THE UNIVERSITY OF CALGARY

SPATIAL AND TEMPORAL TRENDS IN HUMAN MORTALITY IN NEWFOUNDLAND, 1937 - 1971

by

ERIC HUGH LARSON

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF GEOGRAPHY

CALGARY, ALBERTA

JUNE, 1985

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Spatial and Temporal Trends in Human Mortality in Newfoundland, 1937-1971" submitted by Eric Hugh Larson in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor
Dr. Nigel M. Waters
Department of Geography

Dr. Lynn A. Rosenvall
Department of Geography

Dr. Herman W. Konrad
Department of Anthropology

June 4, 1985

ABSTRACT

Patterns of cause of death underwent significant change in Newfoundland between 1937 and 1971. Using the International Classification of Diseases (I.C.D.) chapter structure to observe that change in twenty-one electoral districts, it is possible to reconstruct, spatially and temporally, the transition from a pattern of mortality dominated by infectious diseases and child mortality to a pattern dominated by the chronic and degenerative diseases characteristic of an older, "modern" population. This transformation has been characterized by Omran (1971) as an "epidemiologic transition," and his model of that transition is used as a framework for analysis and discussion. The results of canonical correlation analysis and a space-time cluster analysis indicate that a regionally homogenous "pre-modern" pattern of mortality was replaced, between 1946 and 1961, by a "modern" pattern of mortality. That transformation was expressed spatially, with some districts showing modern patterns as early as 1946, while others lagged behind. After 1961, a decline in spatial variation in mortality pattern was observed. The model of epidemiologic transition is tentatively linked to certain models of economic development which are resonant

with an ergodic hypothesis of the interchangeability of space and time.

ACKNOWLEDGEMENTS

The guidance provided by three men in particular has led me to the completion of this thesis. Dr. Nigel Waters, my friend and supervisor, led me through the entire project with encouragement, sound advice and good humor. Dr. Simon Evans of Memorial University encouraged me to study in Newfoundland, gave freely of his time while in Calgary and provided help in many ways from far away Corner Brook. Dr. Phil Gersmehl of the University of Minnesota had nothing to do with this project directly, but everything to do with it indirectly. Others to whom I owe important intellectual debts and whose influence somehow led me here include the late Dr. Bernard Bowron, Dr. Toni McNaron, Dr. Herman Konrad, Mr. Rick Jackson and especially my parents, Marvin and Jean Larson. No one could have hoped for a finer group of teachers and mentors.

Thanks are also owed to-

-Jan Clark, Carey Johannesson and "the guys," who were gracious hosts to me in St. John's during a very hectic time in their lives.

-Ms. Beth Saunders of the Vital Statistics Division of Statistics Canada in Ottawa, whose assistance in data collection was invaluable.

-Dr. Michael Staveley of Memorial University, who pointed out some very important things to me.
-My friends around the Geography Department, especially Joyce Rowden and Karen Marion.
-Nigel McHugh and David Greene, who have been superb friends in every way.
-and most of all, Mary, who kept me going from a long way away with phone calls, care packages and love.

To my parents

TABLE OF CONTENTS

Chapter Page

1. Introduction

1.1 Introduction 1
1.2 Outline of the study 3

2. Background to the Study- Medical Geography

2.1 Introduction 6
2.2 The theory of epidemiologic transition 8
2.2.1 Outline of the theory of epidemiologic transition 10
2.2.2 Measuring the epidemiologic transition . 14
2.3 Mapping mortality. 17
2.3.1 Techniques for disease mapping 19
2.4 Canadian geographic mortality studies. 22

3. Background to the Study- Demographic Change in Newfoundland

3.1 Introduction 24
3.2 Population growth and patterns of resource exploitation 25
3.2.1 Summary. 33
3.3 Recent mortality trends. 35
3.3.1 Infant mortality 35
3.3.2 Maternal mortality 36
3.3.3 The age structure of mortality 39
3.4 Changes in cause of death patterns 42
3.5 Conclusion 42

4. Data Collection and Analysis

4.1 Introduction 45
4.2 Data sources 46
4.2.1 Areal units. 47
4.2.2 Data quality and limitations 50
4.3.3 Organization of the data for analysis. . 54
4.3 Methods for the analysis of the data 56
4.3.1 Canonical correlation analysis 57
4.3.1.1 Data considerations. 58
4.3.1.2 The canonical correlation procedure. 62

4.3.1.3 Interpreting canonical correlation analysis 65
4.3.1.4 Summary- canonical correlation analysis 74
4.3.2 Cluster analysis 76
4.3.3 Discriminant analysis. 79
4.4 Summary. 81

5. Results of Data Analysis

5.1 Introduction 82
5.2 Canonical correlation analysis of the cause of death data. 82
5.2.1 Objectives for the canonical correlation analysis 83
5.2.2 Results of the canonical correlation analysis 86
5.2.3 Interpretation of the canonical correlation analysis 87
5.3 Cluster analysis of cause of death patterns. .100
5.3.1 Objectives for the cluster analysis. . .101
5.3.2 Results of the cluster analysis. 102
5.3.2.1 The seven cluster solution . . .106
5.3.2.2 The four cluster solution. . . .111
5.3.2.3 The three cluster solution . . .115
5.3.2.4 The two cluster solution 117
5.4 Conclusions. 119

6. Discussion and Conclusions

6.1 Introduction 123
6.2 Epidemiologic transition as a temporal process. 124
6.3 The epidemiologic transition in space. 127
6.4 Epidemiologic transition in space-time 128
6.5 Limitations of the study 130
6.6 Conclusions. 132

REFERENCES. 135

LIST OF TABLES

Table Page

4.1 Chapter structure of the International Classification of Diseases. 52

5.1 Results of canonical correlation, 1937 - 1941 . . 88

5.2 Results of canonical correlation, 1941 - 1946 . . 89

5.3 Results of canonical correlation, 1946 - 1951 . . 90

5.4 Results of canonical correlation, 1951 - 1956 . . 91

5.5 Results of canonical correlation, 1956 - 1961 . . 92

5.6 Results of canonical correlation, 1961 - 1966 . . 93

5.7 Results of canonical correlation, 1966 - 1971 . . 94

5.8a Cluster profiles, seven cluster solution 108

5.8b Cluster membership, seven cluster solution . . .108

5.9a Cluster profiles, four cluster solution. 112

5.9b Cluster membership, four cluster solution. . . .112

5.10a Cluster profiles, three cluster solution. . . .116

5.10b Cluster membership, three cluster solution. . .116

5.11a Cluster profiles, two cluster solution. 118

5.11b Cluster membership, two cluster solution. . . .118

LIST OF FIGURES

Figure	Page
3.1 Population of Newfoundland, 1750 - 1979	30
3.2 Infant mortality rates, 1921 - 1970	37
3.3 Maternal mortality rate, 1936 - 1971.	38
3.4 Age structure of total mortality, 1937 - 1981 . .	40
3.5 Cause structure of total mortality, 1937 - 1971 .	43
4.1 Electoral districts used in the study	48
4.2 Layout of data matrices	55
5.1 Dendrogram of the general cluster solution.	103-104
5.2 Error sums of squares of the cluster solution . .	107

CHAPTER ONE

Introduction

1.1 Introduction

The notion that geographic variation in patterns of disease can be used to understand something of the relationship between humans and their environment is one which dates at least to the time of Hippocrates. His "Airs, Waters and Places" was used as a medical text until well into the nineteenth century when the germ theory of disease began to dominate medical thinking (Dubos, 1980).

> ...for two thousand years, the physicians under ["Airs, Waters and Places'"] influence regarded it as a law of nature that, just as every country posesses its own plant and animal kingdoms, in the same way it possesses a characteristic disease kingdom (Dubos, 1980, p. 39).

The germ theory of disease quickly became the dominant understanding of disease in individuals since that theory was very successful in systematically leading researchers to successful treatments and cures for disease. The spatial analysis of patterns of disease diminished in importance as medical science focused increasingly on treating and preventing specific diseases in individuals.

In the latter part of this century there has been a renewed interest in the spatial analysis of disease

patterns and disease environments. Interest has come from medical geographers (May's (1958) work in describing a multitude of disease ecologies was fundamental to the creation of modern medical geography), and also from demographers and epidemiologists. As medicine turns more of its attention towards environmental links to health and disease it is inevitable that some of that effort will go into the comparison of rates of disease and death both interregionally and intraregionally. New and better atlases of disease and mortality are being published continously. Recent examples include a series of atlases of Canadian cancer mortality, general mortality and urban mortality published by Statistics Canada (1980a, 1980b, 1984). The existence of disease mapping sections at the Centers for Disease Control in Ottawa and Atlanta also suggests the utility of spatial analysis of disease patterns (Nisen, 1980). Disease mapping, which began with John Snow's map of cholera in London (Howe, 1970; Pyle, 1979), has reached new levels of both graphical and statistical sophistication. The ancient idea of "disease kingdoms" and the demonstrated utility of that concept, however, remains the driving force behind modern studies of geographical variation in disease incidence.

If the dimension of time is added to a study of spatial variation in disease patterns, it is possible to

describe, in terms of health, the underlying structure of demographic change. Space, or perhaps more accurately, spatial variation, becomes a measure of historical process (Johnston, 1983; Blaut, 1961; Harvey, 1969). In this study, space is used to measure the rate, type, and direction of change in the disease environment as indicated by demographic variables. When overall patterns of disease and death are considered in space-time in this way, it is important to recognize that the concern is not with explaining disease in any individual; the germ theory of disease has proved far more useful in that arena. Instead, it is expected that the changing structure of disease patterns over time and through space will show something of the variation in the relationship between humans and their environment in those dimensions.

1.2 Outline of the study

Preliminary investigation of recent Canadian mortality trends suggested that the pattern of human mortality in Newfoundland had undergone significant change in the past fifty years. A cause of death pattern dominated by infant and child mortality and infectious diseases was replaced rapidly by a pattern dominated by chronic and degenerative disease and over-65 mortality. The goal of this study was

to investigate and describe the temporal and spatial nature of that transformation.

As background to the study, the relevant literature from medical geography, population geography and demography is reviewed in Chapter Two. Omran's (1971; 1977) model of epidemiologic transition is also discussed. That model describes the mortality transition often associated with the more all-encompassing (and somewhat problematical) theory of demographic transition. Omran's model is used throughout the study as a framework for the analysis and discussion of temporal trends in cause of death patterns in Newfoundland.

Chapter Three includes a brief review of the population history of Newfoundland and a discussion of recent changes in mortality patterns there. Temporal trends in infant mortality, maternal mortality and the age structure of mortality are reviewed in some detail. Finally, regional trends in cause of death patterns are considered. It is clear, from this review, that a shift in the mortality pattern of the type outlined by Omran's model occurred in Newfoundland during the twentieth century.

The remainder of the thesis is concerned with an analysis of the spatial, and spatial-temporal, expression of epidemiologic transition. The data used in the analysis is described in Chapter Four. Canonical correlation and

cluster analysis were the multivariate procedures employed in the analysis of the data; the use of those techniques is also outlined in Chapter Four. Chapter Five is a presentation of the statistical analysis completed for the study, including formal interpretation of the results.

Finally, Chapter Six is a discussion of the broader implications of the study. In general, it was found that epidemiologic transition was not a purely temporal process in Newfoundland. The tranformation of the mortality pattern was expressed spatially in both a clear and significant fashion. Since the analysis was largely exploratory, this study must be viewed as essentially descriptive and empirical. Given its descriptive nature, it would be inappropriate to draw theoretical conclusions from this study, and none is drawn. The study can serve, however, as a useful guide to similar studies which may eventually lead to clear theoretical statements linking geography and the study of population dynamics.

CHAPTER TWO

Background to the Study- Medical Geography

2.1 Introduction

The work of Ralph Audy (1954, 1971) and Melinda Meade (1976, 1977) suggests that patterns of disease and health can be used as indicators of the underlying structure of human-environment-culture relationships. These relationships are expressed, partially, in population dynamics. As Meade noted,

> Basic to [Audy's (1954, 1971)] work is a positive conception of health as adaptability rather than as the "absence of disease." ...when health is considered as adaptability, the concept of disease is transformed from that of a biological entity... it becomes a measure of the maladaptive interactions among the familiar triad of population, environment and culture. (Meade, 1977, p. 383)

If Audy's line of analysis is accepted, disease patterns may show the outlines of both the successful and unsuccessful interactions taking place between a population, its environment and its culture.

Further evidence that change in disease patterns can measure social, cultural, and environmental change has been offered by Omran (1977). In proposing a theory of

epidemiologic transition (described below), he indicates that change in the age and cause structure of mortality associated with the demographic transition in the West has been driven primarily by social, and not medical, forces. Examination of the mortality transition in the United States from a pattern of cause of death dominated by infectious disease to one dominated by chronic disease showed that the transition was well underway before advances in medicine had had much impact. The transition was found to be correlated with the changes in social and economic structure usually referred to as economic development.

Historical data on the age and cause structure of death, despite the many difficulties with such data (see Chapter 4), constitute one of the best available data bases for understanding and reconstructing past patterns of population-environment-culture interaction. If this data is analysed spatially, as well as temporally, it becomes possible to study the changing regionalization of patterns of death. If the idea is accepted that change in disease patterns reflects social change as well as epidemiologic change, a spatial analysis of those patterns will show, to some extent, the changing regionalization of social and environmental change.

In this study, spatial-temporal trends in mortality in Newfoundland between 1937 and 1971 are described based on analysis of demographic and cause of death data. While the primary goal of this thesis is to describe those trends accurately, an underlying theme is that shifts in population dynamics reflect shifts in the structure of human-environment relationships.

2.2 The theory of epidemiologic transition

The theory of epidemiologic transition proposed by Omran (1971) is related to the theory of demographic transition and perhaps can be understood as a subset of demographic transition theory. Analysis of the changing structure of cause of death patterns is the focus of epidemiologic transition theory whereas the concern of demographic transition theory has been the drop in death rates and lagged drop in fertility rates which has characterized the recent population history of the West (Trewartha, 1969).

Omran noted (1971, p. 10) that it was some of the perceived weaknesses in demographic transition theory which led him to the epidemiological transition approach to population dynamics. Demographic transition theory (also known as Thompson-Notestein theory) has been the subject of controversy in the literature for many years. The source

of that controversy is demographic transition theory's inability to explain both the lagged drop in fertility observed in the West and the forces which cause a population to move from one stage to another. Consequently, much of the recent discussion (usefully summarized by Woods (1982)) has focused on explaining fertility transtion. The study of the epidemiologic transition, on the other hand, involves looking at the structure of mortality transition and is, perhaps, a necessary precursor to understanding the forces behind fertility transition and the movement from one demographic stage to another.

In looking for a deeper understanding of the history of population dynamics, Omran turned to epidemiology, the study of disease and death in populations. Epidemiologic history, when analysed within an appropriate historical framework, can lead to a useful description of population dynamics in a particular place and, it is hoped, shed some light on the forces which create demographic change.

> Inasmuch as patterns of health and disease are integral components of population change, epidemiology's reservoir of knowledge about these patterns and their determinants in population groups serves not only as a basis for prediction of population change but also as a source of hypotheses that can be further tested to correct, refine and build population theory. (Omran 1971, pp. 509-510)

One of the goals of this study is to explore the use of a spatial-temporal model of epidemiologic transition. The results of analysis completed for this study (presented in Chapter 5) show that spatial analysis, at least at the electoral district level of areal aggregation, can be useful for understanding the history of epidemiologic and demographic change in Newfoundland.

2.2.1 Outline of the theory of epidemiologic transition

Epidemiologic transition theory holds that, over a long period of time, a poulation's cause of death structure shifts from being dominated by infectious disease to being dominated by chronic and degenerative disease (Omran, 1971; Omran, 1977; Pyle, 1979). That this type of transition has taken place in the West is well understood in a general sense. It is often discussed in writings on demographic transition theory and is often implicit in other discussions (Trewartha, 1969; Woods, 1982). What has not been so clearly understood in the past is the exact nature of that shift and its value in helping to create a more comprehensive understanding of population dynamics.

Omran suggests that the mortality transition can be divided into three stages.

1) An age of pestilence and famine characterized by very high and fluctuating mortality rates.

Life expectancies are between twenty and forty years. Infectious and parasitic diseases are the leading causes of death.

2) An age of receding pandemics characterized by a decline in mortality rates (and substantially less fluctuation). Fertility rates remain high. Mortality rates are progressively less influenced by epidemics and famines. Life expectancies increase to 30 to 50 years. Early and late phases of this stage are recognized. In the late phase in particular, childhood diseases decline as important causes of death as does tuberculosis. Infectious diseases are still the leading cause of death, but cancer, heart disease and other diseases characteristic of an older and more urban population become more important as causes of death.

3) An age of degenerative and man-made diseases characterized by continued declines in mortality rates, low fertility and very significant improvement in the life expectancies of children and young women. Heart disease, cancer and cerebrovascular diseases are the leading causes of death. Tuberculosis rates drop off dramatically. Life expectancies may reach

seventy years (Omran, 1971, pp. 516-517, p. 533; 1977 p. 5).

There are many variations on these themes especially in terms of the rate at which any particular population moves from one stage to another. There are also differences in the forces which drive epidemiologic transition in different populations. The impact of medical technology, for example, is probably far greater in many third world nations which are experiencing a "delayed" version of the epidemiologic transition, than in the West, where social transformation was a more important catalyst to epidemiologic change. Pyle (1979) has described epidemiologic transiton in somewhat more environmentalist terms, identifying patterns of disease and population characteristics linked to "agrarian" cultural influences (infectious disease, high fertility, high child mortality) and patterns of disease linked to "industrial and post-industrial" cultural influences (low fertility, low child mortality, high rates of chronic and degenerative diseases).

Most discussions of mortality transition in the West link the transition most closely to social, rather than medical, change. In the third world, where the "delayed" model of epidemiologic transition is probably the best description of mortality change, medical and public health

measures may be the main drivers of the transition. In the "accelerated" model, which best describes that pattern of change in Japan, for example, mortality transition linked to social change was already underway but was accelerated by the impact of medical knowledge and technology (Omran, 1971; 1977). Study of mortality transition in the United States (Omran, 1977) shows that mortality decline and the emergence of "modern" mortality patterns was not strongly linked to medical advances. Increased life expectancies accompanied by declines in death rates and infectious diseases were attributed to improvement in living standards as measured by nutrition, personal income, housing standards and personal hygiene practices. Improvement in midwifery practices and increased use of isolation of patients with infectious diseases were probably significant as well (Omran, 1977). The argument that improvement in general living standards contributed more to mortality transition than medical practices is strengthened by comparison of the historical mortality patterns of whites with the pattern for non-whites.

> Whites have always been better off [in the U.S.] with regard to housing, living standards, social and economic levels, nutrition, access to medical care and other cultural and demographic characteristics. [For whites] the transition started earlier and moved faster. ... Non-whites still have higher death rates from infectious diseases than whites and lower rates of

> degenerative diseases. (Omran, 1977 pp. 33-34)

Finally, one of the factors in epidemiologic transition which is probably very important, but impossible to measure, should be noted. Some diseases, for reasons not well understood, have gone into ecologic decline over the centuries. Dubos (1980) notes that the virulence of syphillis had declined significantly by the sixteenth century, long before the cause of the disease was understood and, obviously, long before the age of antibiotics. Tuberculosis, the leading cause of death in the western world during the nineteenth century was apparently in decline in terms of both severity and number of people affected before there was any systematic treatment of the disease, or isolation of its victims. Other infectious diseases which seem to have gone into decline for reasons unrelated to medical or public health measures include plague and diptheria (Omran, 1977; Dubos, 1980).

2.2.2 Measuring the epidemiologic transition

There are many possible approaches to the measurement of demographic change. In this study, trends in infant and child mortality, maternal mortality and the age structure of mortality are examined at the provincial level. Cause

of death patterns are explored intra-regionally for evidence of the space-time characteristics of epidemiologic change. Taken together, along with some background understanding of the population and settlement history of Newfoundland, analysis of those trends will make it possible to address the two primary research objectives of this study. First, it will be possible to describe with some precision the spatial and temporal characteristics of changing mortality patterns in Newfoundland between 1937 and 1971. Second, the results of that analysis can be used to ascertain whether or not space is a useful dimension for observation of epidemiologic change at the intra-regional level of data aggregation.

It was noted above that the epidemiologic transition represents a profound shift in the mortality risk for children and young women; analysis of life tables for males who survive childhood show only slight increases in life expectancy over the past century (Dubos, 1980; Omran, 1977). The decline of infectious diseases as major killers translates demographically into large declines in infant and child mortality since children are at greatest risk from those diseases. Infant mortality in particular is recognized as one of the better indexes of the general health of a population (Rice, 1983, p. 36).

The trends in child and infant mortality can also be measured through examination of changes in the age composition of mortality. As mortality risk for children declines, the contribution that child mortality makes to overall mortality declines sharply. This drop-off in child mortality (and parallel rise in over sixty-five mortality) is very pronounced in Newfoundland between 1937 and 1971.

The improvement in the health of young women reflects both lower general fertility rates resulting in less exposure to the health risks of maternity, and improvement in prenatal care and midwifery practices. One of the better measures of this improvement is the maternal mortality rate, which is a ratio of the number of deaths related to childbirth to the number of births in a given year. Trends in maternal mortality, infant mortality and the age structure of mortality in Newfoundland are discussed in some detail in Chapter Three.

Analysis of change in the cause structure of mortality can yield some of the most valuable information concerning the mortality experience of a regional population. Unfortunately, the data used in this portion of the study presents many difficulties and the results must be viewed with some caution. The problem of misdiagnosis of cause of death is, of course, always present. This is particularly true in Newfoundland where, especially in the early years

of the study, death certificates were often completed by people without medical training, usually clergymen. Changing medical taxonomy and diagnostic sophistication also presents problems. Changes in the International List of Causes of Death and Disease (I.C.D.) over the years make caution necessary especially when comparing numbers of deaths in certain classes of disease as is done in this study. Occasionally data will have to be moved from its recorded class to another one to make the data comparable through time. Despite these problems, vital statistics, including recorded cause of death, are the best source of information available for reconstructing and trying to understand historical mortality patterns. When cause of death data is examined in conjunction with demographic data which inspires more confidence (such as infant mortality, child mortality and general mortality data), and with some understanding of the population's history, cause of death data becomes a rich source of information, particularly for gaining insight into changes in the interaction of population, environment and culture (Meade, 1977).

