



Université d'Ottawa • University of Ottawa



Université d'Ottawa • University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES

FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

Adrijana CORLUKA

AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M. Sc.(Epidemiology)

GRADE - DEGREE

Department of Epidemiology and Community Medicine

FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

Creation of Database Linking Contextual Small-Area Characteristics to
Successful Aging : Contributions from GIS Science

I. McDowell

DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

CO-DIRECTEUR DE LA THÈSE - THESIS CO-SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

C. Nair

A. Spasoff

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES
SUPÉRIEURES ET POSTDOCTORALES

DEAN OF THE FACULTY OF GRADUATE
AND POSTDOCTORAL STUDIES

CREATION OF A DATABASE LINKING CONTEXTUAL SMALL-AREA
CHARACTERISTICS TO SUCCESSFUL AGING:
CONTRIBUTIONS FROM GIS SCIENCE

By

Adrijana Corluka

Thesis submitted to
the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the
Master of Science degree in Epidemiology

Department of Epidemiology and Community Medicine
University of Ottawa
Ottawa, Ontario
October 2003

© 2003 A. Corluka



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-494-01452-0

Our file Notre référence

ISBN: 0-494-01452-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

ABSTRACT

While medical geography has long contributed to epidemiological studies, applications of GIS in health research are now only beginning to be realized. Individual-level variables such as genetics or lifestyle do not fully explain the phenomena of health and disease - social and physical environments play a role in determining the health of populations as well. Using individual-level data from a cohort of close to 5000 elderly Canadians, GIS was used to create a spatial database of neighborhood socio-demographic and economic characteristics, based on proximity and containment analysis, to aid in understanding how environmental context influences successful aging in Canada. The work done for this thesis resulted in the creation of the first national combined spatial and aspatial database composed of demographic, socio-economic and GIS-derived local contextual spatial data linked to individual successful aging outcome data via postal code.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Ian W. McDowell (Epidemiology) and Dr. Michael Sawada (Geography) whose advice and knowledge were instrumental in the completion of my work. Were it not for Dr. McDowell's course 'Social Aspects of Epidemiology' this thesis topic would not have been conceptualized. Dr. McDowell's logical and highly conceptual thinking, as well as his superior editing and writing skills made this experience highly educational. Dr. Michael Sawada's enthusiasm for GIS, his patience and skill as a teacher, and his encouragement in pursuing new academic experiences have made this a positive and memorable experience.

I would like to acknowledge the members of the Department of Epidemiology and Community Medicine of the University of Ottawa, most notably Fay Draper and Dr. Rama Nair, for their help and continuous support, which has been unforgettable.

A special thank you goes to Dr. Robert Spasoff, whose dedication to his students and the discipline of epidemiology are greatly inspiring.

I am indebted to my fellow epidemiology colleagues Mandy Fisher-Weselak, Michelle Turner, Christine Navarro and Tanya Guenther for their help, support and willingness to listen.

Many thanks go to Rebecca Zalatan for her sage advice, and to the members of the Laboratory of Paleoclimatology and Climatology (Department of Geography, University of Ottawa), who gave me a window to the world of physical geography.

This study would not have been possible without the generous provision of the spatial data from DMTI Spatial Inc., the availability of the Census 1996 data from the Data Liberation Initiative of Statistics Canada via the University of Ottawa, and individual-level data from the Canadian Study of Health and Aging.

I'd like to thank Dr. Elizabeth Kristjansson for supplying the outcome data and offering very helpful advice.

This research was in part financially supported by a Studentship from the Health Information Partnership of Eastern Ontario, as well as the Canadian Study of Health and Aging.

I am especially grateful to my family for understanding my commitment to the task and their generous support. They provided me with the foundation upon which my goals were based and enabled me to succeed.

Ottawa, Ontario
October, 2003
A.C.

ABSTRACT	2
ACKNOWLEDGEMENTS	3
LIST OF FIGURES	7
LIST OF TABLES	8
1. INTRODUCTION	9
1.1 GEOGRAPHY AND HEALTH	9
1.2 SOCIAL DETERMINANTS OF HEALTH AND THE PHYSICAL ENVIRONMENT	11
1.3 CONTEXTUAL AND COMPOSITIONAL EFFECTS IN EPIDEMIOLOGY	15
1.4 SUCCESSFUL AGING	20
1.5 GEOGRAPHIC INFORMATION SYSTEMS (GIS)	21
2. CONCEPTUAL FRAMEWORK	24
2.1 RESEARCH QUESTIONS	26
2.2 STATEMENT OF OBJECTIVE OF THE THESIS	26
3. DEFINING AREAS: BACKGROUND	27
3.1 POSTAL CODE STRUCTURE	27
3.2 CANADIAN CENSUS, 1996	30
3.2.2 <i>Census data</i>	35
3.2.3 CENSUS DATA AT THE CENSUS TRACT LEVEL	37
3.2.3.1 <i>Rules for census tract delineation</i>	37
4. DATA SOURCES	39
4.1 POSTAL CODE CONVERSION FILE + (PCCF+)	39
4.2 SOCIO-ECONOMIC SPATIAL DATA	41
4.2.1 <i>Enhanced Points of Interest Datafile</i>	41
4.2.2 <i>Unique Enhanced Postal Code File</i>	42
4.2.3 <i>Multiple Enhanced Postal Code File</i>	43
4.2.3.1 <i>Area covered by postal codes located at three or more points</i>	45
4.2.4 <i>Vector vs. Raster GIS</i>	47
4.3 CANADIAN STUDY OF HEALTH AND AGING (CSHA)	49
4.3.1 <i>Definition of Successful Aging</i>	50
5. METHODS	53
5.1 POSTAL CODE CONVERSION FILE + (PCCF+)	53
5.2 ARCGIS 8.2	53
5.3 SPATIAL DATA PREPARATION	55
5.3.1 <i>Proximity: Nearest essential services</i>	56
5.3.2 <i>Containment: Number of community establishments and services</i>	56
5.3.3 <i>Containment: Economic characteristics</i>	58
5.4 CREATING THE BUFFERS	59
5.5 ENHANCED POINTS OF INTEREST DATAFILE	60
5.6 SPATIAL JOINING	64
5.7 STATISTICAL ANALYSIS SOFTWARE	66

5.7.1 <i>Linking the datasets</i>	66
5.7.2 <i>Analysis</i>	66
6. RESULTS	68
6.1 PCCF+ GEOCODING RESULTS	68
6.1.1 <i>Missing and Incomplete Postal Code Data</i>	68
6.1.2 <i>Postal codes mapped to more than one location</i>	70
6.2 AREA COVERED BY POSTAL CODES: CREATION OF CONVEX HULLS	72
6.2.1 <i>Issues in using MEP for analysis of proximal establishments</i>	75
6.3 TRACKING PARTICIPANTS WHEN LINKING DATASETS	75
6.4 GIS VARIABLES IN DATASET	79
6.5 ANALYSIS	80
6.5.1 <i>Descriptive Statistical Analysis</i>	80
7. DISCUSSION	93
7.1 OBJECTIVE: CREATION OF DATABASE	93
7.1.1 <i>Limitations</i>	93
7.2 OBJECTIVE: GEOGRAPHY AND EPIDEMIOLOGY	95
7.2.1 <i>Geographies of scale</i>	97
7.2.2.1 Postal Codes	97
7.2.2.2 Residential areas	99
7.2.2.3 Spatial Autocorrelation	100
7.2.2 <i>Querying via SIC</i>	101
7.3 OBJECTIVE: PRELIMINARY STATISTICAL ANALYSIS	103
7.3.1 <i>Contextual effects discussed</i>	106
7.3.2 <i>Causality and social context's direct effects on health</i>	107
7.3.3 <i>Future Directions</i>	109
7.3.3.1 Collective Measures	109
7.3.3.2 Biophysical Environment Data	111
8. CONCLUSION	114
REFERENCES	116
GLOSSARY	124
APPENDIX I - 1996 CENSUS VARIABLES	130
APPENDIX II - BUFFERING POSTAL CODES AT A 500 METRE RADIUS	136
APPENDIX III - ENHANCED POINTS OF INTEREST	137
APPENDIX IV - PROXIMITY: NEAREST ESSENTIAL SERVICES: SIC QUERIES	141
APPENDIX V – CONTAINMENT: NUMBER OF COMMUNITY ESTABLISHMENTS AND SERVICES: SIC CODES	142
APPENDIX VI – CONTAINMENT: ECONOMIC CHARACTERISTICS - INDUSTRIAL FEATURES DETAILED	144
APPENDIX VII - CODES FOR THE INDUSTRIAL FEATURES	147

APPENDIX VIII - SPATIAL JOINING	148
APPENDIX IX - DOCUMENTATION OF SAS CODE WRITTEN FOR DATA LINKAGE	150
APPENDIX X - POTENTIAL VARIABLES CREATED AND USED FROM GENERAL SOCIAL SURVEYS OF DIFFERENT CYCLES	158

LIST OF FIGURES

<i>Figure 1.6.1 Inter-relations between group-level variables</i>	<i>19</i>
<i>Figure 3.1.1 First postal code characters in provinces or districts across Canada.....</i>	<i>28</i>
<i>Figure 3.1.2 Example of FSA locations, Saint John, New Brunswick.....</i>	<i>29</i>
<i>Figure 3.2.1 Hierarchies in the geographic organization of the census</i>	<i>32</i>
<i>Figure 4.2.3.1 Map of a postal code corresponding to multiple points.....</i>	<i>46</i>
<i>Figure 4.2.3.2 Postal code of figure 4.2.3.1 mapped using the UEP file.....</i>	<i>47</i>
<i>Figure 5.2.1 Example of an ArcMap screen.....</i>	<i>54</i>
<i>Figure 5.5.1 Query window in ArcGIS used to formulate queries</i>	<i>61</i>
<i>Figure 6.1.1.1 Tracking the results of the geocoding.....</i>	<i>69</i>
<i>Figure 6.2.1 Histogram of Convex Hull PC Areas (Range: 0-1000 km²)</i>	<i>73</i>
<i>Figure 6.2.2 Histogram of Convex Hull PC Areas (Range: 0-100 km²)</i>	<i>74</i>
<i>Figure 6.2.3 Histogram of Convex Hull PC Areas (Range: 1000-3000 km²)</i>	<i>74</i>
<i>Figure 6.3.1 CSHA Participants across Canada, mapped using ArcGIS</i>	<i>76</i>
<i>Figure 6.3.2 Tracking participants in the linking of the database</i>	<i>78</i>
<i>Figure 7.1.1.1 Map of Edmonton, Alberta indicating CSHA participant postal code locations overlain on census tracts.....</i>	<i>96</i>
<i>Figure 7.1.1.2 Map of St. John's, Newfoundland indicating CSHA participant postal code locations overlain on census tracts.</i>	<i>97</i>

LIST OF TABLES

<i>Table 1.5.1 Types of group-level variables.....</i>	<i>17</i>
<i>Table 3.2.1.1 Potential levels of census data aggregation.....</i>	<i>34</i>
<i>Table 6.1.1.1 Inconsistencies in clinical and screening postal codes</i>	<i>70</i>
<i>Table 6.1.2.1 CSHA postal codes linked to Multiple Enhanced Postal Code file</i>	<i>71</i>
<i>Table 6.4.1 ‘Proximity: Nearest essential services’ variables in the dataset</i>	<i>79</i>
<i>Table 6.4.2 ‘Containment: Number of community establishments and services’</i>	<i>79</i>
<i>Table 6.4.3 ‘Containment: Economic characteristics’ variables in the dataset</i>	<i>79</i>
<i>Table 6.5.1 Number of men and women in database successfully aging</i>	<i>80</i>
<i>Table 6.5.2 Testing normal distribution of distance variables</i>	<i>81</i>
<i>Table 6.5.3 Distances from home to selected service facilities, by sex.....</i>	<i>82</i>
<i>Table 6.5.4 Distances from home to selected service facilities, by successful/unsuccessful aging CSHA participants.....</i>	<i>83</i>
<i>Table 6.5.5 Distances from home to selected service facilities, by successful/unsuccessful aging for females</i>	<i>85</i>
<i>Table 6.5.6 Distances from home to selected service facilities, by successful/unsuccessful aging for males</i>	<i>86</i>
<i>Table 6.5.7 Distance Variables: Spearman Correlation Matrix for Successful and Unsuccessful Aging Combined.....</i>	<i>87</i>
<i>Table 6.5.8 Odds ratios for probability of successful aging by distance variables.....</i>	<i>88</i>
<i>Table 6.5.9 Count Variable Distribution for Successful Aging</i>	<i>90</i>
<i>Table 6.5.10 Income Distribution for Successful Aging</i>	<i>91</i>
<i>Table 6.5.11 Income Distribution for Successful Aging by Unique Census Tract</i>	<i>92</i>

1. INTRODUCTION

1.1 Geography and Health

As the study of the structure, features and spaces people inhabit, geography has had a long tradition in medicine and studies in health. In 400 B.C. Hippocrates taught that the environment wherein an individual lives is of utmost importance in treating patients, and that whoever wishes to investigate medicine properly should only proceed after having examined the 'airs, waters and the place where the disease occurs' (1).

Epidemiology has had a relatively shorter tradition in geography – disease maps were only first generated after the yellow fever epidemics of the late 18th and early 19th centuries and the cholera outbreaks of the 19th century (1). Perhaps the most well-known instance of geographical techniques being used in an epidemiological study is John Snow's 1854 dot map of cholera around the Broad Street water pump in London. The clustering of cholera cases in the area surrounding the well supported Snow's hypothesis that cholera was a water-borne disease, with the pump the local source of infection.¹ The

¹ Interestingly, John Snow was not actually the first person to draw maps showing the disease around the Broad Street Pump. Brody et al. (2) indicate that it might have been a London sewer engineer by the name of Cooper who was the first to plot the cholera cases on the map. Snow did not draw a map until December 1854; the first spot map was produced in September 1854 by Edmund Cooper. Cooper's investigation resulted from public complaints linking the sewers to the cholera outbreak. Rumours had circulated that sewer works had disturbed the soil of an ancient pit where bodies had been buried during the plague of 1665, and that the process had freed or generated noxious gases that caused the cholera. Cooper (a sewer engineer) plotted the locations of the sewers and the residences where deaths from cholera had occurred. From his study of the map, he had concluded that the houses closest to the sewer openings through which the gases were vented had no greater number of deaths than houses not so situated. The number of deaths on Broad Street appeared to be equally divided between the parts of the street being served by a new and old non-connected sewer. Cooper's plan accounted for 316 deaths (all those recorded in the registrar general's weekly returns to Sept. 9), far more than the 83 deaths that Snow had investigated during the first week of September. Brody et al. also write that it appears from Cooper's report that Cooper, unlike Snow, used his map as an analytical investigative tool. They go on to say that just as Snow had an agenda in drawing his maps (to implicate a contaminated pump), the sewer commission had an agenda – to clear the gully holes and sewer excavations of suspicion. Cooper called attention to the number of deaths on Broad

collaboration between geography and epidemiology continues to this day, with concepts and techniques increasingly being shared between the two disciplines over the past decade.

The most obvious use of geographic methods in epidemiology is in studying infectious diseases (e.g., HIV/AIDS) (3-5), or vector-borne diseases (e.g., malaria) (6-9). The study of the distribution of diseases such as AIDS or the influenza virus, globally or locally, frequently involves examining and tracking the diffusion of the disease, in an attempt to determine origin and control or eliminate the spread of disease. Often, attempts to understand why certain diseases seem to only occur in certain places and not others has led to new insights into the nature of the disease itself (e.g., environmental and socio-economic causes/factors) (10).

The role of place in shaping people's health experiences hasn't been confined to infectious or vector-borne diseases. Rather, the idea that location influences health has spread to epidemiology in general, and social epidemiology (the study of the role and distribution of social determinants of health and disease) in particular. Increasingly, it is being shown that lifestyle and genetics do not account for all aspects of health and it has been recognized that where you live matters (11). Debate has focused on the relative importance of people and/or place characteristics (12), with arguments being made for research focusing on places as actually experienced by people and as a context for their lives (12-16). Since the early 1990s there has been renewed interest in the ecological

Street, but apparently neither he nor any of the other commissioners noticed any peculiar concentration of spots around the location of the infamous pump. Simply plotting deaths on a map did not lead the commission to Snow's interpretation of the facts.

perspective within social epidemiology, medical geography and medical sociology, and consideration of place has become increasingly important (15). This perspective is especially relevant in planning population health interventions, as intervening at a contextual level rather than at the individual level may be more effective in improving health (17). Identifying characteristics of places that influence health provides a focal point for environmental modification (18).

1.2 Social Determinants of Health and the Physical Environment

Some researchers believe that improvements to health and population well-being will come from the growing public health movement which emphasizes the role of community and environmental health determinants (19;20). Determinants of health can be grouped into three categories (21): endogenous (including hereditary and acquired determinants of health), exogenous (physical environment, lifestyle factors, and the social environment) and health care/preventive factors (physical and mental health care, and health protection, collective prevention and health promotion).

The past decade has seen a resurgence of interest in exogenous health determinants, and the social determinants of health in particular. Childhood environment, the work environment, unemployment, patterns of social relationships, social exclusion, food and nutrition, addictive behaviour, as well as other factors, have been shown to be related to the causation of ill health, and are determinants of differences in disease rates within and between societies (22). These differences in disease rates between and within populations comprise health inequalities.

Physical and built environments also have a role to play as exogenous health determinants; they are clearly linked to socio-economic status. It is well known that housing conditions are associated with health outcomes: where we live and the house we live in are important determinants of our exposure to health risks. This relationship works independently of the causes of death and illness, which have changed over the last century, while the same gradient between socio-economic status and health has persisted. But what aspects of 'housing conditions' are linked to poor health outcomes? Research in this area declined after major problems associated with poor housing (crowding, poor ventilation, and lack of plumbing and clean water) were addressed through 'slum' clearance and development of housing projects (19). Despite these efforts, there remain a wide spectrum of diseases associated with poor housing conditions, including lead poisoning, injuries and mental illness (23). Respiratory diseases, especially, have been the focus of many studies linking poor housing with medical conditions such as asthma (24-30). Interestingly, research is showing that social determinants of health also come into play when associating different aspects of housing with health - one recent study showed that home ownership was negatively associated with mortality, whereas the presence of plumbing and heating were not associated with mortality (31).

Cohen et al.'s study on neighbourhood physical conditions and health found that neighbourhood physical deterioration, as measured by presence of boarded-up vacant housing units, was associated with premature mortality from all causes and from several specific causes, and with morbidity from sexually transmitted diseases after control for other known socio-economic correlates of these outcomes (19). The authors linked social and physical environments in their impact on health, concluding that boarded-up housing

may be related to mortality risk because of its potential adverse impact on social relationships and opportunities to engage in healthful behaviours. They suggest that homeownership is likely to contribute to neighbourhood stability and stronger social controls and recommend that neighbourhood physical conditions deserve further consideration as a potential global factor influencing health and well-being (19).

A randomized study has been undertaken to link better housing and neighbourhoods to better health. 'Moving to Opportunity for Fair Housing,' initiated in 1994 by the U.S. Department for Housing and Urban Development, was a randomized, controlled trial in which families from public housing in high-poverty neighbourhoods were moved into private housing in near-poor or non-poor neighbourhoods, with a subset remaining in public housing (32). Within five public housing authorities (in Baltimore, Boston, Chicago, Los Angeles, and New York City), randomly selected experimental groups of households with children received housing counselling and vouchers that had to be used in areas with less than 10 percent poverty. Families chosen for the experimental group received tenant-based rental assistance that helped pay their rent, as well as housing counselling to help them find and successfully use housing in low-poverty areas. Two control groups were included to test the effects of the program: one group already receiving rental assistance and another just coming into the rental assistance program. Over 4500 families have participated in the experiment. Better mental health is cited as the main success of the program: the better the neighbourhood, the more the improvement. It was concluded that allowing residents to move from highly concentrated areas of poverty leads to wider opportunities for themselves and their children.

Other studies have also found that housing and neighbourhood quality had an impact on satisfaction with the local physical environment and perceptions of safety, and social and psychological well-being (33). Cohen et al. suggest that physical conditions of a neighbourhood are not merely a consequence of social structures; rather they are likely to be in dynamic relationship with social structures and may facilitate or constrain co-operation, supervision, and feedback, all of which are critical to the adoption of low-risk health behaviours (19). It has been found that for the elderly, the use of green outdoor common areas can predict both the strength of neighborhood social ties and sense of community. Compared with individuals who had very little contact with green outdoor spaces, older adults who had greater exposure to such spaces experienced greater involvement in neighborly activities, reported stronger social relationships with friends and neighbors, and tended to be more familiar with residents in their building (34).

Neighbourhood residents may interpret physical cues of low quality neighbourhoods as the absence of social controls, which could compromise residents' feelings of order, stability, and social connectedness; thus run-down and deteriorating structures in neighbourhoods might be associated with concerns of safety (33). As a neighbourhood declines, its ability to provide psychological comfort lessens, resulting in greater emotional insecurity (33). For the elderly in particular, housing issues are of fundamental importance to their health and independence (35).

These studies suggest that the variables constituting characteristics of the physical and social environment such as housing conditions (e.g. boarded-up housing, poorly maintained dwellings), as well as neighbourhood income measures and greenness of an area could be important determinants of health and successful aging. These should

therefore be included in any contextual analysis that aims to relate social determinants to health.

The central aim of this thesis is to create a database containing information on the social and built environment, and to link this to the subset of participants in the Canadian Study of Health and Aging (discussed in future sections) who had or had not aged successfully. The database includes data on the period of construction of the dwelling, whether it was owned or rented, and whether minor or major repairs had to be done, among others.

1.3 Contextual and compositional effects in epidemiology

Predictors of health outcomes, such as social determinants of health, often contribute their effects at levels other than at the individual level. The effects of determinants of health at higher levels can be difficult to isolate without taking into consideration contextual and compositional effects.

One of the most important topics in modern geographical epidemiology involves analyzing the role of contextual and compositional effects on health (18). *Compositional explanations* draw our attention to the **characteristics of individuals** concentrated in particular places (12); *contextual explanations* draw our attention to **area or group properties**² of the local physical and social environments themselves. A third dimension, *collective explanations*, consists of the socio-cultural and historic features of communities, and is considered by some to be incorporated in the contextual explanation.

² Macintyre et al. (12) define contextual properties as ‘opportunity structures’ in the local and physical environment – existing structures which provide opportunities for use or service; for example, because too few public parks are provided, or it is difficult to access those that exist, children in deprived areas do not have the opportunity to go outside to play.

Since the collective properties of local residents are part of the context facing any individual living in that place, Macintyre et al. (12) view collective explanations as being part of contextual explanations, and would therefore not be considered distinct from contextual explanations.

Compositional effects represent the more traditional epidemiologic interpretations of ecological-level relationships (36). According to Diez-Roux (37), when group differences in an outcome are attributable to the **differences in the characteristics of the individuals** that make up the groups, they are said to result from **compositional effects**. When group differences are attributable to the **effects of group-level variables or properties**, they are said to result from **contextual effects**. Contextual effects are most often used to refer to the effect of a *derived group-level variable* on an individual-level outcome, after controlling for that variable at the individual level. However, Diez-Roux explains, the term '**contextual effects**' is also generally used to refer to the effects of **group-level variables**, and can apply to any situation involving lower level (e.g., individuals) units nested within higher level units (e.g., communities) (37).

Group-level variables can be categorized into four main types of variables: derived, environmental, integral and structural. Table 1.5.1 summarizes the types of group-level variables most often seen in epidemiologic research and when referring to contextual effects. Understanding the different types of group-level variables is useful for conceptualizing the level at which different variables act. These were summarized from Diez-Roux's glossary for multi-level analysis (37).

TABLE 1.5.1 TYPES OF GROUP-LEVEL VARIABLES

<i>Type of Group Variable</i>	<i>Definition</i>	<i>Examples</i>
Derived	- Group-level variable constructed by mathematically, summarizing the characteristics of individuals in the group - Some cannot be represented at the individual-level and specifically refer to group-level construct while other derived variable do have an individual-level counterpart - Also known as contextual variables, analytic variables and aggregate variables	Median income refers specifically to a group-level construct. Average income of a census tract is the group-level variable derived from income of individual living in the census tract.
Environmental	- Not derived by aggregating individual-level characteristics - Have group-level and individual analogues, but are used exclusively as proxies for individual-level exposures	Days of sunlight in community Air Pollution
Integral	- Are not summaries of characteristics of individuals in the group - Refers to group constructs and have no individual-level analogues - Also known as primary variables or global variables	Types of political or economic systems Population density
Structural	- Refers to relations or interrelations between members of a group - Sometimes a sub-type of integral variables	Social support networks

Derived variables are a type of group-level variable constructed by mathematically summarizing the characteristics of individuals in the group (37). Some derived variables cannot be represented at the individual-level and specifically refer to a group-level construct (e.g. median income of a census tract), while other derived variables do have an individual-level counterpart (e.g. average income of a census tract and income of an individual living in the census tract). Aggregate variables fall under the 'derived variables' category. Aggregate data are constructed by combining information from lower-level units (e.g. individuals) to create data for higher-level units (e.g. the group). Data from provinces, for example, can be aggregated to create national-level data. In the case of this thesis, census tract data were used. Data at this level are aggregated from information collected from individuals.

Derived variables are closely interrelated with integral variables. Integral variables are also a type of group-level variable, but differ in that they have no individual-level counterpart. Examples of integral variables include social disorganization or population density.

Structural variables refer to relations or interactions between groups and are sometimes considered a subtype of integral variables (37). Examples of structural variables include social networks.

For environmental variables, Diez-Roux (37) explains that in the context of ecological studies and multi-level analysis, they have been used to refer to group-level measures of physical or chemical exposures. Though they do have group level and individual level analogues, these variables are not derived by aggregating characteristics

of individuals and are used exclusively as proxies for individual level exposures, rather than as indicators of a group level property.

Figure 1.6.1 attempts to illustrate the inter-relations between the various group-level variables.

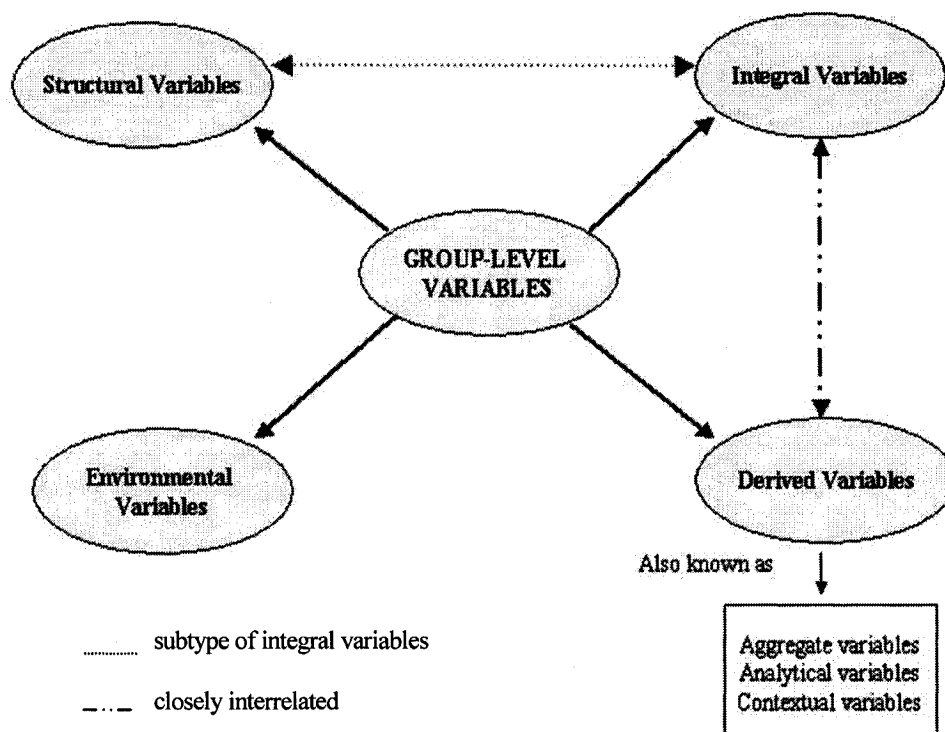


Figure 1.6.1 Inter-relations between group-level variables

Diez-Roux (14) suggests that recognizing the need to distinguish the effects of context and composition when examining area effects on health has increased reporting on various analytic methods. Contextual and multilevel analyses require data sets including both neighbourhood-level and individual-level predictors, which allow examination of neighbourhood or area effects after individual-level confounders have

examination of neighbourhood or area effects after individual-level confounders have been controlled for; these datasets would also allow for the study of individual-level factors as modifiers of area effect (14).

This thesis consists of the creation of a database that includes individual, as well as area-level characteristics, for the purpose of facilitating further research into the complexity of individual- and area-level influences on health. For this thesis, the data were linked to a subset of individuals in the Canadian Study of Health and Aging (discussed in future sections). This subset consisted of 5073 individuals from the original 10 263 in the study who were included in the successful aging component. These comprised people in the 1996 wave of the study who were living in the community and who were free of cognitive impairment. The 5073 individuals were classified as successfully or unsuccessfully aging as defined in Section 1.4.

1.4 Successful Aging

Successful aging was chosen as an outcome well suited to looking at which contextual variables are important in determining health. As successful aging is influenced by a multiplicity of determinants of health (38) it is argued that successful aging cannot be described or understood without the kind of transdisciplinary thinking that recognizes these influences on the health of elderly persons.

Definitions of successful aging vary widely, though are typically based on scores for physical, social, and psycho-cognitive functioning and on feelings of well-being (39). Generally, individuals who are aging successfully are classified as living in the community with no disability in activities of daily living (ADLs), have excellent or good

self-rated health, and are cognitively intact, as illustrated by a high score on the Mini Mental State Examination (MMSE) (40). Definitions of successful and healthy aging may be very similar, and the distinction between the two remains ill-defined. McDowell, Kristjansson and Aylesworth (41) argue that rather than use the two terms interchangeably, healthy aging should be defined in terms of physical and mental capacities, plus some level of social interaction; successful aging extends the notion of healthy aging to incorporate active lifestyles and a sense of emotional well-being and contentment into the definition.

The successful aging definition used is the following:

Successful aging extends the theme of healthy aging to include what the person does with their health. A person who is aging successfully is in control, physically and mentally, of their life. They live independently, maintain a positive outlook on their life and would be judged by others as contributing to the social good. It is an adaptive concept in which compensations can occur, such that success can occur despite imperfect health. Thus, someone may be aging successfully despite experiencing a physical disability (41).

1.5 Geographic Information Systems (GIS)

In keeping with the transdisciplinary thinking associated with successful aging and epidemiology, innovative technology developed in the field of geography was used in this study to gain more information about how the individual's context may contribute to successful aging.

