2004

Using GIS and spatial statistics to explore and model demand for emergency medical services in the city of Sudbury, Ontario

Marc Lefebvre
Wilfrid Laurier University

0-612-92274-X

Using GIS and Spatial Statistics to Explore and Model Demand for Emergency Medical Services in the City of Sudbury, Ontario

By
Marc Lefebvre
Honour's Bachelor of Arts (Geography), Laurentian University, 1998

THESIS
Submitted to the Department of Geography and Environmental Studies
in partial fulfilment of the requirements
for the Master of Arts of Geography
Wilfrid Laurier University
2004

Abstract

The purpose of this research is to examine the nature of the relationship between EMS ambulance call volume and demographic, socioeconomic and geographic (urban structural) forces in the City of Sudbury, a medium sized city of approximately 100,000 persons in Ontario, Canada. As in past research in the area of EMS demand, linear regression is used to model this relationship. However, unlike previous work, spatial autocorrelation inherent in real world data is addressed to mitigate violation of the assumption of independence required for classic regression.

Using a Geographical Information System (ArcView 3.2) EMS data are geolocated onto a spatial framework for which 1996 census data are available. A spatial analysis program (SpaceStat 1.90) is used to operationalize a spatial model, and perform a battery of spatial diagnostics.

After exploring the data with an aspatial stepwise regression model and identifying spatial autocorrelation in the explanatory variables, a mixed aspatial/spatial stepwise regression model is used. The variables "Percent People Living Alone" and its spatially lagged version, a lagged version of "Percent of Apartment Dwellings" and lagged "Percent of People Aged 20 to 64" account for 52% of the variation in EMS calls per 1,000 persons.

Clearly, demand for Emergency Medical Services varies greatly from place to place within a community. And this variety is, in part at least, related to underlying demographic and socioeconomic realities. Strategic deployment of resources based on these realities could assure provision of more effective and efficient Emergency Medical Services. Also, injury prevention and health promotion programs could be targeted more precisely to groups and areas in need through the help of EMS demand analysis.

Acknowledgements

I would like to first and foremost thank my advisor Dr. Barry Boots (Wilfrid Laurier University) for his patience and guidance throughout this project. I have learned so much from him. Also, my thesis committee member, Dr. Bob Sharpe (Wilfrid Laurier University) has always been interested and enthusiastic about my work. His energy has often served to recharge my own efforts. My readers, Dr. Rob Feick (University of Waterloo), and Dr. Sean Doherty provided thoughtful insights and invaluable feedback during the defence of this thesis. Dr. Pam Bryden (Wilfrid Laurier University) acted as Chair of the defence and proxy for Dr. Doherty. I thank them all sincerely for their time and effort. Grant Simpson and Jo-Anne Horton from Wilfrid Laurier University's Geography and Environmental Studies Department also provided important technical and administrative support.

A special thanks to the Sudbury Base Hospital Program without whose data this work simply could not have been undertaken. Dr. Rob Lepage, Sylvie Salminen, Pam Labelle, and Dan Gregoris have my gratitude for offering and providing their database. This thesis focusses a great deal on the limitation of that database, however, I must emphasise these are due to resource constraints and certainly not a lack of willingness and effort on their part to collect the best possible data. Also, without the approval of the Sudbury Regional Hospital's Research Ethics Committee this work would not have been possible. Any insights gleaned from this thesis is attributable to their appreciation of the importance of research in their community.

I must also gratefully acknowledge Mr. Tim Beadman and Joe Nicholls, who, while I was recently employed as a Manager with the Greater Sudbury EMS, approved two leaves of absence in order for me to complete this work. I really needed the extra time away from the hectic business of EMS administration.

My family, particularly my mother and father have never stopped supporting me and encouraging me through my academic efforts. I am so grateful to them.

My fiancee, Steph took so much of her time to help with countless revision of my many drafts. If there is any polish to this thesis it is probably thanks to her. More importantly, her constant encouragement and support are what kept me going when I was weary.