2.3 Mapping mortality

The cartographic presentation of disease and mortality incidence was one of the earliest and most powerful tools of medical geography. Though disease mapping is not an

important part of this study, a short review of its development is appropriate. The history of disease mapping reflects, to a great degree, the history of medical geography, especially the development of increasingly sophisticated statistical techniques for evaluation of medical geographic data.

Maps of yellow fever incidence in New York in the late eighteenth century, and of cholera in England during the early nineteenth century, are some of the earliest examples of useful disease maps (Howe, 1970). Snow's map of cholera incidence in London (published in 1855) suggested strongly that an outbreak of the disease was linked to a particular water pump (Howe, 1970; Learmonth, 1975). Disease mapping continues to be important in epidemiologic research. As Howe remarks,

> Disease maps are constructed to show facts, to show spatial distributions with an accuracy that cannot be attained on pages of description or statistics... They record observations in succinct form; They aid in analysis; they stimulate and aid in the formulation of working hypotheses... (Howe, 1970, p.16)

The development of better, more easily understood maps of disease and mortality continues to be a major concern of medical geographers and others interested in the spatial aspects of disease. Several of the more relevant disease mapping procedures are discussed below.

2.3.1 Techniques for disease mapping

Snow's map of cholera in London was a dot map, each dot representing one case of cholera. Mapping the incidence of disease in this way can be a powerful method for identifying possible environmental links to disease, understanding the rate and direction of disease diffusion, and identitfying areas of endemic disease (Pyle, 1979). Disease maps have been useful in the identification of foci for epidemiological research and as guides to the planning of disease control programs.

When the concern is with general rates of disease and mortality in a population it is often not possible, practical, or even desirable, to map individual cases or deaths. In the search for spatial patterns, it is often useful to aggregate the data in areal units and use choropleth mapping techniques to display the disease rate information. It is also the case, particularly in historical studies, that location of cases or deaths will not be reported as occurring at a precise location, but only within a census tract, county, or other reporting unit. Demographic cartograms can be useful for preventing distortion of disease and mortality rate information created by different physical size of reporting areas. Their use has been discussed by Howe (1970).

In this study, the data are aggregated at an electoral district level (reported deaths based on the deceased's usual place of residence). Openshaw (1977) and Davies (1984) have both pointed out that extreme care must be taken in interpreting results from one level of areal aggregation as being applicable at another. Strictly speaking, the results and conclusions drawn in this study can only be considered as relevant at the electoral district level (Davies, 1984).

Much of the literature on disease mapping has been concerned with the cartographic presentation of Standardized Mortality Ratios or other standardized death rates. Standardization techniques remove the effect of age structure of a population on the death rates, insuring some sort of comparability through time and space. One of the more common techniques involves converting death rates (in whatever form) to standardized "z" scores and categorizing the areal units in the study based on those scores. In other words, classification is based on how many standard deviations the score for a particular unit is from the mean for the region (Armstrong, 1969). This technique can also be used to look at incidence rates, rates of infant mortality and other general rates.

A related technique, probability mapping, involves comparing actual death rates in subregions with overall

regional rates and using the Poisson distribution to identify areas with significantly high rates and those with significantly low rates (Choynowski, 1968; McGlashan and Harington, 1976). Though a very attractive method for mapping disease incidence, its utility in studies of general mortality and historical change in mortality patterns is doubtful. Since it does not show variation that is statistically non-significant, much interesting (though "non-significant") information may be lost. For this reason, the mapping of "z" scores has proved to be a more popular technique, as it shows both significant and non-significant variation (Statistics Canada, 1980; Armstrong, 1969; Howe, 1980; McGlashan and Harington, 1976).

Another procedure, which has implications for a study of epidemiologic transition, is Pyle's (1979) technique of mapping factor scores of areal units from a factor analysis of leading causes of death. It was noticed that certain patterns of disease tend to cluster in the same factor space. Two principle (orthogonal) factors were interpreted as representative of infectious patterns of disease and chronic patterns of disease. By mapping areal units, in this case countries, based on their factor scores through time, an attempt was made to identify regions dominated by one pattern of disease or another and regions in apparent

transition. The time series of maps shows the progression of some countries from a pattern dominated by infectious disease to a pattern dominated by chronic and degenerative disease.

2.4 Canadian geographic mortality studies

There have been very few geographic studies of mortality patterns in Canada, particularly at the intra-regional level. Studies done in Canada have been primarily at the national level such as Field's (1980) study, or at a very local level such as Massam's (1980) study of spatial variation of mortality in Montreal. Innes (1980) has examined the patterns of mortality associated with several causes of death in Ontario at the local and provincial level. None of these stuides, however, involves a significant historical component, probably because of the difficulties of working with such data.

Field's work suggests that Atlantic Canada may be in some sort of health transition, at least insofar as this would be revealed by mortality patterns.

> The most diverse patterns [of mortality] are encountered in the Atlantic region, with areal units more or less equally distributed between the low, medium, and high mortality classes. (Field, 1980 p. 49)

The only geographic work done on mortality patterns in Atlantic Canada is Barret's (1972) undergraduate thesis

which is an examination of trends in Standardized Mortality Ratios for the electoral districts of Newfoundland between 1935 and 1966. Barret found mortality rates decreasing overall through that period with an especially large drop between 1951 and 1956. Higher rates of general mortality were observed in the more urbanized districts, though the risk for women in urban areas was less than in rural ones. Barret made no attempt to examine cause of death patterns.

Comparison of mortality data from the Canadian provinces suggests that there has been a significant shift in the pattern of death in Newfoundland. Statistics for infant mortality and several causes of death show that Newfoundland was "out of synch" with the average national pattern until into the 1960's, when provincial trends come more into line with national ones (Basavarajappa, 1976; Statistics Canada, 1976). Field's (1980) work indicated that regional mortality patterns tend to dominate the intra-regional ones, but he noted that the breadth of the class intervals used in regional studies may mask significant intra-regional variation (p. 49). This study is an attempt to describe and understand the intra-regional variation in mortality observed in Newfoundland between 1937 and 1971.

CHAPTER THREE

Background to the Study- Demographic Change in Newfoundland

3.1 Introduction

The meaning of change in a population's cause of death structure over time and through space can be understood only if there is an understanding of the broader demographic context in which that change takes place. The history of population growth, structure and distribution, along with the history of settlement, combine with social and economic history to create both the backdrop for, and the measure of, epidemiologic change. Significant changes in the health of a population (and, by implication, the nature of the complex interaction between environment, population and culture (Audy, 1954)) will show up as change in demographic indicators. Change will be observed in infant mortality, child mortality, the age structure of general mortality and fertility as well as in the pattern of cause of death. The variability of these indicators in time and space will indicate variability of health patterns in those dimensions. If the notion that health measures the interaction of humans and their environment is accepted, we have, in the study of demographic and

epidemiologic history, tools for constructing a partial history of man-environment interaction. The specific inferences which can be drawn based on that history are limited in scope and are best combined with more traditional, and more intensive, human ecology studies, if the meaning of epidemiological change is to be understood.

Full consideration of the human ecological context of epidemiological transition in Newfoundland is far beyond the scope of this thesis. The more modest goal of this study is examination and description of the process of epidemiologic transition within the demographic context of the population history of Newfoundland. This chapter is concerned with the settlement and population history of Newfoundland as background to a more specific study of spatial and temporal changes in cause of death patterns. No attempt is made to reconstruct the complex history of the peopling of Newfoundland; Rowe (1980) and Mannion (1977) present excellent introductions to that rich study. Instead, a brief discussion of Newfoundland settlement history is presented, followed by a more specific discussion of recent mortality trends.

3.2 Population growth and patterns of resource exploitation

Europeans had probably been fishing off the coast of Newfoundland many years before its official discovery in

1497 by John Cabot. The history of European settlement of the island, however, is considerably shorter. The exploitation of the fishery did not require permanent settlement and the fishery was the main source of European interest in Newfoundland.

The French, for example, used only the harbor at Plancentia on the Avalon peninsula, and that only as a rendevous point for their fishing fleet. French fishermen, because of their access to large supplies of salt, were able to store their fish wet, using the salt as a preservative. Once back in France, the fish was dried. This type of exploitation of the fish resource became known as the Bank fishery (Staveley, 1970). English fishermen, without access to large supplies of salt, were required to go ashore, usually on the east coast of the Avalon peninsula, to sun dry their fish before returning home. Staveley remarks that,

> The conflicting practice of the French and English in fisheries technologies... were crystallised within the ranks of the English themselves by a parallel conflict which was concerned with the control of the fishery- the basic problem was whether the fishery was to be dominated by transient summer fishermen employed by English base merchant houses (known as the ship fishery), or whether it should be based on the plantation of independent settlers (the shore fishery). It is from the amalgam of these various influences (French or English enterprise: ship, shore or bank

> fishery; merchant or settler) that we may draw the essential threads of Newfoundland's historical geography, especially in terms of settlement and economy. (Staveley, 1970, p. 3)

Several attempts were made at colonization in the early part of the seventeenth century, most notably the "plantation" at Cupids on Conception Bay, which was founded in 1610. Other settlements were established south of St. John's. Piracy, wars, illness and the obstructionism of English West Country merchants, who had interests in preventing the development of a fishery controlled by Newfoundlanders, all contributed to the failure of those settlements. The unsteady start of permanent settlement in Newfoundland was also related to the relative ease with which a would-be settler could change his mind about settling and return to England. Rowe points out that,

> If [the settler's] experience with permanent settlement proved distasteful or unprofitable, he could "pull up stakes" and resume work with the migratory fishery, return to England or go to New England without any great financial loss. (Rowe, 1980, p. 122)

Estimates of the size of the permanent population of Newfoundland at the beginning of the eighteenth century vary widely. Rowe (1980) puts the 1730 figure at 2702 residents (almost all English) spread between Notre Dame Bay and the Burin peninsula, the so-called "English Shore." Handcock (1977) estimates the figure at closer to 1000.

By 1730 the French were not a significant factor in settlement. Their one large settlement at Placentia was given up as a result of the Treaty of Utrecht in 1713. The French, however, did retain rights to the offshore fishery from Cape St. John around the Great Northern Peninsula to Cape Ray near the current site of Port aux Basque. That length of coastline became known as the French Shore. The English laid claim only to the English Shore and agreed not to settle along the French Shore or in any other way interfere with the French fishery. The French, for their part, agreed not to settle permanently along the French Shore, erecting only temporary structures which were necessary for the summer fishing season. Though English settlements were eventually established along the French Shore as the seal fishery developed, the existence of French rights to that coastline retarded, to some extent, English expansion past Cape St. John.

Until approximately 1730, most settlements in Newfoundland were located along Trinity or Conception Bay or on the shore just south of St. John's. There were also a few small English settlements in Notre Dame Bay. The levels of population during the century after the establishment of the first colony in 1610 fluctuated substantially and no clear growth trends emerged. However, a pattern of resource exploitation had begun to develop

which would lead to the entrenchment of the outport settlement pattern. The prosecution of the cod fishery was the dominant economic activity, supported in other parts of the year by subsistance agriculture, the fur trade, boat building and (especially along the northeastern shore) the seal fishery.

The development of an outport settlement pattern was pushed along by increased immigration to Newfoundland. The first major wave of immigration came from southern Ireland in the 1720s and 1730s. By 1750, the permanent population had reached 7,000, including over 900 women. By 1800, the population of the island was over 20,000 and still growing (Rowe, 1980). The history of Newfoundland's population growth is summarized in Figure 3.1. Throughout the early nineteenth century, new economic opportunities were being presented to residents and immigrants in the form of the seal fishery, the salmon fishery and the fur trade. These opportunities, coupled with substantial population growth led to the spread of settlements along the shores of Bonavista Bay, Trinity Bay, and Notre Dame Bay. The factors that led to the spread of settlement along the northeast coast were, according to Sanger (1977),

> (1) increased pressure on resources in the Avalon peninsula occasioned by the rapid increase in the number of permanent settlers there during the early part of the 19th century; and (2) the growing awareness that the income

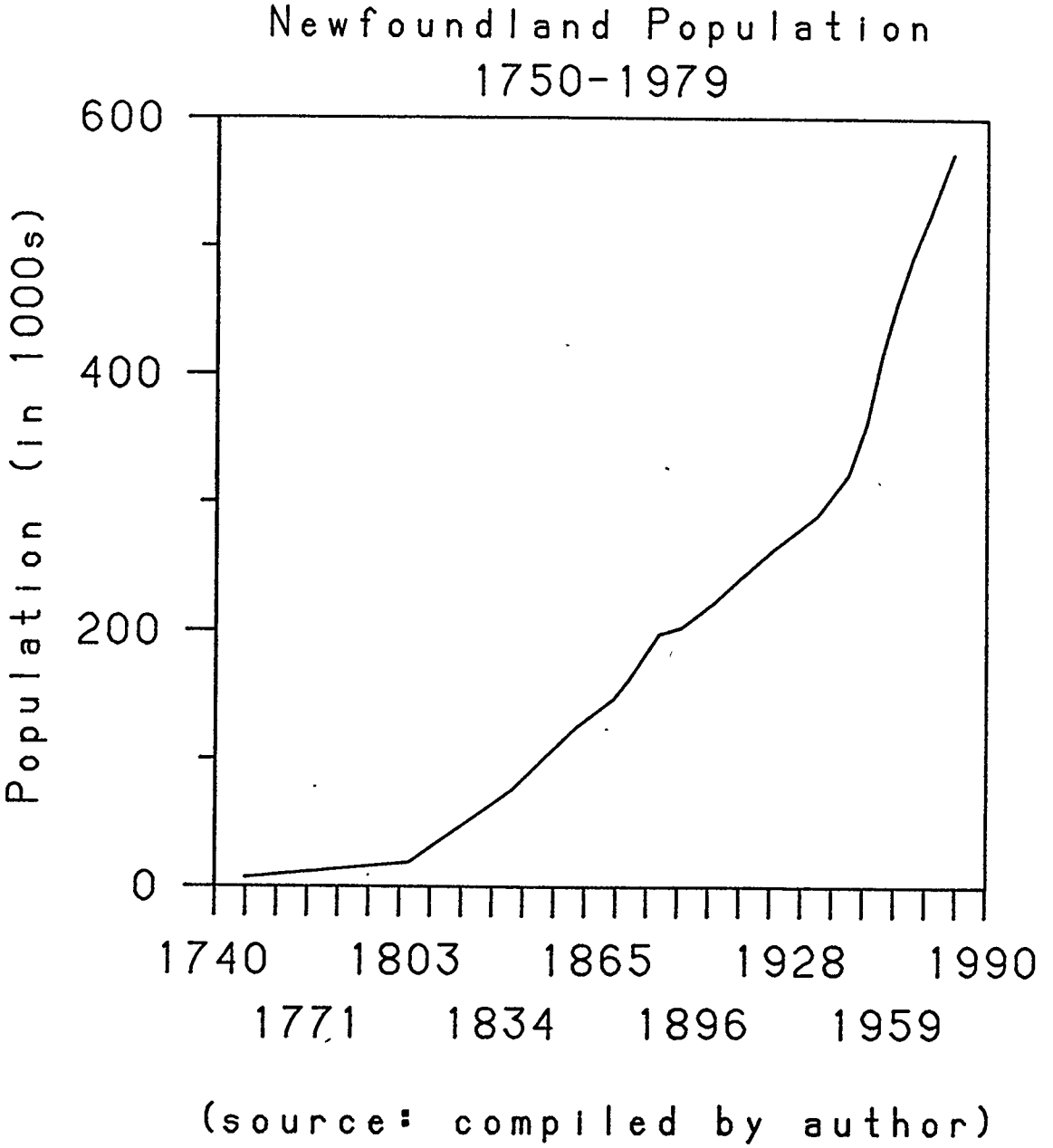

Figure 3.1 Population of Newfoundland, 1750 - 1979

> earned by the participants in the recently launched large-vessel seal fishery could also be supplemented by a landsman operation on the northeast coast. These factors acted as a fresh incentive for people to move and settle in areas close to the migratory route of the harp seal. (Sanger, 1977, p. 141)

Besides the pressure on resources created by an expanding population and increased economic opportunity, the early nineteenth century also saw a shift in the interests of the English West Country merchants to prosecution of the ship fishery on the Grand Banks. There is argument amongst historians and historical geographers about the degree to which West Country merchants actually retarded the spread of settlement; in any event, whatever restraint had been imposed was removed in the early 1800s. The existence of the French Shore continued to restrict settlement to some extent in those areas while the inhospitable environment and the restricted opportunity for a shore-based fishery along the South Coast limited development there.

Once the primary importance of the cod fishery, and later the seal fishery and fur trade, is clear, the development of a widely dispersed, isolated outport settlement pattern is easily understood. Rowe has summarized it well.

> ...settlement in these small, isolated communities took place in a logical

> geographic direction. The Avalon Peninsula was settled, then Trinity Bay, then Bonavista Bay, the Strait Shore, Notre Dame Bay and so on. Similarly on the South West Coast, people moved for economic and social reasons- to have more room, to be near good fishing grounds or to form religion-based communties. ...[Harbors could] accommodate only a specific number of stages and wharves and no more...(Rowe, 1980, pp. 128-129)

Staveley's (1977) work supports this position. He has calculated average population increase and decrease in various districts of Newfoundland measured at different points in the nineteenth century. In the second half of the century especially, the greatest increases took place around Trinity Bay and Bonavista Bay and in the Fogo and Twillingate districts located in Notre Dame Bay. Growth rates were generally quite low in St. John's and Conception Bay, and in some Avalon districts there was a decrease in population.

The history of the settlement of the west coast of Newfoundland is a variation on themes established during the settlement of the English Shore. Catholic Acadians from Cape Breton were some of the first settlers on the west coast. Together with a number of French immigrants, they began the settlement of St. George's Bay and the Port au Port peninsula in the early nineteenth century. Somewhat later, (after 1860) the Bay of Islands and Bonne

Bay were settled by a slightly more diverse group (predominantly Catholic) which included not only Acadians but Labrador fishermen, Irish immigrants, and loggers from New Brunswick as well (Mannion, 1977). The salmon fishery and the fur trade, and eventually, a logging industry, were the commercial forces which drove settlement on the west coast. The cod fishery, so important to the English Shore settlement, was controlled largely by the French, the west coast being part of the French Shore. The west coast settlements were quite illegal but the French seldom made an issue of the settlements as long as the residents did not try to prosecute an offshore cod fishery. In fact, Mannion (1977) points out that the two groups generally cooperated with each other. The settlers exploited the fur, timber, and agricultural resources along with the fisheries of the bays while the French exploited the fishery in the Gulf of St. Lawrence. Conflict arose only when the settlers intruded upon the French cod fishery (Mannion, 1977, p. 249).

3.2.1 Summary

Through the eighteenth and nineteenth centuries, a way of life based mainly on the exploitation of primary resources developed in Newfoundland. The spatial expression of this economy was a pattern of widely

dispersed isolated outport settlements along the six thousand mile coast of Newfoundland. Newfoundland's isolation made the development of an industrial economy somewhat improbable and so the outport-based economy has persisted as the most important type of economy on the island until recently (see, for example, Brox, 1972). As that economy has broken down, many aspects of what Omran (1971) has called the "modernization complex" have manifested themselves, including increased urbanization and "modern" patterns of disease, mortality, and demographic structure. The spatial dynamics of that change, if understood in historic regional context, can provide insight into not only the spatial dynamics of health change but the spatial dynamics of change in man-environment interaction.

Any attempt to understand the meaning of recent change in demographic trends must be framed by an understanding of the basic political, economic and social forces which shaped the patterns of settlement in the region. The crucial themes in the historical geography of Newfoundland have concerned the exploitation of the various fisheries and other primary resources and the pattern of outport settlement which developed as a consequence. It is in this context that demographic change must be evaluated so that

one does not lose sight of the rich human history contained in the rates and percentages which measure that change.

3.3 Recent mortality trends

In this section, twentieth century trends in infant mortality, child mortality and the age structure of general mortality in Newfoundland are discussed. It was noted in Chapter Two that the epidemiologic transition favors women and children over men (Omran, 1971, p. 512). This suggests that any inquiry into epidemiologic transition should be particularly concerned with the groups most affected. Infant, child and maternal mortality is therefore the concern of much of the discussion below.

3.3.1 Infant mortality

The decline in infant mortality associated with the epidemiologic transition reflects a decline in the number of deaths from prematurity, the bronchitis, pneumonia influenza complex, and diarheal diseases, all still major killers of infants and children in the third world. A decline in diseases such as scarlet fever and diptheria and other so-called "diseases of childhood" is important here as well. When the trends in infant mortality risk are broken down, it is clear that the most improvement over time is in mortality in the post-neonatal period, that is, after the first month of life. Rates of death from

congenital anomalies, which usually cause death during the first month of life have not declined nearly so much (Omran, 1977, p. 21).

Figure 3.2 illustrates the twentieth century trends in infant mortality in Newfoundland, Ontario, and Canada. Infant mortality is calculated as a ratio of infant deaths to the number of live births during the year. Both Statistics Canada and (before Newfoundland joined Confederation) the Department of Public Health and Welfare in Newfoundland have maintained excellent statistics on infant mortality during the twentieth century. Figure 3.2 was compiled from both sources.

Probably the most important and interesting trend to note is the convergence of the rates for Ontario, Canada, and Newfoundland. This is a pattern which can be seen in standard mortality ratios for various diseases, child mortality, and overall age standardized death rates. The decrease in inter-regional diversity suggested by the graphs may indicate a kind of homogenization which can be expected as a result of the emergence of a "modern" pattern of disease and mortality.