Geographic data analysis has been shown to be a useful analytic tool for researchers interested in understanding spatial relationships between points or areas as well as researchers interested in the role of contextual effects (18). A major, and considerably recent, technical and methodological contribution to epidemiology from the

area of geography is that of geographic information systems science (GIS-Sci). Many of the concepts, approaches, methods and techniques of GIS-Sci are operationalized within geographic information system (GIS) software. These systems have the ability to deal with the attributes of geographic features and the relationships between geographic features and objects and thus to support complex analysis (1). An attribute is a piece of information that identifies what the feature is, describes it, and/or represents some magnitude associated with the feature. Each geographic feature has one or more associated attributes. The attributes of a census tract, for example, might include its median income, its population, the average level of education and its coordinate position on the map.

Data about real-world objects are stored in a database and dynamically linked to a digital map representing abstractions and generalizations of real world objects that can be displayed graphically. When the data in the database change, the map is updated to reflect the changes. In general, people use a GIS for four main purposes: data creation/editing, data display, analysis, and output. Each was used in this thesis.

A GIS typically stores many pieces of information about each location in a database, where each type of information is stored as a separate layer (42). Each layer contains features having similar attributes and organized by theme (e.g., provinces, social services, etc.). By overlaying these separate layers in a map and geographically linking them, relationships between different phenomena can be shown. The GIS overlay operation is similar to an overhead projector, with a series of transparencies laid upon it (43). Each transparency, or layer, contains and displays different data. One layer, for example, could contain the median income of every neighbourhood within a city.

Another layer could contain information on the location of toxic waste dumps within that same city. To determine whether poorer neighbourhoods are more likely to be found closer to polluted areas, these layers can be overlaid and analyzed spatially. Using GIS, analyses can attempt predictions as a function of the relationships between attributes, as well as attempt simulations and modeling of possible scenarios (1).

Applications of GIS in diverse areas of health research are only now beginning to be realized. Several studies have used GIS to study environmental impacts, such as air quality, on health (44;45). In addition, this technology and associated concepts of GIS-Science are being increasingly used in combination with administrative and survey data to assess availability of public health services, access to services and utilization (46-49). Using demographic information in conjunction with geographic information systems and public health is also becoming common (44;50). Through linking geographic data to descriptive data, the user is able to see patterns showing how phenomena vary by location (42). Maps generated by GIS can show the distribution of disease and their ecological relationships with cultural patterns and physical environments. The automated mapping and visualization provided within GIS systems is effective for conceptualizing causes or generating hypotheses. Displaying the locations of outliers and influential values on maps and showing variation in values over space and time can add a great deal to epidemiologic research (51).

Currently, the potential of GIS-Science capabilities is not being fully exploited in health research. This thesis attempts to help fill that gap, and works through the methodology and limitations of creating a database linking participants in a national cohort study with spatial data and Census data.

2. CONCEPTUAL FRAMEWORK

In their research on features of local areas that might promote or damage health, Macintyre, Ellaway and Cummins (12) have used five features of an area as an organizing framework:

1. ***Physical features of the environment shared by all residents of a locality:*** includes air and water quality and are likely to be shared by neighbourhoods across a wider area.
2. ***Availability of healthy environments at home, work and play:*** areas vary in their provision of decent housing, secure, safe parks, etc. These environments may not affect everyone living in an area in the same way that air and water quality do; they may affect the employed more than the unemployed, the elderly versus the non-elderly, etc.
3. ***Services provided, publicly or privately to support people in their daily lives:*** these include education, transport, health and welfare services. Again how these affect people may depend on personal circumstances (e.g., public transportation may matter more if one doesn't have a car).
4. ***Socio-cultural features of a neighbourhood:*** these include the political, economic, ethnic and religious history of a community: norms and values, the degree of community integration, levels of crime, threats to personal safety, and networks of community support.
5. ***The reputation of an area:*** how areas are perceived by their residents, by service planners and providers, by banks and investors, may influence the infrastructure of the area, the self-esteem and morale of the residents, and who moves in and out of the area.

Though not specific to successful aging, this framework provides a constructive guideline for thinking through how features of the environment are organized and what types of features of the physical and social environment may influence health.

As Macintyre et al. explain, the first three of these categories can best be seen as material or infrastructural resources. These consist of features of the physical and social environment which may promote or damage health, either directly or indirectly, through

the possibilities they provide for people to live healthy lives. The first type of feature, physical features of the environment shared by all residents in a locality, would impact health directly. Data on air and water quality for the area under study could not be obtained at the time of the data collection for this thesis but constitutes an objective for future research.

For the second feature type, indicators of the availability of healthy environments could potentially be gathered from various environmental data sources, however these were not included in the database.

The third feature type, services provided, was derived using geographic information systems. GIS was used to calculate distances to establishments providing what are considered to be essential services for an elderly individual's day-to-day living, such as banks, hospitals and grocery stores, as well as to count the amount of services and establishments provided within walking distance of the individual's residential location. These services serve as indicators of 'opportunity structures' which is what Macintyre et al. consider socially constructed and socially patterned features of the physical and social environment (12). Previous studies have already concluded that proximity to needed services is important to the well-being of the elderly (52).

Socio-cultural features of a neighbourhood, the fourth feature type, were obtained at the census tract level, from the 1996 Canadian Census.

Preliminary work was done on gathering data on the reputation of an area from the General Social Survey coordinated by Statistics Canada. Though the data for this type of feature could not be obtained in time for this thesis, this may be also be a direction for future studies and is addressed in the discussion.

Based on this conceptual framework, the purpose of this thesis is to create a database that links individual-level with area-level characteristics of the individual's socio-economic and physical environment in an attempt to establish a base for future studies into successful aging.

2.1 Research Questions

- 1) What are various approaches and issues involved in creating a database linking individual, contextual and spatial data at the national spatial scale?
- 2) What contributions can geographic information systems science make in the creation of this database?

2.2 Statement of objective of the thesis

- A) To create a database composed of individual and area-level data. This database will contain demographic, socio-economic, and spatial data. These data will be linked via postal code to individual outcome data from participants in the Canadian Study of Health and Aging (CSHA) who have aged successfully or unsuccessfully.
- B) To determine the methods for creating a database using geographic information systems and the methodological issues of a study involving data at different levels of geography.
- C) To undertake preliminary analyses of the individual and area-level data in relation to successful aging.

3. DEFINING AREAS: BACKGROUND

There are many ways to define areas in Canada; each exists for a different purpose, and not necessarily to facilitate epidemiological research. These areas include, among others, enumeration areas, municipalities and provinces. The choice of area is fundamental; it defines the context in which an individual is situated, it influences the choice and meaning of dependent variables and is important when making comparisons amongst varying health outcomes in different areas. Income inequality, for example, is a relative measure dependent on areas being compared, and patterns of disease frequency may vary depending on the geographic level being looked at. It is therefore crucial to understand the basic structure of the areas used in this thesis in order to review possible consequences of such choices on the results of epidemiological studies.

3.1 Postal Code Structure

Canadian postal codes are composed of six alternating alpha-numeric characters (e.g. A#A #A#). The first three characters represent the forward sortation area (FSA), which identifies a major geographic area in an urban or rural location. It is the unit required for the primary sorting of mail.

The first character of the forward sortation area segment identifies one of the 18 major geographic areas, provinces or districts (53). These can be seen in Figure 3.1.1, beginning from Newfoundland and Labrador ('A') and heading west towards British Columbia ('V') and the Yukon Territory ('Y').



Figure 3.1.1 First postal code characters in provinces or districts across Canada

Taken from the 'Canada Postal Guide', Section B, Chapter 3 – Addressing, Figure B.3.9, found at the Canada Post website: <http://www.canadapost.ca/personal/tools/pg/manual/b03-e.asp#c013>.

The second character of the forward sortation area is an important component of mail preparation as it identifies either:

- an **urban postal code**: numerals 1 to 9 (K1J). Urban postal codes are generally serviced by letter carrier or community mailboxes; or
- a **rural postal code**: numeral 0 (zero) (A0A). Rural postal codes are serviced by rural route drivers and/or postal outlets (53).

The third character of the forward sortation area segment (K1J) in conjunction with the first two characters, defines an exact area of a city or town or other geographic area. An example of this can be seen in Figure 3.1.2, which depicts the city of Saint John, New Brunswick.

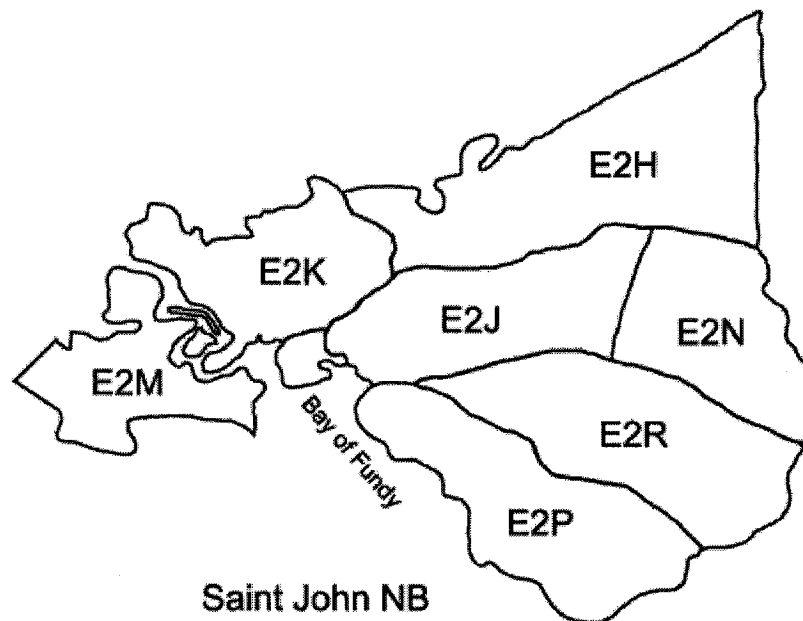


Figure 3.1.2 Example of FSA locations, Saint John, New Brunswick

Taken from the 'Canada Postal Guide', Section B, Chapter 3 – Addressing, Figure B.3.10, found at the Canada Post website: <http://www.canadapost.ca/personal/tools/pg/manual/b03-e.asp#c013>.

The second half of the postal code (#A#) is composed of the local delivery unit (LDU), which comprises the smallest delivery unit within a forward sortation area (53). There are many LDUs in each FSA. In areas with carrier delivery, groupings of LDUs form each carrier's delivery route, called a postal walk (PW), which is larger than an enumeration area but smaller than a census tract (census areas are discussed below). The LDU allows for more specific sortation within a forward sortation area. In urban areas, the last three digits may indicate a specific city block (one side of a street between two intersecting streets), a single building or, in some cases, a large-volume mail receiver. In rural areas, the last three digits, together with the forward sortation area, identify a specific community.

The postal code system is a geographic system designed by Canada Post solely to facilitate the delivery of the mail. It is quite different from the geographic systems used

by Statistics Canada. Generally, the forward sortation area (FSA), in urban areas, is about the size of four to six census tracts, but does not coincide with census tract boundaries (54). Postal geography is constantly changing, unlike census geography, which changes less frequently, from census to census.

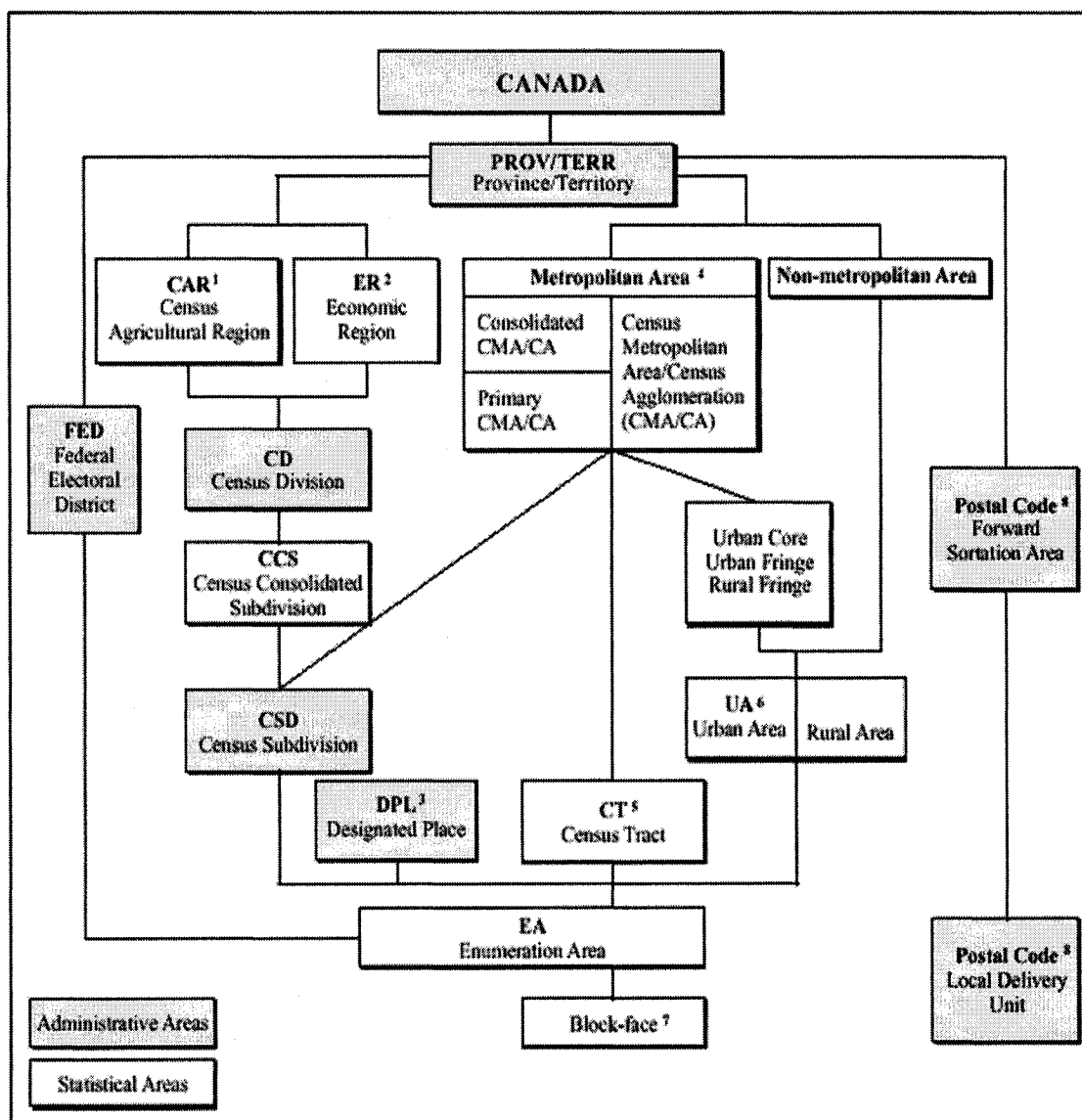
3.2 Canadian Census, 1996³

The Canadian census (officially called the Census of Population and Housing) collects data on every person in Canada based on where they live. The basic unit of measurement is the dwelling, but the census is organized in three different hierarchies of geographic units: national, metropolitan and postal code. These are depicted in Figure 3.2.1. In the national hierarchy, the geographic unit covers the entire country, while in the metropolitan (urban) hierarchy, each level of geographic unit applies only for urban centres. The postal code hierarchy exists for users requiring census data by postal code geography. The geographic hierarchies include several levels, some of which nest completely within the next larger level, while others that do not.

According to the 1996 Census Dictionary, the constitutional basis for the census originates from the requirement to apportion federal electoral representation based on population counts (54). One of the levels in the national hierarchy is the federal electoral district (FED), a federal Member of Parliament's riding. Enumeration areas are defined to respect the FED boundaries and the FEDs add together to form provinces and territories. Many provinces are already divided into official areas for regional and local

³ The 1996 Census is being used in this study, rather than the 2001 Census, as 1996 represents the mid-point of the Canadian Study of Health and Aging.

government purposes, such as counties, regional districts, regional municipalities, municipalities, townships and Indian reserves (54).



¹ Census agricultural regions in Saskatchewan are made up of census consolidated subdivisions.

² Economic regions in Ontario are made up of municipalities (census subdivisions).

³ Currently there are no designated places in Prince Edward Island, Quebec, Yukon Territory and Northwest Territories.

⁴ Five CMAs/CAs cross provincial boundaries.

⁵ All CMAs and only CAs with urban core population of 50,000 or more at the previous census have census tracts.

⁶ Five UAs cross provincial boundaries.

⁷ Only in areas covered by street network files (SNFs).

⁸ The postal code is captured as provided by the respondent on all the questionnaires for 1996. Although shown and treated as part of the geographic hierarchy, strictly speaking, it is not a geographic unit and, therefore, there is no exact relationship between postal codes and enumeration areas.

Figure 3.2.1 Hierarchies in the geographic organization of the census

From the left to the right side of the figure: national hierarchy, metropolitan hierarchy and the postal code hierarchy.

Taken from the 1996 Census Dictionary (54), page 173, Figure 20.

Census data are published for a number of standard geographic areas. These areas are either administrative or statistical, and can be located within Figure 3.2.1.

Administrative areas are defined, with a few exceptions, by federal and provincial statutes. These include:

- Provinces and territories
- Federal electoral districts (FEDs)
- Census divisions (CDs)
- Census subdivisions (CSDs)
- Designated places (DPLs)
- Postal codes

Statistical areas are defined by Statistics Canada as part of the spatial frame used to collect and disseminate census data. These include:

- Census agricultural regions (CARs)
- Economic regions (ERs)
- Census consolidated subdivisions (CCSs)
- Census metropolitan areas (CMAs)
- Census agglomerations (CAs)
- Consolidated census metropolitan areas
- Consolidated census agglomerations
- Primary census metropolitan areas (PCMAAs)
- Primary census agglomerations (PCAs)
- Census tracts (CTs)
- Urban core, urban fringe and rural fringe
- Urban areas (UAs)
- Rural areas
- Enumeration areas (EAs)

There are several levels at which census data are released and can be analysed. The smallest area of aggregation for which data were released in the 1996 Census was the enumeration area, with a population of approximately 150, while provincial populations number between 134 600 and 10 754 000 (Table 3.2.1.1).

TABLE 3.2.1.1 POTENTIAL LEVELS OF CENSUS DATA AGGREGATION

Census area	# of people per level	# of Census areas, Canada (1996)	Comments
Province	134 600 -10 754 000 ¹	10	useful for analyses of health policies
Territory	30 800 – 64 400 ¹	2 ¹	useful for analyses of health policies
Census division		288	‘county’
Census sub-division	300 – 3 million	5984	‘municipality’
Census Tract	2500-8000 (avg. 4000)	4223	homogeneous
Enumeration Area	150	49 361	may be too small to provide stable results

¹ population from the 1996 Census Handbook (55)

In trying to decide which level would most appropriately capture the context of an individual's residential area, several levels were considered. Census sub-divisions were found to be too large, given that they represent municipalities, which can become rather populous and non-indicative of ‘neighbourhood.’ Enumeration areas are established by survey administrators to optimize the efficiency of Census surveying, and contain roughly 150 people, which was considered to be rather small. The level at which it was decided to be most appropriate to represent socio-economic contextual variables was the census tract level. In addition to having a long history of being the unit of choice for analysing neighbourhoods, census tracts are designed with the intention of representing homogenous and adjacent neighbourhood units useful for aggregated analysis (54). In addition, regarding confidentiality and data suppression, census data are often suppressed for areas with low population counts, such as enumeration areas; this is less of a concern

at the census tract level. According to the *1996 Census Dictionary* (54), data for geographic areas with a population of less than 40 persons are suppressed. However, if the data are released at the postal code level or forward sortation area (FSA) level, those areas with populations below 100 persons are suppressed. If the data contain an income distribution, those areas with populations below 250 persons are suppressed. In all cases, suppressed data are included in the appropriate higher aggregate subtotals and totals.

3.2.2 Census data

The 1996 Census variables included in the database were selected based on their relevance, however remote, to health and conceptualized from review of the socio-economic literature. The reasoning for the broad inclusion of variables was that this provides more flexibility and opportunity for future users of the database. In addition, it is always easier to disregard a variable in the analysis rather than to re-create the database or subsequently add additional variables. The categories of variables extracted from the 1996 Census data can be found in Table 3.2.2.1. The data were obtained at the census tract level from the University of Ottawa library through the Data Liberation Initiative of Statistics Canada. The data (in dBASE form) were read into Excel, and the variables of interest were manually selected and appended in a new file. This file was read into SAS and linked to CSHA participants via census tract.

A complete list of the 185 census variables included in the database can be found in Appendix I.

TABLE 3.2.2.1 CENSUS DATA – TYPES OF VARIABLES

<i>Category</i>	<i>Description</i>
Age, Sex and Marital Status	contains total population for 1991 and 1996, land area in sq. km, and population by age groups, gender, and legal marital status
Families - Number, Type and Structure	contains counts of: census families in private households by family size; husband-wife families by family structure; lone parent families by parent gender
Structural Type of Dwelling and Household Size	contains counts of occupied private dwellings by structural type, and numbers of private households by household size
Labour Market Activities	contains population counts by labour force activity
Household Activities	contains population counts by hours of unpaid care to seniors
Education	contains the population counts by school attendance, and distribution of highest level of schooling
Mobility and Migration	contains population counts by mobility status for place of residence 1 year ago and 5 years ago
Sources of Income, Earnings, Total Income and Family and Household Income (I)	contains employment income by composition of total income (includes government transfer payments)
Sources of Income, Earnings, Total Income and Family and Household Income (II)	contains family income by family structure, incidence of low income, and household income by household size
Occupied Private Dwellings and Housing Costs	contains information on dwelling characteristics, household type and household characteristics

3.2.3 Census data at the census tract level

This thesis is primarily concerned with data aggregated to the level of the census tract, which falls into the metropolitan hierarchy as discussed in Section 3.1. Statistics Canada created census tracts (CTs) to equal neighbourhood-like areas of 2,500 to 8,000 people (though preferably close to 4,000) within all CMAs and CAs that contain an urban core with a population of 50,000 or more in the previous census (54).

3.2.3.1 Rules for census tract delineation

The CT initial delineation rules are ranked in order of priority and are taken from the 1996 Census Dictionary published by Statistics Canada (54):

1. CT boundaries must follow permanent and easily recognizable physical features. However, street extensions, utility or transportation easements, property lines and municipal limits may be used as CT boundaries if physical features are not in close proximity or do not exist.
2. The population of a CT should range between 2,500 and 8,000, with a preferred average of 4,000. CTs in the central business district, major commercial and industrial zones, or peripheral areas can have populations outside of this range. The maximum population of 8,000 facilitates delineation of homogeneous tracts.
3. The CT should be as homogeneous as possible in terms of socio-economic characteristics such as similar economic status and social living conditions.
4. The CT shape should be as compact as possible.
5. CT boundaries respect census metropolitan area, census agglomeration, primary census metropolitan area and primary census agglomeration as well as provincial boundaries. However, CT boundaries do not necessarily respect census subdivision boundaries.

According to the 1996 Census dictionary, the CT boundaries generally follow permanent physical features such as major streets and railway tracks and attempt to approximate cohesive socio-economic areas. Though its boundaries generally remain

constant, a CT might be split in a following census, so it can be re-aggregated to equal earlier boundaries for comparability; this characteristic, however, means that CTs do not necessarily follow CSD or CD boundaries (54). This lack of nesting occurs most frequently when neighbouring municipalities adjust their boundaries between censuses. Only at the external outline of a CMA or a CA does a CT boundary have to follow that of a CSD or CD. In practice, however, most CTs do not nest perfectly within CSDs and CDs.⁴

⁴There are a number of EAs, block-faces and some CTs and CSDs that contain no population. This reflects the distribution of population and the fact that many areas such as airports, industrial parks and remote areas, do not have residents. For a variety of reasons, non-populated areas are sometimes designated with their own boundaries even though they have no population.

4. DATA SOURCES

4.1 Postal Code Conversion File + (PCCF+)

Geocoding is the process of assigning geographic coordinates to a postal code and is often a required step in small-area health analyses. Ensuring the accuracy of geocoding is one of many steps needed to produce valid and reproducible results. Researchers have demonstrated that small geocoding errors, even those that affect less than one percent of a total dataset, can have a discernible impact on analytic results (56).

As discussed in previous sections, postal geography and census geography do not match perfectly. Most postal area boundaries and routes have very little correspondence with census boundaries even though they usually build up from the same base in urban areas (the block-face) (54). In addition, the many delivery complexities, such as post office boxes, community mail boxes (“super-boxes”), heavy volume mail users and rural routes, make it difficult to always be able to fit postal geography into Statistics Canada’s geography. It was for this reason that, to assist users, Statistics Canada created the linkage file, Postal Code Conversion File, which links postal code locations to census geography.

This file was obtained through the Data Liberation Initiative of Statistics Canada via the University of Ottawa library. Created by Russell Wilkins of the Health Analysis and Measurement Group at Statistics Canada, the PCCF+ file consists of a SAS control program and a series of reference files derived from the May 2002 Statistics Canada Postal Code Conversion File (PCCF) and the June 1996 Weighted Conversion File (WCF). Based on the postal code being input, the PCCF+ assigns a full range of

geographic identifiers, including to which census tract the postal code corresponds, as well as the postal code's latitude and longitude.

For this thesis, the PCCF+ was used to assign postal codes to census tracts (thus assigning the individuals in the CSHA to their census tracts based on their postal codes). It was also used to identify any errors in the postal codes (due to possible incorrect coding), and to correct any inconsistencies between the postal codes recorded at the screening of CSHA participants and the postal codes recorded at the subsequent clinical examination.

According to the PCCF+ manual (57), records with postal codes which serve more than one enumeration area - including most rural postal codes and several classes of urban postal codes - are assigned geographic codes based on a population-weighted random allocation among the possible codes. This produces an unbiased allocation of events in relation to the resident population. 'Geographic codes' refers to the codes assigned to provinces, census divisions, census tracts and all other areas represented geographically.

According to the PCCF+ manual, due to the nature of the postal code conversion files, a few classes of valid postal codes cannot be assigned full geographic identifiers corresponding to a place of residence or business (57). In such cases, as well as for postal codes that do not match exactly to the Postal Code Conversion File or Weighted Conversion File, the first two or three characters of the postal code are used to try to assign partial geographic identifiers to the extent possible. This takes care of many situations where the last one, two, or three characters of the postal code are invalid, but the first two or three characters are valid. Since many business and institutional

addresses are clearly identified in the PCCF+, this helps in determining whether the postal code corresponds to the participant's place of residence (or business), or whether there was a keying or reporting error.

4.2 Socio-economic Spatial Data

The ultimate purpose of using the GIS data was to derive the characteristics of services and establishments within an individual's residential area that are presumed relevant to health. However, to prepare the spatial data for incorporation into the dataset and subsequent analysis required much manipulation from the data's original form and is discussed in the Methods section.

The data featuring the services of interest come from GIS data provided by DMTI Spatial Inc. DMTI Spatial Inc. is a Canadian company that develops and markets a number of datasets compiled at the national level. The data obtained from DMTI included the Unique Enhanced Postal Code File (UEP), the Multiple Enhanced Postal Code File (MEP), and the Enhanced Points of Interest File (EPOI).

4.2.1 Enhanced Points of Interest Datafile

The 'Enhanced Points of Interest' (EPOI) ASCII file from DMTI was the principal source of the GIS data of interest for this thesis. This file had been created by geocoding the phone books and yellow page directories of each province and territory. Consisting of a national database of nearly one million Canadian businesses, buildings and recreational facilities (58), the EPOI file contains the longitudes and latitudes of business and services for each province and territory, coded using the Standard Industrial

Classification for Establishments. This database allows users to see and analyze selected point of interest data in a given geographic area, and was intended for applications in business, tourism and planning.

The establishments and services in the EPOI file were coded using the Standard Industrial Classification – Establishments (SIC-E) 1980,⁵ which is a system for classifying establishments according to their primary activity (59).

Coverage was nationwide and current as of May 2002, with each location coded in unprojected longitude and latitude, under the North American Datum 1983 (NAD83). A datum refers to a geographic coordinate system, while a map projection uses mathematical formulae to relate spherical coordinates on the globe to flat, planar coordinates. The North American Datum 1983 (NAD83) is based on earth and satellite observations, and has its origin at the earth's centre of mass. Where the origin of a geographic coordinate system is located affects the surface location of all longitude and latitude values. The NAD83 datum is used as a standard in North America.

4.2.2 Unique Enhanced Postal Code File

The community-dwelling participants of the CSHA were linked to the EPOI spatial data via postal code. The postal code file used for this particular linkage came from the 'Unique Enhanced Postal Code File' (UEP). DMTI's Unique Enhanced Postal Code File contains all of the six digit postal codes in Canada (n= 822 265), but gives only one position point per postal code (60). This is the dominant location or most

⁵, The SIC-E is being phased out and the North American Industry Classification System (NAICS) Canada 2002 (current classification system at Statistics Canada) is becoming more commonly used.

representative location for that particular postal code, and is determined by the largest number of addresses.

The Unique Postal Code file is a subset of the Multiple Enhanced Postal Code (MEP) file, described below. The UEP is more commonly used than the Multiple Enhanced Postal Code File as it allows for a one-to-one relationship for postal codes (i.e. there is at most one point in the layer for each postal code), while the Multiple Enhanced Postal Code file has a one-to-many relationship for postal codes. Because in some cases there was more than one person with the same postal code, GIS and the UEP were used to summarize the number of unique postal codes in the CSHA dataset. There were some cases of postal codes not matching the postal codes located in the UEP. This is discussed in the Results section.

4.2.3 Multiple Enhanced Postal Code File

The ‘Multiple Enhanced Postal Code File’ (MEP), like the Unique Enhanced Postal Code File, contains the location of all the postal codes in Canada. I had begun the initial GIS analysis with the MEP as it had been included in the GIS data obtained from DMTI Spatial Inc. Only after I had been using it for some time was it realized that it was the wrong file for the purposes of the thesis. The experiences with the MEP file, however, gave new insight into postal code structure, the range of area covered by postal codes, and that postal codes can be located in more than one place. In new suburban areas, for example, postal codes are now linked to a community mailbox. ‘Community mailbox’ is the Canada Post term for the large steel box where mail is delivered in new subdivisions and where each locked compartment serves one street address. According

to DMTI Spatial Inc. (60), these mailboxes can service both odd and even sides of the same street, or different streets, within a 300 m radius of the community mailbox.

Where it is necessary to represent a postal code in more than one location, the Multiple Enhanced Postal Code file would be used. The MEP file contains specific coordinates for each postal code that is associated with multiple positions, for a total of 977 738 postal code locations; of these 977 738, 822 265 are the unique dominant postal code locations found in the Unique Enhanced Postal Code file. For the example of the community mailbox, the MEP file will have a postal code record for each street range served by the community mailbox with the same postal code but with multiple locations (60). So the postal code is mapped not only to the location of the community mailbox, but also to the street location(s) of the most representative point of that particular postal code. This results in a many-to-one relationship for postal codes (i.e., many spatial coordinates exist for a single postal code). Rural locations are also commonly represented as multiple position postal codes (60). Often, mapped rural postal codes represent the location of the place where the mail is sorted and not the final place of delivery. This can mean that one postal code, with multiple positions, may represent several small rural towns (60).

