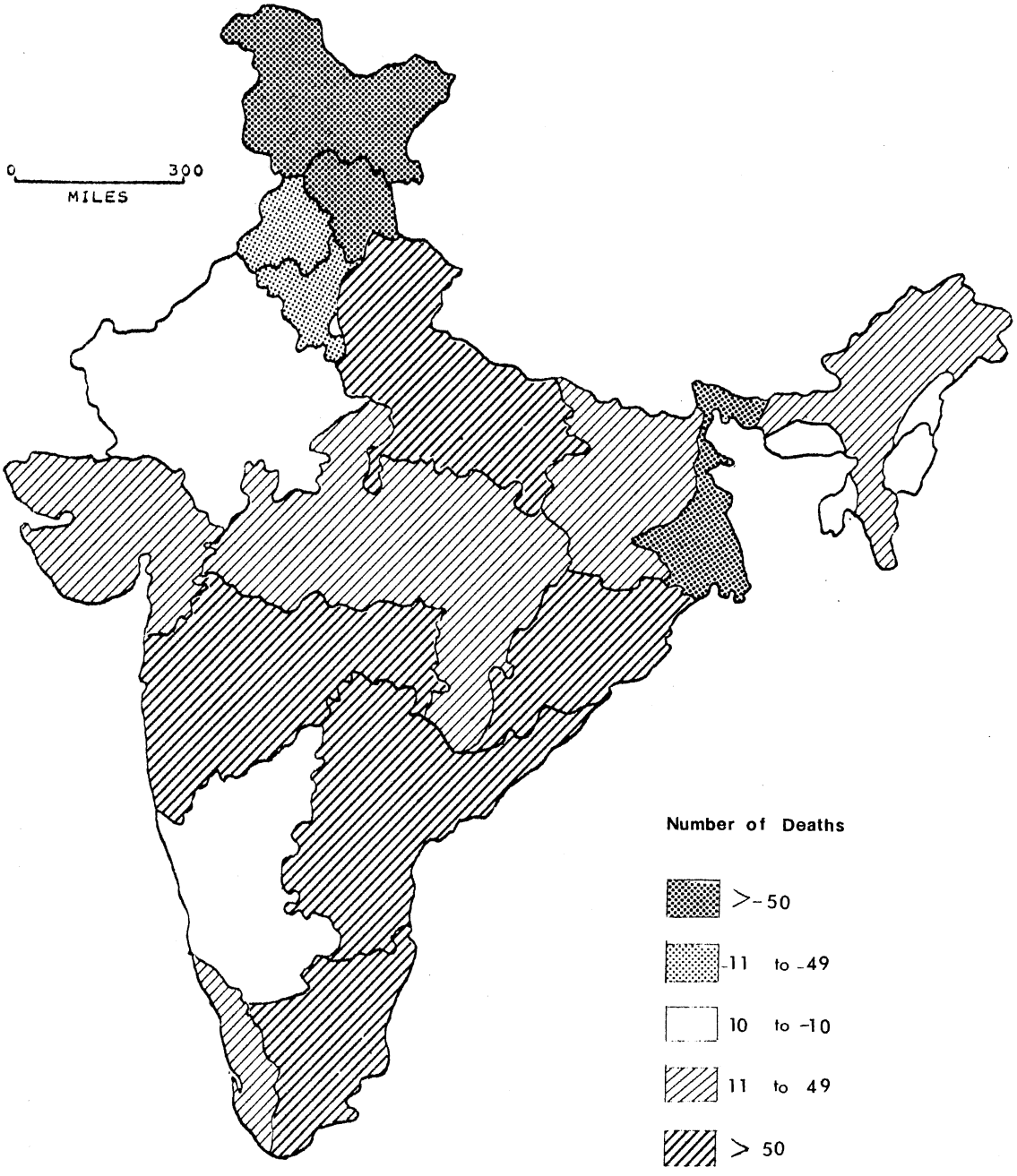


Figure 14. Residuals for t_3 - By State

September. The locations that receive the most rainfall are located along the windward slopes of the Western Ghats in the states of Maharashtra, Mysore and Kerala and the Khasi Hills of Assam. The full effect of the moisture laden winds moving off the Bay of Bengal is also felt along the Indo-Gangetic valley staring in early June. As the winds are funneled westwards along this area their effect diminishes. This results in the steady decrease in the length of the rainy season and the amount of precipitation from the southeast (Calcutta) to the northwest (Delhi). By the end of September, the southwest monsoon is in full retreat. This is followed by a dry, cool season with the eastern two-thirds of the peninsular part of India receiving its rainfall from the northeast monsoon from around late October until around November-December.

Cholera, has marked tendencies to reach epidemic proportions during the monsoon months as has been seen in the history of the disease. The deaths from cholera also seems to rise in the endemic areas soon after the monsoon, as seen in the earlier segments of this chapter. The Indo-Gangetic valley states of Bihar, Bengal and Uttar Pradesh showed marked tendencies to peak around late June-September, depending on the arrival of the monsoon. This also holds true for the Malabar coast on the west. In contrast the states of the Coromandel coast show tendencies to peak in the winter months of November-December.

Assuming this to be true, and a factor in the rise of deaths by cholera, it is possible to explain the interruption of the diffusion of the disease by migration during t_3. It also helps to explain the problem of the very high residuals for this time period which heralds the beginning of the monsoon.

Again, the Markov process estimates more than the actual deaths for the southern states during t_4 (Table 11, Figure 15) (where t_3 was used as the base vector). In U.P., Maharashtra, Rajasthan, Kerala, Punjab and Himachal Pradesh, the predictions were very close to the actual deaths. For the time period t_5 (Table 12, Figure 16) the process was able to predict the actual deaths fairly close to the actual deaths but not as close as the previous time period t_4. The central states of Gujarat, Madhya Pradesh and Maharashtra; the southern state of Andhra Pradesh along with Punjab and Haryana have negative residuals. In contrast, Mysore, Assam, Nagaland and Manipur have many more deaths than predicted. Bengal has 103 more deaths than predicted. The results of the computations of the residuals and the mapped patterns are to be found in Tables VIII through XII and Figures 12 through 16.

Salient Flows of the Disease

A selection criterion of $P_{ij} = >.0500$ was employed to identify the more significant flows of the disease as related to migration. This was employed in order to restrict the analysis to the salient flows only with regard to the 21 states of the country rather than all the possible 441 flows in a 21 x 21 matrix. This means only the states that have a diffusion probability of .0500 were considered to be significant flows. The results of the selection process are summarized in Table XIII. Flows with a probability less than the selected criteria are too numerous to be listed. These significant flows account for about ten percent of the flows but they explain about 50 percent of the movement.

An examination of Table XIII indicates that the major flows are from the outlying states to U.P. and West Bengal with the more

TABLE XI

RESIDUALS t_4 - BY STATE - BASED ON t_3 AS INITIAL VECTOR

State	Actual Deaths	Predicted Deaths	Residuals
AP	3	39	-36
AS	77	38	39
BI	79	62	17
GJ	18	44	-26
HY	41	62	-21
HP	40	39	1
JK	146	88	58
KE	3	13	-10
MP	44	69	-25
MH	56	37	-1
MN	18	14	4
ME	0	1	-1
MY	59	46	13
NG	19	3	16
OR	25	46	-21
PJ	42	113	-71
RJ	1	6	-5
TN	15	42	-27
TP	0	3	-3
VP	215	215	0
WB	391	283	108