3.3.2 Maternal mortality

The decline in maternal mortality in Newfoundland is summarized in Figure 3.3. Much of the improvement in

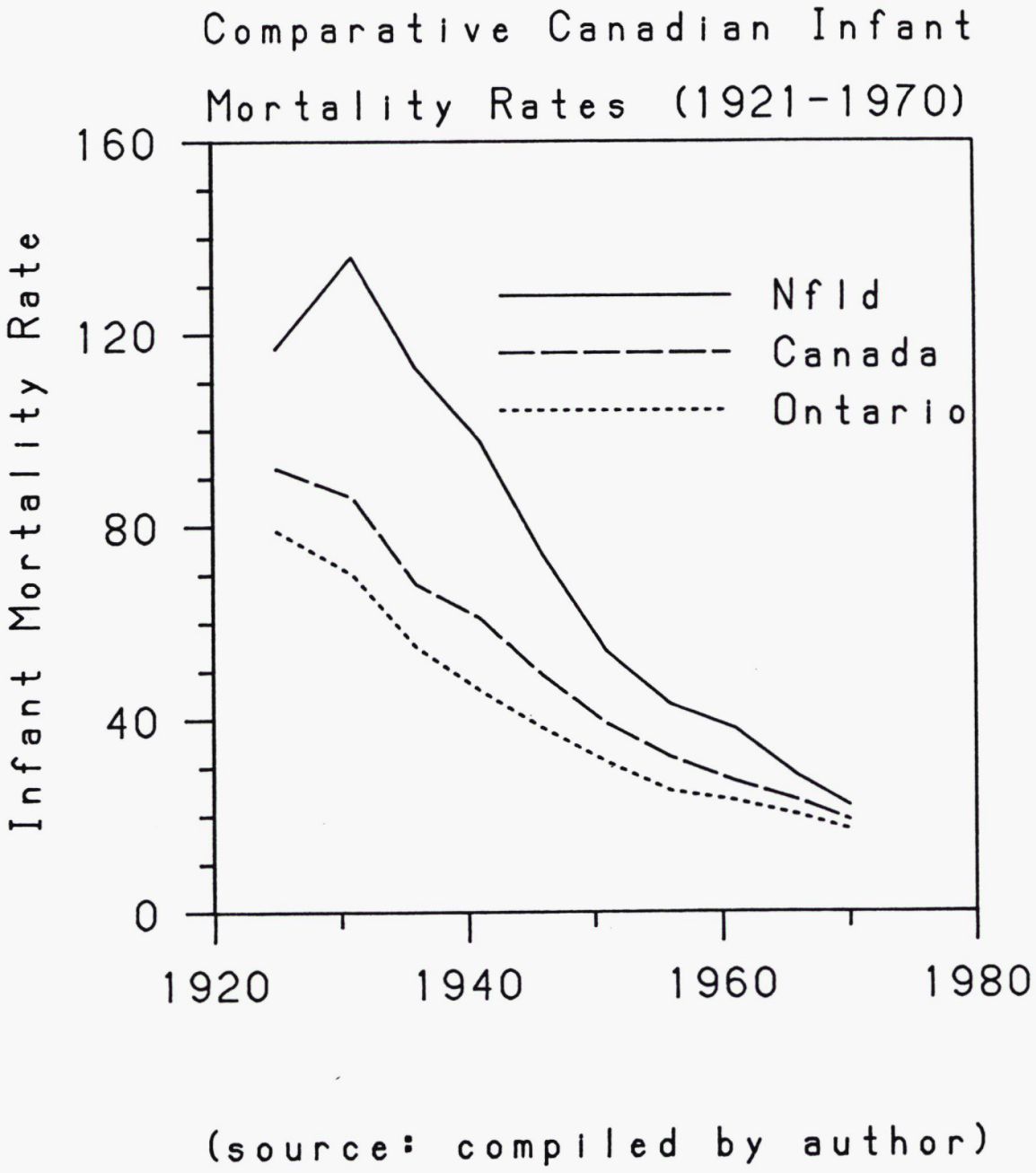

Figure 3.2 Infant mortality rates, 1921 - 1970

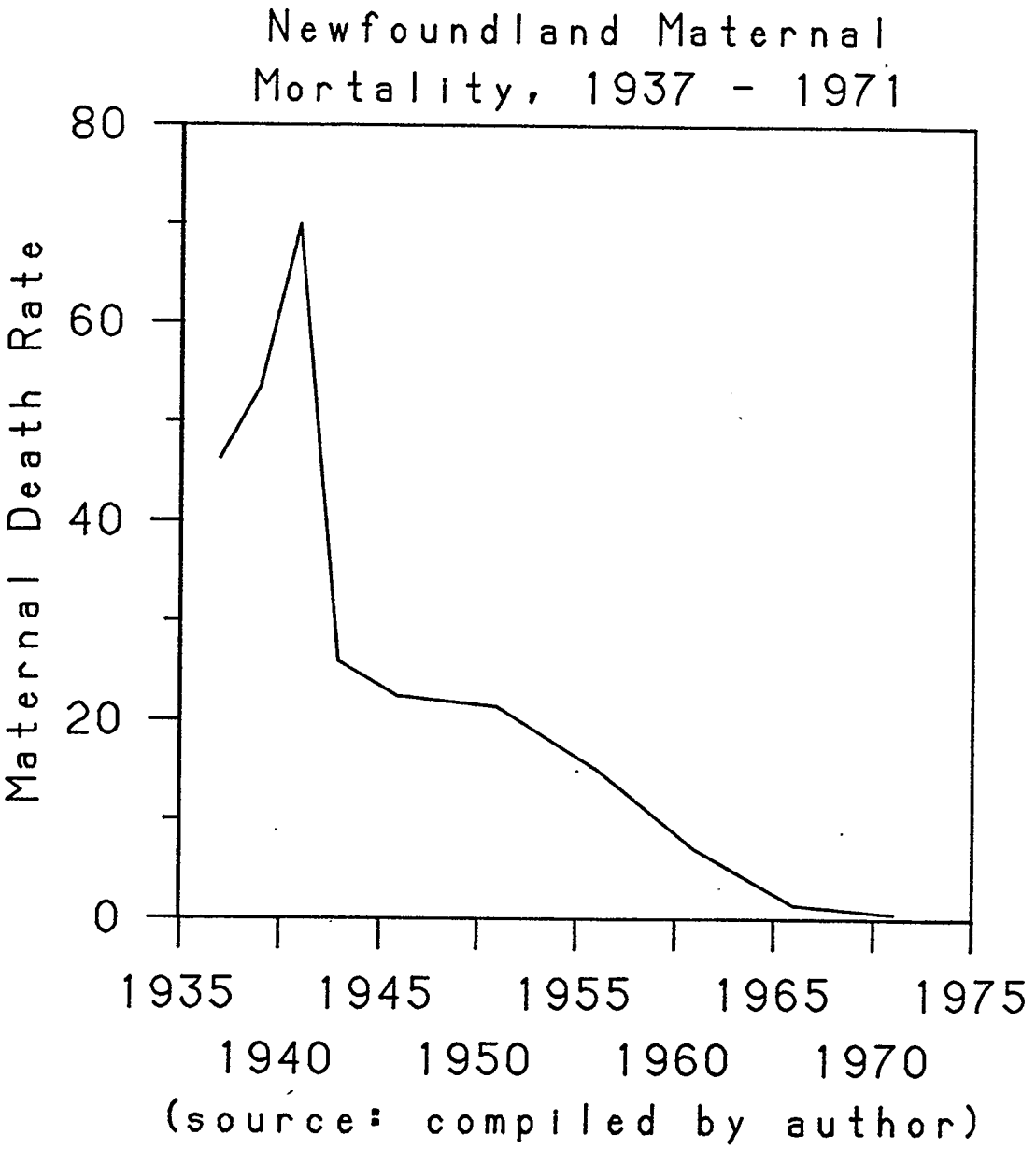

Figure 3.3 Maternal mortality rate, 1936 - 1971

maternal mortality during the study period was probably related to improved midwifery techniques associated with the growth and success of the cottage hospital system during the thirties and the forties (Savoury, 1975). Reports filed by the cottage hospital nurses on their activities (found in the Provincial Archives of Newfoundland) indicate a strong emphasis on the teaching of midwifery in the outports. Barrett's (1972) work suggests that young women of childbearing age also benefited from the trend toward increased urbanization between 1935 and 1966. Though no causal links are established, he notes that the comparison of standard mortality ratios for men and women show higher rates for women in rural areas than for men, while urban areas show increased risk for men and lower risk for women.

3.3.3 The age structure of mortality

Further evidence of significant change in the mortality structure of the population of Newfoundland is found in Figure 3.4. The age structure of mortality is graphed cumulatively from 1937 to 1981. The dashed line indicates the portion of 0-14 mortality attributable to the deaths of children aged 0 to 4 years. A decline in child mortality as a percentage of total mortality is one of the indicators of epidemiologic transition, representing the

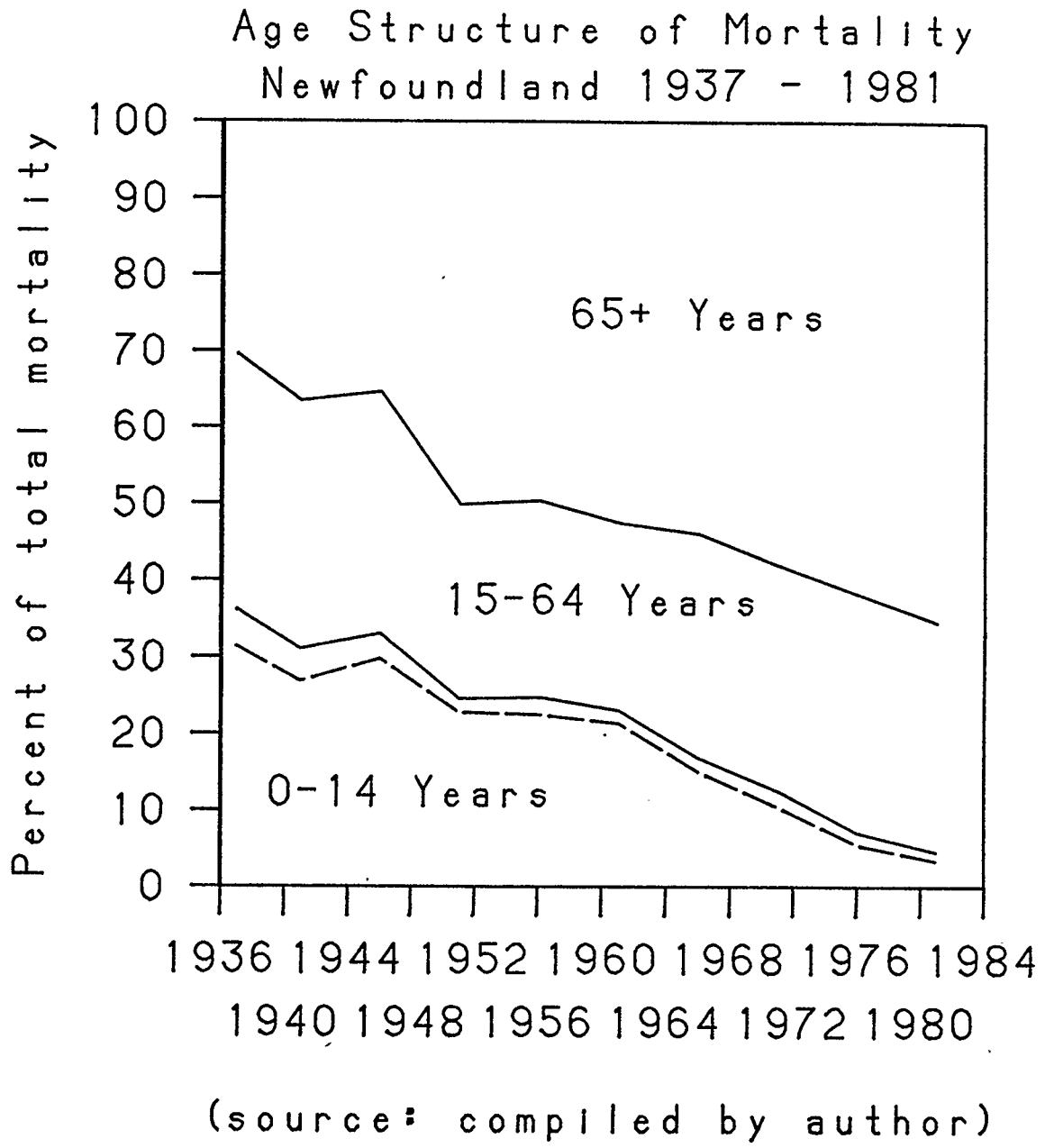

Figure 3.4 Age structure of total mortality, 1937 - 1981

decline of infectious diseases as important causes of death.

The records of the fever hospital in St. John's from the late 1930s and early 1940s (which can be found in the Provincial Archives of Newfoundland), show clearly the pattern of child death. The majority of the patients listed in the record books were children between the ages of a few months and ten years of age. Diptheria, typhoid fever, scarlet fever and combinations of whooping cough and pneumonia were the most frequent reasons for admission. Many died within a few days of being admitted. Figure 3.4 shows a profound change in mortality pattern and suggests a perhaps more important change, the aging of the population.

Changes in the age structure of a population can be shown many ways. For example, Staveley (1977) used Coulson's (1968) method of indexing age structures to graph nineteenth century developments in population age structure in Newfoundland. Figure 3.4 shows another way to examine the aging of the population which may be more relevant to a discussion of an epidemiologic transition. The other side of the drop in child mortality is an increase in over-65 mortality which is shown clearly in the figure. The stability of 15-64 mortality as a percentage of total mortality is confirmation of trend towards a "modern"

mortality pattern dominated by diseases common to an older population.

3.4 Changes in cause of death patterns

The remainder of this thesis is concerned with the analysis of change in the cause structure of mortality in Newfoundland between 1937 and 1971. The temporal development of a modern pattern cause of death at the regional level is summarized in Figure 3.5. The decline in infectious diseases and the rise of chronic and degenerative diseases was rapid and dramatic in Newfoundland. Chapter Four and Chapter Five are devoted to a detailed spatial and temporal analysis of the transformation outlined in Figure 3.5.

3.5 Conclusion

Recent trends in various measures of mortality in Newfoundland suggest that an important shift in population dynamics has taken place there over the past fifty years. The decline in infant and child mortality, the rise in over age sixty-five mortality, and trends in the cause structure of mortality indicate that the population has gone through an epidemiologic transition of the type described by Omran (1971, 1977). Analysis of the intra-regional characteristics of mortality transformation was undertaken to explore the spatial dynamics of that transformation.

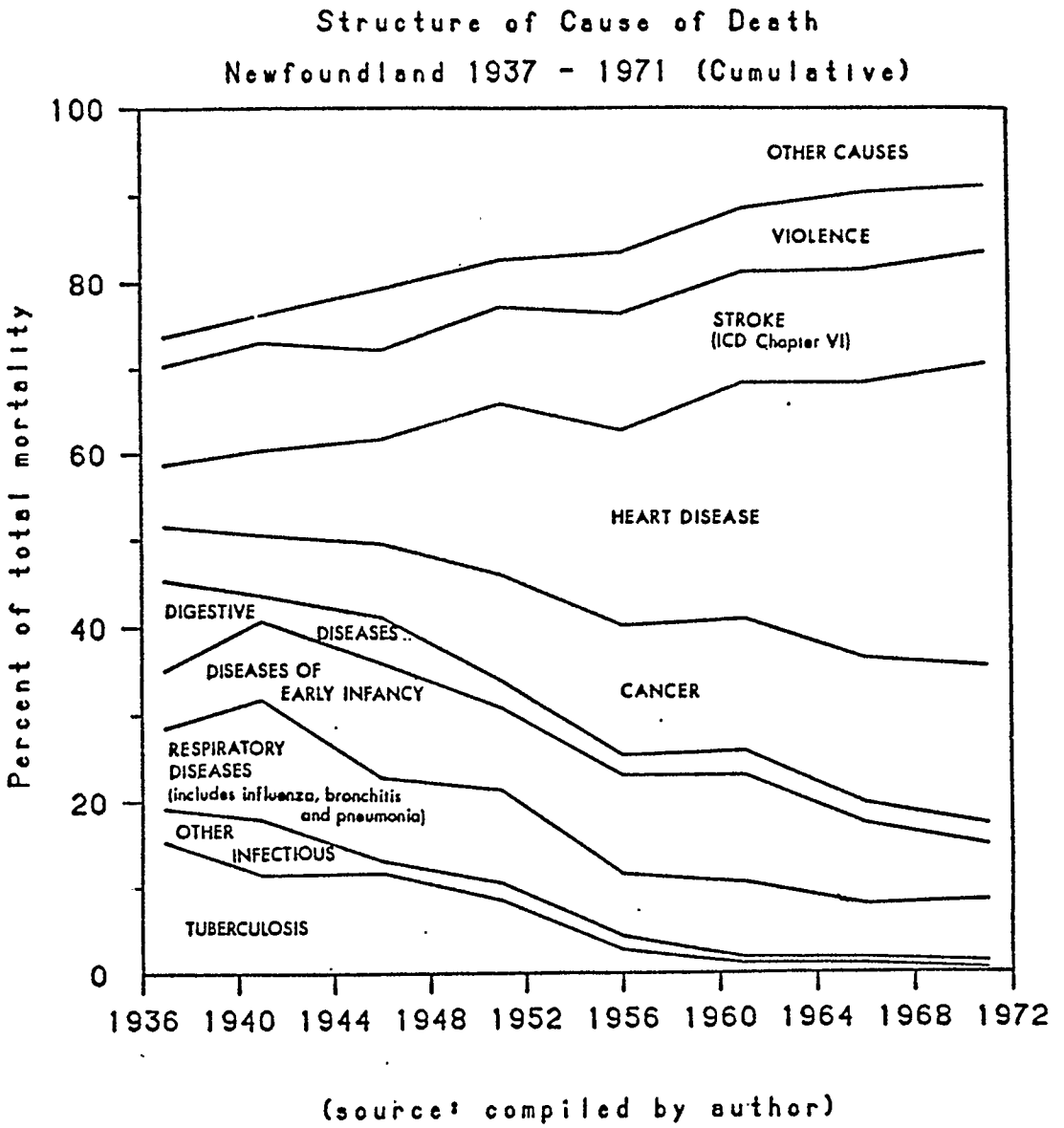

Figure 3.5 Cause structure of total mortality, 1937 - 1971

Results of that analysis, discussed in Chapter Five, indicate that epidemiologic transition in Newfoundland was not a spatially uniform process. As such, the progress of epidemiologic transition in Newfoundland may mirror the spatial dynamics of change in man-environment interaction there.

CHAPTER FOUR

Data Collection and Analysis

4.1 Introduction

An analysis of the dimensions of change in population dynamics must include a discussion of the data used to measure that change, the level of areal aggregation of the data, and the resulting limitations which are imposed on any conclusions drawn from analysis of the data. A description of the statistical methods used for analysis of the data is necessary as well, so that limitations of the study imposed by methodological considerations are clear. The first part of this chapter is a discussion of the data used to explore the dimensions of epidemiological change, in particular the classification of cause of death and the areal units used for reporting mortality. Some of the difficulties and limitations of that data are discussed. The second part of the chapter deals with the statistical techniques used to analyse the data: canonical correlation analysis, cluster analysis, and discriminant analysis. The results of the analysis are presented in Chapter V.

4.2 Data sources

The main sources of data for this study are the mortality statistics generally reported as vital statistics. This data includes information on death rates, cause of death, and the usual place of residence of the deceased. Before Newfoundland joined the Canadian Confederation in 1949, comprehensive reports of vital statistics were published by the health officials of the Government of Newfoundland. These reports included the usual information on births, marriages and deaths, including a fairly comprehensive compilation of data on cause of death and age at death. Cause of death was reported by specific cause and district of usual residence. Causes were grouped into the chapters of an early version of what is now known as the International Classification of Diseases (I.C.D.) (WHO, 1977). These reports are available at the Centre for Newfoundland Studies at Memorial University in St. John's.

The Province of Newfoundland and Labrador continued to publish these reports after Newfoundland joined Confederation, however the quality of the information supplied in the reports declined significantly after the early 1950's, especially in terms of the detail provided concerning cause of death. In many of the reports, summaries of cause of death information were dropped

altogether. For data on cause of death by district from 1949 to 1971, data provided by the Vital Statistics Division of Statistics Canada in Ottawa were used. The data consisted of breakdowns of cause of death by I.C.D. chapter by age and district of residence.

4.2.1 Areal units

Vital statistics in Newfoundland were reported, from 1937 to 1971, in several variations of the electoral districts which were established for the 1935 Census of Newfoundland. These twenty-four districts (twenty-three on the island, one for Labrador) are the basic areal units used in this study. It was necessary to consolidate the original twenty-four districts into twenty-one districts because of changes made to the reporting districts during the period covered by the study. Specifically, the data for Trinity North and Trinity South were combined as were data for Bonavista North and Bonavista South and the Placentia West and Placentia and St. Mary's districts. Figure 4.1 shows the boundaries of the original districts. Unpublished information on the changes to the reporting districts was provided by the Vital Statistics Division of Statistics Canada which enabled the author to sort out a series of confusing changes in the areal reporting units. In particular this information was used to remap data

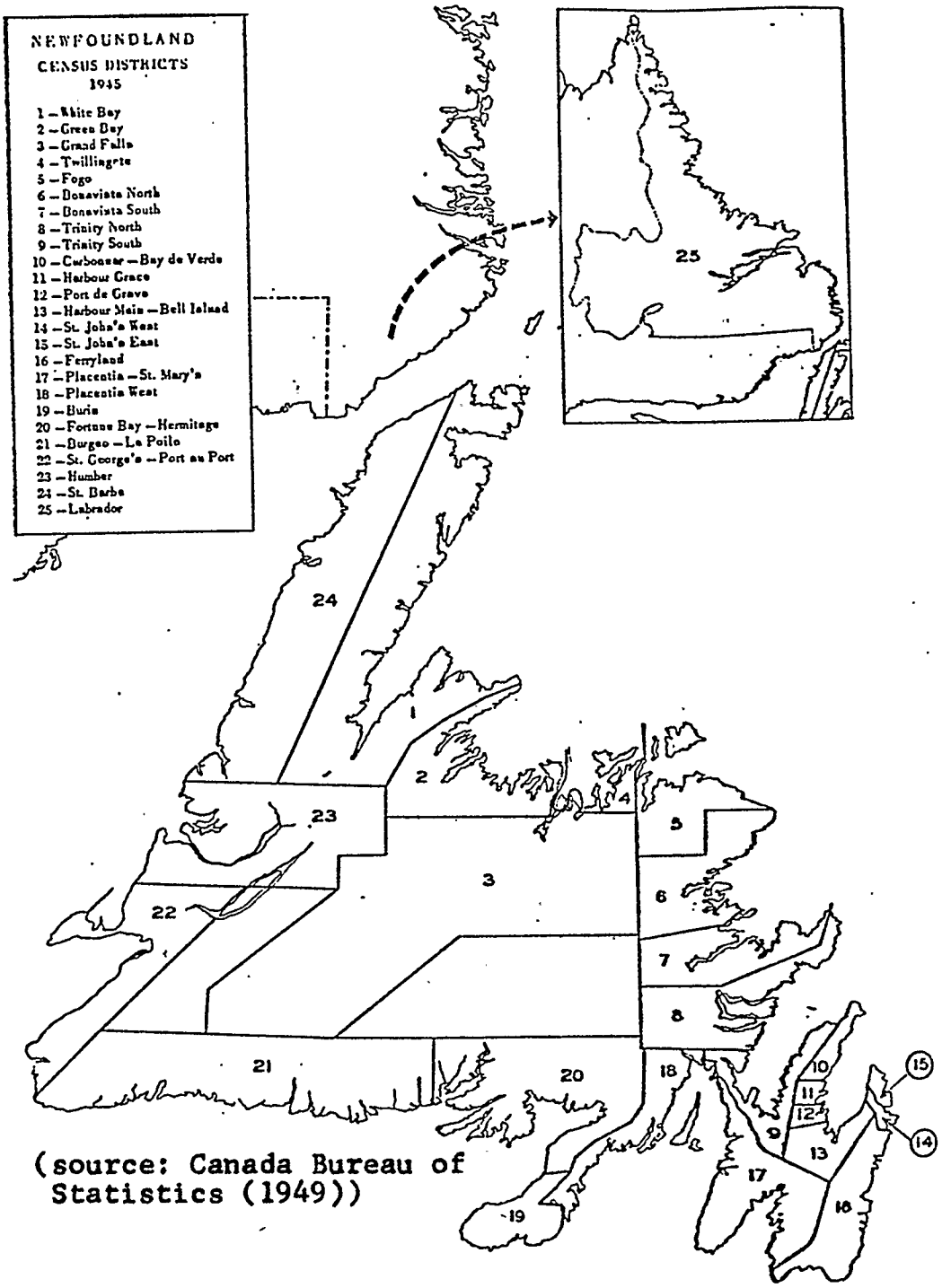

Figure 4.1 Electoral districts used in the study

reported in districts created after 1949 back into the modified "original" districts.