Preliminary analyses showed that the postal codes contained within the multiple-enhanced postal code file did not have a one-to-one relationship with all of the CSHA participant postal codes, so the Unique Enhanced Postal Code file was used to link CSHA participant postal codes to their geographic locations. This is discussed in the Results section.

4.2.3.1 Area covered by postal codes located at three or more points

Concerns arose from working with the MEP regarding the precise location of the CSHA participant's postal code, seeing as 1023 CSHA postal codes mapped to more than one location (discussed later in Section 6.2). In order to further explore the many-to-one relationship of postal codes, the size of the area covered by postal codes which could be mapped to three or more locations was determined using ArcGIS. The reason for looking at postal codes corresponding to three or more points is that at least three points are needed to create a polygon, whose area can then be measured. There were 464 postal codes that mapped to three or more places. The area of a polygon was calculated by creating a convex hull of the postal code in GIS. A convex hull of a set of points (in this case the locations of the postal codes are represented as points on a map) is the smallest polygon containing the mapped postal code locations. It can be visualized as the shape made by a rubber band wrapped around 'outside' points, yet still including any points found within the polygon.

Figure 4.2.3.1 depicts a convex hull created from the location points of a postal code in the Ottawa region. The outside edges of the convex hull/polygon are created by lines connecting each of the outside points (where each point corresponds to the same postal code). Additional postal code locations can be situated within the convex hull/polygon as well. At the widest point of the polygon, the distance between postal code locations can range between eight kilometres (east to west) and 10 kilometres (north to south).

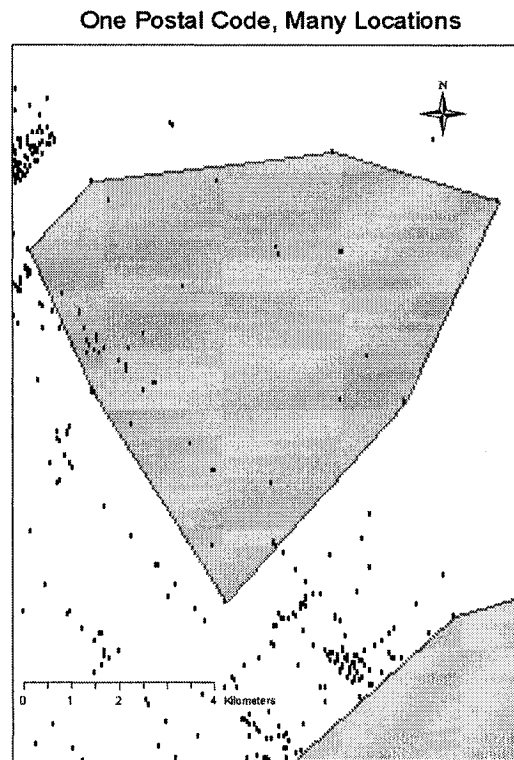


Figure 4.2.3.1 Map of a postal code corresponding to multiple points
 The polygon corresponds to the convex hull which was created using ArcGIS.

The obvious difficulty in using the Multiple Enhanced Postal Code file with its many-to-one relationship is in knowing where in this large area the participant is located. It is also impossible to tell, without more information on the participant (e.g. the actual coordinates of their place of residence), which point corresponded most closely to the participant's location.

Figures 4.2.3.1 and 4.2.3.2 contrast the results of using the MEP and UEP files respectively, to map postal code points.

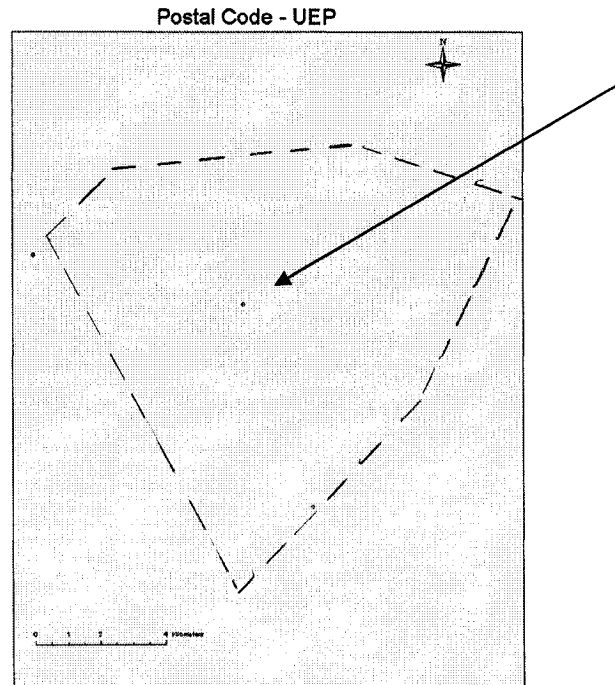


Figure 4.2.3.2 *Postal code of figure 4.2.3.1 mapped using the UEP file*

Figure 4.2.3.2 displays the single most representative point of the location of the same postal code depicted in Figure 4.2.3.1.

4.2.4 Vector vs. Raster GIS

There are two main types of GIS software systems and according to the research question under study, a choice is required as to the most appropriate system. Two GIS software systems exist because there are two approaches to representing space in a GIS: the field view and discrete view of reality. These are embodied within raster or vector GIS systems, respectively.

The vector or discrete model represents observable 'real-world' features such as points, lines and areas. Using various combinations of (X, Y) coordinates, a line may represent a road, and a connected set of lines may create a polygon that represents a city

block or a census tract. The vector approach tends to produce digital spatial data that is both efficient and compact, which is a primary advantage of a vector GIS system. The vector model is, however, not well suited to the representation of field (continuous) data.

Unlike representing reality as discrete features, the raster model represents space as subdivided by a continuous surface of cells, as if covered by a grid. The location of each object is defined by the row and column position of the cells it occupies (1). The spatial units are the cells, each with a unique location. Since each cell can have a different digital value, and cells can be very small (a few metres across), they can represent an enormous complexity of spatial variability; they are easily overlaid on each other as surfaces (1). Satellite images of earth, for example, are digitally transmitted, stored and analyzed as rasters. The advantage to using the raster model is that it is good for representing continuous data and data of great spatial variability (e.g., temperature, precipitation, density and intensity), which it can represent using a grid system. The disadvantage of the raster model is that data are not stored very compactly and thus requires a lot of memory for storage. In addition, the raster data structure contains little information on topological relationships; whereas vector models can spatially link geographic features (e.g., can link a road [line] to the nearest grocery store [point]), the raster model is not well suited to such operations.

Given that this thesis dealt with the discrete locations of participants in relation to surrounding discrete spatial objects (e.g., stores, hospitals) and required the calculation of distances between points, the vector approach was the most appropriate to use.

4.3 Canadian Study of Health and Aging (CSHA)

The individual-level successful aging outcome data were drawn from the Canadian Study of Health and Aging (CSHA). To examine the influence of community characteristics on successful aging, analyses used a previously defined subset of participants drawn from community sample from Cycles 1 and 2 of the study.

The CSHA, a 10-year longitudinal study, was originally designed to study the epidemiology of dementia amongst the elderly, and a wide range of data was collected in order to help identify dementia risk factors and describe patterns of providing care for afflicted individuals. The CSHA has also described patterns of disability, frailty and successful aging, and has recorded utilization of health services for different diagnostic groups. The CSHA involved 10 263 people aged 65 or over, sampled from 36 communities across Canada. Representative samples were drawn from the community and from institutions, and participants were assessed at 5-yearly intervals: in 1991, 1996, and for a final time in 2001.

The sample was designed to provide regional and national prevalence estimates for dementia by age and sex. Thirty-six sampling areas were used in a stratified cluster design with optimal allocation. The study involved 9008 people from the community and 1255 from long-term care institutions, all aged 65 or older. The community-sampling frame was based on computerized records of the provincial universal health insurance plans, with the exception of Ontario. A random sample, stratified by age, was drawn in each study area.

Eighteen study centres were grouped administratively into five regions of Canada: the Atlantic region (4 provinces), Quebec, Ontario, the Prairies (3 provinces) and British

Columbia. The study design provided prevalence estimates for the five regions of Canada.

The Canadian Study of Health and Aging incorporated community interviews and a clinical examination to diagnose dementia. For the community interview, participants who lived in the community and who form the study sample for this thesis, were interviewed in their homes. The main purpose of the community interview, which took place in the home of the community participant, was to administer the Modified Mini-Mental State exam (3MS) cognitive screening test to identify those who appeared to have cognitive impairments that would merit fuller clinical examination. The community interview also collected social and demographic information, and assessed disability and healthy aging.

The clinical examination consisted of a combination of medical and neuropsychological tests administered to the patient; after the physician and neuropsychologist had independently made preliminary diagnoses, they met with the nurse to reach a consensus diagnosis. These diagnoses formed the basis for estimating the prevalence and incidence of dementia. In addition to dementia, however, the study covered other topics including successful aging.

4.3.1 Definition of Successful Aging

The definition and criteria for 'successful aging' come from the as yet unpublished paper by McDowell, Kristjansson and Aylesworth (41). The successful aging questions were administered to all those who scored above 65 on the Modified Mini-Mental State Examination (61). This threshold was chosen to exclude those who

suffered from moderate to severe cognitive impairment. Data were taken from self-report questionnaires and a number of scales were created.

For the creation of the Mental health scale, depression and anxiety were assessed from the SF-36 (62;63) and cognition was recorded using the Modified Mini-Mental State exam (3MS) (61). Psychological well-being was recorded with the Ryff⁶ scale (64-66) and life satisfaction was measured with the Delighted-Terrible⁷ scale (67-69). Engagement with Life was measured with a four-point response scale, participants having been asked “How often in the past summer did you...?” with items including visiting friends, gardening, walking, etc. This was called the Activities scale.

For physical function measurements, Activities of Daily Living (ADL) (70) and the Timed ‘Up and Go’ (TUG) test (71;72) were scaled. Self-rated health was recorded on a five-point scale: in response to the question “How would you say your health is these days?” answers ranged from “Excellent” to “Very poor”. The Pain scale was constructed from self-reported answers to the statement “A little bit of interference” with mood, mobility and activities. Extensive preliminary work was done on scale development and exploratory factor analysis on individual scales to assess item quality and dimensionality.

The operational definition of successful aging was based on a factor analysis of scores from the ADL, TUG, Pain, Self-rated health, and Mental health scales, as well as the Ryff, Terrible-Delighted and Activities scales. Individuals not living in the

⁶ The Ryff scale, created in 1989 presents an alternative approach to studying successful aging that is based on the convergence of lifespan developmental theories, clinical theories of personal growth, and mental health perspectives. Six criteria of well-being are identified: self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life, and personal growth (64).

⁷ The Delighted-Terrible scale is a subjective scale of well-being validated by Andrews and Withey in 1976.

community were excluded. Cut points for the operational definition of successful aging classified 20% of the community-dwelling sample as having aged successfully, with the remaining 80% rated as not having aged successfully. The choice of 20% was arbitrary, and arose because the purpose of McDowell et al.'s work was to compare alternative formulations of healthy and successful aging. One formulation used a criterion approach in which, to be 'healthy,' an elderly person had to report no disabilities, no chronic health problems, no pain, no depression, etc. Applying this definition to the CSHA sample selected 20% as being healthy. In order to compare this 'criterion' approach (termed 'healthy aging') with the factor-analytic approach (termed 'successful aging'), McDowell et al. chose the top 20% of individuals from the factor analytic approach, which has been followed in the present study.

The factor analyses that created the successful aging outcome were undertaken by Dr. Betsy Kristjansson, while the operational definitions and theories were contributed by Dr. Ian McDowell, Dr. Betsy Kristjansson and Richard Aylesworth.

5. METHODS

In the Methods section, the use of the Postal Code Conversion File, followed by the methods used in preparing the spatial data for linking with the successful aging dataset, will be outlined. This section concludes with how the area level data (from the 1996 Census) and spatial data were linked to the individuals in the Canadian Study of Health and Aging and how the dataset was compiled.

5.1 Postal Code Conversion File + (PCCF+)

The CSHA participants were first linked to their census tract via their postal code, using the Postal Code Conversion File from Statistics Canada. By converting the postal codes to their census tracts, I was later able to link them with census data at the census tract level and analyse them statistically.

5.2 ArcGIS 8.2

For this project, the GIS software used was ArcGIS 8.2, produced by ESRI Inc. The applications of ArcGIS 8.2 used in this thesis were ArcMap (the mapping and data manipulation component of ArcGIS), ArcCatalog (the file management component), and ArcToolbox (the geoprocessing and data conversion component).

Figure 5.2.1 is an example of the screen one would see when opening ArcMap, the mapping and data manipulation component of ArcGIS. This map contains cities, rivers and provinces of Canada, as can be seen in the layers listed in the data frame on the left.

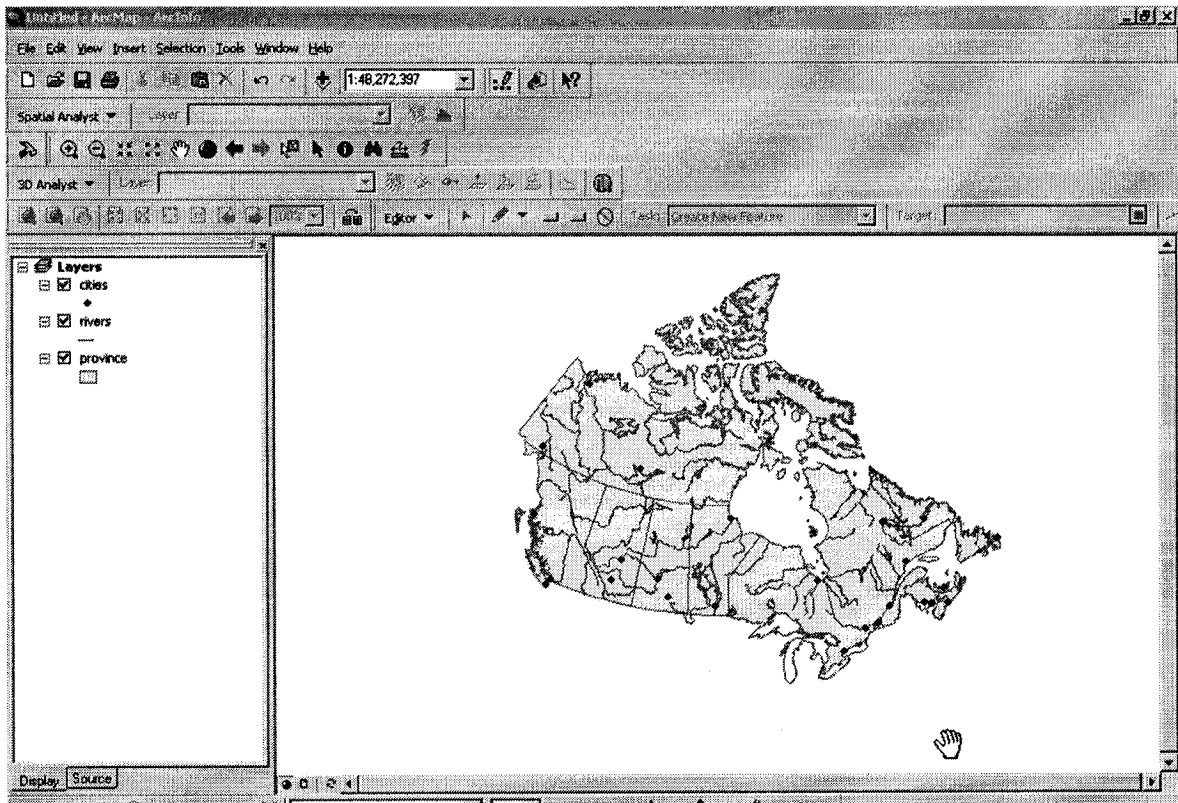


Figure 5.2.1 *Example of an ArcMap screen*

There are highly technical aspects involved in GIS, especially as pertaining to the underlying rules governing GIS methodology. It is important, for example, that the layers all have the same map projection. A projection is a mathematical means of transferring information from the Earth's three-dimensional curved surface to a two-dimensional medium - paper or a computer screen. If two different layers did not have the same projection, they would not match mathematically nor geographically and any calculations or analysis to be done on them will either not be possible or be inaccurate. Different projections are used for different types of maps because each projection is particularly appropriate to certain uses. A projection that accurately represents the *shapes* of the continents, for example, will distort their relative *areas*. Since much of the information in a GIS comes from existing maps, a GIS uses the processing power of the

computer to transform digital information, gathered from sources with different projections to a common projection.

In my analysis, I used a modified North American Equidistant Conic projection.⁸

5.3 Spatial Data Preparation

The GIS was used to create a database containing descriptions of the location of infrastructures hypothesized to be related to successful aging (see Section 5.3.1). Two approaches, Proximity and Containment, were used to analyze these data, creating three spatial indicators. These indicators included features in an individual's residential space that are most likely to affect the day-to-day lives of the elderly. The first indicator, called '*Proximity: Nearest essential services*,' determined the distance to the nearest 'needed' facility from each participant's postal code location. The second approach, called '*Containment: Number of community establishments and services*' counted the total number of establishments of interest contained within a defined distance of each participant's postal code. The third indicator, '*Containment: Economic characteristics*' sought to represent the economic characteristics of the area in which the CSHA participant lives. It summarizes the types of businesses and economic activity in the neighbourhood.

⁸ The standard parallel chosen depends on which region of the Earth is to be mapped – since a relatively narrow latitudinal range of Canada was being mapped (given that most CSHA participants were located close to the southern Canadian border), the standard parallels were defined as: standard_parallel_1= 45.000000, standard_parallel_2= 52.000000 and latitude of origin= 40.000000, with a central meridian= - 96.000000. To preserve correct distances and area (which were the two main concerns), an equidistant conic projection is needed, which requires the two standard parallels. A conic projection, which has a function of latitude, was selected as the scale factor (73), making this projection suitable for depicting Canada, with its broad extent in longitude (width is longer than its 'height' for the areas of interest).

5.3.1 Proximity: Nearest essential services

Given the importance of services in frameworks used to study the physical and social environment's influences on health, and that proximity to needed services is among the environmental variables most consistently related to elderly well-being (74), GIS was used to isolate the data needed to answer the question:

How close is the nearest 'needed' facility (e.g., grocery store, pharmacy, etc.) to the CSHA participant and is this a determinant of successful aging (i.e., does ready access to services play a role in successful aging)?

Because the smallest unit to which the CSHA participant could be localized was the postal code, the postal code location of the CSHA participant was taken as the location of the participant. *Distance to the nearest location* of several types of neighbourhood features from the participant's postal code location were ascertained using GIS. The points of interest were:

- bank
- grocery
- drug store
- department store/general merchandise store
- recreational spaces (such as community centres)
- hospital
- home health care services

This led to the derivation of seven distance variables, further discussed in Section 5.6.

5.3.2 Containment: Number of community establishments and services

The second strategy sought to represent the density of relevant points of interest (again chosen for their relevance to health), in order to answer the question:

Do the number of establishments of interest (e.g., drug stores, senior's centres, etc.) within a 500 m radius of the CSHA participant's postal code location play a role in determining successful aging?

Using GIS, the *number of establishments* of interest within a 500 m radius of the CSHA participant's mapped postal code location were determined, again as a way of deciding whether access to services plays a role in successful aging. A radius of 500 m results in an area of 0.785 km² around that individual's location. In addition to being a conceptually reasonable area for a 'neighbourhood,' this radius was also based on the Troosters et al. (1999) study on the six minute walking distance⁹ in healthy elderly subjects (75). Troosters et al. found that in healthy subjects aged 50-85 years old, walking distance averaged 631 +/- 93 m, and was 84 m greater in the male compared to female subjects. Thus, a buffer zone with a radius of a 500 m around the CSHA participant's postal code location was considered to be an appropriate distance healthy elderly individuals can be expected to walk (see Section 5.4).

In addition to certain points of interest identified in *Proximity: Nearest essential service*, establishments such as cultural spaces and civic and social associations were included in this strategy. Where the first set of variables (banks, grocery stores, etc.) consisted of 'needed' services for day-to-day life, and where distance to these needed services was hypothesized to affect health, the set of variables in the second strategy consisted of features hypothesized to contribute to the vitality of the area of residence of the CSHA participant. It is hypothesized the greater the number of cultural or socializing spaces, for example, the greater the participation of the residents of that area. Moreover, including counts of features such as liquor and tobacco stores or sites where police

⁹ The six minute walking distance is a test commonly used to estimate functional exercise capacity in patients with chronic diseases, including chronic obstructive pulmonary disease (75).

protection is located, may give a better idea of the features found in that individual's residential area.

The points of interest for this method were:

- Organizations of community involvement and participation (civic, social and fraternal associations and membership organizations)
- Cultural spaces (libraries, museum and art galleries)
- Drug stores
- Food stores (includes grocery stores, markets, miscellaneous food stores)
- Health care offices and clinics layer (MD as well as allied health services)
- Home health care services
- Hospitals
- Liquor and tobacco stores
- Police protection
- Religious organizations
- Social services
- Socializing spaces (eating and drinking places)
- Spaces of activity (physical fitness facilities, membership sports and recreation clubs, amusement and recreation services)

5.3.3 Containment: Economic characteristics

The operation *Containment: Economic characteristics* attempts to determine the economic characteristics of the area in which the CSHA participant lives, through counting the number of 'industries' contained within a 500 m radius of that individual's postal code location. In referring to industries or industrial features, 'industry' is used in the broad sense of all economic activity: agriculture, forestry, and fisheries; mining; construction; manufacturing; wholesale trade; retail trade (which includes eating and drinking places, as well as general merchandise stores and food stores); finance, insurance and real estate; transportation, communication and sanitation services; and other services (which included, among others, health services, legal services, and social services).

It is hypothesized that living in more industrialised areas (characterized by, for example, heavy industry, such as manufacturing, which includes petroleum refining as well as industrial and commercial machinery) may not be as conducive to successful aging as living in residential areas or closer to services and retail trade-type environments. The areas of economic activity selected include:

- Agriculture, Forestry and Fishing
- Manufacturing
- Wholesale Trade
- Retail Trade
- Finance, Insurance and Real Estate
- Services
- Public Administration

These are very broad headings encompassing many thousands of particular industries each. Each type of industry (e.g., manufacturing), as well as the specific services, products and industries comprising the type of industry (such as industrial and commercial machinery, chemicals and allied products) was carefully considered in attempting to ascertain which would be even remotely relevant to an elderly individual's residential area. In *Containment: Economic characteristics*, using GIS, a buffer zone of 500 m was again created around the CSHA participant's postal code location, and the frequency of different types of industrial activity calculated in that zone.

5.4 Creating the buffers

In order to determine the frequency/counts of industries/businesses within the study participant's neighbourhood, a buffer with a radius of 500 m was created around the location of the CSHA participant's postal code. Initially, a radius of three kilometres was considered, but this translates into an area of 28.3 square kilometres. Preliminary

analyses indicated that this was too large an area, often including many thousands of industries, particularly in large metropolitan centres. Moreover, this large area occupied a large fraction of most metropolitan cities and was generally not representative of an individual's immediate neighbourhood.

Creating buffers around the postal codes involved using the older and newer versions of the GIS software (ArcView 3.2 and ArcGIS 8.2 respectively). The buffering of the postal code was first done in ArcView 3.2 before importing the data into ArcGIS 8.2 and manipulating and doing the spatial joins there (spatial joins are discussed in greater detail in Section 5.6). ArcView 3.2 was used because the program used to create these buffers was only compatible with ArcView 3.2, and not the more recent ArcGIS 8.2. This buffer program is not a part of the core ArcView package and needed to be downloaded from the ESRI website. More complete details can be found in Appendix II.

5.5 Enhanced Points of Interest Datafile

In order to extract the necessary data (i.e. the variables of interest) from the 'Enhanced Points of Interest' (EPOI) files, GIS was used to query the files and select the variables of interest using the query feature in ArcGIS (Figure 5.5.1). The queries were constructed using SQL,¹⁰ an industry-standard language for creating, updating and, querying relational database management systems. The variables of interest were queried by that variable's Standard Industrial Classification code. A query is a user's request for information, generally as a formal request to a database or search engine. An example of a query would be to type: "SIC" >= '83990000' and "SIC" <'84000000' in the query

¹⁰There is debate over whether SQL is an acronym or whether it is a shortened version of 'sequel.' For those who consider it an acronym, SQL generally stands for Structured Query Language or Standard Query Language.

window. This particular query would pull out the rows corresponding to a standard industrial classification greater than or equal to 83990000 and less than 84000000 in that particular province's points of interest database. The query for a particular service/establishment was applied uniformly across the provinces to maintain consistency.

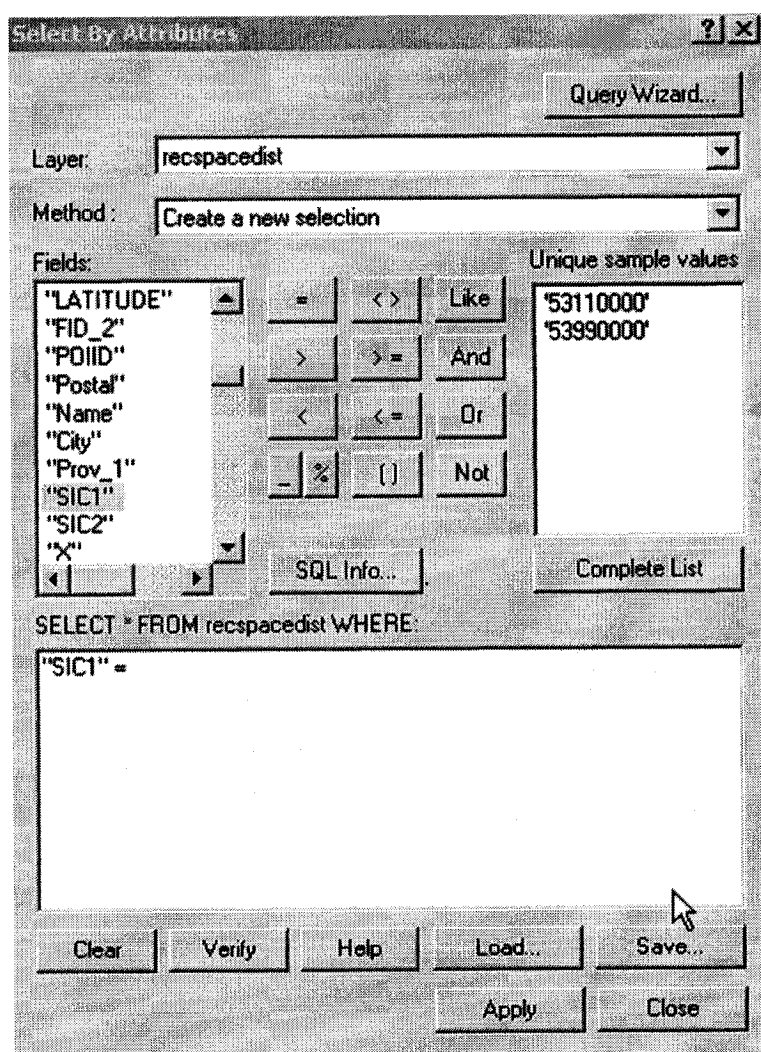


Figure 5.5.1 Query window in ArcGIS used to formulate queries

There is the choice of writing your own query or using the supplied buttons to write the query. An example of a query would be: SIC >= '83990000' and SIC <'84000000'. In this example, the SIC codes correspond to social services.

After careful consideration, twenty-eight features in the Enhanced Points of Interest datafile were deemed most relevant to the health and life of an elderly individual. The full list of features, as well as a brief description of what types of establishments were contained under the feature headings, can be found in Appendix III.

In *Proximity: Nearest essential services* extraction, specific variables such as banks, grocery stores and hospitals, were identified by one SIC code. Other variables, representing aggregations of specific establishments, had to be queried by more than one SIC code. Recreational spaces, for example, represented a combination of codes from the fitness facilities categories (e.g. "SIC1" = '79979900' selects community centres and athletic clubs) and social service categories (e.g. "SIC1" = '83999000' AND "SIC2" = '83220117' selects community centres and senior drop in centres). The queries used to derive the variables in *Proximity: Nearest essential services* can be found in Appendix IV.

The approach of combining various categories of SICs was again used to derive variables for *Containment: Number of community establishments and services*. From the 28 points of interest initially selected, several features were combined to create specific datafiles, in an attempt to create more generalized variables. For the 'health care offices and clinics' data, for example, health and allied services, doctor's clinics, skilled nursing care facilities, nursing and personal care facilities and offices of health care practitioners, along with the DMTI Spatial's health care layer, were combined. DMTI Spatial's health care layer comprised health care facilities including hospitals, long-term care centers,

nursing stations, outpatient clinics and community health centres across Canada. DMTI Spatial's 'tourist' layer was incorporated into the 'cultural spaces' variable, along with art galleries and libraries. The relatively few points in this layer included fishing resorts, science centers and historic sites. These 13 variables, and the SIC queries used to create them can be found in Appendix V.

For the industrial features of *Containment: Economic characteristics*, the vast number and types of industries contained within the file were critically assessed in terms of their possible, even if remote, relevance to CSHA participants. The objective was to be as inclusive as possible in creating a frequency of the types of industries found in the individual's residential location. The data extracted for the industrial features were less specific than for the previous containment strategy and corresponded to the industry division (e.g., manufacturing, services, etc.). The list of industries kept for analysis, as well as the types of business that make up those industries, can be found in Appendix VI. The SIC queries of the selected industries can be found in Appendix VII.

Each query (to extract the point of interest) was run for each of the 10 provinces (since the data were organized by province), resulting in a total of 270 files (seven distance variables for *Proximity: Nearest essential services* (70 files), 13 neighbourhood variables for *Containment: Number of community establishments and services* (130 files), and seven industrial features for *Containment: Economic characteristics* (70 files)). The files corresponding to each of the methods needed to be combined to form a national datafile. This was because the next step would be to link the features to the CSHA participants (as will be seen in section 5.7). To create a national datafile for a particular

point interest, such as social services, the files from the 10 provinces were merged using the GIS merging feature, creating an additional 27 files.

Once all the points of interest were queried, selected and merged to form a national layer, the proximity of the nearest feature (e.g. grocery stores) to the individual's postal code (for *Proximity: Nearest essential services*), as well as the frequencies of particular features within a 500 m radius buffer zone (for *Containment: Nearest community establishments and services* and *Containment: Economic characteristics*), was determined by a GIS procedure known as spatial joining.

5.6 Spatial Joining

Spatial joining formed the primary focus of the GIS data analysis. A spatial join is a type of spatial analysis in which the attributes of features are linked together based on the relative locations of the features that are stored in different spatial data layers. Spatial joining can provide different types of information, such as finding how many participants are located inside a census tract, how many roads intersect a proposed highway, or, in this study, determining the distance to and number of relevant features. The first objective was to determine the distance to the closest of each type of establishment/service for each CSHA participant. In this case, the establishments and postal codes were identified as points – the mapped location point of the social service, for example, and the mapped location point of the CSHA participant's postal code were spatially joined in the GIS, and their proximity was measured. Proximity is calculated by the GIS, based on straight line Euclidean-distance between features.