Finally, this thesis was partially funded through Emergency Preparedness Canada who saw fit to award me the Stuart Nesbitt-White Fellowship in 1998. I am very grateful for their generous financial support and I hope that this work will further the field of Emergency Services.

Table of Contents

Part B

List of Tables

List of Figures

Chapter 1 Introduction

1.0 Introduction

Emergency Medical Services (EMS) provide a much needed link between a community and its health care facilities. Calls for help to Emergency Medical Services require intervention by Paramedics in situations ranging from motor vehicle accidents, falls, violence, heart attacks and strokes, and a number of other medical emergencies. Advancements in paramedic training now mean that life saving techniques have moved from the emergency room to the pre-hospital environment. As Kotzé (1990: 320-321) reflected, "Ambulance personnel now probably play a more important role in pre-hospital care of the sick and injured than ever before. A fact that is sometimes easily overlooked is that the [paramedic] is usually the first suitably qualified person on the scene to deal with medical emergencies." Kotzé's statement from more than a decade ago rings more true today because the role of paramedics in the health care system has grown as increasingly advanced skills are introduced into their scope of practice. Agencies responsible for planning and delivering this health service should have a good understanding of what forces generate their demand in order to strategically place limited resources within their jurisdiction.

Since 1997, legislative changes in Ontario have effectively devolved the responsibility of administration and operations of Emergency Medical Services from the provincial to the municipal mandate. The legislation directed municipalities to provide a service that would be "responsive" to current and future needs based on unique demographic, socio-economic and geographic factors in each community (Ontario Ministry of Health and Long-Term Care, 1998: 3; Ontario Ministry of Health and Long-Term Care & MOH-AMO Land Ambulance Implementation Steering Committee, 1999: 4 and 6).

Messmer (1973) emphasizes that determining the need for ambulance services is an integral part of the planning process. Naturally, strategically locating EMS ambulances requires that those needs be clearly identified and understood. Local governments, responsive to local needs, are expected to offer an effective, yet efficient service (Ball, 1980).

Research into the demand for EMS is certainly not abundant. Nevertheless, early seminal work by Aldrich et al. (1971) was followed by important ecological studies that will be reviewed in the next chapter. Many of these analyses have used classical regression methods to examine the relationship between the demand for EMS and demographic and socioeconomic factors. A few have placed the studies in a geographic context ranging from census tracts to counties; however none have given consideration to how the spatial arrangement of the data may have affected their results.

By using a geographic framework to examine EMS demand but ignoring the spatial characteristics of the variables within it, inferences or conclusions drawn from regression modelling could be invalid, or at least suspect. It is well established that assumptions needed to properly specify a regression equation are typically violated in a geographic context (Cliff & Ord, 1981: 196; Anselin & Griffith, 1988; Haining, 1998:35; Griffith & Layne, 1999:4) mostly due to Tobler's First Law of Geography that holds that "everything is related to everything else, but near things are more related than distant things." Specifically, the assumption of independence between observations is breached. Spatial dependence is the rule for most data in the real world and must be addressed. Results from regression analyses that ignore spatial effects should be treated with suspicion because work has shown that they could be invalid (Cliff & Ord, 1981:196; Anselin & Griffith, 1988; Griffith & Layne, 1999: 71).

Furthermore, conventional regression analysis presumes that relationships between variables are the same throughout the area of interest (Fotheringham, 1997: 89). Spatial researchers recognize that some relationships differ across space and that local relationships can be hidden in a global model (Anselin, 1995; Anselin, 1996: 113; Haining, 1990: 43; Fotheringham, 1997: 89). Local regimes must therefore also be identified to properly investigate the demand for Emergency Medical Services.