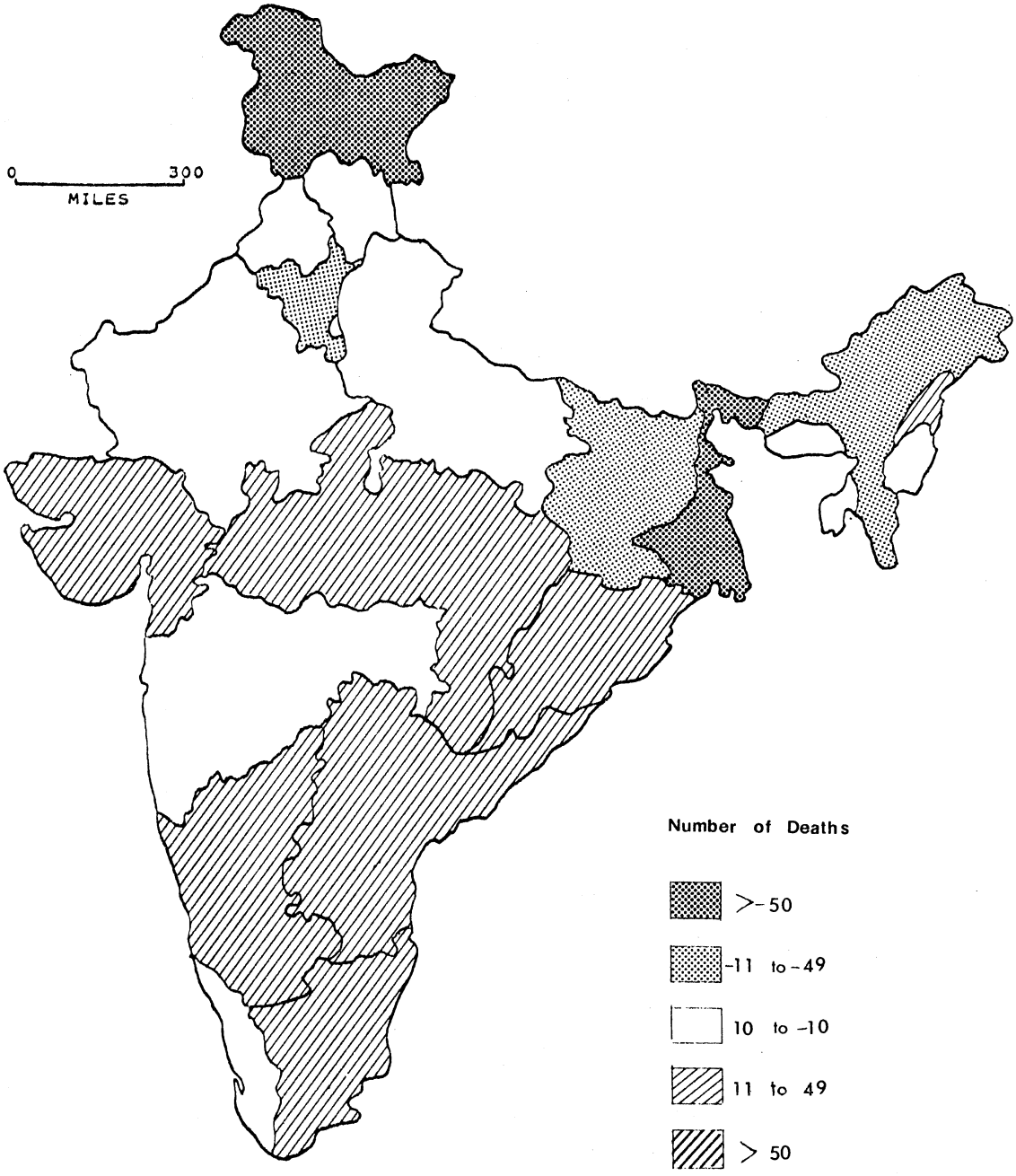


Figure 15. Residuals t_4 - By State - Based on t_3 as Initial Vector

TABLE XII

RESIDUALS t_5 - BY STATE - BASED ON t_3
AS INITIAL VECTOR

State	Actual Deaths	Predicted Deaths	Residuals
AP	17	41	-26
AS	66	32	12
BI	48	94	4
GJ	9	37	-28
HY	21	38	-17
HP	19	19	0
JK	17	24	-7
KE	2	12	-10
MP	23	58	-35
MH	11	47	-36
MN	31	10	21
ME	0	1	-1
MY	67	28	39
HG	29	1	28
OR	40	34	6
PJ	40	59	-29
RJ	1	4	-3
TN	28	35	-7
TP	0	3	-3
VP	168	166	2
WB	265	162	103

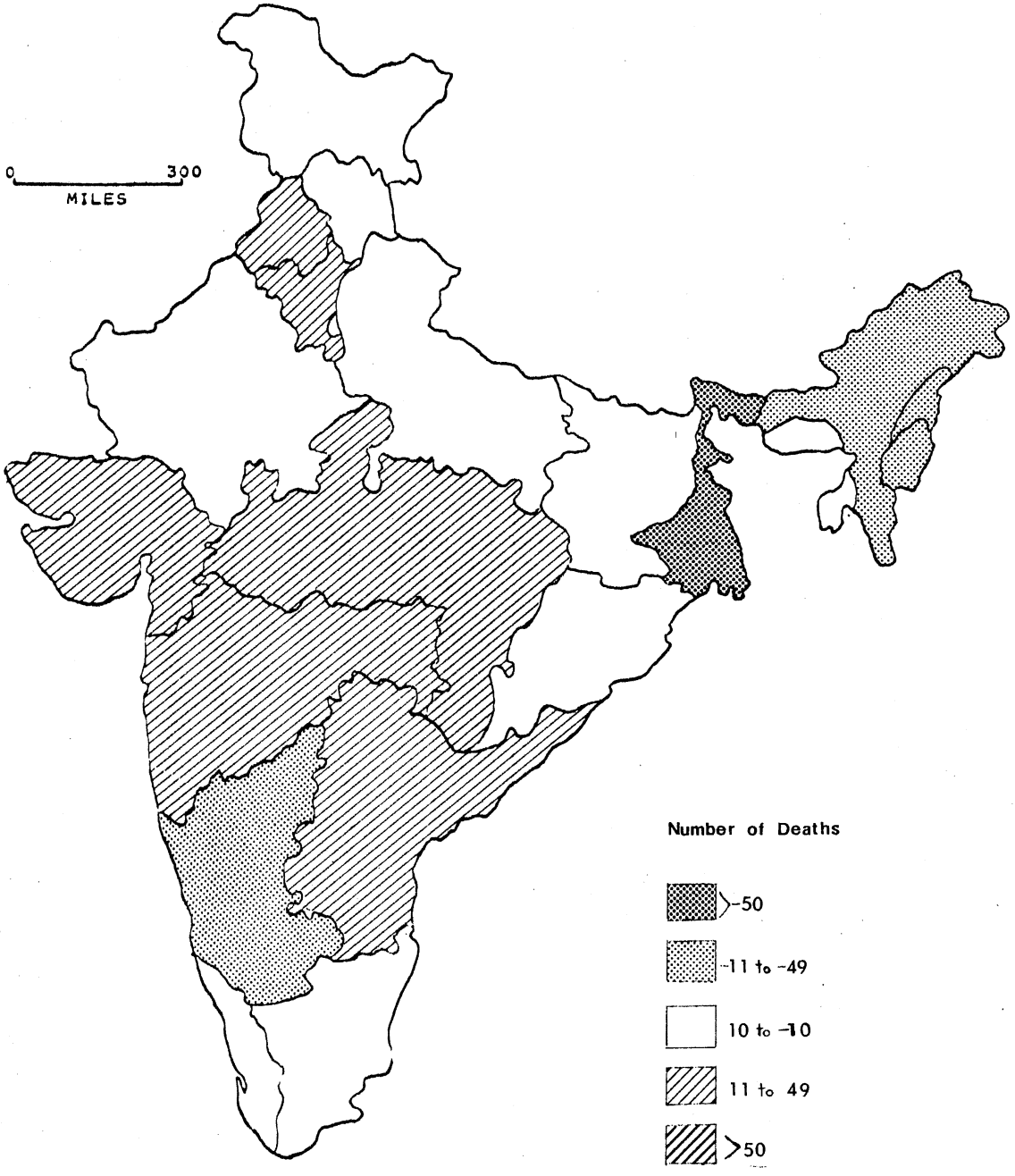


Figure 16. Residuals t_5 - By State - Based on t_3 as Initial Vector

TABLE XIII

SALIENT FLOWS OF CHOLERA RELATED TO MIGRATION
P_{ij} = >.0500

State to State		P_{ij}	State to State		P_{ij}
AP	MY	.0585	ME	MY	.3025
	TN	.0714		UP	.0561
	WB	.0550		WB	.0528
AS	MP	.0850	MY	AP	.0795
	UP	.0899		MH	.0900
	WB	.0636		TN	.0725
BH	TN	.0842		UP	.0914
GJ	MP	.0579	NG	BH	.0748
	MH	.0798		UP	.1198
	PJ	.0557		WB	.1148
MY	PJ	.0580	OR	AP	.0708
				MP	.0656
MP	OR	.1300		MH	.0534
	TN	.0599		UP	.1290
	UP	.0781		WB	.0946
JK	PJ	.2208	PJ	BH	.0504
	UP	.0777		MP	.0512
				PJ	.1174
KE	AP	.0560		WB	.0511
	MP	.0616			
	MH	.0950	RJ	BH	.0544
	UP	.1129		MP	.0511
	WB	.0568		UP	.1585
				WB	.0508
MP	MH	.0783	TN	AP	.0587
	UP	.1560		BH	.1128
	WB	.0555		GJ	.1193
				MH	.0510
MH	AP	.0528		UP	.0772
	GJ	.0500		WB	.0462
	MP	.0762			
	UP	.1215	TP	UP	.1111
				WB	.1292
MN	UP	.1739			
	WB	.1933			