All spatial data layers created through queries as well as the layer of data containing the locations of postal codes corresponding to the CSHA participants were brought into the GIS. For consistent Euclidean-distance calculations within the GIS software, all of these layers were required to be in the same source map projection, the modified North American Equidistant Conic projection.

To spatially join a point of interest (e.g. social services), the map layer containing all social services in the ten provinces is added to the data frame, after having been appropriately projected. In addition, the layer containing the postal code locations of all CSHA participants is also added. The result of this basic GIS operation is a set of distances, covering each feature of interest, for each CSHA participant. A more detailed description of procedures used can be found in Appendix VIII.

Spatial joining was also used to count the numbers of establishments/points of interest within the 500 m buffers created around each CSHA participant's postal code location. In this case, each buffer zone was automatically provided a summary of the numeric attributes of the points that fell inside it (where relevant), and a count field containing the number of points.

An important consideration when working with GIS data is that files can become quite large due to the amount of data contained within each file. Each query, for example, resulted in an output data file (in dBASE format) ranging between 11 kb and 80 MB, depending on the number of establishments corresponding to particular queries and the province. However, the number and size of files created in this thesis was not limited by current computer hardware.

5.7 Statistical Analysis Software

5.7.1 Linking the datasets

The data were linked to create a database in Statistical Analysis Software (SAS) version 8.02. The data could also have been linked in the GIS software, but as various data were already in SAS programs it was easier to link them all in SAS. In addition, the GIS system has limited statistical capabilities, and would not have met the statistical analysis needs for this thesis. The merging is discussed in the Results section.

5.7.2 Analysis

The focus of the thesis is the creation of the database and the compilation of individual-level as well as group-level variables. This will allow opportunities for more complex analyses, such as multi-level and contextual analyses. Multilevel analysis was not appropriate for this thesis as the numbers of CSHA participants in each census tract were too few to satisfy the requirements for a multilevel analysis (a general guideline is that there should be 25 or more observations in each area being compared for a stable multilevel analysis).

Descriptive statistical analyses were done on GIS-derived distance variables to obtain an overview of the data (mean, standard deviation and median). The data were tested for normality using the Kolmogorov-Smirnov test. If the data were normal, parametric tests, such as the t-test, were used. If the data were not normally distributed, nonparametric statistics, such as the Wilcoxon Rank-sum test, were used to determine if there are differences between those that are aging successfully and those that aren't.

Males and females, as well as the successful aging outcomes were compared. In addition, logistic regression was used to determine the odds ratios of selected GIS-derived variables and successful aging.

The distance data were skewed with a few very large values. The variables representing distance (in metres) to essential services were categorized according to the following ranges (in metres):

- <150
- 150-349
- 350-499
- 500-999
- 1000-1999
- 2000-4999
- 5000-7499
- 7500-9999
- 10 000-14 999
- 15 000-19 999
- 20 000-49 999
- >=50 000

These categories were conceptualised as thresholds of distance that would be meaningful in gauging accessibility, and in determining whether there were significant differences between the outcomes. In the preliminary analyses presented in this thesis, I first present mean distance values in metres from the person's postal code location to various facilities hypothesized to promote successful aging and then use these categorized distances in the logistic regression analysis.

6. RESULTS

6.1 PCCF+ Geocoding Results

6.1.1 Missing and Incomplete Postal Code Data

The missing or incomplete data had to be identified to determine whether the CSHA participants' postal codes could be geocoded or used spatially.

Of the 10 263 CSHA participants, 49 had no postal code record. Of these 49, 13 participants resided in the community, leaving 8995 in the community sample (Figure 6.1.1.1). A total of 163 postal codes in the community sample required validation, due to the screening postal code being missing, or because of inconsistencies between the screening and clinical postal codes due to typing errors. The screening postal code is the postal code listed at the screening phase of the CSHA. The clinical postal code is the postal code listed at the clinical examination.

Geographic distances in rural areas are problematic because of the potentially very large distances involved. However, the CSHA did not sample from distant rural areas: 50 kilometres was the maximum that interviewers traveled outside the town. Furthermore, prior to the analysis postal codes were scanned and all those that contained a '0' in the second positions (which indicates a rural dwelling) were eliminated.

Geocoding results showed that 105 postal codes were not linked to post office geography, 44 postal codes were classified as non-residential, nine postal codes belonged to a business building and five postal codes were commercial or institutional.

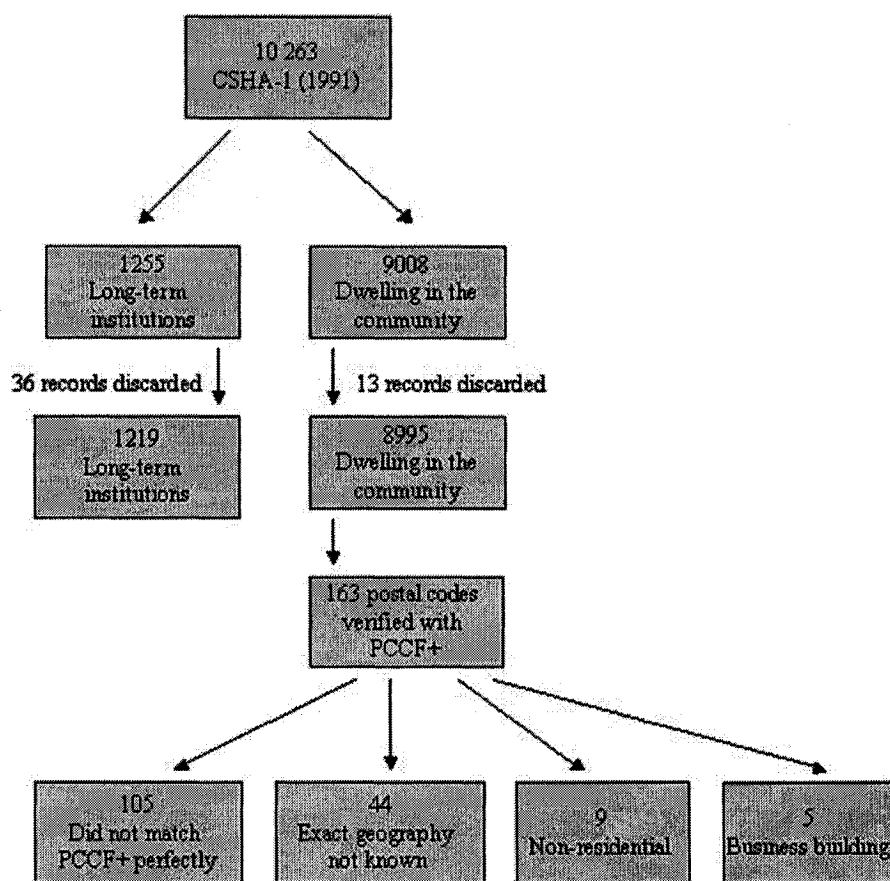


Figure 6.1.1.1 Tracking the results of the geocoding

The 163 postal codes were manually checked against a printout containing both clinical and screening postal codes. Twelve were screening postal codes that did not correspond to their clinical postal code (Table 6.1.1.1).

TABLE 6.1.1.1 INCONSISTENCIES IN CLINICAL AND SCREENING POSTAL CODES

<u>Unique Identifier</u>	<u>clinical postal code</u>	<u>screening postal code</u>
224210	A0A2V0	A0A4H0
205600	B3L2B6	B3G2B6
220206	G7H2Z7	B7H2Z7
208058	H2J2X3	H3J3X3
207431	G8A1L6	L8A1L6
225013	H3T1E6	M3T1E6
210934	N6B3N6	N6R3E2
211864	R2V0Y2	R2V0Y0
212961	S6V5Z2	S6U5Z2
206180	T2M3A7	T2H3A7
225855	B3L4J2	B3L4S2
217711	A1C2Z8	A1C3Z8

In 11 cases, the clinical postal code proved valid and was incorporated into the data. The remaining postal code was geocoded, but the postal code had been retired. Given that this PCCF+ dealt with May 2002 postal codes, and the CSHA subject postal codes were collected in 1991, the postal code was nonetheless kept in the dataset. The postal codes of the remaining 151 subjects which could not be completely validated, were kept in the dataset with the reasoning that these participants may find a match in the GIS geocoding file.

6.1.2 Postal codes mapped to more than one location

Included in the GIS data from DMTI was the Multiple Enhanced Postal Code file, which I had started to use to link the CSHA postal codes. The advantage of using the Multiple Enhanced Postal Code file is that it captures all postal codes, even if the postal code overlaps between different geographic levels. This file is useful if the areas of interest, such as enumeration areas or census tracts, are known and there is a need (such as for marketing purposes) to capture all the postal codes that correspond to the

individuals living in those geographic areas. Of the 8995 community-dwelling participants in the CSHA, there were 7363 unique postal codes. My analyses determined that 1023 out of 7363 CSHA postal codes were mapped to more than one location (i.e., 1023 postal codes had more than one geographic coordinate associated with it). The disadvantage of using the Multiple Enhanced Postal Code file for my study was that I did not know which postal code location belonged to the study participant.

Table 6.1.2.1 summarizes the number of postal CSHA postal codes corresponding to more than one latitude and longitude coordinate. The majority of the 1023 postal codes (54%) could be mapped to 2 locations, while the largest number of points corresponding to a single postal code was 133, most likely a rural postal code.

TABLE 6.1.2.1 CSHA POSTAL CODES LINKED TO MULTIPLE ENHANCED POSTAL CODE FILE

# of points corresponding to individual postal code locations	# of CSHA postal codes (n=7363 unique postal codes)
>1	1023
2	553
3	88
4	66
...	
133	1

The Unique Enhanced Postal Code file corresponded to a one-to-one relationship with the CSHA postal codes, via the single representative postal code. A limitation of using the Unique Enhanced Postal Code file is that some positional inaccuracies might exist with reference to the actual residence of the study participant. A future course of study may include testing the location of the individual's residential location against the

most representative location. However, this would require the individual's address, which was not available for reasons of confidentiality. Given the advantages and disadvantages of both types of postal code files, the use of the Unique Enhanced Postal Code file was felt to be most appropriate for the purposes of the thesis.

6.2 Area covered by postal codes: creation of convex hulls

In order to ascertain the area covered by a postal code, GIS was used to determine the size of the area contained by postal codes that had three or more coordinate points. At least three points are needed to create a polygon, whose area can be measured. The number of CSHA postal codes that had **three or more** coordinates linked to them was 470, which corresponded to 542 participants.¹¹ Of these 470 postal codes, six postal codes could not be mapped, resulting in 464 postal codes whose area was measured. The area of a polygon could be created and calculated in GIS by creating a convex hull for each postal code.

The multiple points corresponding to a single postal code are the vertices of the convex hull. ArcGIS calculates the polygon's area and this corresponds to the area covered by the postal code.

Values were obtained for the convex hull of the 464 postal codes of CSHA participants. The range in area (km²) of the convex hulls was between 0.0000124 km² (12.4m²) (M6B3T6, downtown Toronto) and 2542.3 km² (V0X1W0, rural British Columbia), with a mean area of 92.5 km² +/- 221.8 km² (standard deviation).

¹¹ Though the number of CSHA postal codes corresponding to **two or more** coordinate locations numbered 1023 (out of 7363 unique postal codes), the total number of CSHA participants corresponding to postal codes mapped to **two or more** locations was 1153 of the 8995 (130 participants shared the same postal code).

Histograms were created depicting the frequency of areas covered by the postal codes. The x-axis corresponds to the area (in km^2) covered by the convex hull/polygon of the postal codes. The y-axis represents number of postal codes. Creating the histograms was helpful in seeing the distribution of postal code areas in the CSHA.

Figure 6.2.1, for example, shows that over 350 of the 464 postal codes mapped to 3 or more locations were in the 0 to 100 km^2 range.

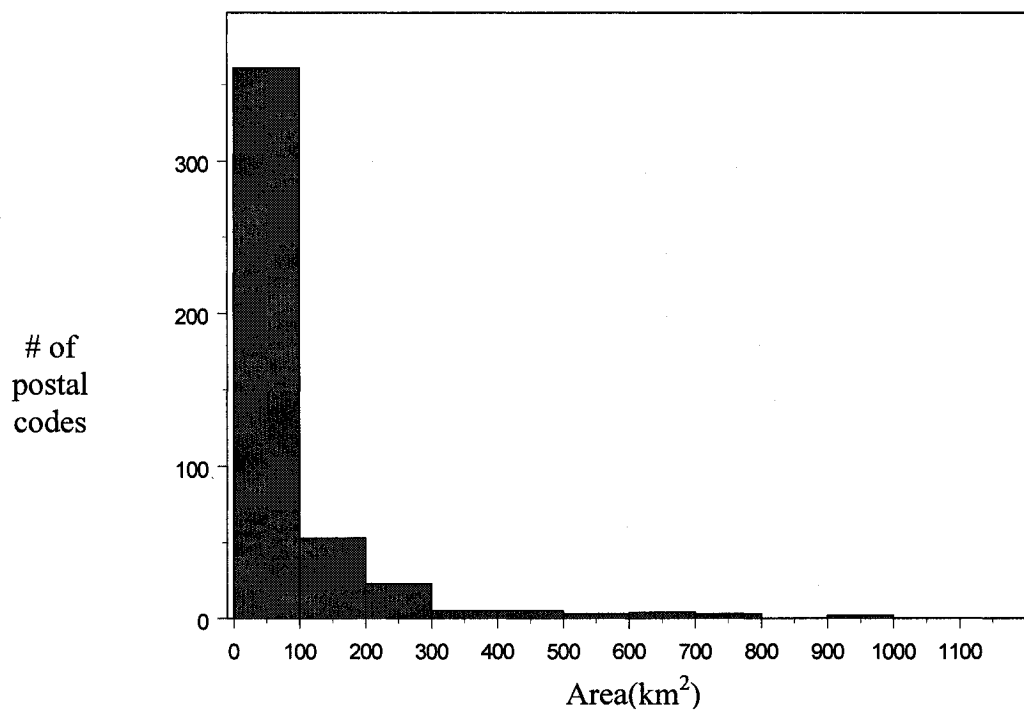


Figure 6.2.1 Histogram of Convex Hull PC Areas (Range: 0-1000 km^2)

Of those postal codes in the 0-100 km^2 category, close to 100 are covered by an area equal to or less than 5 km^2 as can be seen by the histogram in Figure 6.2.2.

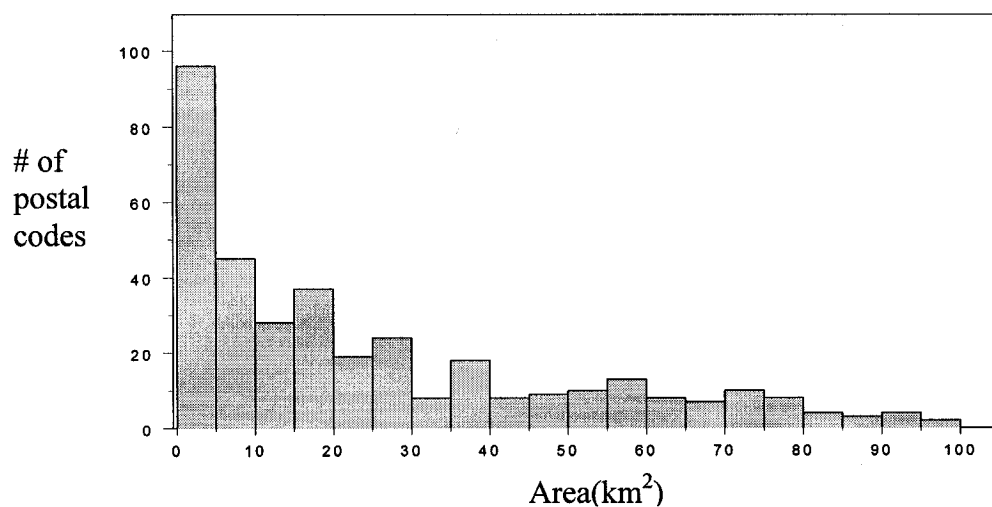


Figure 6.2.2 Histogram of Convex Hull PC Areas (Range: 0-100 km²)

Five postal codes were found to cover areas of over 1000 km² (Figure 6.2.3).

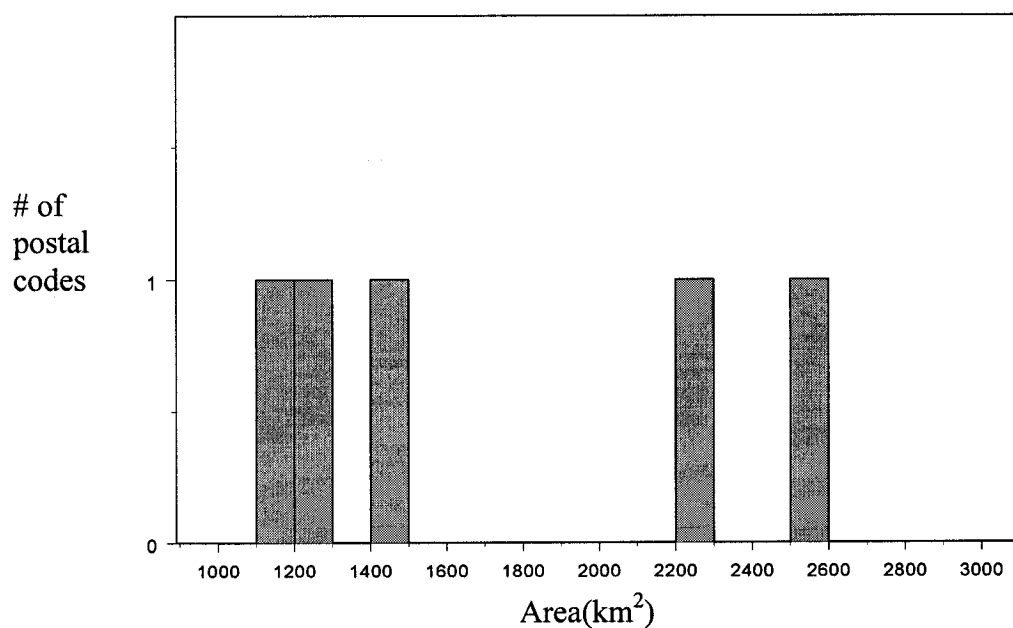


Figure 6.2.3 Histogram of Convex Hull PC Areas (Range: 1000-3000 km²)

These analyses indicate that there exists great variability in the area covered by postal codes, which should be taken into account when working at this level of geography.

6.2.1 Issues in using MEP for analysis of proximal establishments

A 500 m buffer was drawn around each participant's postal code in order to be able to determine how many establishments are located within that buffer and thus within 500 m of the study participant. Drawing a 500 m buffer around a specific postal code which is found in more than one geographic location will result in the buffers of the same postal code overlapping. This would cause an overinflated number of establishment counts to be attributed to the participant corresponding to that postal code. Were I to have used the MEP file to link CSHA participants to the spatial data, the absolute counts of types of industries and points of interest could not have been compared between individuals in the study.

6.3 Tracking participants when linking datasets

From the initial file of 10 214 CSHA Cycle-1 participants, 8995 participants were ascertained to be community-dwelling. The remaining 1219 participants resided in institutions, such as long-term care facilities.

There were 7445 unique postal codes among the 8995 (1550 postal codes were shared by different participants). Of the 7445 unique postal codes that existed for CSHA participants, 82 postal codes (corresponding to 85 individuals) did not match to the DMTI Unique Enhanced Postal Code File, and were not kept for further analysis, resulting in

7363 postal codes in the nearest features file. The CSHA postal codes were first collected in 1991, and the UEP file is from 2002. It is likely that the CSHA postal codes that did not have a match in the UEP had been retired. It was necessary to match the CSHA participants to the UEP since the UEP contained the spatial location of the postal codes, which is needed to map postal code location, and thus CSHA participant location. Therefore, 7363 individual postal codes, corresponding to 8910 individuals, were mapped to their respective locations across Canada using the GIS (Figure 6.3.1). The circles represent individual participants who can be seen clustered around the study centres.

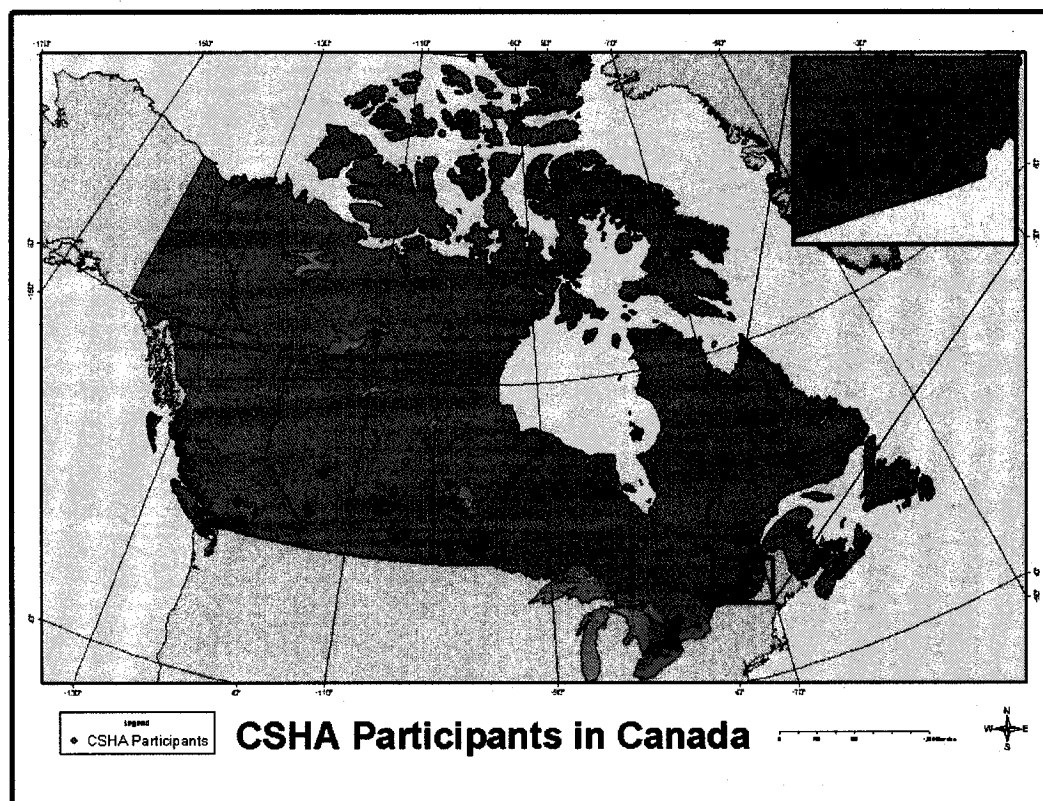


Figure 6.3.1 *CSHA Participants across Canada, mapped using ArcGIS*

The successful aging subset of community-dwelling participants was not available at the time of the GIS work phase, so the GIS analysis was done using 7363 unique postal codes from the community-dwelling CSHA participants. When matched to the CSHA community-dwelling participants (which had already been geocoded with the PCCF+ and had the census tracts associated with them), 5071 out of the 5073 successful aging participants were matched. This is believed to be due to the fact that two of the participants had a postal code that had been excluded or missing in the original 9008 and couldn't be matched (Figure 6.3.2). Linking the GIS data to the successful aging participants resulted in 5014 participants with GIS data linkage. The 57 missing participants corresponded to 54 postal codes that did not exist in the UEP, most likely due to the previously made suggestion that these postal codes had been retired in the 11-year time interval from when the postal codes had initially been collected for the CSHA (1991), and the date of the GIS data (2002).

The census tract data were matched to 4245 out of 5071 participants. Of the 826 missing participant observations, 795 were located rurally, with a census tract number '0.00' and postal code containing '0' as the second character in the forward sortation area. The remaining 31 participants belonged to census tracts where the data had been suppressed, with census tract numbers '999.99.' The rurally located participants are composed of 312 females and 483 males. There are 63 females in rural areas aging successfully, and 249 females living in rural areas not aging successfully. For males living in rural areas, 93 have aged successfully and 390 have not.

For the simple analyses undertaken for this thesis, it was decided to include those participants for whom non-missing data was available from all data sources. There were

4213 individuals to which individual-level data, GIS data and census tract data could be linked and upon which descriptive statistical analyses were run. This does not include those living in rural areas, or for those without census and GIS data.

The SAS code linking participants can be found in Appendix IX.

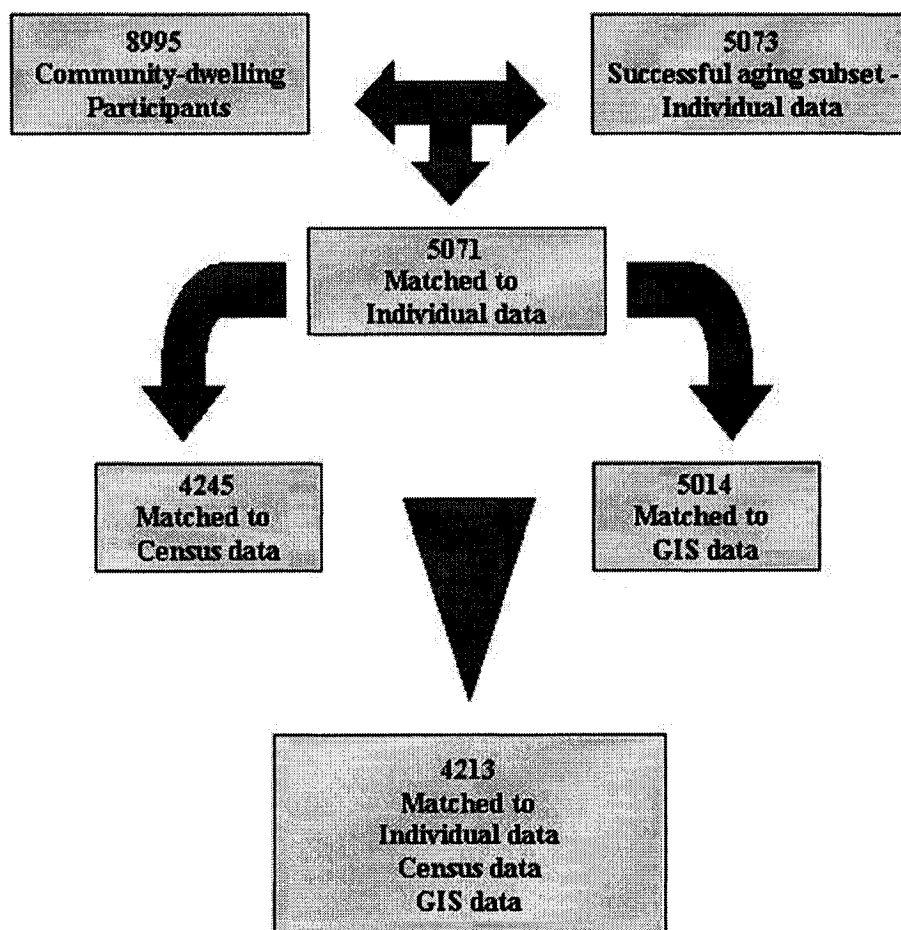


Figure 6.3.2 Tracking participants in the linking of the database

6.4 GIS variables in dataset

The completed dataset contains the spatial variables summarized in Tables 6.4.1, 6.4.2 and 6.4.3.

TABLE 6.4.1 'PROXIMITY: NEAREST ESSENTIAL SERVICES' VARIABLES IN THE DATASET

"Distance to nearest..."	Variable name
Bank	bankdist
Drug store	drugstdist
Grocery store	grcerydist
Home health care services	homecrdist
Hospital	hospdist
Recreation centre	recdist
Store (general merchandise)	storedist

TABLE 6.4.2 'CONTAINMENT: NUMBER OF COMMUNITY ESTABLISHMENTS AND SERVICES'

Count data (Neighbourhood)	Variable name
Organizations of community involvement	COMMINV500
Cultural spaces (libraries, galleries)	CULTSP500
Drug stores	DRUGST500
Food stores	FOODST500
Health care offices and clinics	HLTHCR500
Home health care services	HOMEHC500
Hospitals	HOSP500
Liquor and tobacco stores	LIQRTOB500
Spaces of physical activity	PHYSACT500
Police protection	POLPROT500
Religious organization	RELOGR500
Social services	SOCSEVR500
Socializing spaces (eating and drinking places)	SOCSP500

TABLE 6.4.3 'CONTAINMENT: ECONOMIC CHARACTERISTICS' VARIABLES IN THE DATASET

Count data (Industrial)	Variable name
Agriculture, Forestry and Fishing	FFAGRI500
Manufacturing	MANUF500
Wholesale trade	WHOLETR500
Retail trade	RETAIL500
Finance, Insurance and Real Estate	FINAN500
Services	SERVICE500
Public Administration	PUBAD500

6.5 Analysis

6.5.1 Descriptive Statistical Analysis

Individuals in the final dataset used for analysis consisted of 4213 CSHA participants who had been classified as successfully aging or not, and who could be matched to both non-missing spatial and census data. All statistical analyses were done using Statistical Analysis Software (SAS) version 8.02.

As can be seen in Table 6.5.1 below, 53% of those that aged successfully were female and 47% were male. Of those who had aged unsuccessfully, 37% were male and 63% were female.

TABLE 6.5.1 NUMBER OF MEN AND WOMEN IN DATABASE SUCCESSFULLY AGING

	Successfully Aging (n= 857)	Unsuccessfully Aging (n= 3356)
Males (n= 1656)	406 (47%)	1250 (37%)
Females (n=2557)	451 (53%)	2106 (63%)

In order to decide whether to use parametric or non-parametric tests in the analysis of the data, the Kolmogorov-Smirnov test statistic for normality was applied. The null hypothesis of the test is that data are normally distributed. Table 6.5.2 lists the GIS-derived variables describing distance, which were tested for normality. For every variable, the p-values were significant at $\alpha=0.05$, resulting in consistent rejection of the null hypothesis. This indicates that the distance variables are not normally distributed.

TABLE 6.5.2 TESTING NORMAL DISTRIBUTION OF DISTANCE VARIABLES

Variable (Distance to:)	p-value of result*	Reject Ho at $\alpha = 0.05$?	Are data normally distributed?
Bank	<0.01	Yes	No
Hospital	<0.01	Yes	No
Home health care services	<0.01	Yes	No
Grocery store	<0.01	Yes	No
Merchandise Store	<0.01	Yes	No
Drug store	<0.01	Yes	No
Recreational centre	<0.01	Yes	No

* Kolmogorov-Smirnov test for normality

The mean and standard deviation of the distance variables from *Proximity: Nearest essential services* were ascertained to compare distance to services between males and females. As can be seen in Table 6.5.3, males consistently lived further from these facilities than females, ranging from 311.4 m (distance to grocery store) to 694.6 m (distance to nearest home health care establishment). The large variability in the distribution of the data can be seen in the large standard deviations associated with the mean distance values (Table 6.5.3). The standard deviation is a measure of dispersion that is extremely sensitive to extreme observations. It is recognized that if the data are skewed, the median may be a better representative value than the mean.

As the data are not normally distributed, non-parametric tests were used to ascertain differences between males and females, as well as the successful and unsuccessful. The Wilcoxon Rank-sum test for independent samples assumes that the variable distributions have the same general shape, and evaluates the null hypothesis that the medians of the two populations are identical. The differences between males and

females and their distance to amenities was consistently significant at $\alpha=0.05$, indicating that sex needs to be controlled for.