It should be pointed out that the districts used to report vital statistics do not, in any of their post-Confederation incarnations, correspond to Canadian census districts (which have themselves been modified several times since 1949). The lack of correspondence between census districts and reporting districts for vital statistics precludes, to a large extent, a factorial ecologic study of the social correlates (as measured by census variables) of epidemiological change. It also makes the calculation of accurate, post-Confederation, standardized death rates difficult as the calculation of those rates involves detailed data on the age structure of the population in each district where mortality is measured (Armstrong, 1969). This is the reason that this study uses, in the analysis of cause of death patterns, matrices of percentage of total death linked to each I.C.D. chapter rather than the more conventional approach of analysis of death rates associated with each cause. In 1976, vital statistics reporting units were changed to correspond to Canadian census divisions and subdivisions, making it impossible to extend the current study past that date (Evans, 1985).

4.2.2 Data quality and limitations

In Chapter II it was pointed out that historical cause of death data is often uninspiring in quality. Worse, the quality, or lack of it, is difficult to gauge without access to death certificates, medical records and the help of medical specialists for ascertaining the accuracy of diagnosis. In a regional historical study of overall patterns of mortality such an approach is clearly not practical. For purposes of statistical analysis, the data must be accepted at face value. Conclusions drawn from that analysis, however, must be tempered by an understanding of the possible error in the data set.

Another concern in this study is the system used for registration of deaths in Newfoundland, especially in the early years of the period being studied. A large percentage of deaths were unattended by medical personnel of any kind, and death certificates in those cases were filled out by local clergymen. In the case of many infectious diseases, where symptoms were commonly recognized, one may feel some assurance about the accuracy of cause of death information. One may also assume that persons dying of degenerative disease may have come into contact with the doctors and nurses from the cottage hospitals at some point during their illness and that the diagnosis may have been known to the family of the deceased

and reported to the clergyman. Still, there is no way to pin down a quantitative estimate of error in diagnosis and the reported cause of death must simply be taken as accurate in any single case, and the fact that some error must exist in the data for the population accepted.

The concern about the quality of the cause of death data is mitigated somewhat by the fact that the concern in this thesis is with change in the structure of cause of death at a broad level rather than specific disease rates in specific districts. The I.C.D. Chapter structure (See Table 4.1) is useful for arranging the data when diagnostic accuracy is in doubt. An abbreviated version of the I.C.D. is used to report cause of death in many third world countries today which allows epidemiologists, demographers, and other health researchers to have some idea of disease patterns in places where access to accurate diagnosis (and health care) is often extremely limited (WHO, 1977). Health workers with minimal training are likely to be able to recognize tuberculosis, other infectious diseases, certain cancers and symptoms of heart disease fairly easily while a more thorough diagnosis is not possible without a doctor. This level of information, while perhaps of limited use to a doctor trying to cure disease in individuals, can be extremely useful to those interested in disease and population history. The use of the

Table 4.1 Chapter Structure of the International Classication of Diseases (8th Revision)

Chapter I - Infectious and Parasitic Diseases

Chapter II - Malignant Neoplasms

Chapter III - Allergic, Endocrine and Metabolic Disorders

Chapter IV - Blood Diseases

Chapter V - Mental Disorders

Chapter VI - Diseases of the Nervous System and Sense Organs

Chapter VII - Diseases of the Circulatory System

Chapter VIII - Diseases of the Respiratory System

Chapter IX - Diseases of the Digestive System

Chapter X - Diseases of the Genito-urinary System

Chapter XI - Complications of Preganancy and Childbirth

Chapter XII - Diseases of the Skin

Chapter XIII - Diseases of the Musculoskeletal System

Chapter XIV - Congenital Anomolies

Chapter XV - Certain causes of Perinatal Mortality

Chapter XVI - Symptoms and Ill-defined Conditions

Chapter XVII - Accidents, Poisoning and Violence

(source: WHO (1977))

I.C.D. Chapter structure also helps the researcher avoid difficulties created by changing medical classification. An examination of the history of the I.C.D. shows that while the intra-chapter structure has undergone significant change through the various revisions of the list, the chapter structure itself has remained quite stable with only a few exceptions which are noted below. Haynes (1983) studied the geographical aspects of mortality in Chile using a cause of death classification scheme based primarily on the I.C.D. chapter structure.

In an historical study covering several revisions of the I.C.D. it is important to be conscious of the changes made to the list and the extent to which those changes affect mortality statistics. Two important changes made to the list between 1937 and 1971 made it necessary to "hand massage" some of the data to insure comparability through time. In the Sixth Revision of the I.C.D. (adopted in the early 1950's) influenza was moved from Chapter I (Infectious Diseases) to Chapter VIII (Respiratory Diseases). To adjust for this change, influenza deaths were removed from Chapter I in the 1937, 1941, 1946 and 1951 data matrices (described below) and moved to Chapter VIII. The other change involved the movement of cerebrovascular disease in the Eighth Revision from Chapter VI (Mental and Nervous Disorders) to Chapter VII

(Circulatory Diseases). Since the Eighth Revision was adopted in 1968, only the 1971 data matrix was affected. As a result, Chapter VI was dropped from the canonical analysis of the 1966-1971 pair of data matrices. Another change made in the Eighth Revison was the movement of certain diarheal diseases from Chapter X (Digestive Diseases) to Chapter I (Infectious Diseases). No attempt was made to adjust the 1971 data matrix for this change since diarheal diseases were no longer a major cause of death in Newfoundland by 1971.

4.2.3 Organization of the data for analysis

The cause of death data were assembled into data matrices twenty-one by seventeen, corresponding to the twenty-one modified districts described above and the seventeen chapters of the I.C.D. A matrix was created for each of the years 1937, 1941, 1946, 1951, 1956, 1961, 1966 and 1971. The layout of the data matrices is shown in Figure 4.2. Though standardized death rates are the preferred measure of mortality, the problems with the lack of correspondence between vital statistics reporting districts and census divisions made this very difficult. The nature of the data and the lack of accurate age structure information dictated a different approach. Instead of standardized death rates, the percentage of

Figure 4.2 Layout of data matrices

	INFECTIOUS	CANCER	CEREBROVASCULAR	CIRCULATORY	RESPIRATORY	DIGESTIVE	GENITO-URINARY	MATERNAL	PERINATAL	VIOLENCE
ST. JOHN'S										
HARBOUR MAIN										
PORT DE GRAVE										
HARBOUR GRACE										
CARBONEAR										
TRINITY										
BONAVISTA										
FOGO										
TWILLINGATE										
GRAND FALLS										
GREEN BAY										
WHITE BAY										
ST. BARBE										
HUMBER										
ST. GEORGE'S										
BURGEO										
FORTUNE BAY										
BURIN										
PLACENTIA										
FERRYLAND										
LABRADOR										

total mortality for each I.C.D. chapter was calculated for each district in each time period, creating a series of new data matrices. From these matrices, numerous other versions were created (corresponding to different combinations of variables) for the multivariate statistical analysis described below. In all cases, Chapter XVI (Symptoms and Ill-defined Conditions) data were dropped from the analysis. It was felt that analysis of Chapter XVI data with the rest of the cause of death data would only make the analysis of the data more complex with little, if any, gain in insight into the data. Problems with ipsative data matrices, which sum to a common value across each row, were also avoided in this way.

4.3 Methods for analysis of the data

There were two separate but related goals for the multivariate analysis of the cause of death data. The first was to understand the nature of the changes in the structure of mortality in Newfoundland between 1937 and 1971. The second was to explore the spatial aspects of that change. To address the question of structural change through time, canonical correlation analysis was used to compare successive pairs of cause of death matrices. That is, the 1937 matrix was compared with the 1941 matrix, the 1941 matrix was compared with the 1946 matrix, and so on.

The spatial expression of change in the structure of mortality was examined via cluster analysis, which grouped districts based on the similarity of structure of mortality in those districts. The significance of the clusters was checked with discriminant analysis. The use of these three multivariate techniques is described below, with special emphasis on canonical correlation analysis, since its use is less common in geographical studies than the other two procedures.

4.3.1 Canonical correlation analysis

Canonical correlation analysis is a multivariate statistical technique used to explore the correlation between sets of variables measured on the same observational units (subjects, areas, among others). There is no requirement for an equal number of variables in each data set. Canonical correlation is still relatively new in geographical research, the first example being Berry's (1968) study of the spatial structure of the Indian economy. Since then it has been used in a number of studies including several in medical geography (Monmonier, 1972; Meade, 1979), but it has not achieved the level of use that might have been expected given its potential utility in geographical analysis. There are many reasons for the scarcity of studies utilizing canonical correlation

analysis in geography. Probably the most important are the difficulties involved in interpreting canonical variates and the measurement of statistical significance beyond chi-squared testing of the canonical correlations. This situation has improved considerably in recent years as important advances in both canonical theory and statistical programs used to carry out canonical correlation analysis have been made. In view of these advances, the tendency for geographers to shy away from use of this technique may no longer be justified.

4.3.1.1 Data considerations

Canonical correlation analysis begins with two data matrices, one of q normally distributed "predictor" variables measured on N observational units (such as subjects and areas) and one of p normally distributed "criteria" variables measured on the same N observational units. Whether or not one set of variables can be viewed as dependent and the other as independent seems to be a matter of some confusion in the literature. Pimentel (1978, p. 59)) states flatly that "neither [set] must be considered as predictors or criteria." Clark (1977, p. 7) in his CATMOG on canonical correlation in geographical analysis writes that "neither set of data is given priority in the analysis." Yet, in the same CATMOG and in his

Ph.D. dissertation (1971), Clark analyses urban linkage and structure change in Wales through time. Clearly, one set of variables (the 1958 data) is being looked at, at least implicitly, as predictors of the second set (the 1968 data). More recent work by Briggs and Leonard (1977), Tabachnick and Fidell (1983) and Johnston (1978) indicates that dependence and independence of variable sets depends, for the most part, on how one chooses to look at the data sets. In some cases, these writers suggest, it may be useful to view one set as dependent and one set as independent, while in other cases such an approach may not make much sense.

Another data consideration involves the nature of the numbers which are actually used in the analysis. Early geographical work with canonical correlation frequently involved the analysis of factor or principal components scores (Berry, 1968; Clark, 1977). Briggs and Leonard (1977) indicate that this approach needlessly introduces an extra level of abstraction into a statistical procedure which, in its own right, takes the researcher further away from the original data than he or she might like.

> ...orthogonality and parsimony, [the usual reasons for doing a components analysis first] are achieved within the [canonical] analysis and need not be determined externally. To do so increases the difficulties of interpreting the canonical variates since this must be conducted on the

> basis of relatively imprecise factors rather than more readily understood variables. (Briggs and Leonard, 1977 p. 144-45)

In a type of analysis which bristles with interpretive difficulties, there can be no justification for interpreting canonical variates and scores calculated from a principal components solution instead of variates and scores calculated from the raw data matrices. A possible exception might involve the interpretation of component scores from an oblique component solution along the lines of higher order factor analysis (Nader, 1981). One of the weakness of canonical correlation analysis is that the solutions are strictly orthogonal, though common dimensions may not be orthogonal at all. An oblique components solution might facilitate exploration of the structural dimensions of the data with canonical correlation analysis. To the author's knowledge, this approach has not been explored in the literature. Among other possible problems, one suspects that the degree of statistical abstraction involved would make it difficult to keep track of the "real world" meaning of the canonical variates and their structural implications.

Tabachnick and Fidell (1983) and Pimentel (1978) point out that one of the problems with the canonical procedure is the sensitivity of the results to minor changes in the

data sets. Pimentel (1978, pp. 61-62) suggests that some of this sensitivity may result from inadequate sample size and recommends a minimum sample size of

$$N > p + q + 1$$

though "this often is an inadequate sample size. Samples of 100 individuals might be considered minimal for consistant reliance in results." Sample sizes of 100 may be quite difficult to obtain for many geographical analyses. In the study of spatial structure in a region, the most likely geographical application of canonical correlation, there simply may not be that many nodes or areal units in which the variables have been meausured. All that can be recommended is that the problems and sensitivity which may result from small sample sizes be noted, and that appropriate caution be employed in the interpretation of results. The final point concerning data for canonical analysis is made by Pimentel, who mentions that ipsative measures (measures which sum to some common value across each row of the data matrix) must be avoided. Though percentages are used in the current study, difficulties are avoided because, as discussed above, several categories of cause of death are dropped from the analysis. Consequently, the percentages for each district do not sum to any common value.

4.3.1.2 The canonical correlation procedure

Involved discussion of the mathematical details of canonical correlation analysis is beyond the scope of this thesis. It is, however, important to summarize the basic equations so that the origins of the measures of interset correlation and their relationships with each other are understood. Moreover, understanding the meaning of more recently developed indicators of significance is contingent on understanding the four basic measures of canonical correlation: canonical correlation coefficients, canonical weights, canonical vectors and canonical scores.

Canonical analysis begins with four matrices of intercorrelation of the p and q variables which can be called Rpq, Rpp, Rqq and Rqp. From these matrices, a new matrix, R, is calculated (Tabachnick and Fidell, 1983).

$$R = R_{pp}^{-1} R_{pq} R_{qq}^{-1} R_{qp}$$

This matrix (whose calculation is discussed at some length by Clark (1977)) is used to extract the common dimensions of the two data sets. From this matrix of intercorrelation, the extent to which clusters of vectors (defined by the intercorrelation of the p variables and the intercorrelation of the q variables) occupy the same statistical space can be determined (Clark, 1977). There will be as many (orthogonal) dimensions, or areas of

overlap, as there are variables in the smaller data set. It is doubtful that all of the dimensions, however, will be statistically significant. Significance of the canonical dimensions is determined using Bartlett's chi-squared test. Monmonier (1972) has summarized the canonical exercise in this way

> Given two sets of n and m variables, X and Y , canonical correlation finds two linear transformations
>
> $$U = \sum_{i=1}^{n} u_i X_i = u_1 X_1 + u_2 X_2 +,\ldots, u_n X_n$$
>
> and
>
> $$V = \sum_{j=1}^{m} v_i Y_i = v_1 Y_1 + v_2 Y_2 +,\ldots, v_m Y_m$$
>
> This correlation [between U and V] is the canonical correlation or canonical root. The coefficients u and v are the canonical variates or weights used to indicate the relative importance of the original variable in the linear expression...(p. 1221)

The pairs of combinations of weights in each canonical dimension are referred to as canonical vectors.

A fourth measure of canonical correlation which is produced by most canonical correlation programs is the set of canonical scores which can be used to understand the involvement of each of the observations in the canonical vectors. Canonical scores are calculated by applying the canonical weights from each pair of canonical vectors to a standardized (using z-scores) raw data matrix. The scores will indicate the degree of involvement of each of the

observational units in the common dimensionality measured by the canonical vectors. Analysis of canonical scores when the observational units are areas, especially when several significant canonical correlations are observed, can reveal spatial patterns, if the weights and vectors are interpretable.

Standard canonical correlation analysis then, extracts four measures of correlation between data sets. The canonical correlation coefficients, or canonical roots, measure the common dimensions of the sets of p and q variables. Canonical weights (sometimes called variates) measure the involvement of each of the variables in each of the common dimensions. Each pair of linear combinations of weights is referred to as a canonical vector. Finally, canonical scores on each of the observations measure the participation of each of the observations in each of the canonical dimensions. Only canonical weights, vectors and scores corresponding to significant canonical roots (as measured by Bartlett's chi-sqaured test) are evaluated here.

Interpreting these measures in geographical analysis is not as straightforward as Clark's (1977) summary suggests.

> The structural identity is conveyed by the canonical weights while the spatial patterns are conveyed by the final set

> of measures in canonical analysis, the canonical scores. (p. 9)

This statement suggests that the dimensions revealed to the researcher will be interpretable, which is frequently not the case. The shared dimensionality revealed by canonical correlation analysis represents the best mathematical solution to the problem but not necessarily a comprehensible one. Recent improvements in canonical theory and in the statistical programs which carry out the analysis can improve the interpretability of the results, but the researcher must be prepared for a solution devoid of any "real world" meaning.

4.3.1.3 Interpreting canonical correlation analysis

The first step in interpreting a canonical correlation analysis is determining whether the canonical correlation coefficients are statistically significant. The coefficients are tested against the null hypothesis that the two data sets are unrelated (their correlation is not significantly different from zero). Bartlett's chi-squared test is used but not on the canonical correlations themselves. Instead, a test statistic, Wilk's lambda, calculated from the canonical correlation coefficients is used (Clark, 1977). Tabachnick and Fidell (1983) present a complete discussion of the mathematics involved in

calculating Wilk's lambda. Each canonical correlation (calculated from the linear transformations described above) is tested with the preceeding correlation removed. The first canonical correlation will represent the maximum correlation between the two data sets, the second the next best (orthogonal) solution, and so on. Each correlation is tested with Bartlett's chi-squared statistic at some pre-selected confidence level. Only dimensions which are significant at that confidence level are examined further.

Statistical significance alone, however, does not necessarily indicate that a given coefficient is a good measure of full interset correlation. Because the coefficients are derived via the linear transformations described above it is possible that highly significant, but spurious, canonical correlations will be found. Briggs and Leonard (1977) point out that if just one of the p variables happens to be highly correlated with one of the q variables, a high canonical correlation will be observed,

> ...yet this can hardly represent a high degree of relationship between the two sets. ...Thus, because it is a measure of correlation between canonical variates, the canonical correlation coefficient is a highly unreliable measure of the relationship between the original variables. (p. 134)

What the researcher is interested in is not the value of the canonical correlation, but the extent to which a given

canonical vector represents the original data. This measure is called percent of variance extracted (Briggs and Leonard, 1977; Tabachnick and Fidell, 1983) and is routinely provided by the BMDP6M canonical correlation program. Percent of variance extracted should always be examined along with the values and significance of the canonical correlation coefficients.

While percent of variance extracted measures will show the importance of a canonical variate in explaining the variance within one of the data sets, another measure is needed to discover how much variance in the set of dependent variables is predicted by a canonical variate of the independent set of variables. This measure is referred to as factor redundancy (Briggs and Leonard, 1977, p. 135). For a given variate it is the product of the squared canonical correlation coefficient and the variance extracted from the dependent variables by the variate (of course, if neither set of variables is being viewed as dependent or independent, one can turn the equation around to examine the relationships in either direction).

The interpretation of canonical correlation coefficients is thus dependent on understanding that those coefficients are linear composites of the variables from the original data sets. To interpret the coefficients it is necessary to go beyond simply looking at their values

and respective significance tests and measure variance extracted and factor redundancy. A third measure which should be mentioned before moving on to a discussion of the interpetation of canonical weights and scores is referred to as total redundancy. It is the sum of the factor redundancies of the dependent variable set and "measures the proportion of total variance of the dependent variables which can be predicted by all of the canonical variates of the independent variable set" (Briggs and Leonard, 1977 p. 135). Evaluation of the three measures described above should enable the researcher to understand whether or not the results of a canonical correlation analysis are revealing any important common dimensionality, though problems with the meaning of the canonical dimensions may remain. To address these problems one must turn to interpretation of the canonical weights and vectors.

It was noted above that the canonical weights indicate the relative importance of the variables in each of the canonical vectors. This is true, but it is often not entirely clear what the empirical meaning of the weights is in any given analysis. Clark (1971; 1977) notes that the structural meaning of the weights must begin and end with examining the signs and magnitudes of the weights and trying to make sense of the groupings in each vector. Often, the weights for many variables must be ignored

because they are uninterpretable. More recently there has been substantial improvement in the interpretability of canonical weights. Many authors (Monmonier, 1972; Monmonier and Finn, 1973; Meade, 1979; Briggs and Leonard, 1977; Tabachnick and Fidell, 1983; Johnston, 1978) recommend that the canonical weights be converted so that a new matrix of values is produced which can be understood as the correlation between each of the original variables and the canonical vectors.

> While canonical variates [weights] are like factor loadings... they are not - as are loadings on principal components - correlations between the vectors and the original variables. These correlations can, however, be obtained by correlating the canonical scores with the raw data; this should always be done to facilitate interpretation. (Monmonier, 1972, pp. 1221-1222)

These correlations can be interpreted as standard correlation coefficients. For an example of the degree to which interpretability of canonical vectors is enhanced by the use of correlations instead of canonical weights, see Monmonier and Finn (1973). The weights are shown to be very difficult to work with, often obscuring the meaning of the canonical vectors. When the matrix of correlation with each of the original variables is examined the empirical meaning of the vectors becomes clearer. In much of the literature and in BMDP6M output, the matrix of correlations

of vectors with original variables is referred to as a loading matrix, following the terminology used in components analysis.