significant flows beign from the central and southern states of Andhra Pradesh, Madhya Pradesh, Bihar and Gujarat. The overriding trends reveal the convergence to be centered on the states mentioned above as well as Bihar and M.P. These states represent areas containing large urban centers and since the Markov process hinges on the Gravity Model, the interactions and flows converge at the above mentioned states. An examination of the transition matrix (See Table XV in the Appendix) reveals that although the major convergence is towards the states harboring the disease, there are lesser outer flows away from these centers which are not included in the table because of their multitude.

Salient flows of the disease which were mapped for the two states having the highest deaths show that the disease moves from these areas to the outlying states. The selection criteria for mapping the salient flows was $P_{ij} \geq .0126$. This criteria was chosen in order to show why only the more salient flows from the states and not all possible flows. The flows for the five time periods show that the probability of the disease moving increases with each time period. At t_0, U.P. and Orissa have the highest death rates of 107 and 65 deaths respectively. The probability of the disease moving from either of these states to the extreme south or east is between .0126 - .0300 while the probability of the disease moving to states that are closer is greater--the probability being .0301-.0650. In the case of both these states the disease has a fairly high probability of moving into West Bengal. At t_1, the same states have the highest deaths of 147 and 82 respectively. The probability of movement increases from .0126-.0300 to .0301-.0650 to the neighboring states. The chances of the disease moving into West Bengal is greatly increased from .0301-.0625 to .0025-.0826. Similarly, the disease also shows

probability of moving into Tamil Nadu and Kerala from these states. The movement of the disease is from both U.P. and Mysore. For both these states the disease does have a probability of .0301-.0650 of diffusing to West Bengal while the probability lies between .0126-.0300 of diffusing to outer states and .0301-.0650 to neighboring states.

At t_3 when the endemic area of West Bengal has a fairly large death rate the flows are agin of a greater probability to the neighboring states of Orissa, M.P. and A.P. From Jammu and Kashmir, the flows are of a greater probability to the outlying states of Bengal, U.P., Bihar and Assam with a probability of only .0126-.3000 to the neighboring states of Punjab, H.P. and Haryana. This probably relates to the fact that the transition matrix is based on the Gravity Model. For t_4, the movement of the disease from U.P. and W. Bengal show the same tendency of having a greater probability of moving to the neighboring states with a probability of between .0301 to .0650 and a probability of .0126-.0300 to the outlying states.

In interpreting these salient flows, some caution has to be exercised. First, this accounts only for the greater flows although the insignificant flows do have a substantial affect on the migration process and their additive affect is important. The lesser flows are not presented in this chapter because of their number and the interpretation and understanding could become confusing. The diffusion process as revealed through this analysis indicates the effects to some extent of the migration process on the movement of the disease. Although this study has only tried to gain some insight into the movement patterns of cholera as related to the general migration patterns within the country, an extension of this study using actual migration data might reveal the

true nature of the movement (see Figures 17-21).

Summary of Analysis

The Markov Chain technique which was employed in this study performed quite well despite the inherent data requirements and is a fairly useful tool for both an analysis of diffusion as well as for predictive purposes. Applying the predictions for t_4 and t_5 based on the actual death rates for t_3 show again a fairly close correlation between the actual and the predicted deaths.

Despite the data problems and procedural difficulties, the findings of this analysis are valid within the assumptions of the model. To some extent, the results correlate with reality, but within the assumptions of the Markov Chain the predictions are satisfactory. A summary of the findings are listed below:

1) The deaths from cholera are closely related to a general migration pattern for the whole country.

2) The correlations between the actual and the predicted deaths are statistically significant for t_1 and t_2.

3) During t_3 the correlations decrease in significance, probably because this period signifies the onset of the monsoons in the Ganges valley.

4) The predictions for t_3 are for the most part lower than the actual deaths for practically all the states but more so in the states of U.P. and Bengal.

5) Calibrating the model using t_3 as the base vector results in an improvement in the correlation for t_4 and t_5 unlike that derived from using t_0 as the base vector. But the deaths predicted are