TABLE 6.5.3 DISTANCES FROM HOME TO SELECTED SERVICE FACILITIES, BY SEX

	Combined (n=4213)	Males (n=1656)	Females (n=2557)	Difference in means and medians: males - females	p- value*
Distance in metres to:	mean+/-sd median	mean+/-sd median	mean+/-sd median	mean difference median difference	
Bank	2049.8±3492.7 1033.1	2289.2 ± 3889.3 1102.2	1894.7 ± 3201.1 992.5	394.5 109.7	<0.001
Hospital	2824.5±3697.3 1788.2	3068.3 ± 3708.8 1988.6	2666.5 ± 3681.9 1683.0	401.8 305.6	<0.001
Home health care service	3696.8±4335.6 2180.1	4118.3 ± 4777.7 2439.6	3423.7 ± 4000.8 1993.1	694.6 446.5	<0.001
Grocery store	1519.6±3263.0 657.4	1708.6 ± 3618.7 699.9	1397.2 ± 3004.6 624.0	311.4 75.9	<0.001
Merchandise Store	2885.9±3714.1 1859.8	3143.4 ± 4154.6 1887.5	2719.0 ± 3388.9 1839.4	424.4 48.1	0.005
Drug store	1623.9±3144.9 801.0	1878.8 ± 3642.3 871.8	1458.8 ± 2763.9 754.2	420.0 117.6	<0.001
Recreation centre	7248.9±19 941.9 2594.4	7585.4±19813.9 2821.2	7031.0±20025.2 2474.1	517.4 347.1	<0.001

*based on the Wilcoxon Rank-sum test

It had been hypothesized that individuals living close to services would be more likely to age successfully, due to ready access to needed services. There is a large range in mean distance, from 151.3 m (distance to the bank) to 1135.2 m (distance to the recreational centre) between those that are aging successfully and those that are not (Table 6.5.4). Contrary to expectation, those who are aging successfully tend to live

further away from each of the establishments, except for recreational centres. The largest distance differential between the health outcomes is for proximity to recreational centre, and those who are aging successfully live closer to the recreational centre. The difference between those aging successfully and those not aging successfully again proved significant at $\alpha=0.05$ using the Wilcoxon Rank-sum test for all variables except 'Distance in metres to General Merchandise Store.'

TABLE 6.5.4 DISTANCES FROM HOME TO SELECTED SERVICE FACILITIES, BY SUCCESSFUL/UNSUCCESSFUL AGING CSHA PARTICIPANTS

	Successful (n= 857)	Unsuccessful (n= 3356)	Difference in means and medians: successful - unsuccessful	p-value*
Distance in metres to:	mean+/-sd median	mean+/-sd median	mean difference median difference	
Bank	2170.3 +/- 3731.7 1109.4	2019.0 +/- 3428.9 1007.5	151.3 101.9	0.025
Hospital	3075.9 +/- 3702.3 2029.4	2760.3 +/- 3693.8 1737.7	315.6 291.7	<0.001
Home health care service	4291.8 +/- 4952.5 2556.6	3544.8 +/- 4150.6 2075.2	747.0 481.4	<0.001
Grocery store	1674.3 +/- 3603.4 718.8	1480.1 +/- 3169.5 632.5	194.2 86.3	<0.001
Merchandise Store	3122.8 +/- 4259.1 1880.8	2825.4 +/- 3559.7 1854.3	297.4 26.5	0.098
Drug store	1839.0 +/- 3566.6 927.1	1568.9 +/- 3026.0 766.9	270.1 160.2	<0.001
Recreational centre	6344.7 +/- 15165.5 2804.1	7479.9 +/- 20 983.8 2522.4	-1135.2 281.7	0.001

*based on the Wilcoxon Rank-sum test

It is interesting that those who did not age successfully, on average, lived closer to all services, except for recreational centres. Of course, living in proximity to an establishment does not guarantee that the individual will make use of it. It may also be that those aging successfully may live in places where there are fewer services and establishments, such as suburbs and more residential areas, while those living close to services live in areas of higher density, such as the urban core of cities.

Because sex is a potential confounder, distances to facilities were broken down by sex. Table 6.5.5 and Table 6.5.6 show the distance variable differences for successful and unsuccessful aging for females and males respectively.

TABLE 6.5.5 DISTANCES FROM HOME TO SELECTED SERVICE FACILITIES, BY SUCCESSFUL/UNSUCCESSFUL AGING FOR FEMALES

	<i>Successful</i> (n= 451)	<i>Unsuccessful</i> (n= 2106)	Difference in means and medians: successful - unsuccessful	p-value*
Distance in metres to:	mean +/- sd median	mean +/- sd median	mean difference median difference	
Bank	1912.9 +/- 2967.9 1116.8	1890.8 +/- 3249.5 967.2	22.1 149.6	0.115
Hospital	2874.6 +/- 3707.1 1813.9	2621.9 +/- 3675.8 1663.3	252.7 150.6	0.028
Home health care service	3860.9 +/- 4278.8 2373.2	3330.1 +/- 3933.5 1933.4	530.8 439.8	0.004
Grocery store	1399.4 +/- 2567.6 705.9	1396.7 +/- 3090.7 612.5	2.7 93.4	0.007
Merchandise Store	2838.6 +/- 3742.3 1754.2	2693.5 +/- 3308.7 1856.5	145.1 102.3	0.720
Drug store	1575.4 +/- 2681.0 910.9	1433.8 +/- 2781.4 725.4	141.6 185.5	<0.001
Recreational centre	5593.5 +/- 13 607.1 2650.3	7338.9 +/- 21 137.8 2447.7	-1745.4 202.6	0.026

*based on the Wilcoxon Rank-sum test

TABLE 6.5.6 DISTANCES FROM HOME TO SELECTED SERVICE FACILITIES, BY SUCCESSFUL/UNSUCCESSFUL AGING FOR MALES

	<i>Successful</i> (n= 406)	<i>Unsuccessful</i> (n= 1250)	Difference in means and medians: successful - unsuccessful	p-value*
Distance in metres to:	mean+/-sd median	mean+/-sd median	mean difference median difference	
Bank	2456.2 +/- 4414.8 1078.2	2235.0 +/- 3702.9 1104.2	221.2 26.0	0.320
Hospital	3299.5 +/- 3688.7 2174.5	2993.3 +/- 3713.7 1904.4	306.2 270.1	0.007
Home health care service	4770.6 +/- 5573.3 2861.9	3906.5 +/- 4471.1 2367.9	864.1 494.0	0.009
Grocery store	1979.7 +/- 4465.5 727.4	1620.5 +/- 3169.5 682.5	359.2 44.9	0.042
Merchandise Store	3438.5 +/- 4752.9 1986.9	3047.6 +/- 3938.1 1853.5	390.9 133.4	0.082
Drug store	2131.8 +/- 4328.5 952.1	1796.6 +/- 3387.6 854.7	335.2 97.4	0.010
Recreational centre	7179.1 +/- 16 705.0 3103.2	7717.4 +/- 20 727.9 2772.3	-538.3 330.9	0.047

*based on the Wilcoxon Rank-sum test

Having controlled for confounding by sex, mean distance to bank and to the general merchandise store ceased to indicate that there were significant differences between individuals who had aged successfully or not. Successfully aging women tended to live on average more than a kilometre and a half closer to recreational centres than successfully aging men. The difference in mean distance to recreational centres between

successfully and unsuccessfully aging females was much larger than the difference in mean distance to recreational centres between males.

In order to determine which, if any, distance variables were correlated, a Spearman correlation matrix was constructed (Table 6.5.7). All services were highly correlated with each other ($r \geq 0.60$) except for 'recreational centres' ($0.21 \leq r \leq 0.39$) and all were significant at $\alpha=0.05$.

TABLE 6.5.7 DISTANCE VARIABLES: SPEARMAN CORRELATION MATRIX FOR SUCCESSFUL AND UNSUCCESSFUL AGING COMBINED

	Bank	Hosp.	Homecr.	Grocery	Store	Pharmacy	Recreation
Bank	1.00	0.71	0.60	0.80	0.76	0.80	0.35
Hospital		1.00	0.63	0.79	0.79	0.78	0.30
Homecare			1.00	0.67	0.67	0.71	0.21
Grocery				1.00	0.85	0.90	0.39
Store					1.00	0.84	0.29
Pharmacy						1.00	0.33
Recreation							1.00

In order for the distances to be more meaningful and comparable, they were categorized in the ranges described in the Method section 5.7.2. The probability of successfully aging was modelled using logistic regression, controlling for sex. The odds ratios for the categorized distance variables were analysed and are summarized in Table 6.5.8. The odds ratios are oriented such that a value greater than 1.0 would suggest increasing likelihood of successful aging for those living closer to the facility. If there had been an association, an odds ratio of 1.07 (the odds ratio for the proximity of males to a hospital), for example, can be interpreted so that males living one distance category closer to a hospital were 7% more likely to age successfully than males not living one distance category closer to a hospital. For the same variable, females have an odds ratio

of 0.95. This can be interpreted as an inverse association - women living one distance category closer to a hospital have a 5% less chance of aging successfully than women living one distance category farther from a hospital.

TABLE 6.5.8 ODDS RATIOS FOR PROBABILITY OF SUCCESSFUL AGING BY DISTANCE VARIABLES

	Males (n=1656)	Females (n=2557)	Both (n=4213)
Distance to:	OR (95% CI)	OR (95% CI)	OR (95% CI)
Bank	0.88 (0.68, 1.14)	1.02 (0.83, 1.27)	0.96 (0.81, 1.13)
Hospital	1.07 (0.81, 1.41)	0.95 (0.78, 1.16)	0.99 (0.84, 1.15)
Home health care service	1.14 (0.89, 1.46)	1.07 (0.88, 1.30)	1.09 (0.94, 1.28)
Grocery store	1.04 (0.72, 1.50)	0.84 (0.67, 1.04)	0.89 (0.74, 1.07)
Store (general merchandise)	0.96 (0.60, 1.53)	1.13 (0.89, 1.43)	1.10 (0.90, 1.32)
Drug store	1.00 (0.69, 1.45)	1.10 (0.87, 1.39)	1.01 (0.87, 1.16)
Recreation centre	1.05 (0.79, 1.41)	0.99 (0.83, 1.17)	0.99 (0.86, 1.14)

No variable coefficient estimates were significant at $\alpha=0.05$. As can be seen by the 95% confidence intervals, the null value (1.0) is contained within every interval, indicating that the ratio of cumulative incidence among the exposed and unexposed groups is identical. This means that there is no association observed between exposure and successful aging, even after controlling for sex.

Table 6.5.9 summarizes data on the count of facilities located within 500 m of the person's postal code, by successful aging classification. The average number of

neighbourhood and industrial counts, as well as their medians, indicate that those who had aged unsuccessfully lived close to more establishments and services than those who aged successfully. Each variable was significant at $\alpha=0.05$, except for the 'police protection' variable.

TABLE 6.5.9 COUNT VARIABLE DISTRIBUTION FOR SUCCESSFUL AGING

	<i>Successfully Aging (n= 857)</i>	<i>Unsuccessfully Aging (n= 3356)</i>
Neighbourhood Count Variables:		
Mean +/- Standard Deviation		
Median (Q1, Q3)		
Organizations of community involvement	1.79 +/- 4.1 0 (0, 2)	2.50 +/- 7.1 0 (0, 2)
Cultural spaces	0.95 +/- 3.5 0 (0, 0)	1.23 +/- 4.4 0 (0, 1)
Drug stores	0.55 +/- 1.3 0 (0, 1)	0.72 +/- 1.5 0 (0, 1)
Food stores	1.61 +/- 3.0 0 (0, 2)	2.06 +/- 3.6 1 (0, 3)
Health care offices and clinics	7.33 +/- 17.9 1 (0, 6)	8.00 +/- 17.5 2 (0, 7)
Home health care services	0.07 +/- 0.3 0 (0, 0)	0.12 +/- 0.5 0 (0, 0)
Hospitals	0.16 +/- 0.8 0 (0, 0)	0.22 +/- 1.2 0 (0, 0)
Liquor and tobacco stores	0.26 +/- 0.7 0 (0, 0)	0.36 +/- 1.0 0 (0, 0)
Spaces of physical activity	1.02 +/- 2.3 0 (0, 1)	1.29 +/- 2.7 0 (0, 2)
Police protection	0.13 +/- 0.9 0 (0, 0)	0.19 +/- 1.3 0 (0, 0)
Religious organizations	1.59 +/- 2.5 1 (0, 2)	1.95 +/- 2.8 1 (0, 3)
Social services	0.51 +/- 2.1 0 (0, 0)	0.70 +/- 2.3 0 (0, 0)
Socializing spaces (eating and drinking)	5.13 +/- 11.9 1 (0, 5)	6.65 +/- 14.0 0 (0, 2)
Industrial Count Variables:		
Agriculture, Forestry and Fishing	0.65 +/- 1.3 0 (0, 1)	0.75 +/- 1.3 0 (0, 1)
Finance, Insurance and Real Estate	6.58 +/- 14.2 2 (0, 6)	8.81 +/- 20.7 3 (0, 9)
Manufacturing	1.90 +/- 8.8 0 (0, 2)	2.11 +/- 8.1 0 (0, 2)
Public administration	2.02 +/- 7.1 0 (0, 1)	2.51 +/- 9.3 0 (0, 2)
Retail trade	17.75 +/- 35.9 5 (1, 19)	21.89 +/- 40.8 8 (2, 25)
Services	35.04 +/- 63.1 14 (5, 40)	42.3 +/- 75.1 18 (6, 45)
Wholesale trade	3.08 +/- 6.0 1 (0, 4)	3.69 +/- 6.9 2 (0, 4)

The results suggest that proximity to facilities is not associated with successful aging. An overriding factor may be socio-economic status, in that richer people may live in more residential neighbourhoods that also have fewer business establishments in them. Table 6.5.10 shows that in the census tracts where successful aging individuals live, the median income is \$1110.26 greater than those aging unsuccessfully. In the census tracts linked to the successful aging participants, the average income was \$1524.10 higher than for non-successful aging participants. Univariate non-parametric analyses on median and average income again indicate a significant (at $\alpha=0.05$) difference between successful and unsuccessful aging individuals. The analyses producing the results seen in Table 5.6.10 did not account for participants, either successfully aging or not, living in the same census tract. However, subsequent analyses did control for individuals living in the same census tract, by selecting for unique census tracts within the successful and unsuccessful aging groups. These results are summarized in Table 6.5.11.

TABLE 6.5.10 INCOME DISTRIBUTION FOR SUCCESSFUL AGING

	<i>Successfully Aging (n= 857)</i>	<i>Unsuccessfully Aging (n= 3356)</i>	<i>Δ (\$)</i>	<i>p-value*</i>
Census Variables:				
Mean +/- SD				
Median (Q1, Q3)**				
Median Income (\$)	21 137.14 +/- 5232.69	20 026.88 +/- 5040.46	1110.26	<0.001
	20 581 (17 331, 24 686)	19 482 (16 443, 22 798)		
Average Income (\$)	27 561.13 +/- 8326.53	26 037.03 +/- 7821.50	1524.10	<0.001
	25 513 (22412, 30803)	24 493 (21086, 28905.5)		

* p-value based on the Wilcoxon Rank-sum test

** Q1 represents the first quartile (25th percentile), and Q3 represents the third quartile (75th percentile)

TABLE 6.5.11 INCOME DISTRIBUTION FOR SUCCESSFUL AGING BY UNIQUE CENSUS TRACT

	<i>Successfully Aging (n= 380)</i>	<i>Unsuccessfully Aging (n= 769)</i>	<i>Δ (\$)</i>	<i>p-value*</i>
Census Variables:				
Mean +/- SD				
Median (Q1, Q3)**				
Median Income (\$)	21 778.99 +/- 5395.6	20 676.95 +/- 5160.5	1102.04	0.001
	21 354 (17 804, 25 313)	20 267 (16 972, 23 953)		
Average Income (\$)	27 967.16 +/- 8548.4	26 429.39 +/- 7864.3	1537.77	0.001
	25 980 (22 487, 31 655)	25 006 (21 503, 29425)		

* p-value based on the Wilcoxon Rank-sum test

** Q1 represents the first quartile (25th percentile), and Q3 represents the third quartile (75th percentile)

The results from analyses controlling for census tract overlap between participants as seen in Table 6.5.11 did not differ greatly from results from analyses not controlling for overlap seen in Table 6.5.10. The median income is still approximately \$1100 greater than those aging unsuccessfully, and the average income was approximately \$1500 higher for census tracts in which successful aging participants lived, than in those inhabited by non-successful aging participants.

Results are further discussed in the Discussion section.

7. DISCUSSION

7.1 Objective: Creation of database

The work done for this thesis resulted in the creation of the first combined spatial and aspatial database composed of demographic, socio-economic and GIS-derived local contextual spatial data linked to individual successful aging outcome data via postal code. This has met the first objective set out in this thesis.

The database contains individual-level data from the Canadian Study of Health and Aging (Cycle 1 and 2), and area-level data from DMTI Spatial Inc. and the 1996 Canadian Census at a national scale. The area-level data were linked via postal code (spatial data) and census tract (Census data) to a subset of CSHA participants who had been characterized as having aged successfully or not in the five-year interval between Cycle 1 and 2.

It is only very recently that contextual data other than those collected by Statistics Canada, have been made available at the scale of this study. These are still not complete, thus the assessment of Canadian data sources for this level of study is a contribution in itself.

7.1.1 Limitations

There do exist certain limitations of the database; one concerns the temporality of the data. Though the successful aging outcome is based on cohort data collected over several years, the explanatory variables assembled constitute information collected at one point in time. DMTI's spatial data were current as of Spring 2002, and the Census consisted of data collected on one particular day in 1996. Growing ethical concerns over

maintaining participant confidentiality over the time-span of the CSHA constrained the study's ability to include data on change of residence in the dataset. There is no guarantee that participants were still living in the same neighbourhood in 1996 or 2002 as they had in 1991, though comparisons could possibly be made with Cycle 3 postal code data in future studies. Studies in the U.S. have shown that 10 or 20-year-old census data¹² linked to participants, at the census tract level¹³ does not significantly differ (76). Unemployment may vary over time within locales due to regional and other macroeconomic effects on employment levels, but the relative wealth or poverty of specific locales remains stable in the United States, at least over a 10-year period (76). Thus it is hypothesized that in Canada there will not be drastic socio-economic differences in particular census tracts over a period of five to eleven years.

There are also limitations with respect to the socio-demographic data contained in the database created for this thesis as it lacks area-level ethnic or religious data. A Medline literature search on race, racial effects, and ethnicity indicated a lack of published literature in the context of neighbourhood effects on health in Canada. This may be because the contextual effect of ethnicity is not as relevant to Canada as it is in countries such as the United States (77;78), or that the literature review needed to be more extensive. It is recommended that if future studies will compare regions or countries, cultural and historical legacies may be important contextual considerations in the analysis, but are very hard to capture owing to the great diversity in Canada, and to

¹² The U.S. Census is conducted every 10 years; the Canadian Census collects data every five years.

¹³ U.S. postal geography differs from Canadian postal geography. What would be equivalent to our postal code, the zonal improvement plan (ZIP) code, corresponds to approximately 25 000 residents. The census tract population sizes are roughly equivalent between the two countries; the U.S. census tract numbers approximately 5000 people, while the Canadian census tract population size ranges between 2500 and 8000, though is averaged at approximately 4000 people.

differing rates of assimilation. In addition, it may be that rural participants, who were excluded in the analyses for this thesis, may be different from the census tract-dwelling participants and should be analysed as well.

The outcome data itself may be subject to bias. The successful aging data is used as a secondary analysis, using a data set collected for another purpose – to have aged ‘successfully’ was arbitrarily defined as the top 20%, and the 80% ‘unsuccessful’ will be very heterogeneous.

7.2 Objective: Geography and epidemiology

For the second objective of the thesis, geographic information systems (GIS) capabilities were used to derive neighbourhood-level variables and to link geographic features to individuals in the Canadian Study of Health and Aging (CSHA). The methods for creating a database using geographic information systems (more specifically ArcGIS 8.2) were investigated and methodological issues involving data at different levels were explored.

Using GIS, CSHA participants were mapped according to their postal code location (Figure 6.3.1), which was the first time that CSHA study investigators could see the spatial distribution of participants in their study. Figure 7.2.1.1, for example, shows that clustering of participants occurred around one of the CSHA study centres, Edmonton, Alberta. The circles represent the location of the CSHA participants’ postal codes, which serve as a proxy for the residential location of the CSHA participant. The 1996 census tracts that correspond to Edmonton are overlain to show the extent of clustering in the city’s core.

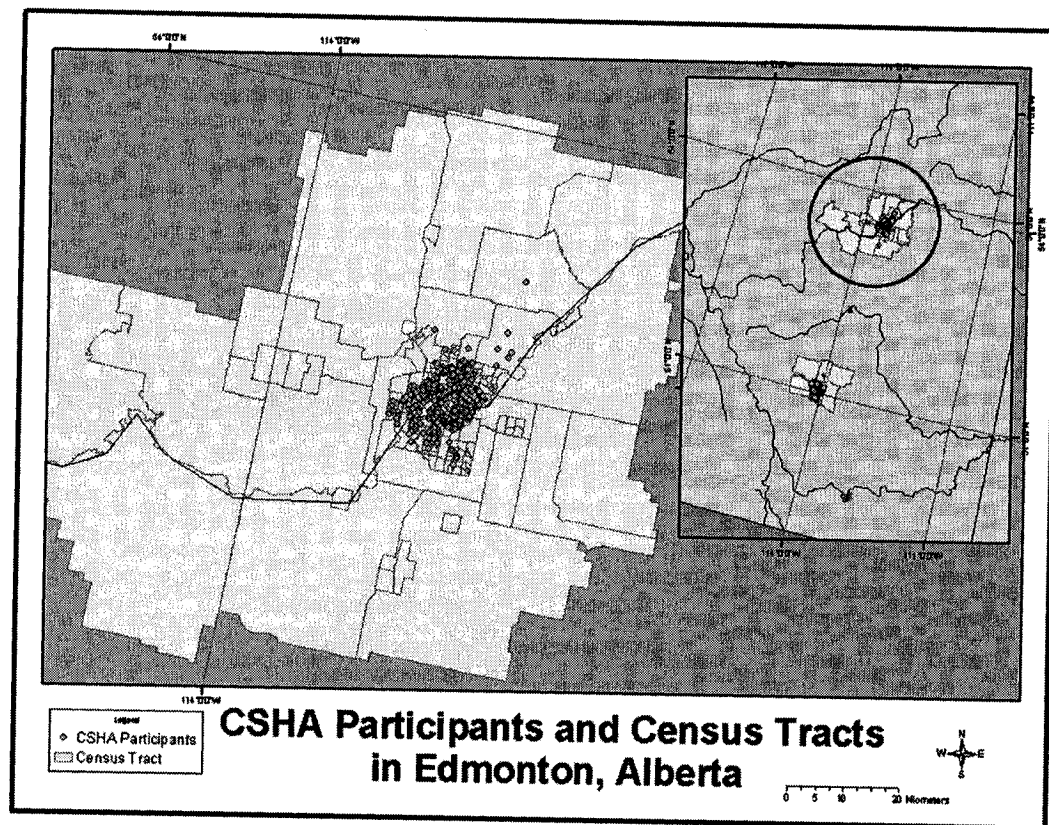


Figure 7.1.1.1 Map of Edmonton, Alberta indicating CSHA participant postal code locations overlain on census tracts.

In other study centres, such as St. John's, Newfoundland, participants were relatively more spread out (Figure 7.1.1.2). The same scale was used in both maps to contrast the differences in clustering, due to the way participants in the two study centres were sampled.

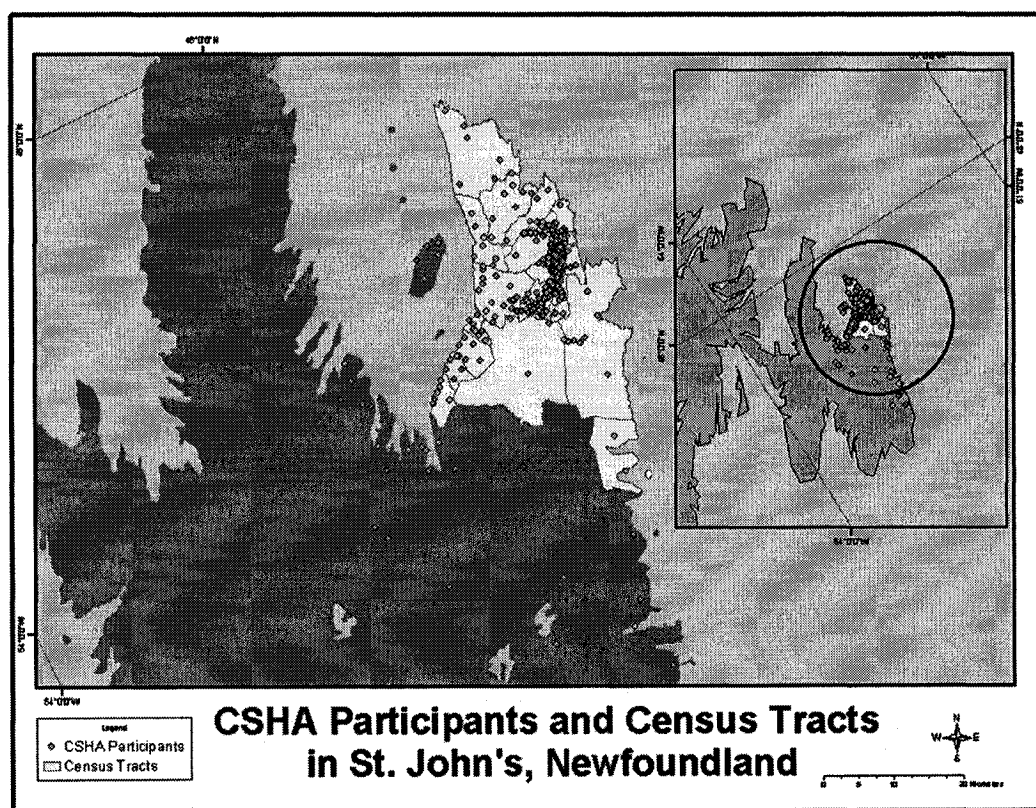


Figure 7.1.1.2 Map of St. John's, Newfoundland indicating CSHA participant postal code locations overlain on census tracts.

7.2.1 Geographies of scale

7.2.2.1 Postal Codes

Due to privacy constraints, the addresses of CSHA participants were not available and this study was required to use postal codes to link individuals to their geographic locations. This required assessing the degree to which postal geography could be used as a means of locating participants, as well as evaluating the problems that postal geography may present with respect to the accurate location of participants within their contextual environment.

From personal communication with Canada Post representatives, it was ascertained that postal codes are assigned according to the amount of mail sent and received; this often correlates with population density, but there are no set numbers or quotas. Geographical borders are set and everyone within that border is assigned that postal code.

The analyses on postal code geography indicated great variability in postal code area, with possible values in this study ranging from 12.4m² to 2542.3 km².¹⁴ In some cases, the postal code area was as large as a census tract. In addition, a unique postal code could be mapped to many places, as seen by using the Multiple Enhanced Postal Code file. Even when the single most representative position reflected the location of the most number of addresses corresponding to that postal code, in reality this postal code position may be several hundred metres from the true location of the individual. The great variability in postal code area, as well as the lack of positioning with respect to the individual limits the use of postal codes in studies.

Another limitation is that the DMTI enhanced points of interest are approximately +/- 20 metres within correct placement from address-based geocoding. Future studies may elect to run sensitivity analyses that randomly offset each participant's postal code according to the distance interval of the square root of the postal code area. The analysis could then be re-run and repeated a number of times to see if the effects produced really exist.

Given these limitations, the differences in outcomes between males and females, and between successful aging and unsuccessful aging participants, may be even more

¹⁴This imprecision systematically arises in rural areas.

pronounced if more precise individual-level data (such as addresses) could be obtained, or if the results found were an artefact of poor positioning.

This component of the thesis was intended to aid informed decisions in future applications when postal geography is used as a means for summarizing the contextual environment surrounding individuals.

7.2.2.2 Residential areas

Census tracts were selected as the level at which census data would be used as they best approximated what might constitute an individual's 'neighbourhood.' However, neighbourhoods are complex environments. They have physical, social, economic, institutional and cultural dimensions, as well as histories and reputations (79). In recognition of how people's lives are shaped by characteristics of their neighbourhoods, Robertson and Weir (79) address the concept that one neighbourhood or geographic space can have many different characteristics depending on the time of day and a person's age, race, or ethnicity. At the same time, a person can live in a neighbourhood for a substantial period of time and never interact with some characteristics.

Residential areas have traditionally been the focus in epidemiological studies on social environment and health. However, limitations may exist in the older notion that the 'geography' of the 'community' is spatially delimited. Macintyre et al. (12) point out that collective and social functioning and practices do not need to be geographically restricted to residential neighbourhoods – collective identity and opportunity may be as much related to other types of spaces such as the family, friends and the workplace.

These cut across the concept of “traditional” geographical space and are only anchored in ‘place’ by common meeting points, such as work or school (12), or interests and recreational past-times.

The complexity of a neighbourhood may make it difficult to fully isolate its contributing effects to health. In addition to spaces of psychosocial construction, future studies might also study non-traditional geographic space as an aspect of an individual’s social environment, such as virtual cyber-communities. Technological advances and the increasing accessibility to the World Wide Web in most parts of the world may make the notion of ‘neighbourhood community’ as rooted in physical geography, irrelevant.

7.2.2.3 Spatial Autocorrelation

In ecological-level studies, scale dependency issues, such as spatial autocorrelation, are a common concern (18). Spatial autocorrelation is almost always present when units of observation (e.g. census tracts) are imposed and exist in spatial relationships with each other (in terms of adjacency, distance, etc.) (1). When used as units of observation, census tracts or municipalities are not truly independent of each other, as most statistical procedures require, since adjacent territories usually influence each other.¹⁵ Median income in adjacent CTs, for example, is expected to be more similar than median income between two CTs that are a greater distance apart. This is due to the Modifiable Area Unit Problem (MAUP) (82;83). In health studies, the modifiable area unit problem occurs when environment-disease interactions appear at one

¹⁵There are numerous statistical procedures designed specifically to adjust for spatial autocorrelation (80;80;81). Yiannakoulis writes that little work has been done to examine the effects of sampling error in the Canadian census and its effects on regression models therefore it is difficult to quantify the effect of spatial autocorrelation in census data (18).

level only to disappear at another level (84). Unless the areal unit is appropriate to the scale of the patterns and processes under study, administrative boundaries will group higher exposed with lower exposed subjects, introducing ecologic or aggregation bias due to misclassification (84;85).

In this thesis, the use of census tract data in this study was selected to minimize the variability within areas, increasing the chances that the individual in the study has more in common with others in his area, thus facilitating study of compositional and contextual effects on health. However, concerns over spatial correlation in this study apply to those participants who share the same census tract. Identical information will be attributed to all individuals who belong to the same census tract and should be taken into consideration when running analyses.