Interpreting the correlations is often a matter of interpreting the canonical vectors as entities, much as one might attempt to label axes in a factor analysis. For example, one may look at the variables from both data sets which load highly with similar signs in a vector and attempt to name that vector based on an understanding of the data sets involved. This has been the approach in many of the geographical applications of canonical correlation analysis. Clark (1977), for example, interprets the two most significant vectors in his study of urban linkage in Wales as structural dimensions denoting "high urban status" and "holiday industry centres" common to both 1958 and 1968.

To some extent, the individual loadings are interpretable as well; low loadings indicate lack of involvement in a particular dimension and high loadings indicate high involvement. Pimentel (1978) suggests that whether one inteprets the vectors or loadings on particular variables is determined by the nature of the data set.

> In some cases it is appropriate to consider a single variable of one set as a predictor of a single variable of the other set. This one to one evaluation might be best when the criteria set is loosely related (e.g.,

> when the set is composed of different species as variables...) However when all variables in a criteria set are joined so closely that scrutinizing patterns of variables is appropriate, the canonical vectors can be translated as patterns of variation predicted by a single predictor of the other set (Pimentel, 1978, p. 62).

Clark (1973) suggests that interpretation of individual variable loadings in a vector may be particularly useful for exploring structural changes in temporal studies in which the data are places, or variables, measured at different points in time. In these types of studies "the analysis isolates those elements within the system which have remained unaltered over time as well as illustrating the ways in which structural changes have occurred." (Clark, 1973, p. 47) If one is interpreting loadings within a vector pair then, variables which load highly and with the same sign in both vectors exhibit a high degree of structural stability from one time period to the next. Significant changes in the loadings on a variable from one time period to another may indicate structural change, though strictly speaking, all that can be said is that the variable in question contributes little to the common dimensionality measured by the canonical vector pair. Interpretation of loadings which shift substantially from one time period to the next is a somewhat tentative

business, though it may be aided by measures of variable redundancy.

Variable redundancy is a measure which offers insights into the participation of individual variables in the total canonical correlation. Such a measure can be important if the researcher is unsure as to whether or not to include a particular variable in either or both of the data sets.

> For a variable in the dependent set, variable redundancy measures the proportion of total variance explained by the canonical model. For a variable in the independent set it measures the relative importance of that variable in contributing to the the explanation of the dependent set. (Briggs and Leonard, 1977, p. 136)

These measures can improve both the quality of the data going into the final canonical correlation analysis and the interpretability of the results. From a table of variable redundancies it is possible to pick out the variables making important contributions to the canonical correlation, though those variables may not be important in all of the significant canonical dimensions. Meade (1979) makes use of variable redundancy measures in her study of the so-called "coastal plain enigma" (of high death rates) in the southeastern United States. Examining variable redundancy makes it possible to identify crucial variables from what might otherwise be a bewildering array of loadings. By reducing the complexity and size of the

loading matrix and identifying important variables, variable redundancy may offer the researcher insights which could be useful in explaining observed patterns. Variable redundancy measures are not provided directly by SPSS or BMDP but can be calculated quite easily from the loading matrices and the squared canonical correlation coefficients provided by BMDP.

The final measure of canonical correlation produced by most canonical correlation programs is the set of canonical scores which measure the participation of each of the observations in the pairs of canonical vectors. If the observations are spatial entities- countries, counties, census tracts- it is possible to use the scores to look for spatial patterns in the canonical dimensions. The scores are calculated by multiplying the standardized raw data for an area times the set of canonical weights for each pair of vectors. So if there are ten observations of five variables in each data set, twenty canonical scores will be produced for the first pair of canonical vectors. By graphing the scores (Clark, 1977) the participation of areal units in the canonical vectors can be observed.

Despite the obvious attraction of canonical scores for purposes of geographical analysis, there has been precious little work by geographers (or others) on the interpretation of canonical scores. Clark (1977, p. 31)

points out that there have been no major studies which have involved the analysis of canonical scores in geography and notes that "there is no overall consensus as to procedures in interpretation [of the scores]." Though geographers are well practiced in mapping scores from factor analysis, which are analogous to canonical scores, there has been no such development in canonical correlation, and analysis of scores remains one of the most obscure aspects of the procedure.

4.3.1.4 Summary- Canonical correlation analysis

Recent improvements in canonical theory and canonical correlation programs have made the canonical procedure more accessible and potentially more useful to geographers. Despite these improvements a great deal of care is necessary in the use of the procedure since the potential for error, ambiguity and misinterpretation of results is still great. As with all statistical analysis, researchers must begin with a thorough knowledge of the data set so that they can decide whether the common dimensionality is in any way meaningful or simply a mathematical solution devoid of "real world" meaning. Additionally, they must be willing to admit that canonical solution is only marginally useful or perhaps uninterpretable.

Canonical correlation analysis is perhaps most useful in geographical analysis for exploring structure and structural change while its utility in discerning understandable spatial patterns is doubtful. In the current study, the spatial significance of the scores is apparently nil; no clear pattern was discernible. Subsequently, canonical correlation analysis was used only for exploration of the aspatial aspects of epidemiological change. In temporal studies its use has been neglected, which is unusual given its ability to measure structural change in time (Johnston, 1978). In fact, temporal studies (of the same variables measured on the same areal units at different points in time) seem to involve considerably less interpretive subjectivity than more conventional applications of the procedure. The increased robustness of the technique in temporal studies is primarily the result of the tendency in such studies to focus on individual variables and their changes in time rather than on subjectively labeled vectors.

Whether one is doing a temporal study of structural change, or a more conventional atemporal study of multivariable structure, the power of this statistical procedure should be apparent to geographers. Many geographical studies involve implicit and explicit assumptions about structure and its meaning; canonical

correlation allows the researcher to look at the actual components of structure and structural change. The numerous measures of interset and intraset correlation provided by canonical correlation analysis make it possible to look at the data sets in several ways. In some cases one may be most interested in the dimensionality measured by the canonical vectors, or in factor and total redundancy, while at other times the most important information can be found in measures of variable redundancy. Canonical correlation analysis involves many ways of seeing data sets and their meaning and it is this versatility that creates both its pitfalls and its substantial power.

4.3.2 Cluster analysis

The term "cluster analysis" refers to a family of statistical procedures used to group objects based on multivariate data. In this study, cluster analysis was used to group Newfoundland electoral districts based on the similarity of their mortality patterns from 1937 to 1966. Once the districts were grouped, the mean mortality pattern of each of the significant clusters was analysed within the framework of the theory of epidemiologic transition to find out whether or not that transition was expressed spatially as well as temporally.

The cluster analysis performed for this study began with the same data matrices as were used in the canonical correlation analysis. Instead of eight separate analyses, however, the matrices were combined so that a single cluster analysis could be carried out. The results of that analysis illustrate the changes in mortality pattern in each district through time. In particular, this type of cluster analysis facilitates the identification of districts which can be identified as demographic innovators and ones which lagged behind the regional average. Districts were grouped using Ward's (1963) clustering method (described below). To insure that each of the variables was weighted equally, the matrix was standardized before analysis of similarity of the districts was undertaken. Measures of similarity which have been used in cluster analysis include correlation coefficients and measures of Euclidean distance calculated in a variety of ways (Everitt, 1974).

One of the more robust hierarchical clustering techniques is Ward's (1963) method (Lorr, 1983). Euclidean distance is used as the similarity measure between groups and grouping is carried out based on the grouping which minimizes the error sums of sqaures at each step. The method is predicated on the notion that when k objects are "clustered" into k groups, one hundred percent of the

information about those groups is available. As the objects are clustered into fewer groups (k-1, k-2,... 1) information is lost at each step. When the objects are clustered into one group, information loss is total. The increase in the value of error sums of sqaures gives the researcher some idea of the percentage loss of information at each step (Everitt, 1974; Lorr, 1983). The calculation of both distance measures and Ward's error sums of sqaures is discussed most clearly by Everitt (1974, pp. 15-17).

The problem which confronts the researcher using this method, or any other clustering procedure, is determining the optimum number of clusters since there is no generally accepted statistical test of cluster significance (Green, 1978). An exercise which can be helpful in selecting the number of clusters is graphing the error sums of sqaures against the number of clusters in a given analysis. Selection may be guided by looking for a cut-off point (Johnston, 1976) where a significant increase in error sums of sqaures is observed. This increase, expressed graphically as a sharp change in slope of the line representing information loss, indicates a significant loss of accuracy and information below a certain number of clusters (Lorr, 1983). Another somewhat more intuitive approach, applicable in studies using raw data (as opposed to, for example, component scores), is examination of the

mean variable scores for the clusters. In this case, knowledge of the data set can guide the researcher to the most useful cluster solution.

In this study, Ward's method was used to cluster analyse the mortality patterns of the twenty-one districts across the twenty-nine years of the study period. The results of that analysis, including the dendrogram, cluster profiles, and a graph of the error sums of sqaures of the solution are presented in Chapter Five. The cluster profiles, and tables showing the grouping of each district through time, are used to determine whether or not the epidemiologic transition found significant spatial expression in Newfoundland.

4.3.3 Discriminant Analysis

Cluster analysis procedures generate a classification of objects based on multivariate observations; the taxonomy is an internal, numerical one. As indicated above, checking the significance of the clusters is difficult. Graphs of error sums of sqaures and examination of mean variable values for the clusters can steer the researcher toward the proper solution, but there is no accepted, rigorous test of the significance of the clusters availiable within the family of cluster analytic procedures. An approach to checking the significance of

the clusters used frequently by urban geographers and others (Johnston, 1978) involves the use of discriminant analysis once a taxonomy has been generated with cluster analysis.

Unlike cluster analysis, discriminant analysis begins with an a priori classification of objects and checks the validity of that classification scheme by classifying objects based on multivariate observations. If objects are classified with a high degree of accuracy into the groups defined a priori, one can feel some confidence about the classification scheme and the significance of the differences between the objects. Discriminant analysis finds linear transformations of the multivariate observations; objects are then placed in groups according to their discriminant scores (useful discussions of discriminant procedures are presented in Nie (1975), Tabachnik and Fidell (1983) and Johnston (1978)). In this study, the classification of Newfoundland mortality patterns by district generated by Ward's clustering method was checked with discriminant analysis, using selected cluster solutions as the a priori grouping. The proportion of "misclassified" districts reflected the adequacy of the taxonomies generated by cluster analysis and so served as a check on the cluster solutions.

4.4 Summary

As mentioned in Section 4.3, there were two objectives which the statistical methodology was designed to meet. The first objective was to describe the nature of the change in the cause structure of human mortality in Newfoundland between 1937 and 1971. A related objective was to pinpoint those changes in time to the extent possible. To meet those objectives, canonical correlation analysis, as described in Section 4.3.1, was used with particular emphasis on canonical loadings and variable redundancy measures to explore the aspatial dimensions of epidemiologic change. The second major objective was to explore and describe the space-time aspects of epidemiologic transition in Newfoundland. This was done using cluster analysis, specifically Ward's hierarchical method. Results were checked with discriminant analysis as described above. The nature and limitations of the data set outlined in Section 4.2 impose some constraints on conclusions drawn from these analyses, and those constraints must be kept in mind when interpreting the results. Results of the canonical, cluster and discriminant analyses are presented in Chapter Five.

CHAPTER FIVE

Results of Data Analysis

5.1 Introduction

The statistical methodology developed for this study and the nature of the procedures involved were reviewed in Chapter Four. In this chapter, the results of the canonical correlation, cluster and discriminant analyses of the cause of death data are presented. A review of the objectives for each analysis is included along with formal interpretations of the results of each analysis. Consideration of the broader implications of those results for the meaning of the entire study is left for discussion in Chapter Six.

5.2 Canonical correlation analysis of cause of death data

Seven separate canonical correlation analyses were carried out. In each analysis, a data matrix of the percentage of death associated with ten I.C.D. chapters (see Chapter Four) in twenty-one Newfoundland districts was compared with a similar matrix for the same districts five years later. (An exception was the 1937-1941 analysis which used a four year interval since data was not available for 1936.) Measures of total redundancy and

variable redundancy were the focus of the analysis. There was no attempt made to name canonical vectors or to analyse canonical scores. In all cases, the analysis was carried out with the BMDP6M canonical correlation program (Dixon, 1983) made available through the University of Calgary's Academic Computing Services department. The BMDP procedure was chosen because it is superior to the other two widely available programs, the Statistical Package for the Social Sciences' (SPSS) CANCORR program and SPSS MANOVA. Loading matrices, redundancy measures, complete univariate statistical summaries and tests for multicollinearity are provided routinely by BMDP while the SPSS programs are not nearly so complete. (See Tabachnick and Fidell (1983) for a summary of the capabilities and limitations of canonical correlation programs.)

5.2.1 Objectives for the canonical correlation analysis

Fundamental to the theory of epidemiologic transition outlined in Chapter Two is the notion of the cause structure of mortality. It is believed that examination of the overall pattern of death by cause for a population will reveal much about the demographic forces at work in that population. It is further asserted that significant change in the cause structure of mortality reflects, to some extent, social, cultural and environmental change (Omran,

1977; Meade, 1976; 1977; 1979). By studying the components of structural change in mortality it should be possible to be more precise, quantitatively and qualitatively, about what is meant by epidemiologic transition and whether or not study of that transition offers us any significant insights into demographic, social or environmental change. Canonical analysis was used in this study in an attempt to explore the component structure of mortality through time. This was the first objective for the canonical analysis.

The second objective for the canonical analysis was to describe the mortality transition which occurred in Newfoundland between 1937 and 1971. This involved trying to pinpoint the time periods of maximum change and those of minimum change. It was expected that high total redundancy measures and high correlation between the data matrices would indicate structural stability, while low total redundancy and low correlations would reflect structural instability. If structural instability was observed, variable redundancy values could be be examined to identify the variables creating the stability, and the ones creating the instability, in the system.

The descriptive and theoretical objectives of the canonical analysis can only be met thorough exploratory data analysis, which canonical correlation lends itself to rather well (Clark, 1977). A problem with using this

technique in data exploration is that it is often difficult, if not impossible, to state falsifiable research hypotheses. Of course, one can test the canonical correlation coefficients for significance using the chi-squared test described in Chapter Four. In this analysis, the only canonical dimensions evaluated were those whose canonical correlation coefficients were found to be statistically significant at a confidence level of 0.90; the null hypothesis is that there is no significant common dimensionality between the data sets. Beyond identification of statistically significant canonical correlation coefficients, however, there are no rigorous tests which can be used by the researcher to evaluate the significance of results. It is up to the researcher to evaluate the structural meaning of the vectors, redundancy measures and loadings. In the sense that non-trivial falsifiable hypotheses cannot be usefully stated in this kind of exploratory data analysis, the results must be viewed as descriptive and essentially non-testable in nature. Quantitative analysis can, however, help focus attention on the most important variables and lead to greater understanding of the data and the meaning of its patterns.

5.2.2 Results of the canonical correlation analysis

The results of the seven analyses are presented in Tables 5.1 to 5.7. Each table includes significant (at a confidence level of 0.90) canonical correlation coefficients, variable loadings for the dependent and independent variable sets in the significant canonical dimensions, and several redundancy measures. The variables used are percentages of total death recorded in ten selected I.C.D. chapters for the two years being considered in each analysis. With the exception of the the 1966-1971 analysis, in which I.C.D. chapter VI was dropped from the analysis (see Chapter Four), the variables used in each independent and dependent set of variables were:

1) Infectious Diseases (Abbreviated INF(year)), corresponding to I.C.D. Chapter I.

2) Cancers (Abbreviated CAN(year)), corresponding to I.C.D. Chapter II.

3) Cerebrovascular Diseases (Abbreviated STR(year)), corresponding to I.C.D. Chapter VI.

4) Circulatory Diseases (Abbreviated (HEAR(year)), corresponding to I.C.D. Chapter VII.

5) Respiratory Diseases (Abbreviated RESP(year)), corresponding to I.C.D. Chapter VIII.

6) Diseases of the Digestive System (Abbreviated DIG(year)), corresponding to I.C.D. Chapter IX.

7) Diseases of the Genito-Urinary System (Abbreviated GU(year)), corresponding to I.C.D. Chapter X.

8) Complications of Pregnancy and Childbirth (Abbreviated MAT(year)), corresponding to I.C.D. Chapter XI.

9) Congenital Anomalies and Certain Causes of Perinatal Mortality (Abbreviated PERI(year)), corresponding to I.C.D. Chapters XIV and XV.

10) Accidents and Violence (Abbreviated VIO(year)), corresponding to I.C.D. Chapter XVII.

Modifications made to some of the figures used are described in Chapter Four.

5.2.3 Interpretation of the canonical correlation analyses

Broadly stated, the question the canonical correlation attempts to answer is: to what extent, and in what way, does the structure of mortality in one year predict the structure of mortality five years later? A second question which can be posed after examining the results of the individual analyses is: what overall trends in the structure of mortality are observed between 1937 and 1971? Finally, it is necessary to consider the limitations and

TABLE 5.1

RESULTS OF CANONICAL CORRELATION, 1937-1941

1937

	Canonical Loadings				
	Canonical Vector No. 1	Canonical Vector No. 2	Canonical Vector No. 3	Canonical Vector No. 4	Variable Redundancy
INF37	-0.013	0.605	0.092	0.282	0.440
CAN37	-0.454	0.337	0.379	-0.371	0.577
STR37	-0.128	-0.155	-0.088	-0.246	0.200
HEAR37	-0.214	-0.189	0.461	-0.627	0.637
RESP37	0.203	-0.037	0.727	0.584	0.857
DIG37	0.762	-0.058	0.091	-0.149	0.608
GU37	-0.154	0.146	0.247	-0.145	0.119
MAT37	0.202	-0.328	-0.119	-0.414	0.312
PERI37	0.575	-0.130	-0.080	0.016	0.352
VIO37	0.298	-0.261	-0.369	0.602	0.611
Variance Extracted	0.137	0.075	0.112	0.158	

1941

INF41	0.489	0.291	0.133	-0.019	0.339
CAN41	-0.263	0.465	-0.032	-0.387	0.416
STR41	-0.435	-0.445	-0.066	-0.419	0.543
HEAR41	0.125	-0.219	0.652	-0.489	0.685
RESP41	0.431	-0.189	-0.358	0.640	0.711
DIG41	0.409	-0.099	0.484	0.297	0.479
GU41	0.136	-0.120	0.428	-0.624	0.557
MAT41	-0.063	0.158	0.171	0.380	0.184
PERI41	0.122	0.747	-0.341	-0.084	0.681
VIO41	0.746	-0.323	0.260	0.122	0.735
Variance Extracted	0.146	0.129	0.120	0.161	
Factor Redundancy	0.146	0.127	0.116	0.145	
Canonical Correlation	0.999	0.992	0.982	0.948	
Total Redundancy	0.534				

TABLE 5.2

RESULTS OF CANONICAL CORRELATION, 1941-1946

1941

Canonical Loadings

	Canonical Vector No. 1	Canonical Vector No. 2	Canonical Vector No. 3	Variable Redundancy
INF41	0.409	-0.026	0.217	0.213
CAN41	-0.091	-0.341	-0.398	0.274
STR41	-0.017	-0.320	0.632	0.487
HEAR41	-0.307	0.407	0.052	0.257
RESP41	0.338	0.455	0.271	0.387
DIG41	0.023	0.273	0.050	0.075
GU41	0.170	0.169	-0.091	0.063
MAT41	-0.052	0.403	-0.049	0.162
PERI41	-0.030	0.092	-0.398	0.162
VIO41	0.040	0.468	-0.031	0.216
Variance Extracted	0.041	0.108	0.085	

1946

INF46	0.332	-0.324	-0.153	0.234
CAN46	-0.622	0.055	0.361	0.514
STR46	0.345	0.072	0.453	0.322
HEAR46	0.037	0.050	0.386	0.147
RESP46	0.398	0.310	0.001	0.251
DIG46	-0.373	0.304	-0.402	0.385
GU46	-0.233	0.351	-0.155	0.177
MAT46	0.505	-0.430	-0.331	0.545
PERI46	0.353	0.023	-0.480	0.347
VIO46	-0.105	0.461	-0.120	0.231
Variance Extracted	0.136	0.082	0.104	
Factor Redundancy	0.136	0.080	0.101	
Canonical Correlation	0.999	0.988	0.983	
Total Redundancy	0.322			

TABLE 5.3

RESULTS OF CANONICAL CORRELATION, 1946-1951

1946

	Canonical Loadings		
	Canonical Vector No. 1	Canonical Vector No. 2	Variable Redundancy
INF46	-0.299	0.223	0.138
CAN46	0.274	0.332	0.182
STR46	0.627	-0.050	0.393
HEAR46	-0.020	0.162	0.026
RESP46	-0.062	-0.080	0.009
DIG46	-0.136	0.144	0.038
GU46	0.264	0.237	0.124
MAT46	-0.268	0.101	0.081
PERI46	-0.282	0.403	0.239
VIO46	0.096	-0.522	0.277
Variance Extracted	0.081	0.071	

1951

INF51	-0.470	-0.232	0.272
CAN51	0.155	0.651	0.440
STR51	0.409	0.133	0.183
HEAR51	-0.015	0.151	0.023
RESP51	0.370	0.175	0.166
DIG51	-0.442	0.330	0.301
GU51	0.194	-0.056	0.040
MAT51	0.102	-0.282	0.088
PERI51	0.102	0.002	0.010
VIO51	-0.290	-0.241	0.140
Variance Extracted	0.088	0.079	
Factor Redundancy	0.088	0.078	
Canonical Correlation	0.997	0.993	
Total Redundancy	0.166		

TABLE 5.4

RESULTS OF CANONICAL CORRELATION, 1951-1956

1951

Canonical Loadings

	Canonical Vector No. 1	Canonical Vector No. 2	Canonical Vector No. 3	Variable Redundancy
INF51	0.341	-0.237	-0.520	0.428
CAN51	0.003	0.012	0.490	0.229
STR51	-0.554	-0.376	0.483	0.668
HEAR51	-0.390	0.087	-0.006	0.158
RESP51	0.008	0.377	0.194	0.176
DIG51	0.209	0.145	-0.055	0.065
GU51	-0.219	0.158	-0.006	0.071
MAT51	0.355	-0.390	0.192	0.311
PERI51	0.510	0.312	0.335	0.461
VIO51	0.695	0.268	-0.098	0.562
Variance Extracted	0.153	0.071	0.094	