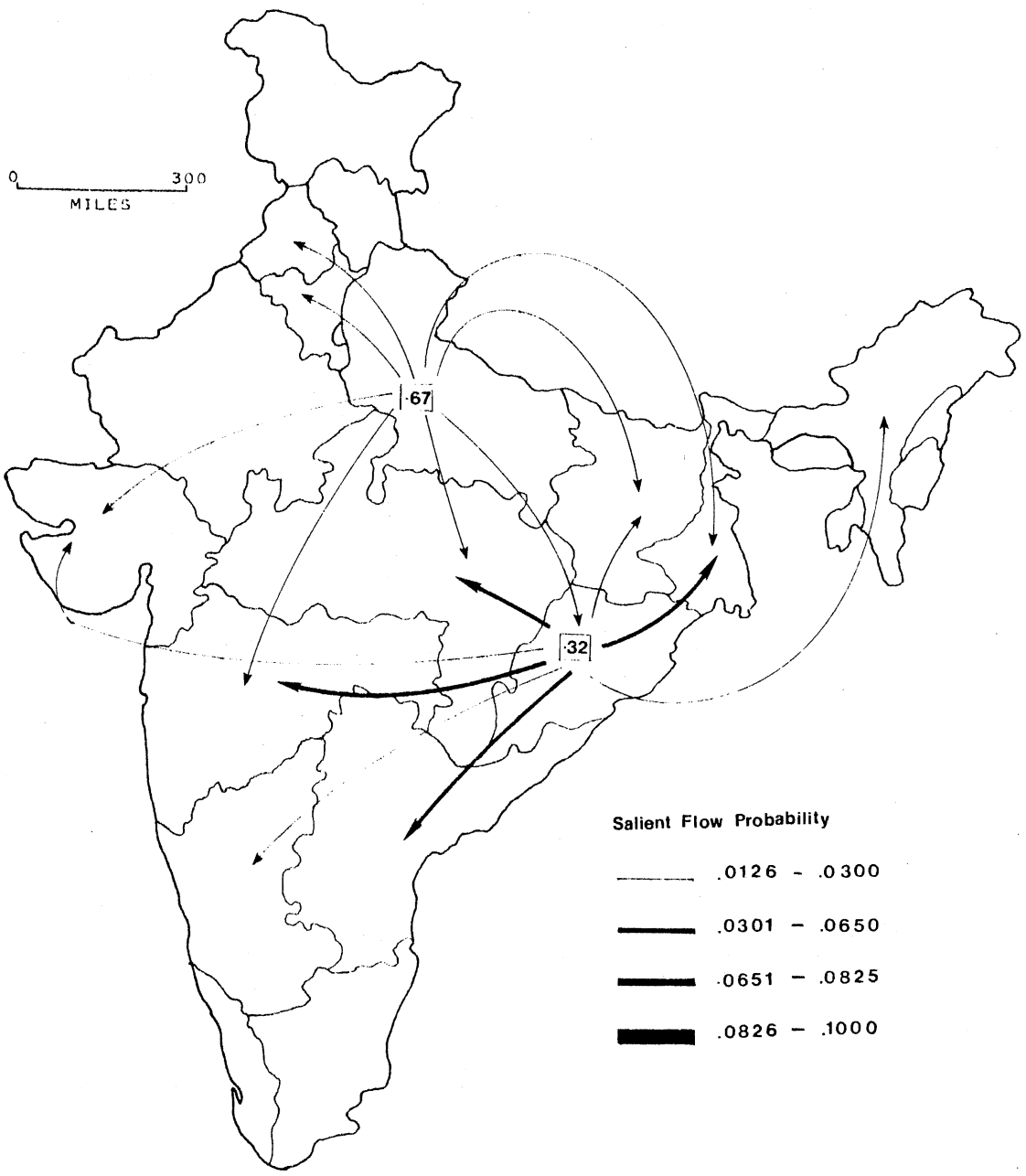


Figure 17. Salient Flows of the Disease for t_0

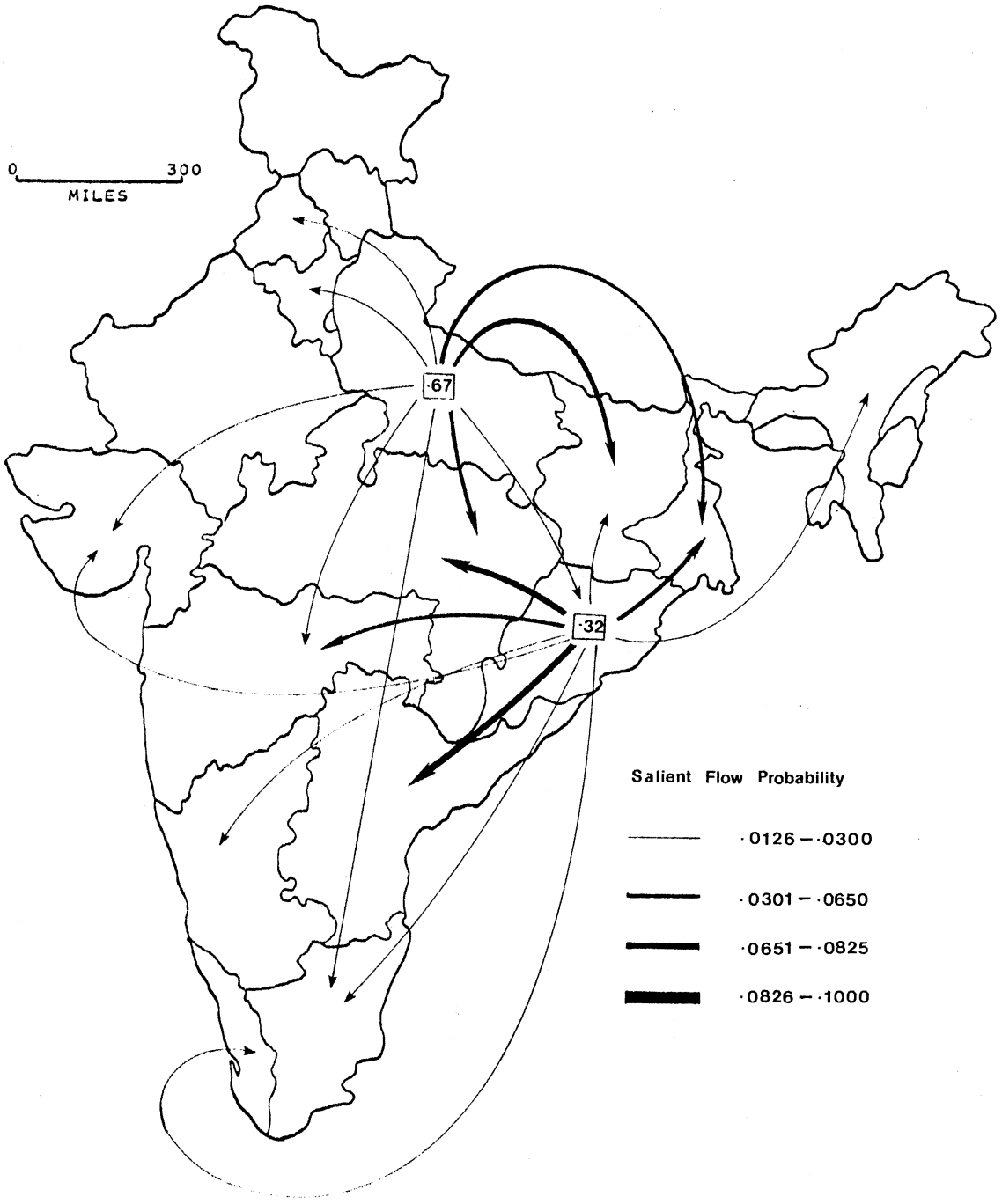


Figure 18. Salient Flows of the Disease for t_1

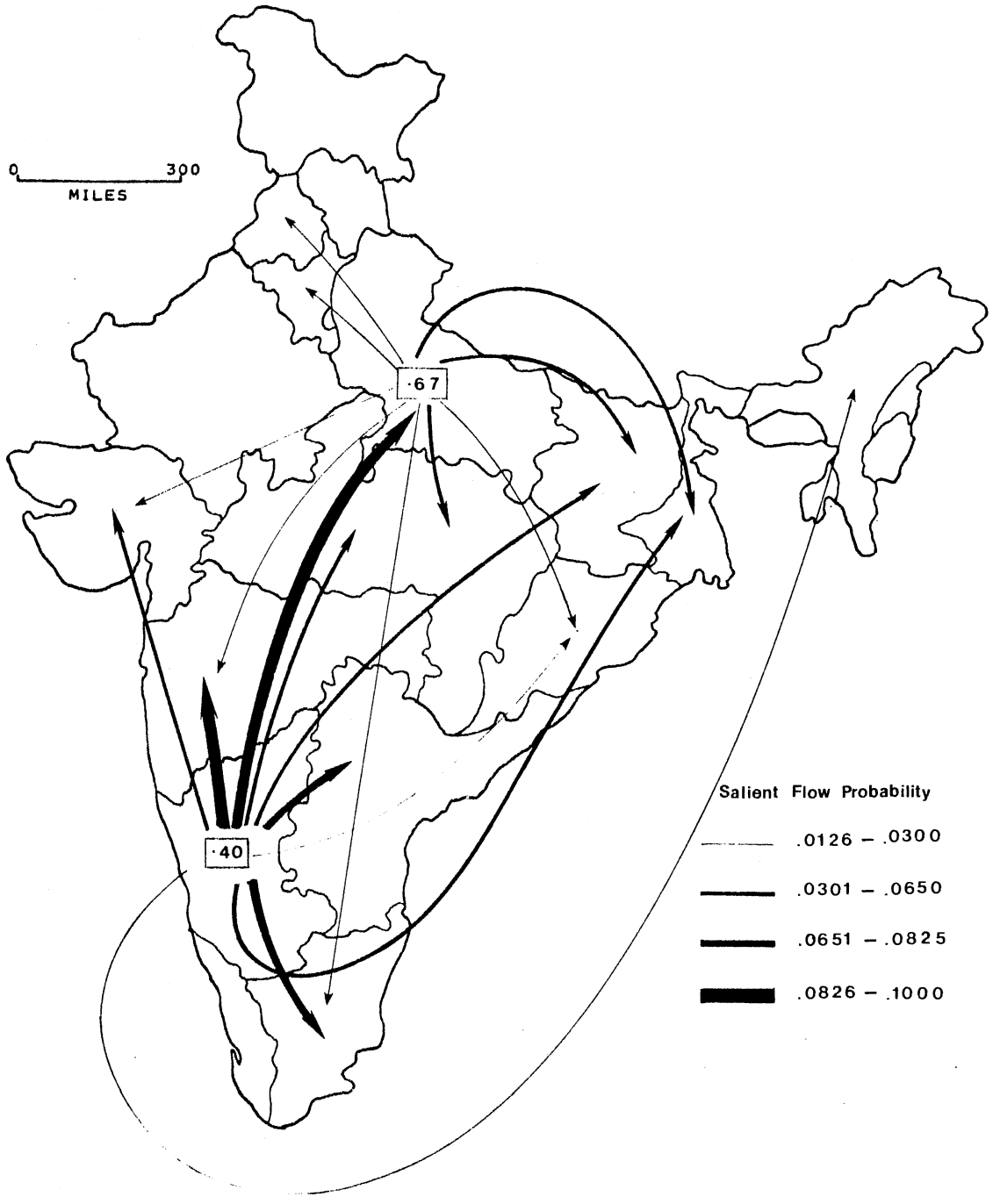


Figure 19. Salient Flows of the Disease for t_2

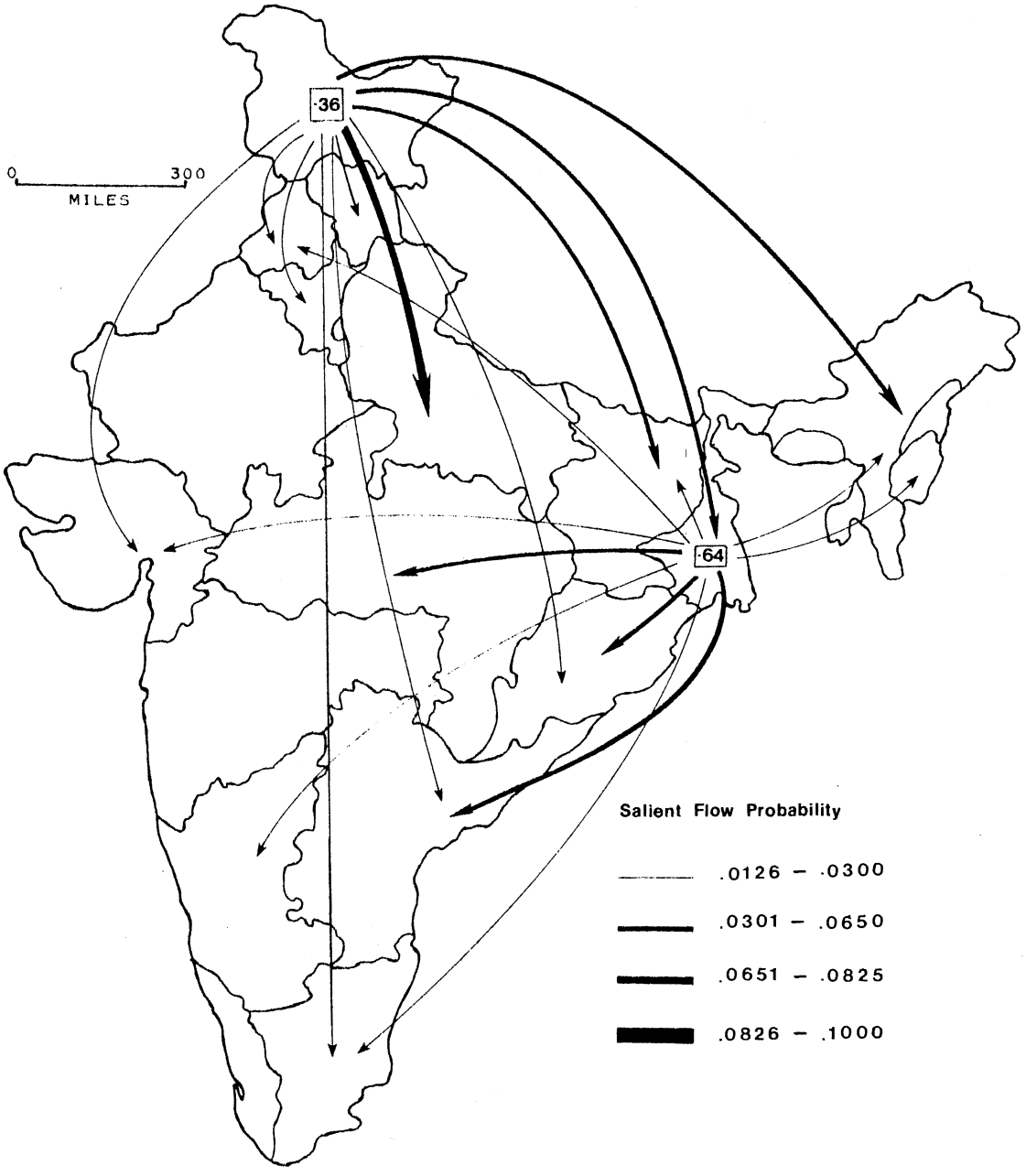


Figure 20. Salient Flows of the Disease for t_3