7.2.2 Querying via SIC

A methodological limitation initially seen in the Methods section (regarding community centres being coded under different categories) is the imprecision of coding of the establishments and points of interest. Using the primary and sometimes secondary standard industrial classification codes associated with the points of interest did not always allow for a more precise identification of the particular service or establishment. It was also observed that the SICs were not coded consistently across all the provinces, so queries were constructed to be as inclusive and wide ranging in order to capture the data that may be relevant but found in other categories (e.g. community centres being classified as both ‘fitness facilities’ as well as ‘social services’). The services corresponding to the codes differed slightly from province to province, perhaps due to

differences in the phone and yellow page directories from which the establishments were coded. Given that the scope of this database is national and that there are hundreds of thousands of points of interest, it was difficult to check that every selected feature was completely relevant to the individual, though attempts were made to be as vigilant as possible.

Despite the rather broad generalization of the coding, it may be hypothesized that the existence of all types of services (directly relevant or not) could collectively benefit the community. Future studies could use the broad generalization of services to create an index of the various services in the neighbourhood, and determine whether the indicator helps predict successful aging.

Were the study conducted on a smaller scale, such as investigating a single community or region for example, the specific type of social service and other points of interest could be more precisely identified.

Another methodological limitation is that the proximity from CSHA participant postal code location to needed amenities was measured in a straight line. Given the complexity of the urban and suburban structures, it is unrealistic to expect that the topography of the individual's residential location would allow them to directly access the points of interest 'as the crow flies.' Future GIS analysis could incorporate street network files to assess means and facility of access. However, our consistent failure to find a link between distance and successful aging suggests this refinement may not be productive.

7.3 Objective: Preliminary Statistical Analysis

Descriptive analyses of the individual and area-level data in relation to successful aging, the third objective of this thesis, was undertaken primarily on the GIS-derived distance variables, but also on the GIS-derived count variables as well as on certain census variables. This was done as a first step to ascertain in part the viability of pursuing advanced statistical and modelling methods.

The questions in *Proximity: Nearest essential services* and *Containment: Nearest community establishments and services* were posed to help guide the selection of the variables determining what variables would be selected. For *Proximity: Nearest essential services*, the proximity of the nearest ‘needed’ facility was determined using spatial joins in GIS. It was hypothesized that individuals living closer to needed services would in fact age more successfully than individuals living farther away from needed services. Results show that there are indeed differences between successfully and unsuccessfully aging participants, but the mean distances suggest the inverse of what had been hypothesized. Individuals who aged successfully, for example, live approximately 1.7 kilometres from the nearest grocery store, while those unsuccessfully aging live approximately 194 metres closer to a grocery store. The sole exception to this trend is proximity to recreational centres where those that aged successfully live on average one kilometre closer.

Results from preliminary analyses of the count data also indicate that those aging successfully had fewer services and industrial features in the 500 m radius around their postal code location (the sole exception being police protection services) than those aging unsuccessfully.

An explanation for the hypothesis not being met might be the influence of income. Mustard et al. (86) argue that urban populations are strongly stratified on the basis of the cost of housing with the ability to afford housing determined by income, which means that most individuals or families rent or buy only so much housing as they can afford.¹⁶ If this is the case in successful aging, it may mean that individuals in this study who are aging successfully tend to live further away from services because they don't need to live close to them. Individuals who can afford a car and drive can easily access needed services and do not need to be located in close proximity to them. Those that aged unsuccessfully may have chosen to live closer to services to be able to more easily access them on foot or via public transportation. Again it may be that those that aged unsuccessfully are only able to afford housing in a cheaper, high-density area, which is usually urban versus suburban, and which will have a greater concentration of services due to population density. People might not be able to afford to live further out, where they might need a car to access amenities.

In outlying principally residential areas, there may be a greater number of community and senior centres, due to demographic characteristics of the areas (e.g., more families and older couples living in the suburbs, individuals living in suburban areas earning higher income, etc.).

The correlation matrix (Table 6.5.5) indicates that all distance variables, except for distances to recreational facilities, are highly correlated, indicating that services tend to be clustered and distance to services tends to be similar.

¹⁶ In geographic settings with substantial rural populations, there may be less of a sorting of households by housing cost, indicating that there is greater within-neighbourhood income variance in rural communities than in urban communities (86).

Logistic regression was used as a first approximation to determine probable predictors of successful aging among the categorized distance variables. Though the wide confidence intervals associated with the odds ratios included the null value indicating no effect, it was hypothesized that gender may be confounding the results. After controlling for gender, the 95% confidence intervals corresponding to male and female odds ratios still contained the null value, although a slight gender effect could be seen. It remains unclear why this should be.

Advanced statistical methods, such as multilevel analysis or contextual analysis, that include individual level and group level predictors, may be required to fully explain these preliminary results.

Multilevel analysis and contextual analysis are similar in that they both allow for the analysis of how group level and individual level variables, as well as their interactions, are related to individual outcomes. Multilevel analysis is used for data with nested sources of variability (such as individuals and individual income nested within their census tracts and median or average income of that census tract). According to Diez-Roux (37), this analytic method allows for the simultaneous examination of the effects of group level and individual level variables on individual level outcomes, while accounting for the non-independence of observations within groups. Multilevel models (also called hierarchical models) can be used to draw inferences regarding the relation of group and individual level variables to individual level outcomes, and inferences can also be made regarding inter-group variation, whether it exists and to what extent it is accounted for by group and individual level characteristics (37).

In contextual analysis, group-level (often aggregate) variables are included together with individual-level variables in standard regressions with individuals as the units of analysis. This allows for simultaneous investigation of how individual level and group level variables relate to individual outcomes. Unlike multilevel models, however, contextual models do not usually account for residual correlation and do not allow for the examination of inter-group variability or of the factors associated with inter-group variability (37).

The inclusion of individual-level data, along with group-level data in the database will facilitate contextual and multilevel analyses in future studies. Multilevel analysis was not appropriate for this thesis as the numbers of CSHA participants in each census tract were too few to satisfy the requirements for a multilevel analysis. The only way this could have been undertaken would be to merge the census tracts so that they would contain the appropriate number of observations. This, however, would create much more heterogeneous areas which would have reduced the differences between them, therefore removing any effect.

7.3.1 Contextual effects discussed

Ecob and Macintyre (87) argue that it is important to distinguish between compositional and contextual explanations for area patterning in order to understand better the geographical variations and to design policies to reduce health damaging behaviour.

In keeping with the definition where compositional explanations refer to the characteristics of individuals concentrated in particular places and contextual

explanations refer to the group properties and opportunity structures (such as services) in the local physical and social environment, compositional variables included in this database may include certain Census variables. The pattern of income distribution in a census tract, for example, refers to the income characteristics of the individuals located within a particular census tract. A contextual explanation may refer to the average or median income. Care must be taken to clearly define the compositional and contextual characteristics of interest a priori.

Contextual variables included in this database may also include many of the Census variables, some of which had been derived by aggregating individual-level data, such as average level of education attained in a census tract. Other contextual variables are integral variables, such as median income or proportion of individuals renting their dwelling. In addition, the GIS containment-based datasets produced here could also be considered as contextual variables as they represent the number of features located within a 500 m buffer zone around the location of the CSHA participant's postal code.

Contextual effects may either impact on all residents equally, or more on some types of residents than others (males compared to females, for instance) (87). Contextual effects can only be detected in studies that involve both within-population and between-population comparisons (or both individual-level and population-level factors) (88).

7.3.2 Causality and social context's direct effects on health

The past few years have seen a renewed interest in the social determinants of health and how our social environment influences health, as health and disease are influenced by individual factors and also factors at the group or community level (89).

However, interpreting meaningful associations between ecological-level risk factors, like neighbourhood income, and broad inclusive health outcomes, like neighbourhood mortality rate, is difficult given the requirements of cause-effect inference as it is defined in the epidemiological tradition (18).

According to Hennekens and Buring (90), assessing causality of an association extends beyond the validity of the results of any single study and includes consideration of other epidemiologic data as well as the biologic credibility of the hypothesis. In addition, two additional criteria for causation include an appropriate time sequence of the association and the presence of a dose-response relationship (90).

In social epidemiology, these criteria become difficult to meet. The social dimensions of the environment we live in include the groups to which we belong, the neighbourhoods in which we live, the organization of our workplaces, and the policies we create to order our lives (91). It can be seen that environmental factors influence the health of populations and individuals, and also that it is possible to identify specific elements of the social environment that impact health; these include community socio-economic status, social structures (e.g. income inequality) and the quality of the environment (richness of life in the neighbourhood, or phenomena such as crime occurrence, local resources or social cohesiveness) (91).

An examination of the pathways linking a component of our social environment, social networks, to health outcomes yields a complex web of interlinking mechanisms – biological, psychological and bio-psychological – that cascades from the macro to the micro, from the upstream to the downstream, to generate potentially powerful influences on health and well-being across the life course (92).

The stage of life at which the effects of environment on health are most influential is widely debated. Is it in the early years? Is it in middle age or later? Are they equally important? Regardless of timing it is known that large-scale social upheavals and transitions profoundly disrupt patterns of social organizations established in earlier life (92). Geographical relocation related to urbanization, housing policy or employment, large-scale social change or depression, and job stress all represent environmental challenges that tear at social networks, which in turn has deleterious consequences on health (92).

Direct biological effects of the social environment are difficult to isolate. There is some speculation that social isolation, disintegration, and disconnectedness are chronically stressful conditions to which increase the rate of aging of the organism, thereby influencing mortality and life expectancy (92). Isolation would then also be associated with age-related morbidity and functional decline (93) and consequently unsuccessful aging.

7.3.3 Future Directions

7.3.3.1 Collective Measures

As an illustration of social context mentioned earlier, as well as a possible future direction for analyses of these data, two measures of collectivity increasingly being studied in the context of public health are social capital and social cohesion. Social cohesion refers to the presence of strong social bonds and an absence of latent social conflict in society (94). Social capital, a subset of social cohesion, is defined as those features of social structures, such as interpersonal trust and norms of reciprocity and

mutual aid, which act as resources for individuals and facilitate collective action (94). Increasing literature on social capital indicates its relevance to health in general (20;94-102), and to successful aging in particular (96). Cannuscio et al. (96) have found that as the population ages, increasing numbers of seniors are living alone, and the availability of social capital will become an important ingredient in successful aging.

Indicators of social capital¹⁷ can be constructed from data from the Canadian General Social Survey, a survey conducted nationally by Statistics Canada. I had made several attempts to obtain Canadian aggregate-level data, either at the enumeration-area or census tract for the General Social Survey (GSS).¹⁸ These attempts were met with little success, primarily due to concerns over issues of confidentiality given the comparatively small sample size of the annual survey (which increased to 25 000 annually in 1998 from 10 000). Overall indicators of social capital could be created from questions found on the survey, such as those concerning personal risk and victimization (e.g. What is the level of crime in your neighbourhood compared to other areas in Canada? Do you feel safe walking alone in your neighbourhood at night?). Specific indicators can be created to include social activities and time use (involvement in political activities? church groups?), and reciprocity (How often do you feel that by helping others, you simply give back what you have received from them?). The list of

¹⁷ In a study assessing the relationship between social capital at the state-level and self rated health Kawachi et al. (101) aggregated social capital indicators which had been obtained from the National Opinion Research Center's General Social Surveys. Indicators included levels of interpersonal trust (% of citizens responding 'Most people can be trusted'), norms of reciprocity (% of citizens responding 'Most people are helpful'), and per capita membership in voluntary organizations (94). These particular indicators, along with voter participation and perception of personal risk, tend to make up the primary indicators of social capital.

¹⁸ The GSS is held every year, and some core questions are asked year after year while other questions were asked during different cycles of the survey, depending on the theme of the survey of that year. The responses from the same questions in different survey years were to have been aggregated, thus increasing sample size and increasing the probability of obtaining sufficient numbers of responses corresponding to the areas in which CSHA participants lived.

variables I had selected to use, and a sample of the original questions they had come from on the General Social Survey can be found in Appendix X. Including data from the GSS would improve the database and create additional research possibilities to researchers interested in studying the social environment of the elderly. Due to a lack of success in obtaining the necessary data from Statistics Canada, this course of action was delayed to future studies.

It is important to note that there has been some criticism over the importance placed on social cohesion and social capital as related to health. Macintyre et al. (12) argue that focus on social cohesion, social capital and perceived position in social or economic hierarchies has ignored many other aspects of shared social functioning such as ethnic, regional and national identity, religious affiliation, political ideologies, kinship systems, domestic division of labour, among many others.

This critique reaffirms the multidimensionality of factors influencing health, and reinforces the importance of considering and creating a thorough conceptual framework with which to study the health outcome of interest.

7.3.3.2 Biophysical Environment Data

Major advances in the study and prediction of vectored diseases and epidemic surveillance in this century is coming from the ongoing development and application of GIS methods (1). However, developments in GIS have not been solely confined to disease surveillance. GIS has benefited and been improved from important technical innovations in the field of remote sensing and improvement in analytic methods for remotely sensed data.

Remote sensing is the sensing of the Earth's surface from space, via satellites,¹⁹ by making use of the properties of electromagnetic waves emitted, reflected or diffracted by the sensed objects, for the purpose of improving natural resources management, land use and the protection of the environment. The imagery from remote sensing is ideally used to monitor and detect land cover changes that occur frequently in urban and peri-urban areas (103). Remotely sensed data can resolve a total environment into its components through an interpretation of the different facets of the same area - the result is a number of map layers, such as vegetation or houses, each representing a subspace with its own characteristics (104). This is important in medical geography and its study of how the context and composition of space influence health.

Three major types of environmental data can be extracted from Landsat Thematic Mapper (TM) data (104): 1) land use and land cover, 2) surface temperatures and 3) Normalized Difference Vegetation Index (NDVI).²⁰ These variables can be obtained directly from the Landsat TM data by using a computer-assisted image-processing-GIS approach.

¹⁹ Satellites have provided an enormous amount of data in health studies involving remote sensing, especially the American LANDSAT (Land Remote-Sensing Satellite). Photography and digital data in multiple bands of the electromagnetic spectrum can be used to differentiate types of land cover and usage to trace pollution in waterways; infrared heat sensing can differentiate thermal plumes in water and air (and thus surface conditions and cover) (1).

²⁰ The NDVI is derived from multi-spectral data obtained through biophysical remote sensing via satellites orbiting the earth. This measure consists of a ratio transformation accentuating the contrast between the visible spectrum (0.4-0.7 μ m), which more strongly reflects energy from soils and litter, and the near infrared spectrum (0.7-1.1 μ m), which more strongly reflects energy from healthy green vegetation (104). For Landsat Thematic Mapper data, NDVI is computed from TM band 4 (0.76-0.90 μ m) and TM band 3 (0.63-0.69 μ m) (104). The value varies from -1 to +1 as greenness increases, and is a ratio computed by using the IDRISI, a type of GIS program (104). Areas which have been built up have negative NDVI values and appear in darker tones on the image, whereas areas with green vegetation have high positive NDVI values which appear as very light tones on the image.

The NDVI is of particular interest. It is one of the most often-used indicators of greenness in biophysical geography,²¹ and greenness has been shown to be important for many health outcomes, including health among the elderly (34). In addition, the NDVI's positive correlation with per capita income, median home values, and education attainment indicates that it relates the biophysical environment to the socio-economic environment so well it may be used as a surrogate of the socio-economic condition of an area – some have also found it to be a useful quality of life measure (104).

Though the NDVI was not calculated or used in this thesis, its potential use in health studies, along with other remote sensing data, is illustrated for future studies and for the creation of biophysical and environmental variables that can be incorporated into the existing database.

²¹ The NDVI has also been used in health to study vector-borne diseases (105).

8. CONCLUSION

The ultimate goal of epidemiologic research is to identify causal pathways or mechanisms of disease (91). Attempting innovations and using methods and tools developed in other disciplines can contribute to this goal. In this thesis, GIS was used in conjunction with epidemiologic as well as socio-demographic and spatial data to go beyond traditional mapping, and was used for data manipulation, analysis and spatial joining of features. There is still great potential for GIS utilization in health studies and certainly more study is necessary in GIS applications to community health studies. There are also methodological implications of data at multiple levels and how they influence health, which it is anticipated that future users of the database created in this thesis will pursue and contribute to further advances in epidemiology and successful aging.

With the aging of populations and the rise in health care costs, research agendas in gerontology have increasingly come to include studies of healthy, or successful, aging (41). With respect to policy research on successful aging, it has been argued that public or organizational policy decisions on resource allocation to enhance successful aging cannot be made with anything but an integrated health determinants (biopsychosocial) model (38). Inui (38) believes that the questions asked in such policymaking will not reside within one or another of the categorical determinants of health when the choices contemplated are often tradeoffs among completely different resources, such as bus passes, medications, city parks, senior citizen centres, or exercise programs. Given the complexity and multiplicity of pathways contributing to health status, incorporating physical environmental data, such as the normalized difference vegetation index, and collective measures, such as social capital or other psycho-social collective constructs, to

the already existing database, would make it more complete. Having such data in a single database (which currently includes individual outcome data, socio-economic factors and other group-level variables) is an important step in facilitating research on the multiple factors and pathways influencing successful aging and other health outcomes. Studying these data will result in more completely understanding what factors influence successful aging, and may contribute greater insight when making policy decisions regarding increasing successful aging.

REFERENCES

- (1) Meade M, Earickson R. Medical Geography. 2nd ed. New York City: Guilford Press, 2000.
- (2) Brody H, Rip MR, Vinten-Johansen P, Paneth N, Rachman S. Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. *Lancet* 2000; 365:64-68.
- (3) Tanser F, Lesueur D, Solarsh G, Wilkinson D. HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. *Tropical Medicine & International Health* 2000; 5(1):40-46.
- (4) Kistemann T, Munzinger A, Dangendorf F. Spatial patterns of tuberculosis incidence in Cologne (Germany). *Soc Sci Med* 2002; 55(1):7-19.
- (5) Latkin C, Glass GE, Duncan T. Using geographic information systems to assess spatial patterns of drug use, selection bias and attrition among a sample of injection drug users. *Drug & Alcohol Dependence* 1998; 50(2):167-175.
- (6) Molyneux DH. Vector-borne infections in the tropics and health policy issues in the twenty-first century. *Transactions of the Royal Society of Tropical Medicine & Hygiene* 2001; 95(3):233-238.
- (7) Yang G, Zhou X, Malone JB, McCarroll JC, Wang T, Liu J et al. GIS prediction model of malaria transmission in Jiangsu province. *Chung-Hua Yu Fang i Hsueh Tsa Chih [Chinese Journal of Preventive Medicine]* 2002; 36(2):103-105.
- (8) Bergquist NR. Vector-borne parasitic diseases: new trends in data collection and risk assessment. *Acta Tropica* 2001; 79(1):13-20.
- (9) Martin C, Curtis B, Fraser C, Sharp B. The use of a GIS-based malaria information system for malaria research and control in South Africa. *Health & Place* 2002; 8(4):227-236.
- (10) Boulos MK. Medical Geography . www.soi.city.ac.uk/~dk708/pg1_1.htm 2003.
- (11) Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001; 55(2):111-122.
- (12) Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002; 55:125-139.
- (13) Kearns RA. Place and Health: Towards a reformed medical geography. *Professional Geographer* 1993; 45(2):139-147.

- (14) Diez Roux AV. Investigating neighborhood and area effects on health. *American Journal of Public Health* 2001; 91(11):1783-1789.
- (15) Macintyre S, Ellaway A. Ecological Approaches: Rediscovering the Role of the Physical and Social Environment. In: Berkman LF, Kawachi I, editors. *Social Epidemiology*. Oxford: Oxford University Press, 2000: 332-348.
- (16) Macintyre S, Ellaway A, Hiscock R, Kearns A, Der G, McKay L. What features of the home and the area might help to explain observed relationships between housing tenure and health? Evidence from the west of Scotland. *Health & Place* 2003; 9:207.
- (17) Schwartz S, Diez Roux AV. Commentary: Causes of incidence and causes of cases - a Durkheimian perspective on Rose. *Int J Epidemiol* 2001; 30:435-439.
- (18) Yiannakoulis NW. Child pedestrian injury in an urban environment: risk in space and time. Department of Earth and Atmospheric Sciences, University of Alberta, 2000.
- (19) Cohen DA, Mason K, Bedimo K, Scribner R, Basolo V, Farley TA. Neighborhood Physical Conditions and Health. *American Journal of Public Health* 2003; 93(3):467-471.
- (20) Hancock T. People, partnerships and human progress: building community capital. *Health Promotion International* 2003; 16(3):275-280.
- (21) Spasoff RA. *Epidemiologic Methods for Health Policy*. New York City: Oxford University Press, 1999.
- (22) Marmot MG, Wilkinson RG. *Social Determinants of Health*. Oxford: Oxford University Press, 1999.
- (23) Krieger J, Higgins DL. Housing and health: time again for public health action. *American Journal of Public Health* 2002; 92(5):758-768.
- (24) Chatkin MN, Menezes AM, Victora CG, Barros FC. High prevalence of asthma in preschool children in Southern Brazil: a population-based study. *Pediatric Pulmonology* 2003; 35(4):296-301.
- (25) Crater DD, Heise S, Perzanowski M, Herbert R, Morse CG, Hulsey TC et al. Asthma hospitalization trends in Charleston, South Carolina, 1956 to 1997: twenty-fold increase among black children during a 30-year period. *Pediatrics* 2001; 108(6):E97.
- (26) Aylin P, Morris S, Wakefield J, Grossinho A, Jarup L, Elliott P. Temperature, housing, deprivation and their relationship to excess winter mortality in Great Britain, 1986-1996. *Int J Epidemiol* 2001; 30(5):1100-1108.

- (27) John U, Greiner B, Hensel E, Ludemann J, Piek M, Sauer S et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Sozial- und Praventivmedizin* 2001; 46(3):186-194.
- (28) Dedman DJ, Gunnell D, Davey SG, Frankel S. Childhood housing conditions and later mortality in the Boyd Orr cohort. *J Epidemiol Community Health* 2001; 55(1):10-15.
- (29) Crain EF, Walter M, O'Connor GT, Mitchell H, Gruchalla RS, Kattan M et al. Home and allergic characteristics of children with asthma in seven U.S. urban communities and design of an environmental intervention: the Inner-City Asthma Study. *Environmental Health Perspectives* 2002; 110(9):939-945.
- (30) Zock JP, Jarvis D, Luczynska C, Sunyer J, Burney P, European Community. Housing characteristics, reported mold exposure, and asthma in the European Community Respiratory Health Survey. *Journal of Allergy & Clinical Immunology* 2002; 110(2):285-292.
- (31) Saul C, Payne N. How does the prevalence of specific morbidities compare with measures of socioeconomic status at small area level? *Journal of Public Health and Medicine* 1999; 21:340-347.
- (32) Leventhal T, Brooks-Gunn J. Moving to opportunity: an experimental study of neighborhood effects on mental health. *American Journal of Public Health* 2003; 93(9):1576-1582.
- (33) Austin DM, Furr LA, Spine M. The effects of neighborhood conditions on perceptions of safety. *Journal of Criminal Justice* 2002; 30:417-427.
- (34) Kweon BS, Sullivan WC, Wiley AR. Green common spaces and the social integration of inner-city older adults. *Environment and Behavior* 1998; 30(6):832-858.
- (35) Harrison L, Heywood F. Health begins at home: the housing-health interface for older people. *Rev Environ Health* 2000; 15(1-2):149-167.
- (36) Shouls S, Congdon P, Curtis S. Modelling inequality in reported long term illness in the UK: combining individual and area characteristics. *J Epidemiol Community Health* 1996; 50:366-376.
- (37) Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health* 2002; 56:588-594.
- (38) Inui TS. The need for an integrated biopsychosocial approach to research on successful aging. *Annals of Internal Medicine* 2003; 139(5 Pt 2):391-394.

- (39) von Faber M, Bootsma-van der Wiel A, van Exel E, Gussekloo J, Lagaay AM, van Dongen E et al. Successful aging in the oldest old: who can be characterized as successfully aged? *Archives of Internal Medicine* 2001; 161(22):2694-2700.
- (40) Jorm AF, Christensen H, Henderson A, Scott, Jacomb PA, Korten AE et al. Factors associated with successful ageing. *Australasian Journal on Ageing* 1998; 17(1):33-37.
- (41) McDowell I, Kristjansson B, Aylesworth R. Healthy and successful aging: conceptual and empirical distinctions. Unpublished 2003.
- (42) Lee A. Maps reveal patterns hidden in databases: part 1. *Information Outlook* 1999; 3(12):18-20.
- (43) Brindley P, Maheswaran R. My favourite software: Geographic Information Systems. *Journal of Public Health Medicine* 2003; 24(2):149.
- (44) Jerret M, Burnett R, Kanaroglou P, Eyles J, Finkelstein N, Giovis C et al. A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A* 2001; 33(6):955-973.
- (45) Brainard J, Jones A, Bateman I, Lovett A. Modelling environmental equity: access to air quality on Birmingham, England. *Environment and Planning A* 2002; 34(4):695-716.
- (46) Scott P, Temovsky C, Lawrence K, Gudaitis E, Lowell M. Analysis of Canadian population with potential geographic access to intravenous thrombolysis for acute ischemic stroke. *Stroke* 1998; 29(11):2304-2310.
- (47) Love D, Lindquist P. The geographical accessibility of hospitals to the aged - a geographic information systems analysis within Illinois. *Health Services Research* 1995; 29(6):629-651.
- (48) Phillips R, Kinman E, Schnitzer P, Lindbloom E, Ewigman B. Using geographic information systems to understand health care access. *Archives of Family Medicine* 2000; 9(10):971-978.
- (49) Miranda M, Dolinoy D, Overstreet M. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. *Environmental Health Perspectives* 2002; 110(9):947-953.
- (50) Brainard J, Jones A, Bateman I, Lovett A. Modelling environmental equity: access to air quality on Birmingham, England. *Environment and Planning A* 2002; 34(4):695-716.
- (51) Hjalmar U. Environmental Epidemiology and GIS: Commentary on "GIS in Public and Environmental Health: Visualisation, Exploration and Modelling" by Anthony Gatrell.

http://geog.queensu.ca/h_and_e/healthandenvir/Finland%20Workshop%20Papers/HJALMARS.DOC; 2003.

- (52) Chapman NJ, Beaudet M. Environmental predictors of well-being for at-risk older adults in a mid-sized city. *Journal of Gerontology* 1983; 38(2):237-244.
- (53) Canada Post. Canada Postal Guide and Reference Tools. We 2003;Section B, Chapter 3.
- (54) Statistics Canada. 1996 Census Dictionary. Statistics Canada 1999;(92-351-UIE).
- (55) Statistics Canada. 1996 Census Handbook. Statistics Canada 2002;(92-352-XIE).
- (56) Burra T, Jerret M, Burnett RT, Anderson M. Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario. *The Canadian Geographer* 2002; 46(2):160-171.
- (57) Wilkins R. PCCF + Version 3J User's Guide. Statistics Canada 2002.
- (58) DMTI Spatial Inc. Enhanced Points of Interest: File Description. DMTI White Paper Series 2003.
- (59) Statistics Canada. Standard Industrial Classification - Establishments (SIC-E) 1980. Statistics Canada 2003.
- (60) DMTI Spatial Inc. Understanding Postal Code Files. DMTI White Paper Series 2003.
- (61) Teng EL, Chui HC. The Modified Mini-Mental State (3MS) Examination. *Journal of Clinical Psychiatry* 1987; 48:314-318.
- (62) Ware JE, Jr., Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston, Mass.: The Health Institute, 1993.
- (63) Ware JE, Jr., Donald CA. The MOS 36-item short-form health survey (SF-36). *Medical Care* 1992; 30:473-483.
- (64) Ryff CD. Beyond Ponce de Leon and life satisfaction: new directions in quest of successful ageing. *International Journal of Behavioral Development* 1989; 12(1):35-55.
- (65) Ryff CD, Essex MJ. Psychological well-being in adulthood and old age: descriptive markers and explanatory processes. *Annual Review of Gerontology and Geriatrics* 1991; 11:144-171.
- (66) Clarke PJ, Marshall VW, Ryff CD, Rosenthal CJ. Well-being in Canadian seniors: findings from the Canadian Study of Health and Aging. *Canadian Journal on Aging* 2000; 19(2):139-159.

- (67) Andrews SM, Withey SB. Social Indicators of Well-Being: America's Perception of Life Quality. New York: Plenum Press, 1976.
- (68) Staats SR, Stassen MA. Age and present and future perceived quality of life. *International Journal of Aging and Human Development* 1987; 25(3):167-176.
- (69) McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 2 ed. New York: Oxford University Press, 1996.
- (70) Spector WD, Fleishman JA. Combining Activities of Daily Living with Instrumental Activities of Daily Living to measure functional disability. *Journal of Gerontology: Social Sciences* 1998; 53B:S46-S57.
- (71) Podsiadlo D, Richardson S. The Timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatric Society* 1991; 39:142-148.
- (72) Rockwood K, Awalt E, Carver D, MacKnight C. Feasibility and measurement properties of the Functional Reach and the Timed Up and Go tests in the Canadian Study of Health and Aging. *Journals of Gerontology: Series A: Biological Sciences and Medical Sciences* 2000; 55A(2):M70-M73.
- (73) Iliffe JC. Datums and Map Projections: for remote sensing, GIS and surveying. Caithness, Scotland: Whittles Publishing, 2000.
- (74) Chapman NJ, Beaudet M. Environmental predictors of well-being for at-risk older adults in a mid-sized city. *Journal of Gerontology* 1983; 38(2):237-244.
- (75) Troosters T, Gosselink R, Decramer M. Six minute walking distance in healthy elderly subject. *European Respiratory Journal* 1999; 14(2):270-274.
- (76) Geronimus AT, Bound J. Use of Census-based Aggregate Variable to Proxy for Socioeconomic Group: Evidence from National Samples. *Am J Epidemiol* 1998; 148(5):475.
- (77) Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol* 2001; 30(4):668-677.
- (78) Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. *J Epidemiol Community Health* 1991; 45(1):35-42.
- (79) Robertson J, Weir K. Using geographical information systems to enhance community-based child welfare services. *Child Maltreatment* 1998; 3(3):224-234.