1956

INF56	0.547	-0.250	-0.340	0.470
CAN56	-0.091	0.125	-0.106	0.033
STR56	-0.358	-0.056	0.215	0.174
HEAR56	-0.344	-0.049	0.299	0.205
RESP56	0.150	0.196	-0.098	0.069
DIG56	0.483	0.266	-0.257	0.364
GU56	0.291	0.120	-0.186	0.130
MAT56	0.208	0.161	-0.449	0.259
PERI56	0.380	0.643	0.076	0.559
VIO56	0.736	-0.231	0.293	0.674
Variance Extracted	0.162	0.070	0.066	
Factor Redundancy	0.162	0.069	0.063	
Canonical Correlation	0.999	0.996	0.975	
Total Redundancy	0.294			

TABLE 5.5

RESULTS OF CANONICAL CORRELATION, 1956-1961

1956

Canonical Loadings

	Canonical Vector No. 1	Variable Redundancy
INF56	0.069	0.004
CAN56	-0.190	0.035
STR56	-0.686	0.466
HEAR56	-0.305	0.092
RESP56	0.621	0.382
DIG56	0.384	0.146
GU56	0.076	0.005
MAT56	0.331	0.108
PERI56	0.663	0.435
VIO56	0.565	0.316
Variance Extracted	0.201	

1961

INF61	0.100	0.009
CAN61	-0.649	0.417
STR61	-0.378	0.141
HEAR61	-0.250	0.061
RESP61	0.436	0.188
DIG61	0.327	0.105
GU61	0.135	0.018
MAT61	-0.125	0.015
PERI61	0.642	0.408
VIO61	0.629	0.392
Variance Extracted	0.177	
Factor Redundancy	0.175	
Canonical Correlation	0.995	
Total Redundancy	0.175	

TABLE 5.6

RESULTS OF CANONICAL CORRELATION, 1961-1966

1961

Canonical Loadings

	Canonical Vector No. 1	Variable Redundancy
INF61	-0.285	0.081
CAN61	0.295	0.086
STR61	0.254	0.064
HEAR61	0.337	0.113
RESP61	-0.024	0.001
DIG61	0.222	0.049
GU61	0.434	0.187
MAT61	-0.071	0.005
PERI61	-0.110	0.012
VIO61	-0.148	0.021
Variance Extracted	0.062	

1966

INF66	-0.157	0.024
CAN66	0.138	0.018
STR66	0.406	0.064
HEAR66	0.503	0.252
RESP66	0.152	0.230
DIG66	0.265	0.070
GU66	0.092	0.008
MAT66	-0.448	0.200
PERI66	-0.196	0.038
VIO66	-0.315	0.098
Variance Extracted	0.090	
Factor Redundancy	0.089	
Canonical Correlation	0.998	
Total Redundancy	0.089	

TABLE 5.7*

RESULTS OF CANONICAL CORRELATION, 1966-1971

1966

Canonical Loadings

	Canonical Vector No. 1	Canonical Vector No. 2	Canonical Vector No. 3	Variable Redundancy
INF66	0.040	-0.087	-0.686	0.423
CAN66	0.504	0.321	-0.135	0.356
HEAR66	0.575	-0.403	0.353	0.577
RESP66	0.206	0.252	-0.306	0.181
DIG66	-0.182	-0.439	0.376	0.334
GU66	0.246	-0.384	-0.278	0.263
MAT66	-0.527	0.652	0.280	0.731
PERI66	-0.583	0.033	0.284	0.398
VIO66	-0.141	0.293	0.117	0.111
Variance Extracted	0.151	0.132	0.121	

1971

INF71	0.028	0.432	0.036	0.176
CAN71	0.025	-0.063	0.155	0.159
HEAR71	0.471	-0.333	0.328	0.410
RESP71	-0.138	0.032	-0.450	0.197
DIG71	-0.186	-0.017	0.085	0.041
GU71	0.079	-0.710	-0.268	0.538
MAT71	-0.265	0.004	-0.157	0.088
PERI71	-0.509	0.235	0.197	0.334
VIO71	0.001	0.736	-0.240	0.508
Variance Extracted	0.067	0.156	0.059	
Factor Redundancy	0.065	0.145	0.052	
Canonical Correlation	0.980	0.965	0.939	
Total Redundancy	0.262			

* No I.C.D. Chapter VI data included

ambiguities of the results. The canonical solutions presented here appear to be sensitive to minor changes in the data, a problem common in canonical analysis, especially when the sample size is low (Pimental, 1978). It is for this reason that the results presented here must be viewed with some caution and in conjunction with the results of the cluster and discriminant analyses.

The amount of "noise" in the data is probably high since some of the I.C.D. categories used in the analysis accounted for a low percentage of total mortality. In such cases, random variation may have a significant influence on results. In some cases, the random component in the data is probably partially responsible for total redundancy figures being rather lower than might be expected. Because of the high noise level, evaluation of the stability or instability of the system must be made in conjunction with the results achieved in the other canonical analyses rather than against some arbitrarily selected level of "explanation."

The highest total redundancy figures of the seven analyses were obtained in the 1937-1941 and 1941-1946 analyses (0.534 and 0.322 respectively). It will be recalled from Chapter Four that total redundancy is a measure of the proportion of total variance of the dependent variables predicted by the canonical model

(Briggs and Leonard, 1977). The structure of mortality in 1937 predicted 53.4 percent of the variance in the structure of mortality in 1941, a substantial figure given the low sample size and high noise level of the data. Structural stability, indicating little change in the overall pattern of mortality, can be reasonably inferred for the period 1937 to 1946 on the basis of the total redundancy values.

In contrast to the 1937-1946 period, the canonical analyses of the data for 1946-1951, 1956-1961 and 1961-1966 indicate rather severe structural instability in human mortality. Total redundancy values of 0.166, 0.175 and 0.089 respectively were observed. A total redundancy measure of 0.294 for the 1951-1956 period suggests a brief stabilization and a figure of 0.262 for the 1966-1971 analysis suggests a restabilization of the structure. The evidence from the canonical analysis alone, however, cannot be considered conclusive.

It was hoped that examination of variable redundancies would make it possible to achieve greater precision in the interpretation of the results. Variable redundancy (discussed in Chapter Four) in the independent set of variables can be viewed as the contribution that each variable makes to explaining the dependent set of variables. In the dependent set, variable redundancy

indicates the proportion of each variable's variance which is explained by the canonical model (Briggs and Leonard, 1977). The variables in this analysis are unweighted, thus it is possible that variables making relatively trivial contributions to total mortality could make significant contributions to the canonical model and, in doing so, mask instability in more important categories. Examination of variable redundancy measures serves, then, as both a check on the quality of the model and as a guide to the component structure of the canonical model.

Variable redundancy values in the 1937-1941 analysis are, for the most part, quite high in both the dependent and independent set of variables. In the dependent set (the 1941 data) only maternal mortality has less than thirty percent of its variance predicted by the canonical model. Very low maternal death rates probably inject a significant random component into the values observed for this variable, possibly accounting for the low variable redundancy. The 1941-1946 model shows less stability; heart disease and infectious diseases show lower variable redundancy values than might be expected in a stable structure. Overall, however, the variable redundancy measures suggest stability, and the total redundancy of the model (0.322) suggests that the 1946 variable redundancies for heart disease and infectious disease may not be

completely incompatible with the model. To a large extent, the variable redundancies support the intepretation that 1937 to 1946 was a period of relative stability in the structure of human mortality in Newfoundland. Examination of the mortality data for Newfoundland as a whole (see Figure 3.5) during the period also supports that interpretation.

The interpretation that the period from 1946 to 1966 was a period marked by a large degree of structural instability is also supported by the variable redundancy measures, though the relative stability of the 1951-1956 period poses some interpretive difficulties. Examination of the raw data for this period suggests that this was a period of significant change in mortality structure. Indeed, the relatively high (for this study) total redundancy and somewhat peculiar variable redundancy measures in the the 1951-1956 analysis raises concerns about the over-sensitivity of the canonical correlation procedure. However, the generally low total redundancy values and varying variable redundancy figures indicate little or no persistent stability in the structure of mortality during that period, a conclusion supported by examination of the raw data for that period.

It was hoped that the canonical analysis would show the emergence of new mortality patterns in some detail,

showing, through variable redundancy measures, the emergence of new sources of stability in the structure of mortality. This proved to be unobservable in the results of the canonical correlation analyses. Observation of the mean disease patterns in the clusters of districts produced by the cluster analysis (discussed below) turned out to be more instructive than the results of the canonical analysis in showing the details of disease patterns.

The objective of pinpointing time periods of stability and instability was, for the most part, fulfilled by the canonical analysis. Additionally, the canonical analyses suggest that structural instability persisted longer than might be suspected from examination of data aggregated at the provincial level. The provincial level data suggest that a new equilibrium had been reached by about 1960. The canonical analysis, however, indicates that, at least in parts of Newfoundland, significant structural change was taking place until at least 1971. This interpretation, however, was not entirely supported by the cluster and discriminanat analyses, discussed in section 5.3 below.

The objective of identifying precisely, through time, the components of structural change must be regarded as largely unmet. The work of Haynes (1983) indicates that the use of I.C.D. chapters as variables is a valid approach to the study of differential mortality. Consequently, the

author is not inclined to suspect the validity of the classification of causes of death as the source of the failure to identify the components of change in the cause structure of mortality. That there was significant change in the cause structure of mortality in Newfoundland was established in Chapter Three and confirmed to a large extent by the canonical analyis itself. The failure to identify the components of change with canonical correlation analysis is, therefore, probably due to a high noise level in the data and a marginally acceptable sample size (Pimentel, 1978).

5.3 Cluster analysis of cause of death patterns

A single cluster analysis was carried out using the cause of death data for seven time periods aggregated into a single data matrix. 1971 data were dropped from the cluster analysis because changes in the I.C.D. structure made comparison of the 1971 data with data from the previous seven time periods problematical (Moriyama, 1966); there was no way to insure comparability with the 1937 to 1966 data. The grouping was completed with the CLUSTAN 1C cluster analysis package using Ward's hierarchical method. Four cluster solutions were selected for further analysis: seven groups, four groups, three groups and two groups. Discriminant analysis was employed to check each

solution for accurate classification and ability to discriminate amongst significantly different districts. The stepwise discriminant analysis was performed using the SPSS (Release 9.1) DISCRIMINANT program.

5.3.1 Objectives for the cluster analysis

The primary objective for the cluster analysis was to describe, as parsimoniously as possible, the space-time aspects of epidemiologic transition in Newfoundland between 1937 and 1966. To accomplish this, a cluster analysis of a data matrix 147 districts (seven observations from 1937 to 1966 on twenty-one districts) by ten I.C.D. variables was carried out. The I.C.D. variables were the same as those used in the canonical correlation analysis described above. This approach made it possible to compare directly observations from one year with observations from all other years. For example, if the mortality pattern for the district of Humber in 1946 was most similar to the mortality pattern in Fortune Bay in 1951, those two districts would be clustered together, regardless of the time difference. By charting each district through time according to its assigned cluster, it is possible to describe, based on profiles of mean variable values for each cluster, the progress of epidemiologic transition

across the twenty-one districts and through the twenty-nine years covered by the analysis.

The second major objective for the cluster analysis was to ascertain the degree to which the epidemiologic transition is expressed spatially. If the transition in Newfoundland was primarily a temporal process, very little district to district variation in cluster assignment in a given year would be expected. That is, one would expect all districts in 1937 to be assigned to the same cluster, while in 1966, all districts would again be assigned to a single, but different, cluster. On the other hand, if the transition were entirely spatial in expression, one would expect a group of districts to have the same cluster assignment through all seven time periods. It is obviously reasonable to expect that the process of change is both spatial and temporal; the important question revolves around which dimension is dominant at the intra-regional level examined in this study.

5.3.2 Results of the cluster analysis

The general cluster solution obtained with Ward's hierarchical method is presented in dendrogram form in Figure 5.1. Each district in each year is identified by a district abbreviation and the last two digits of the year of the observation. The dendrogram shows a very

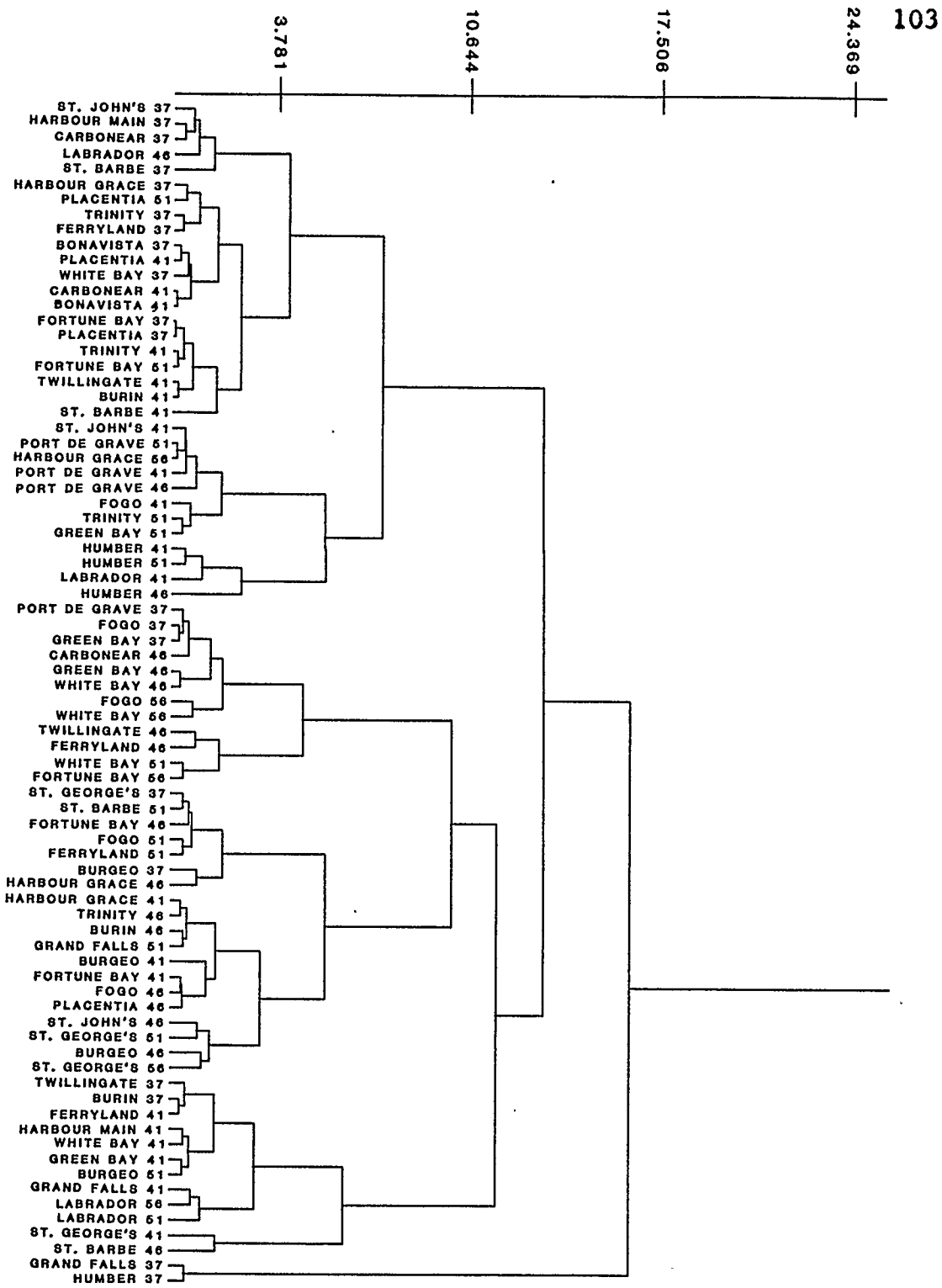

Figure 5.1 Dendrogram of the general cluster solution

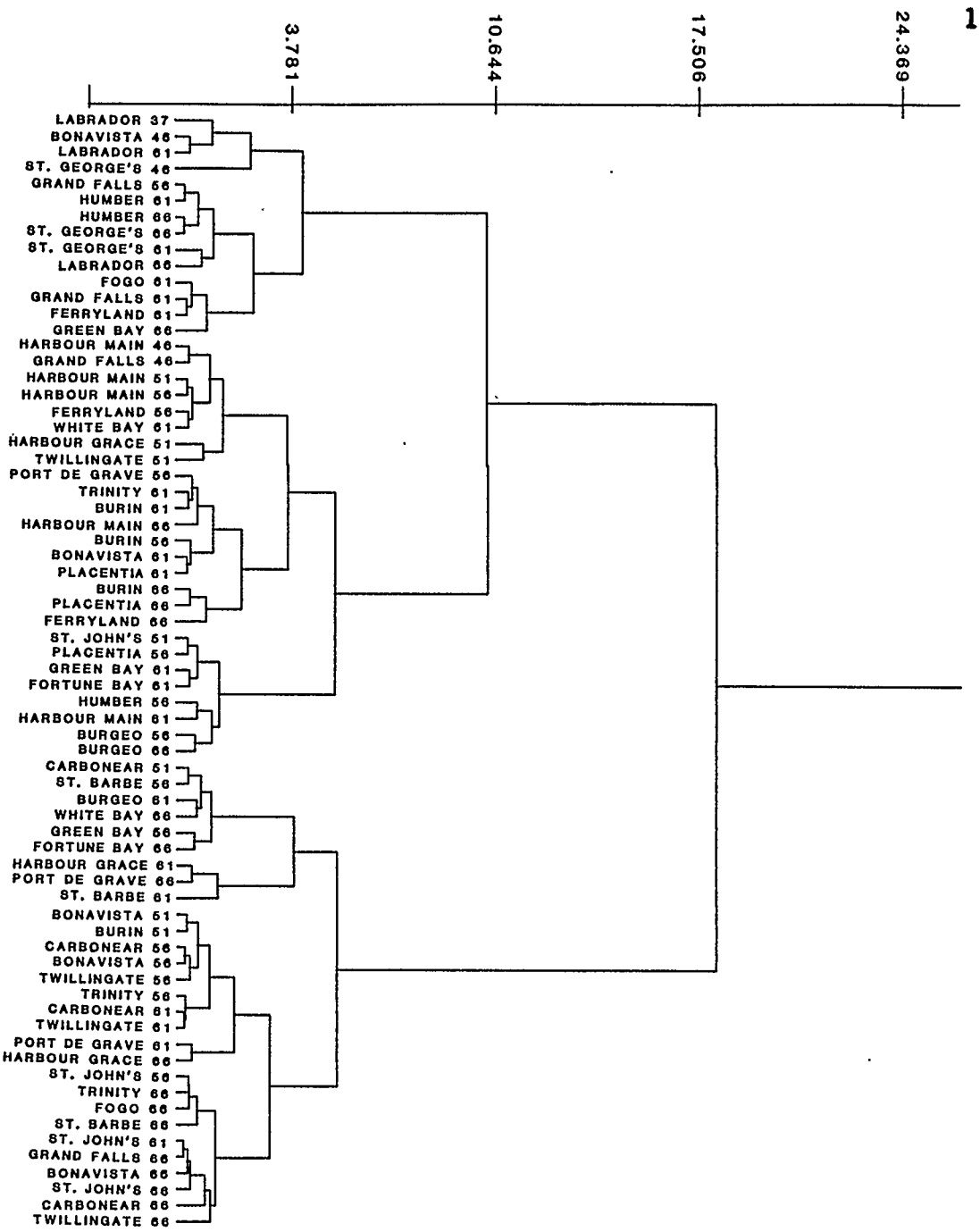

Figure 5.1 (continued)

satisfactory grouping in the sense that there is almost no "chaining" of single outlier districts into large, otherwise well-established clusters (Johnston, 1976). The only exceptions are the 1937 mortality patterns for the Humber and Grand Falls districts. Very high rates of death from digestive diseases in those districts in 1937 made them very distinct, as the dendrogram shows clearly. In evaluation of the various cluster solutions discussed below, those districts were left ungrouped and classified using the discriminant functions established by the other 145 cases. No other modifications were made to the cluster solutions in preparing the classifications for evaluation with discriminant analysis.

Without any formal analysis, three separate classification schemes are suggested by the dendrogram: two groups, three groups and four groups. The two group solution stands out the most. The error sums of sqaures for two groups is approximately 18 while the one group "solution" (100 percent information loss) is approximately 62. This translates to only 29 percent information loss for the most parsimonious cluster solution. Empirically, this result indicates that there were at least two strongly developed, and very different, mortality patterns extant in Newfoundland between 1937 and 1966. The three and four group solutions involve even less information loss,

approximately 25 percent and 21 percent, respectively. Information loss at various numbers of clusters is summarized in Figure 5.2 which summarizes the trend in error sums of squares for the general cluster solution. Ultimately, the optimum solutions were not chosen based on "cut-off" points since the graphs did not clearly identify any such solutions. Instead, selection was guided by examination of cluster profiles, theoretical considerations, interpretability and the results of discriminant analysis. The details of four separate cluster solutions are discussed below. Cluster profiles and charts of the cluster membership of each district are presented for two, three, four and seven cluster groupings. Discriminant analyses were run for each of the groupings and the results of that analysis are also discussed.