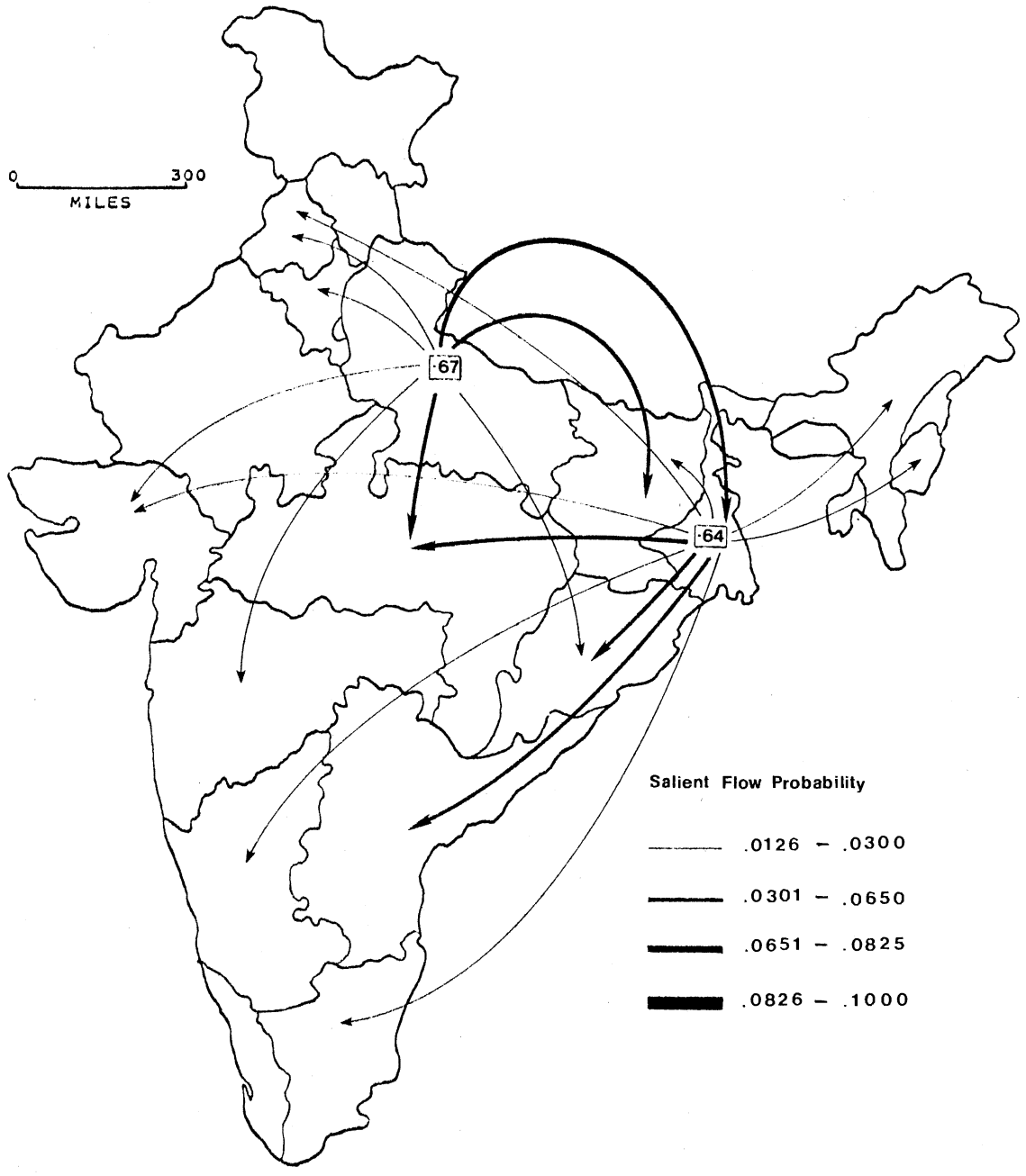


Figure 21. Salient Flows of the Disease for t_4

definitely lower than the actual deaths for the states of U.P. and Bengal.

6) Isolating the salient flows yielded about ten percent of the total flows which explained about 50 percent of the diffusion patterns. The flows converge in the states of U.P., Bengal, Bihar and Madhya Pradesh.