- (80) Fotheringham AS, Brunson C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. New York: Wiley, 2002.
- (81) Getis A, Griffith DA. Comparative Spatial Filtering in Regression Analysis. *Geographical Analysis* 2002; 34(2):130-140.
- (82) Jelinski DE, Wu JG. The modifiable areal unit problem and the implications for landscape ecology. *Landscape Ecology* 1996;129-140.
- (83) Qi Y, Wu JG. Effects of Changing spatial resolution on the results of landscape patterns analysis using spatial autocorrelation indices. *Landscape Ecology* 1996; 11(1):39-49.
- (84) Albert DP, Gesler WM, Levergood B. Spatial Analysis, GIS, and Remote Sensing Applications in the Health Sciences. Chelsea, Michigan: Ann Arbor Press, 2000.
- (85) Greenland S. Divergent biases in ecologic and individual-level studies. *Statistics in Medicine* 1992; 11(9):1209-1223.
- (86) Mustard CA, Derksen S, Berthelot JM, Wolfson M. Assessing ecologic proxies for household income: a comparison of household and neighbourhood level income measures in the study of population health status. *Health & Place* 1999; 5(2):157-171.
- (87) Ecob R, Macintyre S. Small area variations in health related behaviours; do these depend on the behaviour itself, its measurement, or on personal characteristics? *Health & Place* 2000; 6(4):261-274.
- (88) Diez Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health* 1999; 88:216-222.
- (89) Syme SL. Foreword. In: Berkman LF, Kawachi I, editors. *Social Epidemiology*. New York: Oxford University Press, 2000.
- (90) Hennekens CH, Buring JE. *Epidemiology in Medicine*. Philadelphia: Lippincott, Williams and Wilkins, 1987.
- (91) Yen IH. The Social Environment and Health: A Discussion of the Epidemiologic Literature. *Annual Review of Public Health* 1999; 20:287-308.
- (92) Berkman LF, Glass T. Social Integration, Social Networks, Social Support, and Health. In: Berkman LF, Kawachi I, editors. *Social Epidemiology*. 2000: 137-173.
- (93) Berkman LF. The changing and heterogeneous nature of aging and longevity: a social and biomedical perspective. *Annual Review of Gerontology and Geriatrics* 1988; 8:37-68.

- (94) Kawachi I, Berkman LF. Social Cohesion, Social Capital, and Health. In: Berkman LF, Kawachi I, editors. *Social Epidemiology*. New York: Oxford University Press, 2000: 175-190.
- (95) Baum F, Palmer C. 'Opportunity structures': urban landscape, social capital and health promotion in Australia. *Health Promotion International* 2002; 17(4):351-361.
- (96) Cannuscio C, Block J, Kawachi I. Social Capital and Successful Aging: The Role of Senior Housing. *Annals of Internal Medicine* 2003; 139:395-399.
- (97) Gooden BI. Social capital, stress and the health of rural African-Americans in central Virginia. University of North Carolina Chapel Hill, 1998.
- (98) Hawe P, Shiell A. Social capital and health promotion: a review. *Soc Sci Med* 2000; 51(6):871-885.
- (99) Hendryx MS, Ahern MM, Lovrich NP, McCurdy AH. Access to health care and community social capital. *Health Services Research* 2002; 37(1):87-103.
- (100) Hyyppa MT, Maki J. Individual-level relationships between social capital and self-rated health in a bilingual community. *Prev Med* 2001; 32(2):148-155.
- (101) Kawachi I, Kennedy BP, Glass R. Social capital and self-rated health: a contextual analysis. *American Journal of Public Health* 1999; 89(8):1187-1193.
- (102) McKenzie K, Whitley R, Weich S. Social capital and mental health. *British Journal of Psychiatry* 2002; 181:280-283.
- (103) Zha Y, Gao J, Ni S. Use of normalised difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* 2003; 24(3):583-594.
- (104) Lo CP, Faber BJ. Integration of Landsat Thematic Mapper and Census Data for Quality of Life Assessment. *Remote Sens Environ* 1997; 62:143-157.
- (105) Kitron U, Kazmierczak JJ. Spatial analysis of the distribution of Lyme disease in Wisconsin. *Am J Epidemiol* 1997; 145(6):558-566.

GLOSSARY

Attribute: An attribute is a piece of information describing the feature on the map. The attributes of a census tract, for example, would include its median income, its population and the average level of education. An attribute is a characteristic of a geographic feature described by numbers and characters, with attribute values including categories, ranks, counts, amounts and ratios.

Attribute table: A table containing the attributes (data) of a particular layer or shapefile. To explore the attributes of a layer on a map, open its attribute table. Once open, you can select features and find features with particular attributes. More than one table can be opened at a time (e.g. an attribute table for administrative boundaries can be open and viewed at the same as the attribute table for cities).

Biomass: The total mass of living matter within a given unit of environmental area.

Block-face: usually refers to one side of a street between two consecutive intersections or major intersecting geographic features such as railway tracks. In effect, it is a small strip of houses. For large apartment buildings, separate block-face designations are assigned if the apartment building constitutes a complete enumeration area (EA). In fact, a very large apartment building can be subdivided into more than one EA, and each EA will have its own block-face.

Buffer: A zone of a specified distance around a feature or features. Buffers are useful for proximity analysis (for example, finding all stream segments within 300 feet of a proposed logging area).

Census consolidated subdivision (CCS): a special aggregation of census subdivisions which provides a level of geography between the CSD and CD which facilitates data analysis. In the rural context, the CCS is a grouping of smaller municipalities, usually contained within a larger municipality.

Census division (CD): the general term applied to areas established by provincial law which are intermediate geographic areas between the municipality and the province levels. Census divisions represent counties, regional districts, regional municipalities and other types of provincially legislated areas. In Newfoundland, Manitoba, Saskatchewan and Alberta, provincial law does not provide for these administrative geographic areas. Therefore, census divisions have been created by Statistics Canada in cooperation with these provinces for the dissemination of statistical data. In the Yukon Territory, the census division is equivalent to the entire Territory.

Census subdivision (CSD): the general term applying to municipalities (as determined by provincial legislation) or their equivalent (for example, Indian reserves, Indian settlements and unorganized territories). In Newfoundland, Nova Scotia and British

Columbia, the term also describes geographic areas that have been created by Statistics Canada in cooperation with the provinces as equivalents for municipalities for the dissemination of statistical data. According to the national hierarchy, census subdivisions add together to form census divisions. The CDs form provinces and territories.

Census tract (CT): created by Statistics Canada to equal neighbourhood-like areas of 2,500 to 8,000 people (preferably close to 4,000) within all CMAs and CAs that contain an urban core with a population of 50,000 or more in the previous census. The CT boundaries generally follow permanent physical features such as major streets and railway tracks and attempt to approximate cohesive socio-economic areas. One unique feature of CTs is that their boundaries are generally held constant from one census to the next, so that CTs are comparable over time, although they may be split.

Centroid: a centre point of a defined area and not necessarily the most representative point of the postal code.

Collinear: Passing through or lying on the same straight line

Convex hull: is the smallest polygon for which each point is either on the boundary or in the interior of the polygon. The convex hull of a set of points can be visualized as the shape made by a rubber band wrapped around 'outside' points.

Coordinate System: same as **Geographic Coordinate System**.

Coverage: the area covered.

Data frame: in ArcMap, a frame on the map that displays layers occupying the same geographic area. More than one data frame may be found on the map open in ArcMap, depending on how one wants the data organized. For example, one data frame might highlight a study area and another might provide an overview of where the study area is located.

Database: A database is a collection of data that is organized so that its contents can easily be accessed, managed, and updated.

Datum: defines the position of the spheroid relative to the center of the earth and establishes the coordinate system. A datum provides a frame of reference for measuring locations on the surface of the earth. It defines the origin and orientation of latitude and longitude lines. Whenever you change the datum, or more correctly, the geographic coordinate system, the coordinate values of your data will change. The North American Datum for 1983 (NAD83) is the datum for map projections and coordinates throughout North America.

Designated places: refer to areas created by provinces to provide services and to structure fiscal arrangements for submunicipal areas which are often within unorganized areas. The concept of a designated place generally applies to small communities for

which there may be some level of legislation, but where the communities fall below the criteria established for municipal status, that is, they are “submunicipal” or unincorporated areas. The areas recognized as designated places may not represent all places having the same status within a province.

Euclidean distance: The straight line distance between two points.

Forward Sortation Area (FSA): represents a specific area within a major geographic region or province. The forward sortation area provides the basis for the primary sorting of forward mail.

Fractal: an object having a fractional dimension; one which has variation that is self-similar at all scales, in which the final level of detail is never reached and never can be reached by increasing the scale at which observations are made.

Geocoding: to geocode is to provide geographic coordinates for an address so that it can accurately be placed on a map.

Geodatabase: A relational database that stores geographic data. More precisely, the geodatabase is an object-oriented data model introduced by ESRI that is used to store spatial and attribute data and the relationships that exist among them. The geodatabase provides tools for creating “smart” geographic features and enforcing database integrity. A geodatabase can store feature classes, feature datasets, nonspatial tables, and relationship classes.

Geo-referencing: In order to ensure that all maps in a GIS database overlay accurately, the data set is 'geo-referenced' to a common coordinate system. In many countries the Universal Transverse Mercator (UTM) projection is commonly used to define coordinates in the GIS.

Geographic Coordinate System: a reference system using latitude and longitude to define the locations of points on the surface of a sphere or spheroid. A coordinate system specifies the units used to locate features in two-dimensional space, and the origin of those units.

Idrisi: raster-based GIS program.

Landcover: Land cover corresponds to a (bio)physical description of the earth's surface. It is that which overlays or currently covers the ground. This description enables various biophysical categories to be distinguished — basically, areas of vegetation (trees, bushes, fields, lawns), bare soil, hard surfaces (rocks, buildings) and wet areas and bodies of water (watercourses, wetlands).

LANDSAT: Land Remote-Sensing Satellite. The Landsat satellite has a repeat cycle of 16 days but its sensor is fixed to view vertically and so cannot be programmed to image

scenes from different orbits. The Landsat image size is 185x170 km. The satellite has two sensors, Thematic Mapper (TM) and Multispectral scanner (MSS).

Latitude: the first component of a spherical coordinate system used to record positions on the earth's surface. Latitude indicates the angular distance north or south of the earth's equator measured through 90 degrees.

Layers: an important part of using GIS to create custom geographic visual information is the idea of layering. This is the idea that layers of data, corresponding to a geographic layer of Earth, can be assembled with unique themes from different data. These layers can be applied individually or in concert to create thematic maps with multiple characteristics. An example of using layering to exclude data would be overlapping two layers (containing data on residential population and ambulance coverage) and excluding the overlapping data to reveal what houses in a rural area were not covered by ambulance service.

Line Data: Examples of line data include road networks, utility lines, stream drainages, and fault lines.

Local Delivery Unit: is identified by the last three characters of the postal code and allows for more specific sortation within a forward sortation area. In urban areas, it may indicate a specific city block (one side of a street between two intersecting streets), a single building or, in some cases, a large-volume mail receiver. In rural areas, the last three digits, together with the forward sortation area, identify a specific rural community.

Longitude: the second component of a spherical coordinate system used to record east-west positions on the earth's surface, measured in degrees as the arc or position of the earth's equator intersected between the meridian of a given place and the prime meridian, which runs through Greenwich, England.

Modifiable areal unit problem: The idea that if the boundaries of the spatial units one deals with are changed, the results of the spatial analysis will differ also.

Nearest Neighbour Analysis: This analysis determines if a set of points is dispersed (far apart) or clustered (close together), or if the points are simply random (some close together, some far apart). This is useful to our understanding of physical, biological, or social phenomena, as there is usually a reason for the point pattern, be it dispersed, clustered or random.

Normalized Difference Vegetation Index (NDVI): The NDVI is derived from multi-spectral data obtained through biophysical remote sensing via satellites orbiting the earth. This measure consists of a ratio transformation accentuating the contrast between the visible spectrum (0.4-0.7 μ m), which more strongly reflects energy from soils and litter, and the near infrared spectrum (0.7-1.1 μ m), which more strongly reflects energy from healthy green vegetation. The value varies from -1 to +1 as greenness increases. Areas with green vegetation have high positive NDVI values which appear as very light tones

on the image; areas which have been built up tend toward negative NDVI values and appear in darker tones on the image.

Point Data: Examples of point data include location of wells, post office, man holes, stream gauges, bird nesting sites or control points.

Points of Interest: establishments/services/economic activities of relevance to thesis.

Polygon Data: Examples of polygon data include land use, vegetation cover, electoral districts, soil types, and zoning.

Projections: A projection is a mathematical means of transferring information from the Earth's three-dimensional curved surface to a two-dimensional medium - paper or a computer screen. Different projections are used for different types of maps because each projection is particularly appropriate to certain uses. For example, a projection that accurately represents the shapes of the continents will distort their relative sizes.

Since much of the information in a GIS comes from existing maps, a GIS uses the processing power of the computer to transform digital information, gathered from sources with different projections to a common projection.

Raster GIS: stores map features in raster or grid format and generalizes the location of features to a regular matrix of cells. Raster GIS data structures are preferred for digital elevation modeling, statistical analysis, remotely sensed data, simulation modeling and natural resource applications. Using a raster system, continuous phenomena can be well represented (raster model – good for continuously varying features like temperature).

Remote Sensing: the sensing of the Earth's surface from space by making use of the properties of electromagnetic waves emitted, reflected or diffracted by the sensed objects, for the purpose of improving natural resources management, land use and the protection of the environment.

Shapefiles: store geographic features and their attributes. Geographic features in a shapefile can be represented by points, lines, or polygons (areas).

Spatial Autocorrelation: occurs when units of observation (e.g. census tracts) exist in spatial relationships with each other (in terms of adjacency, distance).

Spatial Joining: A spatial join is a type of spatial analysis in which the attributes of features are joined together based on the relative locations of the features. Spatial joining can provide different types of information, such as finding what's inside a polygon, what intersects a feature, and finding the nearest feature. The nearest feature is defined as the feature that is geographically closest to another one. In this case it would be points plotted geographically. Proximity is based on straight line distance between features.

Spheroid: when used to represent the earth, a spheroid is the three-dimensional shape obtained by rotating an ellipse about its minor axis.

Standard Industrial Classification (SIC): a system for classifying establishments according to their primary activity. It is used to facilitate the collection, tabulation, presentation and analysis of production and related data.

Structured Query Language (SQL): is a standard interactive and programming language for getting information from and updating a database. Queries take the form of a command language that lets you select, insert, update, find out the location of data, and so forth. There is also a programming interface. To retrieve information users execute queries to pull the requested information from the database using criteria that is defined by the user. I have also seen it referred to as 'Standard Query Language'.

Tessellation: the process of dividing an area into smaller, contiguous tiles with no gaps in between them.

Thematic Mapper: One of two sensors on the LANDSAT. Landsat TM data has 30m resolution in six spectral bands ranging from Blue to Middle Infrared and 120m resolution in one Thermal Infrared band. TM data has been acquired since 1982. Due to a lack of onboard recording capacity TM data cannot be acquired in some parts of the world.

Thiessen polygons: Voronoi polygons are also known as **Thiessen** polygons. The edges of the polygons are arranged in such a way that any location within the polygon is closer to that polygon's point sample (where there is only one) than any other point sample. The tessellation is calculated using the Delauney triangulation. As a result, all cells in the polygon are given the same value as the point sample it contains, which makes the polygons "exact predictors".

Topology: Topology is a mathematical procedure used to determine feature spatial relationships and properties. Topology makes most types of geographic analysis possible.

Vector GIS: Geographic Information Systems which store map features in vector format store points, lines and polygons with high accuracy. They are preferred in urban applications where legal boundaries and the analysis of networks are important. Using a vector system, discrete phenomena can be well represented by points, lines, and polygons (vector model – good for discrete features like locations or paths). Some drawbacks of a vector system are that it is expensive, and requires greater computer time and processing and storage capabilities than the raster system.

Vectored disease: disease transmitted by a vector, usually an insect (e.g. malaria, transmitted via the Anopheles mosquito, which is the vector of the disease; malaria is an insect-vector disease).

APPENDIX I - 1996 CENSUS VARIABLES

Age, Sex and Marital Status:

This file contains total population for 1991 and 1996, percent change in population 91-96, land area in sq. km, and population by age groups, gender, and legal marital status.

Field Heading	Description
C91totpop	Population, 1991 (100% data)
C96totpop	Population, 1996 (100% data)
C96popchng	Population percentage change, 1991 – 1996
C96lndarea	Land area in square kilometres, 1996
C96totpops	Total population by sex and age groups (100% data)
C96males	Male, total
c96females	Female, total
c96pp50_54	50-54 "Male and Female, total"
c96pp55_59	55-59 "Male and Female, total"
c96pp60_64	60-64 "Male and Female, total"
c96pp65_69	65-69 "Male and Female, total"
c96pp70_74	70-74 "Male and Female, total"
c96pp75_79	75-79 "Male and Female, total"
c96pp80_84	80-84 "Male and Female, total"
c96pp85_	85+ "Male and Female, total"
c96pp15_	Total population 15 years and over by legal marital status (100% data)
c96single	Never married (single)
c96married	Legally married (and not separated)
c96separat	Separated, but still legally married
c96divorce	Divorced
c96widowed	Widowed

Families - Number, Type and Structure:

This file contains counts of: census families in private households by family size; husband-wife families by family structure; lone parent families by parent gender.

Field Heading	Description
C96family	Total number of census families in private households by family size (20% sample data)
C96famhw	Total husband-wife families by family structure (20% sample data)
C96famhwm	Total families of now-married couples
C96nochwm	Total without sons and/or daughters at home
C96wichwm	Total with sons and/or daughters at home

C96famlp	Total lone-parent families by sex of parent (20% sample data)
C96nfla	Living alone
C96famper	Number of family persons
C96avgpfam	Average number of persons per census family
C96pp65_	Total number of persons 65 years and over (20% sample data)
C96nfwr65_	Living with relatives
C96nfwnr65	Living with non-relatives only
C96nfla65_	Living alone
C96fp65_	Number of family persons 65 years and over

Structural Type of Dwelling and Household Size:

This file contains counts of occupied private dwellings by structural type, and numbers of private households by household size.

Field Heading	Description
C96totdwl	Total number of occupied private dwellings by structural type of dwelling (20% sample data)
C96sngdet	Single-detached house
C96semidet	Semi-detached house
C96rowhse	Row house
C96aptdedu	"Apartment, detached duplex"
C96apt5st	"Apartment building, five or more storeys"
C96apt_5st	"Apartment building, less than five storeys"
C96otsngat	Other single attached house
C96movedwl	Movable dwelling

Labour Market Activities:

This file contains population counts by labour force activity.

Field Heading	Description
C96lfpop	Total population 15 years and over by labour force activity (20% sample data)
C96lfin	In the labour force
C96lfem	Employed
C96lfun	Unemployed
C96lfno	Not in the labour force
C96lfpr	Participation rate
C96lfer	Employment-population ratio
C96lfur	Unemployment rate

Household activities:

This file contains population counts by hours of unpaid care to seniors.

Field Heading	Description
C96happcs	Population 15 years and over by hours of unpaid care to seniors (20% sample data)
C96hanocs	No hours of care to seniors
C96ha_5cs	Less than 5 hours of care to seniors
C96ha5cs	5 to 9 hours of care to seniors
C96ha10cs	10 or more hours of care to seniors
C96hamppcs	Males 15 years and over by hours of unpaid care to seniors (20% sample data)
C96hamnocs	No hours of care to seniors
C96ham_5cs	Less than 5 hours of care to seniors
C96ham5cs	5 to 9 hours of care to seniors
C96ham10cs	10 or more hours of care to seniors
C96hafppcs	Females 15 years and over by hours of unpaid care to seniors (20% sample data)
C96hafnocs	No hours of care to seniors
C96haf_5cs	Less than 5 hours of care to seniors
C96haf5cs	5 to 9 hours of care to seniors
C96haf10cs	10 or more hours of care to seniors

Education:

This file contains the population counts by school attendance, and highest level of schooling.

Field Heading	Description
c96sap1524	Total population, 15 to 24 years by school attendance (20% sample data)
c96sano	Not attending school
c96edpop15	Total population 15 years and over by highest level of schooling (20% sample data)
c96edg9	Less than grade 9
c96edg913	Grades 9 to 13
c96edg913n	Without secondary school graduation certificate
c96edg913y	With secondary school graduation certificate
c96eduni	University
c96edunind	Without degree
c96eduniyd	With bachelor's degree or higher

Mobility and Migration:

This file contains population counts by mobility status for place of residence 1 year ago and 5 years ago.

Field Heading	Description
c96mb5	Total by mobility status 5 years ago (20% sample data)
c96mb5nonm	Non-movers
c96mb5move	Movers
c96mb5nmig	Non-migrants
c96mb5mig	Migrants
c96mb5mint	Internal migrants
c96mb5trap	Intraprovincial migrants
c96mb5terp	Interprovincial migrants
c96mb5ext	External migrants

Sources of Income, Earnings, Total Income and Family and Household Income (I):

This file contains employment income by composition of total income (includes government transfer payments).

Field Heading	Description
C96avgei	Average employment income (\$)
C96stdei	Standard error of average employment income (\$)
C96compinc	Total - composition of total income % (20% sample data)
C96comemp	Employment income %
C96comgovt	Government transfer payments %
C96comoth	Other %
C96inc1	\$ 1,000 - \$ 2,999
C96inc3	\$ 3,000 - \$ 4,999
C96inc5	\$ 5,000 - \$ 6,999
C96inc7	\$ 7,000 - \$ 9,999
C96inc10	\$10,000 - \$11,999
C96inc12	\$12,000 - \$14,999
C96inc15	\$15,000 - \$19,999
C96inc20	\$20,000 - \$24,999
C96inc25	\$25,000 - \$29,999
C96inc30	\$30,000 - \$34,999
C96inc35	\$35,000 - \$39,999
C96inc40	\$40,000 - \$44,999
C96inc45	\$45,000 - \$49,999
C96inc50	\$50,000 - \$59,999
C96inc60	\$60,000 and over
C96avginc	Average income \$

C96medinc	Median income \$
C96stdinc	Standard error of average income \$

Sources of Income, Earnings, Total Income and Family and Household Income (II):

This table contains family income by family structure, incidence of low income, and household income by household size.

Field Heading	Description
C96fainc	Total census families (20% sample data)
C96fain0	Under \$10,000 Integer
C96fain10	\$ 10,000 - \$19,999
C96fain20	\$ 20,000 - \$29,999
C96fain30	\$ 30,000 - \$39,999
C96fain40	\$ 40,000 - \$49,999
C96fain50	\$ 50,000 - \$59,999
C96fain60	\$ 60,000 - \$69,999
C96fain70	\$ 70,000 - \$79,999
C96fain80	\$ 80,000 - \$89,999
C96fain90	\$ 90,000 - \$99,999
C96fain100	\$100,000 and over
C96faavgin	Average family income \$
C96famedin	Median family income \$
C96fastdin	Standard error of average family income \$
C96fam	All census families (20% sample data)
C96famavin	Average family income \$
C96famsdin	Standard error of average family income \$
C96ecfam	Total - economic families (20% sample data)
C96ecfamli	Low income
C96ecfamot	Other
C96ecfamin	Incidence of low income %
C96poppr	Total - population in private households (20% sample data)
C96popprli	Low income
C96popprot	Other
C96popprin	Incidence of low income %

Occupied Private Dwellings and Housing Costs: This file contains information on dwelling characteristics, household type and household characteristics.

Field Heading	Description
c96dwlg	Total number of occupied private dwellings (20% sample data)
c96avgrms	Average number of rooms per dwelling
c96avgbed	Average number of bedrooms per dwelling

c96avgdval	Average value of dwelling \$ [This is perceived by the owner/tenant]
c96owndwlg	Owned
c96rentdwl	Rented
c96bandhou	Band housing [<i>Native American bands</i>]
c96dwlgmai	Regular maintenance only
c96dwlgmin	Minor repairs
c96dwlgmaj	Major repairs
c96dwlg46	Period of construction, before 1946
c96dwlg60	Period of construction, 1946-1960
c96dwlg70	Period of construction, 1961-1970
c96dwlg80	Period of construction, 1971-1980
c96dwlg90	Period of construction, 1981-1990
c96dwlg96	Period of construction, 1991-1996
c96privhhd	Total number of private households by household type (20% sample data)
c96onefam	One-family households
c96multfam	Multiple-family households
c96nonfam	Non-family households
c96hhdsnup	Number of persons in private households (20% sample data)
c96hhdsavg	Average number of persons in private households
c96tenant1	Tenant one-family households without additional persons
c96avgrent	Average gross rent \$
c96rent30	Gross rent spending 30% or more of household income on shelter costs
c96ownfam1	Owner one-family households without additional persons
c96avgmaj	Average owner's major payments \$
c96major30	Owner's major payments spending 30% or more of household income on shelter costs

APPENDIX II - BUFFERING POSTAL CODES AT A 500 METRE RADIUS

This constitutes the first step to obtaining counts of the establishments within a 500 m radius of the postal code. A buffer was first created around the specific postal code, and then the shapefile containing the pre-defined point of interest for all 10 provinces was spatially joined to the buffer shapefile.

When creating buffers in ArcGIS, none of the original attribute fields carry over to the new output theme table. This creates a problem because the service/establishment counts associated per postal code within a radius of 500 m could not be obtained (needed for the *Containment: Number of community establishments* and services and *Containment: Economic characteristic strategies*).

The “BufferWithAttribute” script (file name: AS11281.zip) written by Thad Tilton, needed to be downloaded from the ESRI Inc. ArcScripts website (<http://arcscripsts.esri.com/>). This script allows the user to specify an attribute from the input theme table (e.g., postal code name) to appear in the theme table of the output (postal code buffer) theme. All buffering was done in ArcView 3.2 as the script was only available for this version.

The script is loaded in ArcView 3.2 and named WizardBuffer.FINISH. This overrides the existing ArcView System script for creating buffers.

Procedure:

- download and unzip script ‘AS11281.zip’ in ArcView 3.2
- click on the “load textfile” button and find unzipped script file (file extension ‘.ave’) and put it in the script window
- rename the script window “wizardbuffer.finish”
- click on the check mark button on main toolbar
- add CSHA postal codes in View window
- arrange screens so that script can be seen at the same time as ‘View’
- click first on the data window, then on the script window and click OK
- go to “View” and then “properties” and define the projection system from a previously existing shapefile projected in the desired projection
- click on “View” and create the buffer as usual, using the Buffer Wizard

APPENDIX III - ENHANCED POINTS OF INTEREST

Description of Standard Industrial Classification (and some examples of what is included in the classification, notably aberrations)	Primary SIC
Grocery Stores	54110000
• IGA, Loblaws, 7-Eleven, Niagara Grocery, etc.	
Meat & fish markets	54210000
• Includes slaughter houses, delis, seafood products • M&M meat shops, Swiss Chalet, St-Hubert	
Fruit & vegetable markets	54310000
• Byward Fruit Market, farms, stalls, markets • Market Fresh, Herb and Spice Shop, Lapointe Fish Ltd	
Candy, nut & confectionery stores	54410000
• The Fudgery, Rocky Mountain Chocolate Factory, etc.	
Miscellaneous Food Stores	54990000
• Natural food stores, restaurants, Second Cup • coffee houses • Bulk Barn, 'The Beer Store', chiropractic care center • Pepsi-Cola Canada Beverages, catering companies • 'Alive Health Centre' • A lot of Chinese restaurants • No fast-food chains (like Burger King, McDonald's) • Not convenience stores	
Eating Places	58120000
• Franchises (A&W, Boston Pizza, KFC, McDonald's) • Cafes, delis	
Drinking Places (Alcoholic Beverages)	58130000
• Pool 'palaces', Knights of Columbus • Club Age d'Or, Catholic Social Centre • Blarney Stone Pub, Chers Place • Royal Canadian Legion Hall, etc.	
Drug Stores and Proprietary Stores	59120000
• PharmaSave, PharmaPlus	

- some Canada Posts, Uniprix, Jean Coutu, etc.

Liquor Stores 59210000

- In some of the smaller provinces, some discrepancies exist:
- in NB Labatt breweries are listed, no liquor stores
- Quebec includes SAQ and depanneurs.
- In ON, Liquor Control Board of Ontario (LCBO) AND 'The Beer Store'.

Tobacco Stores and Stands 59930000

- no stores in PEI, therefore PEI not included

Physical Fitness Facilities 79910000

- Kung Fu, Boxing Clubs, Club Fit, Tai Chi
- Hockey Impact Training, karate schools, golf courses, etc.

Membership Sports And Recreation Clubs 79970000

- Community leagues, golf clubs, lacrosse association
- Centre for Creative Arts, pools, soccer clubs
- Canadian Wilderness Survival school
- gymnastics club, squash, etc.

Amusement and Recreation Services >79990000

- Includes swimming pools, billiards, bingo halls
- kickboxing, tour operators (heli-skiing, etc.)
- karate, curling clubs, boxing clubs, martial arts

Offices and Clinics of Doctors of Medicine >80110000

- Apart from real MDs offices, includes livestock and pet supply
- therapeutic massage clinic
- forensic and medico-legal services
- specialty eye clinics, laser clinics, specialists (urologists)

Offices and Clinics of Health Practitioners 80490000

- Includes physiotherapy, chiropractors
- mental health clinic, naturopathy
- anxiety and phobia clinic, Asia herbs and acupuncture
- Jungian analysis, hospitals, community clinics,
- hearing centers, eye care
- language pathologists

Skilled Nursing Care Facilities 80510000

- No features found in Manitoba
- Includes senior's residences with nursing facilities

Nursing and Personal Care Facilities 80590000

- Includes senior residences
- General Medical and Surgical Hospitals 80620000
- Includes hospitals
 - also included clinique psychiatrique in Quebec
- Home Health Care Services* 80820000
- Includes hospices, therapeutic massage studios
 - nursing stations, Victorian Order of Nurses, Meals on Wheels
 - *NOTE: some services in **hhcserv** are contained in **hallserv**
- Specialty Outpatient Facilities 80930000
- In ON, had feng shui consultations
 - Alcohol abuse recovery centers
 - methadone clinic, pregnancy outreach programs
 - Alzheimers and related disorders clinic
- Health and Allied Services 80990000
- Health unit, midwifery services
 - physiotherapy, massage therapy, ultrasound
 - naturopathy, acupuncture
- Libraries 82310000
- Banff Public Library, Town of Cochrane
 - Ottawa Public Library, Rotary Regional Library, etc.
- Social Services 83990000
- United Way, food banks, rehab centers
 - family and community services
 - Ducks Unlimited Canada, society for disabled
 - ballet school, Gideon bibles, AIDS friends
 - drop-in centers, MS society, Salvation Army
 - some public libraries (Sudbury)
 - unemployment and help centers
- Museums and Art Galleries 84120000
- Western development museum, chapel gallery
 - Rocky Lane School museum
 - Echo Community Centre and Pool, Terrace Art Gallery
- Civic, Social, and Fraternal Associations 86410000
- Masonic order, Royal Canadian Legion
 - Salvation Army, Al Shamal Shriners, Indian band
 - Elks, Sexual Violence Services, gift shop
 - Be Bear Aware, Ducks Unlimited.

Religious Organizations 86610000

- Jehovahs witnesses, Lutheran church
- Mennonite church, Jasper United Church
- Catholic church, Cham Shan temple
- Franciscan friars, etc.

Membership Organizations* 86990000

- Amnesty Int., Chamber of Commerce
- The Friendship Society, BMX clubhouse
- Scouts Canada Comm. Hall, Cancer Society
- senior center, agricultural institute, Alcoholics Anonymous
- Mental Health Association, etc.
- *seems to be a good indicator of membership

Police Protection 92210000

- Police services, drug section, firearms office
- crime stoppers, Surete Quebec
- National Parole Board
- Public works and GSC, Interpol.