5.3.2.1 The seven cluster solution

In addition to the three solutions suggested by the dendrogram, the seven cluster grouping was evaluated. Information loss was quite low (approximately 16 percent) and the districts were well distributed amongst the seven groups (see Table 5.8a). The primary reason for evaluation of the seven group solution, however, was more theoretical than empirical. If the epidemiologic transition was a purely temporal process, one would expect district

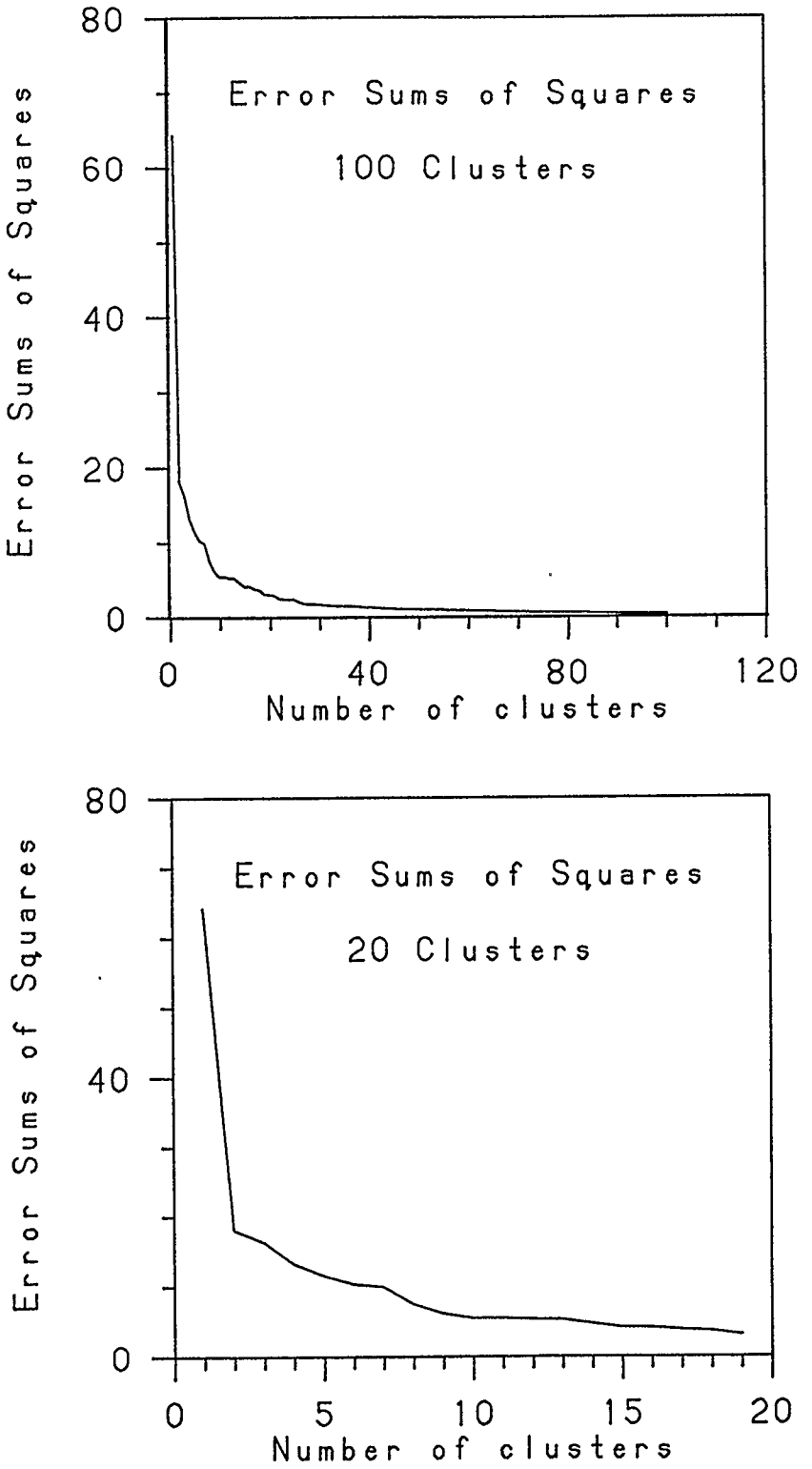

Figure 5.2 Error sums of squares of the cluster solution

TABLE 5.8a

CLUSTER PROFILES,
SEVEN CLUSTER SOLUTION

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Mean
INF	18.3	12.2	13.7	18.4	4.3	4.0	3.2	10.3
CAN	7.2	9.4	8.7	6.0	9.0	15.6	14.8	10.7
STR	12.5	17.5	8.4	12.9	10.2	10.7	17.1	12.8
HEAR	10.0	10.6	10.1	4.8	20.0	22.6	29.0	16.6
RESP	12.3	5.4	9.9	14.5	9.7	10.3	5.9	9.7
DIG	4.2	4.3	4.2	3.8	3.2	2.9	2.4	3.5
GU	3.2	1.5	1.3	1.0	1.0	1.7	2.2	2.0
MAT	0.8	1.7	0.7	3.1	0.3	0.6	0.1	0.8
PERI	7.8	7.5	12.4	9.6	15.2	11.8	9.2	10.2
VIO	3.2	3.8	4.6	4.6	15.0	7.1	5.7	5.9
Number of Districts in the Cluster	31	12	23	12	14	24	31	

TABLE 5.8b

CLUSTER MEMBERSHIP,
SEVEN CLUSTER SOLUTION

	1937	1941	1946	1951	1956	1961	1966
St. John's	1	1	3	6	7	7	7
Harbour Main	1	4	3*	6	6	6	7*
Port de Grave	2	1	1	1	6	7	7
Harbour Grace	1	3	3	6	1	7	7
Carbonear-Bay de Verde	1	1	2	7	7	7	7
Trinity	1	1	3	1	7	6	7
Bonavista	1	1	5	7	7	6	7
Fogo	2	1	3	6*	2	5	7
Twillingate	4	1	2	6	7	7	7
Grand Falls	3*	4	3*	6*	5	5	7
Green Bay	2	4	2	1	7	6	5
White Bay	1	4	2	2	2	6	7
St. Barbe	1	1	4	3	7	7	7
Humber	3*	1	1	1	6	5	5
St. George's-Port au Port	3	4	5	3	3	5	5
Burgeo and LaPoile	3	3	3	4	6	7	6
Fortune Bay-Hermitage	1	3	3	1	2	6	7
Burin	5	1	3	7	6	6	5*
Placentia	1	1	3	3*	6	6	6
Ferryland	1	4	2	3	6	6*	6
Labrador	5	1	3*	4	4	5	5

Discriminant Analysis achieved 93.7% correct classification.

*Districts reclassified by discriminant analysis.

observations from seven time periods to group into seven clusters with each cluster being dominated by district mortality patterns from one time period. Only random variation would create deviant districts. Stated formally, the seven cluster solution was evaluated to test the hypothesis that epidemiologic transition was a spatially uniform process.

The membership of the seven clusters in this analysis shows clearly that a hypothesis of spatial uniformity must be rejected. Table 5.8b, which charts the cluster membership for each district through the seven time periods, shows wide variablity in cluster membership for the various districts in any given year, especially in the first five time periods (1937 to 1956). Discriminant analysis of the grouping achieved 93.7 percent "correct" classification using the groups assigned by the cluster analysis. Districts which were reclassified in the discriminant analysis were assigned to groups suggested by the discriminant analysis and are identified in the charts with an asterisk.

It is important to point out that the numbering of the clusters is arbitrary and no progression from, for example, cluster one to cluster seven should be inferred. The meaning of a district's membership in any particular cluster can only be understood through examination of the

cluster profiles presented in Table 5.8a. Clusters one, two, three and four appear to represent "early" mortality patterns of mortality, dominated by high rates of death from infectious diseases and respiratory diseases. Clusters five, six and seven appear to represent "modern" mortality patterns where high rates of cancer and heart disease are the leading causes of death.

With the cluster profiles in mind, one can return to Table 5.8b and note that there seems to be a clear trend through time toward membership in clusters five, six and seven. This shows, as expected, that the epidemiologic transition has a strong temporal component. The more interesting results, however, have to do with the pronounced variablity of mortality patterns in space at several points in time, indicating that the epidemiologic transition is definitely not a spatially uniform process. The number of clusters in this solution, however, make characterization of some of the clusters quite difficult. The differences in the mean profiles can be noted, but generalization based on those differences occasionally presents serious interpretive problems. For example, significant differences in rates of cerebrovascular disease separate cluster two and cluster three. These differences are noted, but the meaning of those differences for a discussion of the epidemiologic transition is not clear,

especially when one considers the similarity of the cluster profiles on most of the other variables. A more parsimonious grouping may make the space-time aspects of shifts in mortality patterns easier to understand.

5.3.2.2 The four cluster solution

The four cluster solution appears to be more germane to a discussion of epidemiologic transition than the seven cluster solution described above. The cluster profiles presented in Table 5.9a show that two of the clusters, cluster one and cluster four, clearly represent early and modern mortality patterns. Cluster one, with a membership of thirty-five districts mostly from 1937 and 1941, is characterized by the highest rates of infectious and respiratory diseases and low rates of death from cancer and circulatory diseases. Cluster four, whose members are mostly from 1961 and 1966, is characterized by a pattern of low death rates from infectious diseases and high rates of cancer, cerebrovascualar and circulatory mortality. Cluster two and cluster three appear to represent transitional mortality patterns. The profile of cluster two is quite close to the profile for cluster one, though cluster one is characterized by higher rates of infectious and respiratory mortality and a higher rate of cancer mortality. Cluster three, on the other hand, is rather

TABLE 5.9a

CLUSTER PROFILES,
FOUR CLUSTER SOLUTION

	C_1	C_2	C_3	C_4	Mean
INF	18.3	14.6	4.1	3.2	10.3
CAN	7.2	8.2	13.3	14.8	10.7
STR	12.5	12.2	10.5	17.1	12.8
HEAR	10.0	8.7	21.7	29.0	16.6
RESP	12.3	9.9	10.1	5.9	9.7
DIG	4.2	4.1	3.0	2.4	3.5
GU	3.2	1.3	1.5	2.2	2.0
MAT	0.8	1.7	0.5	0.1	0.8
PERI	7.8	10.3	13.0	9.2	10.2
VIO	3.2	4.4	9.9	5.7	5.9
Number of Districts in the Cluster	35	41	37	34	

TABLE 5.9b

CLUSTER MEMBERSHIP,
FOUR CLUSTER SOLUTION

	1937	1941	1946	1951	1956	1961	1966
St. John's	1	1	1*	4*	4	4	4
Harbour Main	1	1*	3	3	3	3	4*
Port de Grave	2	1	1	1	3	4	4
Harbour Grace	1	2	2	3	1	4	4
Carbonear-Bay de Verde	1	1	2	4	4	4	4
Trinity	1	1	1*	4*	4	3	4
Bonavista	1	1	3	3*	4	3	4
Fogo	2	1	2	2	2	3	4
Twillingate	2	1	2	3	4	4	4
Grand Falls	2*	2	2*	3*	3	3	4
Green Bay	2	2	2	2*	4	3	3
White Bay	1	1*	2	2	4*	3	4
St. Barbe	1	1	2	2	3	3	3
Humber	2*	1	1	1	3	3	3
St. George's-Port au Port	2	2	3	2	2	3	3
Burgeo and LaPoile	2	2	2	2	3	4	3
Fortune Bay-Hermitage	1	2	2	1	2	3	4
Burin	2	1	2	4	3	3	3
Placentia	1	1	1*	1	3	3	3
Ferryland	1	2	2	2	3	4*	3
Labrador	2*	1	2*	2	2	3	3

Discriminant Analysis achieved 88.9% correct classification.

*Districts reclassified by discriminant analysis.

close to the cluster four profile, with slightly lower rates of cancer and circulatory mortality than cluster four, and slightly higher rates of respiratory mortality. The discriminant analysis of the four group clustering acheived 88.9 percent "correct" classification.

The space-time expression of epidemiologic transition is apparent in Table 5.9b. All districts start out, in 1937, as members of cluster one or cluster two, eventually ending up in cluster three or cluster four by 1961 or 1966. More important, the table suggests that some districts, such as St. John's and Harbour Main, could be identified as "innovators" (in terms of mortality pattern) with modern mortality patterns emerging as early as 1951. Other districts, such as Fogo, seem to lag behind the average, with modern patterns not emerging until 1961. Table 5.9b also shows that many districts remained in a "late-transitional" pattern in 1966. Districts located around Conception Bay, Trinity Bay and Bonavista Bay show modern mortality patterns somewhat earlier than the more isolated south and west coast districts, a regionalization which echoes Staveley's (1977) work showing a similar regionalization of age structures and fertility in the late nineteenth and early twentieth centuries.

Table 5.9b can be used to identify time periods of maximum change by examination of the number of districts

which switch clusters between any two adjacent time periods. Taking this approach, the period between 1951 and 1961 emerges as a period of great instability and change in mortality patterns in Newfoundland. This is especially true of the 1951 to 1956 period. In contrast, the 1937 to 1951 and 1961 to 1966 periods show more stability. This confirms, to some extent, the results of the canonical correlation analysis described above, at least as far as the earlier period is concerned. The 1961-1966 canonical analysis suggested great instability during those years while the cluster and discriminant analysis indicate substantial stability and the development of a trend towards a spatially homogeneous regional mortality pattern during those years. The results of the cluster analysis inspire considerably more confidence than the canonical correlation analysis. The numerous potential sources of spurious results in canonical analysis have been discussed above. Moreover, the cluster analysis results were confirmed by discriminant analysis at several clustering levels, and a review of the raw data (see, for example, Figure 3.5) suggests that a restabilization of mortality pattern occurred after 1961.

5.3.2.3 The three cluster solution

The three cluster solution was created by merging cluster one and cluster two from the four cluster solution. Subsequently, the mortality profiles of cluster two and cluster three in this grouping correspond exactly to the profiles of clusters three and four in the four group solution. The mortality patterns expressed in the cluster profiles of the three clusters could be named "early," "late-transitional" and "modern." Cluster two must be viewed as an essentially modern pattern (see Table 5.10a) distinguished from cluster three by somewhat higher respiratory death rates and slightly lower cancer and circulatory mortality. Discriminant analysis of the grouping achieved 91.0 percent "correct" classification.

Table 5.10b identifies roughly the same spatial and temporal trends as the the four cluster solution, especially the rapid change between 1951 and 1956. In general, however, the three cluster solution is substantially less satisfactory than the four cluster solution. In this case, more parsimonious expression of patterns involved substantial loss of interpretable information (the early-transitional cluster) and no gain in clarity.

TABLE 5.10a

CLUSTER PROFILES,
THREE CLUSTER SOLUTION

	C_1	C_2	C_3	Mean
INF	16.2	4.1	3.2	10.3
CAN	7.7	13.3	14.8	10.7
STR	12.3	10.5	17.1	12.8
HEAR	9.3	21.7	29.0	16.6
RESP	10.9	10.1	5.9	9.7
DIG	4.1	3.0	2.4	3.5
GU	2.	1.5	2.2	2.0
MAT	1.3	0.5	0.1	0.8
PERI	9.2	13.0	9.2	10.2
VIO	3.9	9.9	5.7	5.9
Number of Districts in the Cluster	72	39	36	

TABLE 5.10b

CLUSTER MEMBERSHIP,
THREE CLUSTER SOLUTION

	1937	1941	1946	1951	1956	1961	1966
St. John's	1	1	1	2	3	3	3
Harbour Main	1	1	2	2	2	2	3
Port de Grave	1	1	1	1	2	3	3
Harbour Grace	1	1	1	2	1	3	3
Carbonear-Bay de Verde	1	1	1	3	3	3	3
Trinity	1	1	1	3*	3	2	3
Bonavista	1	1	2	3	3	2	3
Fogo	1	1	1	2*	3*	2	3
Twillingate	1	1	3*	2	3	3	3
Grand Falls	1*	1	1*	1	2	2	3
Green Bay	1	1	1	1	3	2	2
White Bay	1	1	1	1	3*	2	3
St. Barbe	1	1	1	1	3	3	3
Humber	1*	1	1	2*	2	2	2
St. George's-Port au Port	1	1	2	1	2*	2	2
Burgeo and LaPoile	1	1	1	1	2	3	2
Fortune Bay-Hermitage	1	1	1	1	1	2	3
Burin	1	1	1	3	3*	2	2
Placentia	1	1	1	1	2	2	2
Ferryland	1	1	1	1	2	3	2
Labrador	1	1	1	1	2	2	2

Discriminant Analysis achieved 91.03% correct classification.

*Districts reclassified by discriminant analysis.

5.3.2.4 The two cluster solution

The two cluster solution was created by merging cluster two and cluster three from the three group solution. The two clusters clearly represented mortality patterns which could be identified as "early" and "modern" (cluster profiles are presented in Table 5.11a). Cluster one's membership is dominated by 1937 and 1941 districts while cluster two's is dominated by 1961 and 1966 observations (see Table 5.11b). Discriminanat analysis achieved 96.5 percent "correct" classification.

Even at this level of information loss, the spatial variation in the temporal trend toward modern mortality patterns is preserved with some districts clearly leading the way and others lagging behind. The two cluster solution also confirms the impression that 1951 to 1956 was a period of great change and instability in the epidemiologic environment in Newfoundland, an impression supported by Barret's (1972) work. The interpretation of trend toward a regionally homogeneous mortality pattern is also strongly confirmed, but one must be mindful of the degree of generalization and information loss operating at the two group level. Though the four cluster grouping is clearly the most interesting and interpretable solution, the two cluster solution is useful for verifying the existence of strongly expressed spatial differentiation in

TABLE 5.11a

CLUSTER PROFILES,
TWO CLUSTER SOLUTION

	C_1	C_2	Mean
INF	16.2	3.7	10.3
CAN	7.7	13.9	10.7
STR	12.3	13.3	12.8
HEAR	9.3	24.7	16.6
RESP	10.9	8.3	9.7
DIG	4.1	2.7	3.5
GU	2.1	1.8	2.0
MAT	1.3	0.3	0.8
PERI	9.2	11.4	10.2
VIO	3.9	8.1	5.9
Number of Districts in the Cluster	77	70	

TABLE 5.11b

CLUSTER MEMBERSHIP,
TWO CLUSTER SOLUTION

	1937	1941	1946	1951	1956	1961	1966
St. John's	1	1	1	2	2	2	2
Harbour Main	1	1	2	2	2	2	2
Port de Grave	1	1	1	1	2	2	2
Harbour Grace	1	1	1	2	1	2	2
Carbonear-Bay de Verde	1	1	1	2	2	2	2
Trinity	1	1	1	1	2	2	2
Bonavista	1	1	2	2	2	2	2
Fogo	1	1	1	2*	2	2	2
Twillingate	1	1	1	2	2	2	2
Grand Falls	1*	1	1*	2*	2	2	2
Green Bay	1	1	1	1	2	2	2
White Bay	1	1	1	1	2*	2	2
St. Barbe	1	1	1	1	2	2	2
Humber	1*	1	1	1	2	2	2
St. George's-Port au Port	1	1	2	1	1	2	2
Burgeo and LaPoile	1	1	1	1	2	2	2
Fortune Bay-Hermitage	1	1	1	1	1	2	2
Burin	1	1	1	2	2	2	2
Placentia	1	1	1	1	2	2	2
Ferryland	1	1	1	1	2	2	2
Labrador	1*	1	1	1	1	2	2

Discriminant Analysis achieved 96.5% correct classification.

*Districts reclassified by discriminant analysis.

the process of epidemiologic transition in Newfoundland between 1937 and 1966.

5.4 Conclusions

There were four objectives for the statistical analysis of the cause of death data. The first objective was to describe, with some precision, the components of change in the cause structure of mortality in Newfoundland between 1937 and 1971. Canonical correlation analysis was used in an attempt to identify I.C.D. variables creating both structural stability and instability through time. It was hoped that the results of the canonical analysis would lend greater quantitative and qualitative precision to discussions of epidemiologic transition. Unfortunately, the results of the canonical correlation analysis were not entirely conclusive, probably due to a small sample size and a high noise level in the data. The objective of obtaining a more precise description of temporal change in the structure of mortality through canonical correlation analysis must be regarded as largely unmet. It is possible that part of the reason for the partial failure of the canonical analysis may lie in the degree of intra-regional variation in mortality patterns, which was clearly revealed by the cluster analysis. It may be that the amount of variation was so great, and the sample size so low, that

the regional change in mortality structure was simply not observable through canonical analysis.

Greater success was realized in the application of canonical correlation analysis in meeting the second objective of the data analysis- identification of periods of stability and instability in the cause structure of mortality in the study area. Canonical analysis was used to identify the extent to which the structure of mortality in one year predicted the structure of mortality five years later. Stability in that structure was observed between 1937 and 1946 and great instability was observed between 1946 and 1966. These observations were largely confirmed by the cluster analysis, with the exception of the 1961 to 1966 period. Cluster analysis showed the emergence of a stable structure after 1961. Identification of sources of instability in the 1946 to 1966 period was not successful for reasons described above.

The third objective of the analysis of the cause of death data was was to ascertain whether or not the epidemiologic transition found significant spatial expression in Newfoundland between 1937 and 1966. The 1971 data was dropped from this part of the study because of serious problems with its comparability with data from the other seven time periods. A cluster analysis of data for all twenty-one districts in all seven time periods was

carried out to explore the spatial and temporal pattern in mortality transition. The seven cluster solution showed significant intra-regional variation within single time periods, indicating that epidemiologic transition was not a spatially uniform process in Newfoundland. Discriminant analysis confirmed both the accuracy and statistical significance of the grouping. Four, three and two cluster solutions were also explored and found to be accurate classification schemes.

Having confirmed that epidemiologic transition was expressed in significant spatial, as well as temporal, variation, the fourth objective for the analysis emerged. The final objective was to describe the space-time expression of that transition in Newfoundland. Construction of cluster profiles of mean percentages of total mortality ascribed to the ten I.C.D. chapters used in the analysis was performed in order to identify and describe the mortality pattern which defined each cluster. "Early" and "modern" patterns emerged clearly from the profiles constructed for several cluster solutions. The profiles asociated with the four cluster solution were felt to be the most relevant to a study using the theory of epidemiologic transition as a framework for analysis and discusssion of demographic change. Two cluster profiles were identified as representing "early" and "modern"

mortality patterns while the other two profiles represented transitional stages.

To examine the spatial expression of epidemiologic change, tables showing the progress of the districts from one cluster to another, through time, were constructed. These tables can be used to identify districts whose patterns anticipated epidemiologic transition and districts which lagged behind the average pattern. Identification of periods of stability and instability in mortality pattern is also facilitated by those tables. The results of the cluster analysis suggest that a fairly stable "early" mortality pattern dominated the entire region from 1937 to 1946, followed by a period of rapid change and significant spatial variation in mortality patterns between 1946 and 1961. A trend towards a stable, regionally homogeneous, modern pattern of mortality seems to have developed after 1961.

CHAPTER SIX

Discussion and Conclusions

6.1 Introduction

This study used Omran's (1971) model of epidemiologic transition as a framework for analysis and discussion of spatial and temporal trends in human mortality in Newfoundland between 1937 and 1971. That model is a descriptive model of population dynamics which focuses on the decline of mortality patterns dominated by infectious diseases, and the emergence of patterns dominated by chronic and degenerative diseases. In the Europe and North America, this transition was linked less to medical advances than to the social and economic changes Omran calls the "modernization complex." As well as being viewed as a fundamental force behind demographic change, the epidemiologic transition may be usefully understood as a change in the disease environment to which a population is exposed. The study of the process of epidemiologic transition may be a way to study the process of change in the relationship between a population and its environment.