7) From the endemic area in Bengal, the salient flows are of a greater probability to the neighboring states. This holds true for most of the other states of Mysore, and U.P. In the case of Jammu and Kashmir, the flows are of a greater probability to the outlying states of West Bengal and U.P. and of a lesser probability to the neighboring states of Punjab and Haryana.

8) The larger flows are augmented by counter flows back to the ares of a lesser probability.

The findinsg of this research indicate that the hypothesis (refer to Chapter I, p.) can be accepted. That is, the simulated deaths do have a close association with the actual deaths especially for t_1 and t_2 using t_0 as the base vector. This is also true for t_4 and t_5 using t_3 as the base vector. The purpose of the study established in Chapter I was fulfilled as the relationship between distance and time and the effects of a general migration pattern on the diffusion of cholera was established. The findings are valid within the assumptions of the model and are constrained by the assumptions outlined in Chapter IV and are therefore hypothetical.

CHAPTER VI

SUMMARY, IMPLICATIONS AND FURTHER RESEARCH

Summary

The major objective of this thesis was to obtain some insight into the diffusion patterns of cholera. In order to obtain this insight, the movement of the disease was related to a general migration pattern. The chosen year of study was 1971 as the available data were inconsistent and inaccurate prior to 1971. The endemic area of cholera is Bengal. The purpose of the study was to examine the diffusion patterns of cholera from this endemic center to other parts of India. The area of study was the 21 states of India. Information on the actual deaths from cholera for India in 1971 was obtained from the Vital Statistics of India published by the Registrar General of India.

The History of Cholera in India

The history of cholera in India is fascinating as the endemic foci lies in the country and the disease has been known to have existed there since time immemorial. The spread of the disease from this center to both eastern Asia and Europe and thence on to America was discussed in Chapter III. The six pandemics spreading to many different parts of the world was discussed. By the end of the nineteenth century it had been eradicated from most parts of the world due to increased sanitary

facilities but it continued its rampage in India.

The characteristics of the disease render it difficult for the illiterate people of the country to distinguish it from ordinary diarrhea and food poisioning. This resulted in large numbers of people succumbing to the disease. The death rate from cholera has consequently remained high in the country where the incidence of monsoon also acts as a fine medium for the propagation of the disease. The historical trends of the disease were concluded with present day patterns.

Methodology

To investigate the diffusion of cholera within India, the Gravity model was used to generate a migration flow matrix. The matrix was the then used as input to a Markov Chain model where the transition matrix embodies the process of movement.

The Markov Chain model calculates the probability of any phenomenon moving from state i to state j based upon previous time periods. Death rates from cholera for t_0 were incorporated into the model as the base vector. The Markov Chain process then predicts deaths for t_1 through t_5. The year 1971 was divided into six time periods. Each time period coincides with seasonal changes; hence t_0 consists of the months of Jan-Feb; t_1 (March-April), t_2 (May-June), t_3 (July-August), t_4 (Sept-Oct) and t_5 (Nov-Dec). The resulting predictions were then compared to the actual deaths by computing residuals.

Predictions by the Markov Chain Model

The resulting predictions from the Markov process were fairly close to the actual deaths for each time period. The predictions for t_1 and

t_2 were very good. With the onset of the monsoon the process predicts fewer than the actual deaths in the endemic areas. A calibration of the model using t_3 as the base vector instead of t_0 definitely improved the estimation procedure for the two following time periods. The resulting predictions were close to the actual number of deaths.

The salient diffusion flows indicate a convergence towards the states with large urban centers e.g. U.P. and West Bengal. The disease has a greater probability of moving to neighboring states rather than to outlying states. In the case of Jammu and Kashmir the disease shows a greater probability of moving to the outlying states rather than to the neighboring states of Punjab and Haryana. Within the assumptions of the study and the inherent data problems, the Markov process has been able to predict close to the actual deaths.

Implications

Given the above results, the following question arose: To what extent does the role of pilgrimages -- as migration -- influence the movement of the disease within the general pattern of movement?

General Tendencies

Since the discovery of cholera vibrio by Koch and the immunization process, cholera, still takes a high toll of lives in India. The major areas of concentration historically as well as in 1971 are the Ganges-Brahmaputra delta in West Bengal and major pilgrimage areas like U.P. Although it is known that cholera is a highly contagious disease, the extent of its diffusion as related to migration has never been tested. If monthly migration data were available and the movement of the disease

controlled for different time periods, the results of a similar analysis may prove more strongly the extent of migration effects on the disease.

Pilgrimages, which are a part of the Indian life and their effect on the diffusion of cholera would add an interesting dimension. Controlling the effects of the monsoon and then examining the relationship between cholera and migration and vice versa would probably reveal the exact extent of the effect of each factor on diffusion patterns.

Possible Control Measures

It would appear that although cholera has been eradicated from many other parts of the world where it once held sway, it has not been eradicated at its endemic source. This would indicate that control measures that are existent are not sufficient enough to eradicate the disease.

Better cooperation between village doctors and public health departments who treat the patients are needed. Most poor villagers would probably visit the village doctors to be treated for simple diarrhea instead of proper diagnosis and treatment. This would require a better understanding of the general characteristics of the disease by the village doctors. The same goes for trained doctors. Secondly, severe control measures like immunization practices for the entire population, especially the poor and the pilgrims, before the onset of the monsoons, should be implemented. Better understanding of the movement of the disease is necessary to set up public health facilities and sanitary control measures.

Further Research

Cholera has been the subject of intense study by bacteriologists and similar health professionals with concentration focused on the bacteriological characteristics, immunization, vaccines, etc. A spatial and temporal study to identify where the disease is likely to move and where the increases are likely to be would prove more effective in practicing better control measures. With the reappearance of cholera in places where it had been believed to be non-existent like the U.S.A., there is a definite need to study the movement of the diseases and the factors contributing to it. It has been seen in the history of cholera that the disease has reappeared after each pandemic was thought to be over and it is likely that the disease might develop new strains. Further research in the following areas might be significant:

1) A case study of Bengal -- the endemic foci over a longer period with the unit of study being the various districts and the movement of the disease both interstate and intrastate.

2) A study covering the whole of India but over a longer time period, 10-15 years, with both the effects of monsoon and migration controlled for. Moreover, instead of using January-February as t_0, the cycle could be started at the onset of the monsoon and a period of 12 months thence, so that the movement can be traced from the onset of one monsoon to another.

3) The importance of pilgrimages could be traced, which would be one of the best overviews. There is a considerable increase in the deaths after the pilgrimage months as the disease moves rapidly through close contact.

4) The Monte Carlo simulation model could be used to predict death rates and a comparative study conducted with the Markov Chain analysis to define a more accurate method of future predictions.

Each of these avenues of research could contribute to the overall understanding of the spatial aspects of cholera leading to more effective control programs not only for India but for other parts of the world.

A SELECTED BIBLIOGRAPHY

1. Abler, R., et al., Spatial Organization: The Geographer's View of the World. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1971.

2. Ackerknect, Erwin H. History and Geography of the Most Important Diseases. New York: Hafner Publishing Company, Inc., 1965.

3. Bailey, Norman T. J. The Mathematical Theory of Infectious Diseases and Its Applications. New York: John Wiley and Sons, Ltd., 1967.

4. Blumberg, Baruch S. Genetic Polymorphisms and Geographic Variations in Diseases. New York: Grune and Stratton, 1961.

5. Blumen, Isadore, et al, The Industrial Mobility of Labor as a Probability Process. Ithaca: Cornell University Press, 1955.

6. Brown, L. A. Diffusion Processes and Location. Philadelphia: Regional Science Research Institute Bibliography Series, No. 4, 1968.

7. _____. "On the Use of Markov Chain in Movement Research." Economic Geography, Vol. 46 (1970), pp. 393-403.

8. Bryden, J. L. Vital Statistics of the Bengal Presidency. Cholera Epidemics Reviewed in Relation to Former Epidemics: A Record of Cholera in the Bengal Presidency 1817 to 1872. Calcutta, 1874. In R. Politzer, Cholera. Geneva: World Health Organization, 1959.

9. Burkitt, Denis P. "Geographic Distribution." In D. P. Burkitt and D. H. Wright (ed), Burkitt's Lymphoma. Edinburgh: E. S. S. Livingstone, 1970.

10. Burton, Lloyd E. and Hugh Smith. Public Health and Community Medicine. Baltimore: The Williams and Wilkins, Co., 1975.