APPENDIX IV - PROXIMITY: NEAREST ESSENTIAL SERVICES: SIC QUERIES

Department store:	"SIC1" = '53110000' OR "SIC1" = '53990000'
Bank:	"SIC1" = '60210000'
Recreational spaces:	"SIC1" = '79979900' OR "SIC1" = '83999000' AND "SIC2" = '83220117'
Drug store:	"SIC1" = '59120000'
Grocery store:	"SIC1" = '54110000'
Home health care services:	"SIC1" >= '80820000' AND "SIC1" <= '80830000'
Hospital:	"SIC1" = '80620000'

**APPENDIX V – CONTAINMENT: NUMBER OF COMMUNITY ESTABLISHMENTS AND
SERVICES: SIC CODES**

Codes for the community features of interest within 500 m radius

01 Community involvement and participation

Civic/social/fraternal assoc.	"SIC1" >= '86410000' AND "SIC1" < '86510000'
Membership organizations	"SIC1" >= '86990000' AND "SIC1" < '87000000'

02 Cultural Spaces

Art Galleries and Museums	"SIC1" >= '84120000' AND "SIC1" < '84130000'
Libraries	"SIC1" >= '82310000' AND "SIC1" < '82320000'
Tourism layer	"POIType" = 'TOU'

<u>03 Drug stores</u>	"SIC1" >= '59120000' AND "SIC1" < '59130000'
------------------------------	--

04 Food stores

Fruit/veg market	"SIC1" = '54310000' AND "SIC1" < '54410000'
Meat/fish market	"SIC1" >= '54210000' AND "SIC1" < '54220000'
Grocery store	"SIC1" = '54110000' AND "SIC1" < '54120000'
Misc. food market	"SIC1" >= '54990000' AND "SIC1" < '55000000'

05 Health care offices and clinics

Health and allied services	"SIC1" >= '80990000' AND "SIC1" < '81000000'
MD Clinics	"SIC1" >= '80110000' AND "SIC1" < '80210000'
Skilled Nursing Care Fac.	"SIC1" >= '80510000' AND "SIC1" < '80520000'
Nursing/personal care Fac.	"SIC1" >= '80590000' AND "SIC1" < '80600000'
Offices of HC practitioners	"SIC1" >= '80490000' AND "SIC1" < '80510000'
Healthcare layer	"POIType" = 'HCR'

<u>06 Home Health Care Services</u>	"SIC1" >= '80820000' AND "SIC1" < '80830000'
--	--

<u>07 Hospitals</u>	"SIC1" >= '80620000' AND "SIC1" < '80630000'
----------------------------	--

08 Liquor/tobacco stores

Tobacco stores	"SIC1" >= '59930000' AND "SIC1" < '59940000'
Liquor stores	"SIC1" >= '59210000' AND "SIC1" < '59310000'

09 Physical Activity

Amusement and Recreation

`"SIC1" >= '79990000' AND "SIC1" < '80000000'`

Fitness facilities

`"SIC1" >= '79910000' AND "SIC1" < '79970000'`

Golf

`"POIType" = 'GLF'`

Sports and Recreation

`"SIC1" >= '79970000' AND "SIC1" < '79990000'`**10 Police Protection**`"SIC1" >= '92210000' AND "SIC1" < '92220000'`**11 Social services**`"SIC1" >= '83990000' AND "SIC1" < '84000000'`**12 Socializing spaces**

Eating places

`"SIC1" >= '58120000' AND "SIC1" < '58130000'`

Drinking places

`"SIC1" >= '58130000' AND "SIC1" < '58140000'`**13 Religious organizations**`"SIC1" >= '86610000' AND "SIC1" < '86620000'`

APPENDIX VI – CONTAINMENT: ECONOMIC CHARACTERISTICS - INDUSTRIAL
FEATURES DETAILED

AGRICULTURE, FORESTRY AND FISHING	
01 Agricultural Production Crops 0111 to 0191	
02 Agricultural Production Livestock And Animal Specialties 0211 to 0291	
07 Agricultural Services 0711 to 0783	
08 Forestry 0811 to 0851	
09 Fishing, Hunting, And Trapping 0911 to 0971	
MINING	Did Not Keep
10 Metal Mining 1011 to 1099	
12 Coal Mining 1221 to 1241	
13 Oil And Gas Extraction 1311 to 1389	
14 Mining And Quarrying Of Nonmetallic Minerals, Except Fuels 1411 to 1499	
CONSTRUCTION	Did Not Keep
15 Building Construction General Contractors And Operative Builders 1521 to 1542	
16 Heavy Construction Other Than Building Construction Contractors 1611 to 1629	
17 Construction Special Trade Contractors 1711 to 1799	
MANUFACTURING	
20 Food And Kindred Products 2011 to 2099	
21 Tobacco Products 2111 to 2141	
22 Textile Mill Products 2211 to 2299	
23 Apparel And Other Finished Products Made From Fabrics And Similar Materials 2311 to 2399	
24 Lumber And Wood Products, Except Furniture 2411 to 2499	
25 Furniture And Fixtures 2511 to 2599	
26 Paper And Allied Products 2611 to 2679	
27 Printing, Publishing, And Allied Industries 2711 to 2796	
28 Chemicals And Allied Products 2812 to 2899	
29 Petroleum Refining And Related Industries 2911 to 2999	
30 Rubber And Miscellaneous Plastics Products 3011 to 3089	
31 Leather And Leather Products 3111 to 3199	
32 Stone, Clay, Glass, And Concrete Products 3211 to 3299	
33 Primary Metal Industries 3312 to 3399	
34 Fabricated Metal Products, Except Machinery And Transportation Equipment 3411 to 3499	
35 Industrial And Commercial Machinery And Computer Equipment 3511 to 3599	
36 Electronic And Other Electrical Equipment And Components, Except Computer Equipment 3612 to 3699	
37 Transportation Equipment 3711 to 3799	
38 Measuring, Analyzing, And Controlling Instruments; Photographic, Medical And Optical Goods: Watches And	

Clocks 3812 to 3873	
39 Miscellaneous Manufacturing Industries 3911 to 3999	
TRANSPORTATION, COMMUNICATIONS, ELECTRIC, GAS AND SANITARY SERVICES	
40: Railroad Transportation 4011 to 4013	Did Not Keep
41 Local And Suburban Transit And Interurban Highway Passenger Transportation 4111 to 4173	
42 Motor Freight Transportation And Warehousing 4212 to 4231	
43 United States Postal Service 4311 to 4311	
44 Water Transportation 4412 to 4499	
45 Transportation By Air 4512 to 4581	
46 Pipelines, Except Natural Gas 4612 to 4619	
47 Transportation Services 4724 to 4789	
48 Communications 4812 to 4899	
49 Electric, Gas, And Sanitary Services 4911 to 4971	
WHOLESALE TRADE	
50 Wholesale Trade-durable Goods 5012 to 5099	
51 Wholesale Trade-non-durable Goods 5111 to 5199	
RETAIL TRADE	
52 Building Materials, Hardware, Garden Supply, And Mobile Home Dealers 5211 to 5271	Did Not Keep
53 General Merchandise Stores 5311 to 5399	
54 Food Stores 5411 to 5499	
55 Automotive Dealers And Gasoline Service Stations 5511 to 5599	Did Not Keep
56 Apparel And Accessory Stores 5611 to 5699	
57 Home Furniture, Furnishings, And Equipment Stores 5712 to 5736	
58 Eating And Drinking Places 5812 to 5813	
59 Miscellaneous Retail 5912 to 5999	
FINANCE, INSURANCE AND REAL ESTATE	
60 Depository Institutions 6011 to 6099	
61 Non-depository Credit Institutions 6111 to 6163	
62 Security And Commodity Brokers, Dealers, Exchanges, And Services 6211 to 6289	
63 Insurance Carriers 6311 to 6399	
64 Insurance Agents, Brokers, And Service 6411 to 6411	
65 Real Estate 6512 to 6553	
67 Holding And Other Investment Offices 6712 to 6799	
SERVICES	
70 Hotels, Rooming Houses, Camps, And Other Lodging Places 7011 to 7041	Did Not Keep
72 Personal Services 7211 to 7299	
73 Business Services 7311 to 7389	
75 Automotive Repair, Services, And Parking 7513 to 7549	Did Not Keep

76 Miscellaneous Repair Services 7622 to 7699	Did Not Keep
78 Motion Pictures 7812 to 7841	
79 Amusement And Recreation Services 7911 to 7999	
80 Health Services 8011 to 8099	
81 Legal Services 8111 to 8111	
82 Educational Services 8211 to 8299	
83 Social Services 8322 to 8399	
84 Museums, Art Galleries, And Botanical And Zoological Gardens 8412 to 8422	
86 Membership Organizations 8611 to 8699	
87 Engineering, Accounting, Research, Management, And Related Services 8711 to 8748	Did Not Keep
88 Private Households 8811 to 8811	Did Not Keep
PUBLIC ADMINISTRATION	
91 Executive, Legislative, And General Government, Except Finance 9111 to 9199	
92 Justice, Public Order, And Safety 9111 to 9229	
93 Public Finance, Taxation, And Monetary Policy 9311 to 9311	
94 Administration Of Human Resource Programs 9411 to 9451	
95 Administration Of Environmental Quality And Housing Programs 9511 to 9532	
96 Administration Of Economic Programs 9611 to 9661	
97 National Security And International Affairs 9711 to 9721	
99 No classifiable Establishments 9999 to 9999	Did Not Keep
00 Unclassified 0000 to 0000	Did Not Keep

APPENDIX VII - CODES FOR THE INDUSTRIAL FEATURES

Agriculture, Forestry and Fishing ('ffagri')

"SIC1" >= '01110000' AND "SIC1" < '02910000' OR "SIC1" >= '07110000' AND "SIC1" < '09710000'

Manufacturing ('manuf')

"SIC1" >= '20110000' AND "SIC1" <= '39990000'

Wholesale Trade ('wholetrade')

"SIC1" >= '50120000' AND "SIC1" <= '51990000'

Retail Trade ('retailtr')

"SIC1" >= '53110000' AND "SIC1" <= '54990000' OR
"SIC1" >= '56110000' AND "SIC1" <= '59990000'

Finance, Insurance and Real Estate ('finance')

"SIC1" >= '60110000' AND "SIC1" <= '65530000' OR
"SIC1" >= '67120000' AND "SIC1" <= '67990000'

Services ('services')

"SIC1" >= '72000000' AND "SIC1" <= '73890000' OR
"SIC1" >= '78120000' AND "SIC1" <= '84220000' OR
"SIC1" >= '86110000' AND "SIC1" <= '86990000'

Public Administration ('pubadmin')

"SIC1" >= '91110000' AND "SIC1" <= '97210000'

APPENDIX VIII - SPATIAL JOINING

Once points of interest in a particular province (e.g. social services in Saskatchewan, which includes food banks, community service centres, etc.) have been queried from the file containing all points of interest in that province, they are highlighted in the original attribute table.

The selected points of interest must be exported to a pre-specified folder as a dBASE (.dbf) file. This file, which will contain all social services in Saskatchewan, was then added to ArcMap as a layer. Upon right-clicking the layer in the data frame and selecting 'Display XY coordinates', the data were displayed on the map. This plots the geographic locations of the social services on the map, and was repeated for the remaining nine provinces.

When the point of interest (e.g. social services) has been queried for all ten provinces, another feature of ArcGIS, the GeoProcessing Wizard, was used to combine all 10 layers into a single national layer. It does this by merging all the layers into one; attributes (column headings) are retained if they have the same name. The output from this geoprocessing operation will be in the same coordinate system as the data frame in ArcMap, irrespective of the coordinate system of the input layers. For the most accurate results, the layers used in the geoprocessing operation were in the same coordinate system.

Once finished processing, the newly created merged layer contains data on all the social services in the ten provinces.

This process of querying attributes in each province and merging the results is done to all pre-specified points/establishments that had been selected as potentially most relevant to affecting the health status of an elderly individual.

Before doing the spatial joining, a pre-constructed map of Canada was added to the ArcMap data frame; this was important for visualization of the data. To do the spatial joining of a point of interest (e.g. social services):

- 1) The shapefile containing social services in every one of the ten provinces is added to the data frame, after having been merged using the GeoProcessing Wizard and appropriately projected using the ArcToolbox Projection Wizard.
- 2) The shapefile containing the postal code locations of all CSHA participants is also added. Selecting the 'CSHA postal code' layer in the drop down list, "Join data from another layer based on spatial location" was selected. The GIS, using spatial join operations, will then find and calculate the distance between the closest social service and the location of the CSHA participant's postal code. During this basic GIS operation each CSHA point is given all the attributes of the closest social service, for example, and a distance field showing how close that service is.
- 3) A file name must be specified by which the join will be saved. The spatially joined shapefile is automatically added to the data frame. Open the attribute tables to make sure that the data was joined, as well as to look at the data. This process is repeated, with the new layers continually being added to the just joined and exported layer. This way, the original postal codes from the CSHA participants, as well as the type of and distance to the point of interest are collected as a single data file.

APPENDIX IX - DOCUMENTATION OF SAS CODE WRITTEN FOR DATA LINKAGE

 Author: Adrijana Corluka

Purpose: Documentation of steps taken and SAS code written to create the database.

- the dataset nearest.csha_pc contains all the relevant postal codes of the CSHA participants that were could be matched to the postal codes contained in the GIS data, and mapped.

- the nearest.features dataset contains the distance to particular neighbourhood features nearest to the CSHA participants postal code location (these include distance to grocery store, bank, etc., and are calculated in metres).

-the dataset nearest.indfeatures500 contains counts of particular industrial features (e.g. retail trade, services, etc.) within 500m of the individual's postal code.

- the dataset nearest.poifeat500 also gives counts of even more specific neighbourhood features such as restaurants, cultural spaces, hospitals, etc., within a 500 m radius of the CSHA participant's postal code location.

They are merged into one large dataset, which I named 'total'.

*****;

/*This is the successful aging outcome data provided by Dr. Betsy Kristjansson in August 2003*/

```
data succage;
set success.adria;
run;
```

```
proc sort data=succage; by id; run;
```

```
data csha_pc;
set nearest.csha_pc;
run;
```

```
data cshapc;
set csha_pc;
cma2=cma*10000;
/*converting the character variable into a numeric variable*/
cmactract=cma2 + ct;
/*creating CMACTRACT because CMA and CT are separate variables, and I
need them as one for merging to the census data.
run;
```

```
proc sort data=cshapc;by id; run;
```

```
data comm;
set success.communitynonmissingpc (rename=(unique=id));
run;
```

```

proc sort data=comm; by id; run;

data real (drop=dpl);
merge comm (in=in1) cshapc (in=in2) /*succage (in=in3)*/;
by id;
if in1 and in2 /*and in3*/;
run;
/*(8995 observations)*/

proc sort data=real; by pcode; run;

*****
Trying out matching the 'real'(8995 participants) with the UEP file to
see if there are any missing postal codes:
*****;
data uep;
set success.uep (rename=(postalcode=pcode));
/*had to re-name for merging*/
run;

proc sort data=uep; by pcode; run;

data uep_pc;
merge real(in=in1) uep (in=in2);
by pcode;
if in1 and /*not*/ in2;
run;

proc sort data=uep_pc; by id; run;

proc sort data=real; by id; run;

*****
Creating the data file of successful aging participants
matched to the geocoded information
*****;
data succagepc;
merge real (in=in1) succage (in=in2);
by id;
if in1 and in2;
run;
/*5071 observations*/

proc sort data=succagepc; by cmactract;run;

/* merging the list of healthy/succ aging participants (succage) with
the participants mapped to census tract - 2 missing between community-
dwelling CSHA participants (real) and the 5073 CSHA participants which
were given to me by Dr. Betsy Kristjansson.
Result: 5071 participants are successfully aging*/

*****
Matching the CSHA subset of successful aging individuals
to their census tract data, new dataset called ind_census,
for individual and census data:
*****;

```

```

data censusdata;
set nearest.censusdata ;
cmactract=cmact*1; /*have to make it numeric to match the individual
data*/
run;

/* In the geocoded individual-level data, census tracts were defined as
defined as numeric but in the nearest.censusdata dataset, it is defined
as character. I converted the character variable of the census tract
in the census data dataset into a numeric one:
Note that it is not possible to directly change the type of a variable.
It is only possible to write the variable to a new variable containing
the same data, although with a different type. By renaming and dropping
variables, it is possible to produce a new variable with the same name
as the original, although with a different type.*/

proc sort data=censusdata; by cmactract; run;

proc sort data=succagepc; by cmactract;run;

data ind_census;
merge censusdata (in=in2) succagepc (in=in1);
by cmactract;
if in1 and in2;
run;
/*4245 observations*/

proc sort data=abc; by id; run;

*****
Checking for rural participants and other missings -
merged with the census data to get a better idea of
why 826 were missing (code below):
*****;
/*data aaa;
set succagepc;
if ct =0.00;
run;

proc freq data=aaa;
table succ2;
run;*/

/*This accounts for 795 of the 826 missing participants; the 31
remaining participants' census tracts seem to have had their data
suppressed (i.e. CT='999.99'). The CTs were assigned in the geocoding
step with the PCCF+. Of the 795 rural individuals (with 2nd character
of postal code='0'), 156 successfully aging, 639 not successfully
aging)*/

data bbb;
merge censusdata (in=in2) succagepc (in=in1);
by cmactract;
if in1;
run;
/*5071 observations*/
proc sort data=bbb; by id; run;

```

```

data ccc;
set bbb;
if ct=0.00;
run;
/*795 observations*/

proc sort data=ccc;by id; run;

data ddd;
merge ccc (in=in1) bbb(in=in2);
by id;
if in2 and not in1;
run;
/*4276 observations - all observations except for CT='0.00'*/

data fff;
merge ddd/*4276*/ (in=in1) ind_census/*4245*/(in=in2);
by id;
if in1 and not in2;
run; /* 31 missing obs!!!*/

*****
Trying out matching the 'real' with the UEP file to see if there are
any missings:
*****;
/*data uep;
set success.uep (rename=(postalcode=pcode)); *had to be re-named for
merging*;
run;

proc sort data=uep; by pcode; run;

proc sort data=succagepc; by pcode; run;

data uep_pc;
merge succagepc(in=in1) uep (in=in2);
by pcode;
if in1 and /*not*; in2;
run;*/

*****
5014 observations - 57 participants could not (activate the /*not* by
removing / and *)
be matched to postal codes contained in the GIS data. It appears as
though
57 postal codes had been retired since 1996 and were not in the DMTI
UEP,
and thus were also not in the DMTI GIS data, therefore 57 participants
could
not be linked to GIS data.
Result: 5014 successful aging participants could be linked to the GIS
data
*****;

```

```

data feat (drop= cma ct_name csd_type prcdcsd prfedea city_1 city_2
city_3 city_4 city_5
city_6 fed87 fed96 X Y fid_1 fid_2 fid_3 fid_4 fid_5
fid_6 fid_7 fid_8 fid_1_1 fid_1_1_1 fid_1_1_1_1 fid_1_1_2 fid_1_1_3
fid_1_1_4
poiid poiid_1 poiid_2 poiid_3 poiid_4 poiid_5 poiid_6 latitude
longitude
prov_1 prov_2 prov_3 prov_4 prov_5 prov_6 prov_7 sic1_1 sic1_2 sic1_3
sic1_4
sic1_5 sic1_6 sic2_1 sic2_2 sic2_3 sic2_4 sic2_5
sic2_6 x_1 x_2 x_3 x_4 x_5 x_6 y_1 y_2 y_3 y_4 y_5 y_6 );
set nearest.features (rename=(postalcode=pcode));
/* renaming the variable postalcode to 'pcode' so as to maintain
uniformity)*/
run;

proc sort data=feat;by pcode;run;

*****
Merging the nearest features to the participants in the
succesful/healthy aging study
*****;
data featsuccage;
merge feat (in=in2) succagepc (in=in1);
by pcode;
if in1 and in2;
run;

proc sort data=featsuccage; by pcode; run;

*****
Merging the counts of the different industries within a 500 m radius of
CSHA postal codes to the healthy aging/census tract/distance to nearest
feature data
*****;
data indcount;
set nearest.indfeatures500 (rename=(postalcode=pcode));
/* renaming the variable postalcode to 'pcode' so as to maintain
uniformity)*/
run;

proc sort data=indcount; by pcode; run;

data indfeatsuccage;
merge indcount (in=in2) featsuccage (in=in1);
by pcode;
if in1;
run;

proc sort data=indfeatsuccage; by pcode; run;

/* There are 5014 observations and 63 variables included in
this dataset */

*****
Merging the counts of the different specific points of interest
within a 500 m radius of CSHA postal codes to the

```

```

successful aging/census tract/distance to nearest feature data:
*****;
data poicount (drop=id);
set nearest.poifeat500(rename=(postalcode=pcode));
/* renaming the variable postalcode to 'pcode' so as to maintain
uniformity)*/
run;

proc sort data=poicount; by pcode; run;

proc sort data=indfeatsuccage; by pcode; run;

data indfeatsuccagepoi;
merge indfeatsuccage(in=in1) poicount (in=in2) ;
by pcode;
if in1 and in2;
run;

proc sort data=indfeatsuccagepoi; by pcode; run;

*****
Renaming the dataset 'indfeatsuccagepoi' to indicate that the
GIS data for all three methods (distance to 'nearest,' counts of
points of interest within a neighbourhood context (500 m), counts
of the industrial features within 500 m of individual's postal code)
are linked in this one dataset
*****;
data allgis;
set indfeatsuccagepoi;
run;

/* There are 5014 observations and 76 variables (incl. the successful
aging variables) included in the dataset 'allgis'*/

*****
Categorizing the variables
*****;
/*
data newdist;
set allgis;
if 0<=bankdist<150 then newbank=1;
else if 150<=bankdist<350 then newbank=2;
else if 350<=bankdist<500 then newbank=3;
else if 500<=bankdist<1000 then newbank=4;
else if 1000<=bankdist<2000 then newbank=5;
else if 2000<=bankdist<5000 then newbank=6;
else if 5000<=bankdist<7500 then newbank=7;
else if 7500<=bankdist<10000 then newbank=8;
else if 10000<=bankdist<15000 then newbank=9;
else if 15000<=bankdist<20000 then newbank=10;
else if 20000<=bankdist<50000 then newbank=11;
else if bankdist>=50000 then newbank=11;

if 0<=hospdist<150 then newhosp=1;
else if 150<=hospdist<350 then newhosp=2;
else if 350<=hospdist<500 then newhosp=3;
else if 500<=hospdist<1000 then newhosp=4;

```

```

else if 1000<=hospdist<2000 then newhosp=5;
else if 2000<=hospdist<5000 then newhosp=6;
else if 5000<=hospdist<7500 then newhosp=7;
else if 7500<=hospdist<10000 then newhosp=8;
else if 10000<=hospdist<15000 then newhosp=9;
else if 15000<=hospdist<20000 then newhosp=10;
else if 20000<=hospdist<50000 then newhosp=11;
else if hospdist>=50000 then newhosp=11;

if 0<=homecrdist<150 then newhmcr=1;
else if 150<=homecrdist<350 then newhmcr=2;
else if 350<=homecrdist<500 then newhmcr=3;
else if 500<=homecrdist<1000 then newhmcr=4;
else if 1000<=homecrdist<2000 then newhmcr=5;
else if 2000<=homecrdist<5000 then newhmcr=6;
else if 5000<=homecrdist<7500 then newhmcr=7;
else if 7500<=homecrdist<10000 then newhmcr=8;
else if 10000<=homecrdist<15000 then newhmcr=9;
else if 15000<=homecrdist<20000 then newhmcr=10;
else if 20000<=homecrdist<50000 then newhmcr=11;
else if homecrdist>=50000 then newhmcr=11;

if 0<=grcerydist<150 then newgroc=1;
else if 150<=grcerydist<350 then newgroc=2;
else if 350<=grcerydist<500 then newgroc=3;
else if 500<=grcerydist<1000 then newgroc=4;
else if 1000<=grcerydist<2000 then newgroc=5;
else if 2000<=grcerydist<5000 then newgroc=6;
else if 5000<=grcerydist<7500 then newgroc=7;
else if 7500<=grcerydist<10000 then newgroc=8;
else if 10000<=grcerydist<15000 then newgroc=9;
else if 15000<=grcerydist<20000 then newgroc=10;
else if 20000<=grcerydist<50000 then newgroc=11;
else if grcerydist>=50000 then newgroc=11;

if 0<=storedist<150 then newstore=1;
else if 150<=storedist<350 then newstore=2;
else if 350<=storedist<500 then newstore=3;
else if 500<=storedist<1000 then newstore=4;
else if 1000<=storedist<2000 then newstore=5;
else if 2000<=storedist<5000 then newstore=6;
else if 5000<=storedist<7500 then newstore=7;
else if 7500<=storedist<10000 then newstore=8;
else if 10000<=storedist<15000 then newstore=9;
else if 15000<=storedist<20000 then newstore=10;
else if 20000<=storedist<50000 then newstore=11;
else if storedist>=50000 then newstore=11;

if 0<=drugstdist<150 then newdrugst=1;
else if 150<=drugstdist<350 then newdrugst=2;
else if 350<=drugstdist<500 then newdrugst=3;
else if 500<=drugstdist<1000 then newdrugst=4;
else if 1000<=drugstdist<2000 then newdrugst=5;
else if 2000<=drugstdist<5000 then newdrugst=6;
else if 5000<=drugstdist<7500 then newdrugst=7;
else if 7500<=drugstdist<10000 then newdrugst=8;
else if 10000<=drugstdist<15000 then newdrugst=9;

```

```

else if 15000<=drugstdist<20000 then newdrugst=10;
else if 20000<=drugstdist<50000 then newdrugst=11;
else if drugstdist>=50000 then newdrugst=11;

if 0<=recdist<150 then newrec=1;
else if 150<=recdist<350 then newrec=2;
else if 350<=recdist<500 then newrec=3;
else if 500<=recdist<1000 then newrec=4;
else if 1000<=recdist<2000 then newrec=5;
else if 2000<=recdist<5000 then newrec=6;
else if 5000<=recdist<7500 then newrec=7;
else if 7500<=recdist<10000 then newrec=8;
else if 10000<=recdist<15000 then newrec=9;
else if 15000<=recdist<20000 then newrec=10;
else if 20000<=recdist<50000 then newrec=11;
else if recdist>=50000 then newrec=11;
run;

proc sort data=newdist; by id; run;
*/
*****
Census data
Matching the individual-level data and the GIS data to the census data,
which
had already been read in at the beginning
*****;

proc sort data=ind_census;by id; run;

proc sort data=allgis; by id; run;

data total;
merge allgis (in=in1) ind_census(in=in2);
by id;
if in1 and in2;
run;

*****
To create a permanent dataset of the merged files:
*****;
/*libname nearest 'C:\USERS\Adrijana\EPOI_nearest';

data nearest.Oct8succgiscensus;
set total;
run;*/

```

APPENDIX X - POTENTIAL VARIABLES CREATED AND USED FROM GENERAL SOCIAL SURVEYS OF DIFFERENT CYCLES

Combining variables to increase sample size:

Personal Risk and Victimization

Personal Risk:

GSS 1988 variable A3 + GSS 1993 variable A3 + GSS 99 A1 (total sample = 49 000)

Interpretation: 'GSS 1998 variable A3' refers to variable A3 from the GSS conducted in 1998. The results from this question would be added to the results of other surveys asking the same question in order to increase sample size. In this case, the sample size from the three surveys ('88, '93, '99) would total 49 000. Question A3 was:

A3: Neighbourhood crime level

General Social Survey of Canada, 1988 - Cycle 3: personal risk

LEVEL OF CRIME IN NEIGHBOURHOOD, COMPARED TO OTHER AREAS IN CANADA.

NEIGHBOURHOOD REFERS TO THE GENERAL AREA SURROUNDING THE RESPONDENT'S HOME.

Value	Label	Unweighted Frequency	Missing Value?
1	HIGHER	1088	
2	ABOUT THE SAME	3284	
3	LOWER	6736	
4	DON'T KNOW	582	
9	NOT STATED	8	Missing

Victimization:

GSS 88 A6B + GSS 93 A6 + GSS 99 A3 (total sample = 49 000)

Perception of neighbourhood safety:

GSS 88 A4 + GSS 93 A4 + GSS 99 A2 (total sample = 49 000)

Social activities and time use

Occurrences of political activity:

GSS 1986 EPISO_61 + GSS 1992 EPI610 + GSS 1998 EPI610 (total sample = 30 000)

Occurrences of child/youth/family organization participation:

GSS 86 EPISO_62 + GSS 92 EPI620 + GSS 98 EPI620 (total sample = 30 000)

Occurrences participates in religious meetings/organizations:

GSS 86 EPISO_63 + GSS 92 EPI630 + GSS 98 EPI630 (total sample = 30 000)

Occurrences participates in religious services/prayer, etc.:

GSS 86 EPISO_64 + GSS 92 EPI640 + GSS 98 EPI640 (total sample = 30 000)

Occurrences participates in fraternal/social organization:

GSS 86 EPISO_65 + GSS 92 EPI651 + GSS 98 EPI651 (total sample = 30 000)

Occurrences participates in volunteer work/helping:

GSS 86 EPISO_66 + GSS 92 EPI660 + GSS 98 EPI660 (total sample = 30 000)

Frequencies of religious/church attendance:

GSS 86 ATTEND + GSS 92 K20 + GSS 96 J22 + GSS 98 RELIGATT + GSS 2000 RELIGATT (total sample = 80 000)

Reciprocity:

GSS 1996 F2D + GSS 1996 F2F + GSS 1996 F2H (total sample = 80 000)

Non-combined (individual variables in differing GSS cycles):

General Social Survey of Canada, 1989 - Cycle 4: education and work

[Number of cases: 9338; Number of variables: 285]

F1A Are you interested in current affairs

F1B Are you interested in economic condition

F2A Are you informed about current affairs

F2B Are you informed about economic condition

M1A Involved in charitable, service/volunteer

M1B Involved in neighbourhood, community/school

M1C Involved in religious, church-related group

M1D Involved in social, cultural/ethnic group

M1E Involved in sports/athletic association

M1F Involved in public interest group

M1G Involved in business, professional/work-related group

M1H Involved in political organization

General Social Survey of Canada, 1992 - Cycle 7: Time use

[Number of cases: 9815; Number of variables: 341]

G1 Last 12 months, participate regularly in sports?

G9 Do you belong to a sport club/sport organization?

General Social Survey of Canada, 1998 - Cycle 12: Time use file

[Number of cases: 10,749; Number of variables: 953]

H12A Internet: communicate

L21H Sense of belonging local community

General Social Survey of Canada, 2000 - Cycle 14: Access to and use of Information
Communication Technology

[Number of cases: 25,090; Number of variables: 456]

M26 Perception of trustworthiness of other people

M27 Did you vote in the last election

M28 Past 12 months, talked other people politics

M29 Past 12 months, search for information about political issue

M30 Past 12 months, have volunteered for a political party

M31 Past 12 months, Written letter/call show to express view