It is not claimed that this study is any sort of definitive human ecology of Newfoundland between 1937 and

1971. On the contrary, it is a largely empirical and descriptive study of change in cause of death patterns in that region. It is, however, useful to note the human ecological themes underlying almost any study of changing disease patterns in a population. This is especially true in medical geographic studies of this kind in which the author can make no claim of any specific medical expertise. Human ecology, a traditional concern of geographers and anthropologists, provides a theoretical context for this type of study and a place to start when considering the broader implications of the analysis presented in the preceding chapters.

6.2 Epidemiologic transition as a temporal process

One of the primary research objectives of this study was to describe the temporal changes which took place in the cause of death structure in Newfoundland between 1937 and 1971. To meet this objective, data on the number of deaths in ten I.C.D. categories for twenty-one districts at five year intervals were assembled and analysed. It was established in Chapter Three, and confirmed by more detailed statistical analysis, that a shift in the structure of mortality of the type described by Omran (1977) had taken place in Newfoundland between 1937 and 1971. Infectious and respiratory diseases, especially

tuberculosis, were the the reported cause of approximately thirty percent of all deaths in the region in 1937. By 1971, those diseases accounted for less than ten percent of total mortality. The same period saw a sharp rise in rates of death from cancer and circulatory diseases. This shift was paralleled by a transformation in the age structure of general mortality. Infant and child mortality accounted for thirty-five percent of all deaths in 1937, ten percent by 1971 and five percent by 1981. There was a corresponding rise in over sixty-five mortality in the same period, a very significant shift in light of Newfoundland's rather young population age structure.

In an attempt to understand with more precision the timing and structure of the transition described above, canonical correlation analysis was employed. It was hoped that canonical analysis of the pattern of cause of death at five year intervals would help identify the components of change at different points in time. Unfortunately, the results of the analysis were inconclusive; those I.C.D. chapter variables creating structural stability and instability were not readily identifiable. Results of the cluster analysis suggest that the degree of intra-regional variation in mortality patterns may have contributed to the interpretive difficulties.

The canonical analysis did suggest that rather strong stability existed in the cause structure of mortality between 1937 and 1946, followed by severe instability between 1946 and 1966. These results were confirmed, for the most part, by the cluster analysis, which indicated that the period of greatest instability (i.e. change) in the structure of mortality was between 1946 and 1961. The profiles of the clusters indicate rapid movement during that period from an "early" pattern of death to a "modern" one. Though the cluster profiles were useful for identifying mortality patterns and change in the leading causes of death, they were not as useful as variable redundancy measures might have been in a more satisfactory canonical correlation analysis. Variable redundancy values have the potential of offering a good deal more insight into the dynamics of epidemiologic transition than the somewhat static measures provided by cluster profiles. When suitable data sets are available, it is clear that the capacity of canonical correlation analysis for shedding light on the dynamics of structural change is great. It is a statistical procedure which deserves more attention from geographers.

6.3 The epidemiologic transition in space

The second major objective of the study was to ascertain whether or not the epidemiologic transition was expressed spatially at the intra-regional level. It was found that a spatially homogeneous intra-regional pattern, which existed in the region between 1937 and 1946, was succeeded by a pattern of significant intra-regional variation between 1946 and 1961. Regional homogeneity was re-established after 1961. A spatial-evolutionary model of epidemiologic transition is suggested by these findings, a model which is resonant with developmental models suggested by Williamson (1965) and Labovitz (1965). In short, Williamson and Labovitz both proposed evolutionary models of regional economic development which used intra-regional spatial variation as measures of development. The models suggest that a period of little spatial differentiation in income, technological development and demographic pattern is followed by a period of severe intra-regional dualism. That era is followed, as "modernization" progresses, by an era of declining spatial differentiation within the region.

This study established that, in Newfoundland, epidemiologic transition can be usefully understood as a spatial, as well as temporal, evolutionary process. Results of the cluster and discriminant analyses indicate that the establishment of "modern" disease patterns

proceeded spatially and temporally along the lines suggested by Williamson's (1965) and Labovitz's (1965) models. Though both of those models are based on economic change and economic indicators, it seems clear that an analogous spatial model of epidemiologic change is suggested. It will be recalled from Chapter Two that Omran (1971) linked epidemiologic transition, in the West, to the "modernization complex." The results of this study are very suggestive of a strong link between modernization and epidemiologic transition in Newfoundland. Greater empirical support could be lent to a predictive model of epidemiologic transition by following this study up with one which traced the spatial and temporal spread of modernization in Newfoundland.

6.4 Epidemiologic transition in space-time

The cluster analysis described in Chapter Five was based on the underlying assumption that space and time are interchangeable to a degree, at least as far as the process of epidemiologic transition is concerned. The seven cluster solution indicated that the process was not spatially uniform and therefore not purely temporal in nature. Having established that a spatial component existed in the process, four, three and two cluster solutions were analyzed for their utility in describing the

spatial and temporal progress of epidemiologic transition. Behind this analysis was the assumption that districts with so-called "modern" mortality patterns in 1946 or 1951 were leading the way, exhibiting patterns which would later become a regional average. The cluster analysis simply identified districts which led and districts which lagged behind (using the interpretation of the cluster profiles from the various cluster solutions).

Harvey (1969) points out that the notion of the interchangeability of time and space, the ergodic hypothesis, underlies much geographical research.

> The ergodic hypothesis amounts to assuming that the statistical properties of a time series are essentially the same as the statistical properties of a set of observations of the same phenomenon taken over a spatial ensemble (Harvey, 1969, p. 128).

The statistical procedures used in this thesis may be a powerful tool for investigating the ergodicity of developmental processes. The procedures can be used to test theory, or for exploration of data for evidence of spatial, temporal, or spatial-temporal pattern from which theory may eventually be constructed. In this study, the canonical and cluster analyses were used largely for data exploration, but also as an empirical test of a model of demographic change.

Harvey (1969) has suggested that general theory in geography

> ...will explore the links between indigenous [to geography] theories of spatial form and derivative [from other disciplines] theories of temporal process (Harvey, 1969, p. 129)

No claim is made, in this study, that Omran's model of epidemiologic transition has been clearly linked to any existing body of geographic theory. The study is an empirical investigation of the spatial expression of a largely temporal process in an intra-regional context. As such, it lays the groundwork for a more formal and theoretical approach to the connections between geographical theory and certain aspects of demographic theory.

6.5 Limitations of the study

This study has focused primarily on the results of an exploratory analysis of cause of death data. Because of the exploratory and empirical nature of the analysis, there are some constraints imposed on the types of conclusions which can be drawn from the results. The limitations of this study originate from two sources: data quality and methodological considerations.

Uncertainties about the quality of cause of death data were discussed in some detail in Chapter Four. Concern was

expressed about the system of death registration in Newfoundland, accuracy of diagnosis and the areal units used for reporting deaths. It was necessary in several cases to recategorize some of the data for some districts and years to insure comparability through time. Problems created by changes to the I.C.D. made it necessary to drop 1971 data from the cluster analysis. To some extent, concerns about accuracy and a high noise level in the data are mitigated by the analysis of regional mortality trends presented in Chapter Three. There were no gross deviations among the districts from the outlines of the temporal pattern suggested by the regional data. Yet, it must be recognized that questions about data quality do persist. Appropriate restraint should be exercised, and the data set well understood, before using the results of this analysis to support any theoretical statements.

The limitations imposed on the study by methodological considerations have less to do with the limitations of the procedures themselves than the type of analysis they supported. It is worth reiterating that the methodology was designed for data exploration and not for formal hypothesis testing. Consequently, one must not expect that this study, alone, can make significant contributions to theory. The methodology employed was designed to facilitate a discussion that was quantitatively more

precise than other studies of epidemiologic transition (Omran, 1977; Rice, 1983), but it is still a descriptive discussion, and not a theoretical one. It was shown formally that epidemiologic transition (at the levels of data aggregation used) was not a spatially uniform process through examination of the seven cluster solution as described above (section 5.3.2.1). Moreover, the spatial expression of epidemiologic transtion, as it was observed in Newfoundland, does seem resonant with some evolutionary models of development whose focus is levels of spatial differentiation through time (Williamson, 1965; Labovitz, 1965). While useful and suggestive, such resonance does not constitute a fully supported theoretical model of spatial evolutionary epidemiologic transition. More evidence, and the more precise descriptions which can then be generated, are required before any serious, and sufficiently general, spatial theory of epidemiologic transition can be postulated. The relationship of the findings of this study to geographical and demographic theory were discussed in section 6.4 above.

6.6 Conclusions

It was believed that analysis of the social, cultural and economic correlates of epidemiologic transition in Newfoundland was beyond the scope of this study.

Additionally, changes to census district boundaries during the study period precluded that approach to a large extent. If the notion that epidemiologic transition reflects social and economic change is acceptable, the results of this study may outline the temporal and spatial dimensions of social and economic change. The striking similarity between the spatial-temporal progress of epidemiologic transition and the models of regional economic development outlined by Labovitz (1965) and Williamson (1965) suggests a connection. That similarity, however, does not constitute evidence of an irrefutable relationship between the spatial progress of "modernization" and the spatial progress of the epidemiologic transition. In particular, questions remain about the spatial and temporal levels at which these models are useful for describing and analyzing a process. The methodology outlined in Chapter Four could be of great utility in attempting to answer those questions in future studies.

This study has concerned itself with the quantitative description of epidemiologic transition in Newfoundland. Great attention has been paid to methodological and data considerations and the relationship of the results of the analysis to geographic theory; this is appropriate in an empirical discussion. In the course of such a discussion, however, it is easier than it should be to lose sight of

the human dimensions of the process which is being dissected so thououghly. In concluding, it is worth recalling that the decline in child mortality and infectious disease, the rise of degenerative disease, the aging of the population, and the spatial and temporal expression of those transformations, all reflect a rich human history and a profound shift in the demographic forces shaping the lives of individuals and populations in Newfoundland.

REFERENCES

Armstrong, R.W., 1969, Standardized class intervals and rate of computation in statistical maps of mortality. Annals of the Association of American Geographers, Vol. 59, pp. 382-390.

Audy, J.R., 1954, A biological approach to medical geography. British Medical Journal, April, 1954, pp. 960-62.

Audy, J.R., 1971, Measurement and diagnosis in health. in Paul Shepard and Daniel McKinley (eds.) Environ/Mental: Essays on the Planet as Home. Boston: Houghton Mifflin Co.

Barret, Michael J., 1972, Mortality patterns in Newfoundland, 1935-66: an introductory analysis. Unpublished B.A. Honours thesis, Department of Geography, Memorial University of Newfoundland.

Basavarajappa, K.G., 1976, Mortality Differences in Canada 1960-1962 and 1970-1972. Ottawa: Statistics Canada.

Berry, Brian J.L., 1968, A synthesis of formal and functional regions using a general field theory of spatial behavior. In Brian J.L. Berry (ed.) Spatial Analysis, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., pp. 419-428.

Blaut, J.M., 1961, Space and process. The Professional Geographer, Vol. 13, No. 4, pp. 1-7.

Briggs, Ronald and Leonard, William A., 1977, Empirical implications of advances in canonical theory. Canadian Geographer, Vol. 21, no. 2, pp. 133-147.

Brox, Ottar, 1972, Newfoundland Fishermen in the Age of Industry. Institute of Social and Economic Research, Memorial University of Newfoundland.

Butler, Gary R., 1982, Sacred and profane space: ritual interaction and process in the Newfoundland house wake. Material History Bulletin 15.

Canada Bureau of Statistics, 1949, Eleventh Census of Newfoundland and Labrador (1945). Ottawa: Canada Bureau of Statistics.

Choynowski, Mieczyslow, 1968, Maps based on probabilities. in Brian Berry and D. F. Marble (eds.) Spatial Analysis, Englewood Cliffs: Prentice Hall, pp. 180-183.

Clark, David, 1971, Urban linkage an regional structure in Wales- a systems approach. Unpublished Ph.D. dissertation, University of Wales, Department of Geography, University College of Swansea.

Clark, David, 1973, Urban linkage and regional structure in Wales: an analysis of change, 1958-68. Transactions of the Institute of British Geographers, Vol. 58, pp. 41-58.

Clark, David, 1977, Understanding Canonical Correlation Analysis. Institute of British Geographers, Study Group in Quantitative Methods, Concepts and Techniques in Modern Geography Series, No. 3.

Coulson, M.R.C, 1968, The distribution of population age structure in Kansas City. Annals of the Association of American Geographers, Vol. 68, No. 1, pp. 155-1776.

Davies, Wayne K.D., 1984, Factorial Ecology. Aldershot, England: Gower Publishing Company.

Dixon, W.J., et. al., 1983, BMDP Statistical Software. Berkeley: University of California Press.

Dubos, R., 1980, Man Adapting. New Haven: Yale University Press.

Evans, Simon E., 1985, Personal communication.

Everitt, Brian, 1974, Cluster Analysis. London: Heineman Educational Books ltd.

Field, Neil C., 1980, Temporal trends and spatial patterns of mortality in Canada. In F. Barret (ed.), Canadian Studies in Medical Geography. Geographical Monographs, No. 8. Toronto: Department of Geography, York University, pp. 32-58.

Gould, Peter, 1968, How to shoot the canonical correlation. University of Ohio, mimeograph.

Green, Paul E., 1978, Analyzing Multivariate Data. Hinsdale, Illinois: The Dryden Press.

Handcock, W. Gordon, 1977, English migration to Newfoundland. In John J. Mannion (ed.), The Peopling of Newfoundland, Social and Economic Research Papers, No. 8: St. John's, Institute of Social and Economic Research, Memorial University of Newfoundland, pp. 15-47.

Harvey, David, 1969, Explanation in Geography. London: Edward Arnold Ltd.

Haynes, Robin, 1983, The geographical distribution of mortality by cause in Chile. Social Science and Medicine, Vol. 17, No. 6, pp. 355-64.

Johnston, R.J., 1978, Multivariate Statistical Analysis in Geography. London: Longman Group Ltd.

Johnston, R.J., 1983, Philosophy and Human Geography. London: Edward Arnold Ltd.

King, Paul E., 1979, Problems of spatial analysis in geographical epidemiology. Social Science and Medicine, Vol. 13D, pp. 249-252.

Kitagawa, Evelyn M., 1964, Standardized comparisons in population research. Demography, Vol. 1, No. 1, pp. 296-315.

Kleinman, Joel C., 1977, Age-adjusted mortality indexes for small areas: applications to health planning. American Journal of Public Health, Vol. 67, pp. 834-840.

Labovitz, Sanford, 1965, Territorial differentiation and societal change. Pacific Sociological Review, Vol. 58, No. 2, pp. 70-75.

Learmonth, A.T.A., 1975, Ecological medical geography. Progress in Human Geography, Vol. 7, pp. 202-226.

Lorr, Maurice, 1983, Cluster Analysis for Social Scientists. San Francisco: Jossey-Bass Publishers.

Loveridge, John C., 1970, Forty years of medical service in Notre Dame Bay 1930-1970, also a brief history of Twillingate and Notre Dame Memorial Hospital. Twillingate Chamber of Commerce.

Mannion, John J., 1977, Settlers and traders in western Newfoundland. In John J. Mannion (ed.), The Peopling of Newfoundland, Social and Economic Research Papers, No. 8: St. John's, Institute of Social and Economic Research, Memorial University of Newfoundland, pp. 234-275.

Marshall, Carter L., 1977, Some exercises in social ecology: health, disease, and modernization in the Ryuku Islands. In M. Taghi Farvar and John P. Milton (eds.) The Careless Technology. Garden City, N.Y.: The Natural History Press, pp. 5-18.

Massam, Bryan and Nisen, William G., 1980, An analysis of mortality patterns in Montreal. In F. Barret (ed.), Canandian Studies in Medical Geography. Geographical Monographs No. 8. Toronto: Department of Geography, York University, pp. 59-66.

May, Jacques M., 1958, The Ecology of Human Disease. New York: MD Publications Inc.

McGlashan, N.D., and Harington, J.S., 1976, Some techniques for mapping mortality. South African Geographical Journal, Vol. 58, No. 1, pp. 18-24.

Meade, Melinda S., 1976, Land development and human health in west Malaysia. Annals of the Association of American Geographers, Vol. 66, No. 3, pp. 428-439.

Meade, Melinda S., 1977, Medical geography as human ecology: the dimension of population movement. The Geographical Review, Vol. 67, No. 4, pp. 379-393.

Meade, Melinda S., 1979, Cardiovascular mortality in the southeastern United States: the coastal plain enigma. Social Science and Medicine, Vol. 13D, pp.257-265.

Monmonier, Mark S., 1972, A canonical correlation model relating mortality to socioeconomic factors. International Geography, 22nd International Congress, pp. 1221-1223.

Monmonier, Mark S., and Finn, Fay Evanko, 1973, Improving the interpretation of geographical canonical correlation models. The Professional Geographer, Vol. 25, pp. 140-142.

Moriyama, Iwao M., 1966, The eighth revision of the International Classification of Diseases. American Journal of Public Health, Vol. 56, No. 8, pp. 1277-1280.

Nader, George A., 1981, The delineation of a hierarchy of nodal regions by means of a higher-order factor analysis. Regional Studies, Vol. 15, No. 6, pp. 475-492.

Newfoundland Department of Health, 1939, Report of the births, marriages and deaths for the year 1937. St. John's: Newfoundland Department of Health.

Newfoundland Department of Health, 1945, Report on the births, marriages and deaths for the year 1941. St. John's: Newfoundland Department of Health.

Newfoundland Department of Health, 1949, Report of the births, marriages and deaths in the province of Newfoundland for the year 1946. St. John's: Newfoundland Department of Health.

Nie, Norman H., et. al., 1975, Statistical Package for the Social Sciences. New York: McGraw-Hill.

Nisen, William G., 1980, Two and three dimensional cartographic representation of mortality in Montreal: an argument for computer mapping. In F. Barret (ed.), Canandian Studies in Medical Geography. Geographical Monographs, No. 8. Toronto: Department of Geography, York University, pp. 79-94.

Omran, Abdel R., 1971, The epidemiologic transition. Milbank Memorial Fund Quarterly, Vol. 49, No. 4, pp. 509-538.

Omran, Abdel R., 1977, Epidemiologic transition in the United States. Population Bulletin Vol. 32, No. 2.

Pimentel, Richard A., 1978, Morphometrics: the multivariate analysis of biological data. Biological Sciences Department, California Polytechnic State University.

Pyle, Gerald F., 1976, Introduction: foundations to medical geography. Economic Geography, Vol. 52, No. 2, pp. 95-101.

Pyle, Gerald F., 1979, Applied Medical Geography. New York: V.H. Winston and Sons.

Pyle, Gerald F. and Rees, P.H., 1971, Modelling patterns of disease and death in Chicago. Economic Geography, Vol. 47, pp. 475-488.

Rao, S.L.N., 1973, On long-term mortality trends in the United States, 1850-1968. Demography, Vol. 10, No. 3, pp. 405-419.

Rice, Gwenda H., 1983, Changing mortality patterns in North Carolina 1920 - 1972: a regional analysis. Unpublished Ph.D dissertation, University of North Carolina, Chapel Hill.

Rowe, Frederick W., 1980, A History of Newfoundland and Labrador. Toronto: McGraw Hill-Ryerson Ltd.

Sanger, Chesley W., 1977, The evolution of sealing and the spread of settlement in northeastern Newfoundland. In John J. Mannion (ed.), The Peopling of Newfoundland. Social and Economic Research Papers No. 8: St. John's, Institute of Social and Economic Research, Memorial University of Newfoundland, pp. 136-151.

Savoury, George R., 1975, The cottage hospital system in Newfoundland. Centre for Newfoundland Studies, Memorial University of Newfoundland, mimeograph.

Sneath, Peter H. and Sokal, Robert R., 1973, Numerical Taxonomy. San Francisco: W.H. Freeman and Company.

Spiegelman, Mortimer, 1965, Mortality trends for causes of death in countries of low mortality. Demography, Vol. 2, No. 1, pp. 115-125.

Statistics Canada, 1976, General Mortality 1950-1972. Ottawa: Statistics Canada.

Statistics Canada, 1980a, Mortality Atlas of Canada, Volume 1: Cancer. Ottawa: Statistics Canada.

Statistics Canada, 1980b, Mortality Atlas of Canada, Volume 2: General Mortality. Ottawa: Statistics Canada.

Statistics Canada, 1984, Mortality Atlas of Canada, Volume 3: Urban Mortality. Ottawa: Statistics Canada.

Staveley, Michael, 1970, The evolution of settlement and economy: fundamental themes in the historical geography of Newfoundland. Department of Geography mimeograph, Memorial University of Newfoundland.

Staveley, Michael, 1977, Population dynamics in Newfoundland: the regional pattern, in John J. Mannion (ed.). The Peopling of Newfoundland, Social and Economic Papers No. 8. St. John's, Institute of Social and Economic Research, Memorial Univeristy of Newfoundland, pp. 49-76.

Tabachnick, Barbara G. and Fidell, Linda S., 1983, Using Multivariate Statistics. New York: Harper and Row.

Thornton, Patricia A., 1977, The demographic and mercantile bases of initial permanent settlement in the Strait of Belle Isle. In John J. Mannion (ed.), The Peopling of Newfoundland, Social and Economic Papers No. 8: St. John's, Institute for Social and Economic Research, Memorial University of Newfoundland, pp. 152-183.

Trewartha, Glenn T., 1969, A Geography of Population: World Patterns. New York: John Wiley and Sons, Inc.

Ward Joe H., 1963, Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, Vol. 58, pp. 236-244.

Williamson, J.G., 1965, Regional inequality and the process of national development. In L. Needleman (ed.) Regional Analysis. Harmondsworth, Middlesex, England: Penguin Books Inc.

Wishart, D., 1978, CLUSTAN User Manual (Third Edition). Program Library Unit, Edinburgh University.

Woods, Robert, 1979, Population Analysis in Geography. London: Longman Group Ltd.

Woods, Robert, 1982, Theoretical Population Geography. New York: Longman Group Ltd.

World Health Organization, 1977, Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death (9th Revision), Volume 1. Geneva: World Health Organization.

Zelinsky, Wilbur, 1971, The hypothesis of the mobility transition. Geographical Review, Vol. 61 pp. 219-49.