11. Chorley, K. J. and P. Haggett (eds). Models in Geography. London: Methune and Co. Ltd., 1967.

12. Cockburn, Adrian. The Evolution and Eradication of Infectious Diseases. Baltimore: The Johns Hopkins Press, 1963.

13. Fonaroff, L. S. "Man and Malaria in Trinidad: Ecological Perspectives of a Changing Health Hazard." Annals, Association of American Geographers, Vol. 58 (1968), pp. 526-556.

14. Girt, John L. "The Geography of Non-Vectored Diseases." In John M. Hunter (ed), The Geography of Health and Disease. Chapel Hill, North Carolina: Department of Geography, University of North Carolina, (1974), pp. 46-80.

15. Goodman, L. A. "Statistical Models for the Mover-Stayer Model." Journal of the American Statistical Association, Vol. 56 (1961 (1961), pp. 841-868.

16. Gould, P. Spatial Diffusion. Association of American Geographers, Resource Paper No. 4. Commissions of the College of Geography, Washington, D.C., 1969.

17. Grundy, F. and W. A. Reinke. "Health Practice Research and Formalized Managerial Methods." Public Health Papers, No. 51. Geneva: World Health Organization, 1973.

18. Hagerstrand, T. Innovation Diffusion as a Spatial Process. Chicago: The University of Chicago Press, 1967.

19. Henschen, Folke. The History and Geography of Diseases. New York: Delacorte Press, 1962.

20. Hirsch, August. Handbook of Historical and Geographical Pathology. (3 Vols.) London: New Sydenham Society, 1883 - 1886.

21. Howe, Melvyn G. A World Geography of Human Diseases. London: Academic Press, 1977.

22. _____. Man, Environment and Disease in Britain. New York: Barnes and Noble, 1972.

23. Hunter, John M. (ed). "The Geography of Health and Disease." Papers of the First Carolina Geographical Symposium. Chapel Hill: University of North Carolina, Department of Geography, 1974.

24. _____. "River Blindness in Nangodi, Northern Ghana: A Hypothesis of Cyclical Advance and Retreat." The Geographical Review, Vol. 56 (1966), pp. 398-416.

25. _____, and J. C. Young. "Diffusion of Influenza in England and Wales." Annals, Association of American Geographers, Vol. 61 (1971), pp. 637-653.

26. Isard, W. Methods of Regional Analysis. Cambridge, Massachusetts: M.I.T. Press, 1960.

27. Kemeny, J. G. and J. C. Snell. Finite Markov Chains. Princeton, N. J.: Van Nostrand Co., 1960.

28. Knight, C. G. "The Geography of Vectored Diseases." In John M. Hunter's The Geography of Health and Disease. Chapel Hill, North Carolina: Department of Geography, University of North Carolina, (1974), pp. 46-80.

29. _____. "The Geography of Sleeping Sickness." Annals, Association of American Geographers, Vol. 61 (1971), pp. 23-24.

30. Kwofie, K. M. "A Spatio-Temporal Analysis of Cholera Diffusion in Western Africa." Economic Geography, Vol. 52 (1976), pp. 127-135.

31. Lauver, Gary D. A Stochastic Analysis of the Internal Mobility with Population Projections for the State Economic Areas of Texas. (Unpublished M.S. thesis, Oklahoma State University, Stillwater, 1977.)

32. Learmonth, A. T. A. Patterns of Disease and Hunger: A Study in Medical Geography. North Pomfret, Vermont: David and Charles, Inc., 1978.

33. _____. "Some Contrasts in the Regional Geography of Malaria in India and Pakistan." Transactions of the Institute of British Geographers, Vol. 23 (1957), pp. 37-39.

34. Lenczer, M. and T. Owen, "The Impact of Tropical and Parasitic Diseases in a Non-Endemic Area." Canadian Medical Association Journal, Vol. 82 (1960), pp. 805-811.

35. le Riche, W. M. and J. Milner. Epidemiology as Medical Ecology. Baltimore: The Williams and Wilkens Co., 1971.

36. McGlashan, N. D. "The Medical Geographer's Work." International Pathology, Vol. 7 (1966), pp. 81-83.

37. _____. "The Scope of Medical Geography." South African Geographical Journal, Vol. 47 (1965), pp. 35-40.

38. McKinley, Earl B. A Geography of Disease. Washington: George Washington University Press, 1935.

39. McNamara, C. "A History of Asiatic Cholera." (1874) In R. Pollitzer, Cholera. Geneva: World Health Organization, 1959.

40. May, Jacques M. "Medical Geography: Its Methods and Objectives." The Geographical Review, Vol. 40 (1950), pp. 9-41.

41. _____. The Etiology of Human Disease. New York: M. D. Publications, 1958.

42. _____. Studies in Disease Ecology. New York: Hafner Publishing Company, 1961.

43. Mitchell, Robert L. "Trace Elements in Soil and Factors that Affect Their Availability." In Helen C. Cannon and Howard C. Hopps (eds). Geochemical Environment in Relation to Health and Disease. Boulder: The Geological Society of America, (1972), pp. 4-16.

44. Momiyama, M. S. Seasonality in Human Mortality. Tokyo: University of Tokyo Press, 1977.

45. Morrison, P. A. "Theoretical Issues in the Origin of Population Mortality Models." Environment and Planning, Vol. 5 (1973), pp. 125-134.

46. Pollitzer, R. Cholera. Geneva: World Health Organization Monograph Series No. 43, 1959.

47. Prais, J. "Measuring Social Mobility." Journal of the Royal Statistical Society, Series A, Vol. 118 (1955), pp. 55-56.

48. Prothero, Mansell R. Migrants and Malaria. London: Longmans, Green and Co., Ltd., 1965.

49. Pullen, Roscoe L. Communicable Diseases. Philadelphia: Lea and Feibiger, 1950.

50. Pyle, Gerald F. Applied Medical Geography. New York: John Wiley and Sons, 1979.

51. _____. "The Diffusion of Cholera in the United States in the Nineteenth Century." The Geographical Analysis, Vol. 1 (1969), pp. 59-75.

52. _____. "Disease Ecological Aspects of Rocky Mountain Spotted Fever." In Gerald F. Pyle Applied Medical Geography. New York: John Wiley and Sons, 1979.

53. The Registrar General of India. Vital Statistics of India, 1971. New Delhi: Ministry of Home Affairs (Vital Statistics Division), 1976.

54. Rodenwalt, E. and Helmut Jusatz (eds). Welt-Seuchen Atlas (World Atlas of Epidemics), 3 Vols. Hamburg: Falk, 1952 - 1961.

55. Rogers, Andre. "A Markovian Policy Model of Interregional Migration." Regional Science Association, Vol. 17 (1966), pp. 205-224.

56. Rowntree, Seebohm, B. Poverty, A Study of Town Life. London: Macmillan, 1901.

57. Salkin, M. S. et al. Applications of the Markov Chain Analysis. Stillwater: Ag Experiment Station Research Report, Oklahoma State University, p. 707.

58. Scheil, Joseph B. and A. J. Wepfer. "Distribution of Endemic Goiter in the United States." Economic Geography, Vol. 52 (1976), pp. 116-126.

59. Simmons, James S. et al. Global Epidemiology. (3 Vols.) Philadelphia: J. B. Lipincott. 1944 - 1954.

60. Simon, H. A. and C. P. Bonini. "The Size Distribution of Business Firms." American Economic Review, Vol. 48 (1958), pp. 607-617.

61. Smith, David M. Patterns of Human Geography. New York: Crane, Russack and Co., Inc., 1975.

62. Sorre, Maximilian. "Principes Generaux de la Geographie Medicale." In G. Pyle, Applied Medical Geography. New York: John Wiley and Son, 1979.

63. Sticker, G. "Abhandlungen aus der Seuchengeschite und Seuchenlehre. Band: Die Cholera." In R. Pollitzer, Cholera. Monograph Series No. 43. Geneva: World Health Organization, 1959.

64. Stock, Robert. Cholera in Africa. London: International African Institute, 1976.

65. Traver, J. D. and R. Gurley. "A Stochastic Analysis of Geographic Mobility and Population Projections of the Census Division in the U.S." Demography, Vol. 4 (1965), pp. 134-139.

66. Verhasselt, V. Maps on Cancer Distribution. Brussels: Vrige Universitiet Brussel, Geografisch Instituut, 1975.

67. Voronov, A. G. "The Geographical Environment and Human Health." Soviet Geography, Vol. 18 (1977), pp. 230-237.

68. World Health Organization. The First Ten Years of the World Health Organization. Geneva: World Health Organization, 1958.

69. Yeates, Maurice. An Introduction to Quantitative Analysis in Human Geography. New York: McGraw-Hill Book Co., 1974.

TABLE XIV

1971 POPULATION BY STATE

State	Total Population
AP	43,502,708
AS	14,625,152
BI	56,353,369
GJ	26,697,475
HY	10,036,808
HP	3,460,434
JK	4,616,632
KE	21,347,375
MP	41,654,119
MH	50,412,235
MN	1,072,753
ME	1,011,699
MY	29,299,014
MG	516,449
OR	21,944,615
PJ	13,551,060
RJ	25,765,806
TN	41,199,168
TP	1,556,342
VP	88,341,144
WB	44,312,011

Source: The Registrar General of India. Vital Statistics of India, 1971. New Delhi: Vital Statistics Division, Ministry of Home Affairs, 1976, p. 37.

Figure 22. Center of State Used in the Gravity Model

TABLE XV

FUNDAMENTAL TRANSITION MATRIX DERIVED FROM
THE MARKOV CHAIN TECHNIQUE

State	AP	AS	BI	GJ	HY	HP	JK	KE	MP	MH	MN	ME	MY	NG
AP	0.1119	0.0326	0.0537	0.0600	0.0285	0.0130	0.0047	0.0189	0.0766	0.0669	0.0084	0.0014	0.0448	0.0006
AS	0.0520	0.1030	0.0554	0.0463	0.0291	0.0127	0.0069	0.0161	0.0882	0.0618	0.0092	0.0018	0.0304	0.0017
BI	0.0625	0.0391	0.0607	0.0549	0.0336	0.0139	0.0052	0.0172	0.0738	0.0638	0.0107	0.0017	0.0343	0.0003
GJ	0.0644	0.0354	0.0567	0.0671	0.0373	0.0127	0.0053	0.0181	0.0788	0.0728	0.0088	0.0015	0.0342	0.0006
HY	0.0492	0.0305	0.0570	0.0492	0.1125	0.0119	0.0057	0.0158	0.0700	0.0639	0.0074	0.0012	0.0291	0.0006
HP	0.0670	0.0371	0.0527	0.0534	0.0338	0.0221	0.0056	0.0174	0.0749	0.0630	0.0104	0.0017	0.0346	0.0007
JK	0.0578	0.0433	0.0576	0.0522	0.0436	0.0137	0.0117	0.0171	0.0768	0.0661	0.0090	0.0015	0.0314	0.0007
KE	0.0703	0.0371	0.0543	0.0565	0.0351	0.0136	0.0053	0.0258	0.0797	0.0729	0.0086	0.0015	0.0335	0.0007
MP	0.0626	0.0460	0.0538	0.0528	0.0329	0.0135	0.0053	0.0177	0.0891	0.0705	0.0087	0.0015	0.0333	0.0007
MH	0.0683	0.0369	0.0569	0.0572	0.0351	0.0130	0.0051	0.0187	0.0802	0.0765	0.0087	0.0015	0.0335	0.0007
MN	0.0592	0.3413	0.0512	0.0486	0.0309	0.0155	0.0050	0.0162	0.0741	0.0581	0.0187	0.0017	0.0323	0.0007
ME	0.0758	0.0360	0.0582	0.0572	0.0307	0.0121	0.0049	0.0165	0.0759	0.0739	0.0084	0.0064	0.0600	0.0006
MY	0.0771	0.0355	0.0583	0.0592	0.0330	0.0124	0.0049	0.0175	0.0776	0.0731	0.0086	0.0015	0.0440	0.0006
NG	0.0529	0.0811	0.0577	0.0462	0.0292	0.0134	0.0062	0.0160	0.0789	0.0608	0.0100	0.0018	0.0302	0.0084
OR	0.0721	0.0376	0.0523	0.0525	0.0338	0.0168	0.0052	0.0179	0.0786	0.0663	0.0098	0.0016	0.0354	0.0007
PJ	0.0610	0.0340	0.0574	0.0531	0.0462	0.0132	0.0057	0.0171	0.0755	0.0666	0.0091	0.0014	0.0327	0.0006
RJ	0.0611	0.0344	0.0576	0.0532	0.0455	0.0132	0.0057	0.0172	0.0756	0.0666	0.0092	0.0015	0.0328	0.0006
TN	0.0713	0.0352	0.0621	0.0660	0.0325	0.0124	0.0050	0.0179	0.0755	0.0703	0.0092	0.0015	0.0376	0.0006
TP	0.0637	0.0437	0.0604	0.0531	0.0311	0.0142	0.0053	0.0173	0.0754	0.0631	0.0110	0.0017	0.0342	0.0008
VP	0.0498	0.0340	0.0582	0.0479	0.0453	0.0119	0.0049	0.0158	0.0724	0.0625	0.0089	0.0014	0.0296	0.0006
WB	0.0713	0.0463	0.0466	0.0509	0.0247	0.0185	0.0054	0.0177	0.0790	0.0580	0.0128	0.0021	0.0362	0.0008

TABLE XV (Continued)

State	OR	PJ	RJ	TN	TP	VP	WB
AP	0.0386	0.0558	0.0045	0.0581	0.0031	0.1898	0.1277
AS	0.0334	0.0504	0.0042	0.0401	0.0032	0.2195	0.1347
BI	0.0355	0.0567	0.0046	0.0504	0.0032	0.2254	0.1521
GJ	0.0347	0.0612	0.0048	0.0456	0.0031	0.2348	0.1220
HY	0.0305	0.0670	0.0050	0.0410	0.0029	0.2501	0.0992
HP	0.0460	0.0568	0.0045	0.0487	0.0032	0.2083	0.1580
JK	0.0343	0.0738	0.0048	0.0444	0.0031	0.2332	0.1241
KE	0.0367	0.0583	0.0047	0.0477	0.0031	0.2280	0.1265
MP	0.0362	0.0564	0.0046	0.0435	0.0031	0.2413	0.1262
MH	0.0356	0.0580	0.0047	0.0480	0.0031	0.2322	0.1259
MN	0.0398	0.0554	0.0045	0.0451	0.0034	0.2289	0.1693
ME	0.0346	0.0545	0.0045	0.0550	0.0030	0.2071	0.1245
MY	0.0352	0.0567	0.0046	0.0535	0.0031	0.2197	0.1238
NG	0.0340	0.0513	0.0043	0.0423	0.0032	0.2209	0.1512
OR	0.0434	0.0572	0.0046	0.0465	0.0032	0.2244	0.1402
PJ	0.0345	0.0664	0.0048	0.0456	0.0031	0.2474	0.1243
RJ	0.0346	0.0662	0.0048	0.0457	0.0031	0.2458	0.1252
TN	0.0341	0.0568	0.0046	0.0560	0.0031	0.2185	0.1296
TP	0.0364	0.0554	0.0045	0.0494	0.0038	0.2196	0.1558
VP	0.0320	0.0619	0.0049	0.0416	0.0031	0.2966	0.1166
WB	0.0436	0.0527	0.0042	0.0465	0.0035	0.1755	0.2336

VITA[2]

Sujata Warrier Mitra

Candidate for the Degree of

Master of Science

Thesis: A STOCHASTIC ANALYSIS OF THE DIFFUSION OF CHOLERA IN INDIA IN 1971

Major Field: Geography

Biographical:

Personal Data: Born in Trivandrum, India, March 1, 1955, the daughter of Mr. and Mrs. G. S. Warrier.

Education: Graduated from St. John's Diocesan Girls' High School, Calcutta, India, in July, 1973; received Bachelor of Arts degree in Geography from University of Calcutta, Calcutta, India, in December, 1976; completed the requirements for the Master of Science degree at Oklahoma State University in May, 1980.

Professional Experience: Graduate teaching assistant, Department of Geography, Oklahoma State University, 1978-79.

Professional Organizations: Oklahoma Academy of Science; Southwest Social Science Association.