Integration of GIS and spatial statistical/analysis programs allow us to incorporate Exploratory Spatial Data Analysis (ESDA) into the analysis of phenomena that are inherently spatial, such as the demand for health services (Haining, 1996: 53). In looking at the use of EMS, we can measure the degree of spatial dependence, anticipate its effects, and provide more meaningful interpretation of results. Also, we can identify those regions where we may need to modify our model (Haining, 1996: 63; Fotheringham, 1997:90). Moreover, these tools allow us to determine the scale at which a phenomenon occurs. Whereas previous work has examined EMS demand at the regional level and in some cases the census tract level, today's GIS technology facilitates the analysis at multiple scales including finer scales (the enumeration area) that may be more appropriate for the process in question.

1.1 Purpose

The purpose of this research is to examine the nature of the relationship between EMS ambulance call volume and demographic, socioeconomic and geographic (urban structural) dimensions in the City of Sudbury, a medium sized city of approximately 100,000 persons in Ontario, Canada. As in past research in the area of EMS demand, linear regression is used to model this relationship. However, unlike previous work examining demand in other cities or counties, the spatial nature of the problem is addressed. Explicitly recognizing the unique geography of communities, the Sudbury model is not intended to be generalisable to all municipalities. Rather, the intention here is to convince emergency planners that spatial analysis can be, and needs to be, incorporated if they intend to perform an ecological analysis using census data to examine patterns in their own municipalities.

To reach this goal, a number of objectives need to be met. First, a Geographical Information System (ArcView 3.2) is used to spatially enable ambulance call events on a digital road network. Second, again using GIS, those events are aggregated to a pre-existing spatial framework based on census Enumeration Area (EA) boundaries. Ecological variates from the 1996 Census of Canada are examined in terms of their relationship with EMS demand using a regression equation. Next, using a loosely coupled spatial analysis program (SpaceStat 1.90) and the GIS, a spatial model is devised and a battery of diagnostics are used to mitigate any misspecification of the regression model.

1.2 Thesis Organisation

In order to meet the objectives described above, this thesis is organized into five chapters including this introduction. Chapter 2 provides a background of Emergency Medical Services and reviews past research in EMS demand. It briefly outlines the history of EMS in Ontario in order to emphasize the great impact that recent decisions by the Government of Ontario continue to have on EMS. Furthermore, it is a review of the major literature that has investigated demographic and socioeconomic determinants of EMS demand.

Chapter 3, Methods, is divided into three main parts. The chapter is prefaced with an explanation of the basic conceptual framework and research approach used. Some pitfalls of geographic analysis, ecological fallacy, the modifiable areal unit problem, and spatial autocorrelation are then reviewed. Part A describes the study area, part B introduces the datasets as well as the methods used to prepare them for analysis, and part C describes the analytical methods used.

The study area description in Part A provides some context and outlines the distinctive geography of the City of Sudbury that makes it a convenient study site. In part B, both EMS and ecological variates datasets are described. How these datasets were collected and acquired is then explained. Next, the processing steps needed to spatially enable the data are described, including how the ambulance call data are geolocated on a digital road network using a GIS. Also, the attention that both the EMS events database and the digital road network needed to ensure good address matching results is outlined. How the georeferenced data are aggregated to enumeration areas in preparation for exploratory analysis, regression and spatial modelling is then explained. Part B is concluded with an explanation of how the spatial model is operationalized.

In Part C, the analytical methods section, the many Exploratory Data Analysis (EDA) and Exploratory Spatial Data Analysis (ESDA) techniques used are described. Afterwards, a classical regression model is proposed and the assumptions of that model that are affected by the spatial nature of this study are listed. Finally, a spatial model is posited.

Chapter 4 discusses the results of the analysis. An interpretation of these results with particular attention to the effects of spatial dependence is provided.

Finally, Chapter 5 provides a summary of the work and findings. It also includes suggestions for future study. The conclusion recaps the salient points of this paper and reminds the reader of the potential use of GIS and Spatial Analysis in assessing EMS demands within a community. Clearly, studies in medical services must consider the inherent spatial component of health information (Parker & Campbell, 1998; Gatrell & Senior, 1999: 926). It is hoped that this work will advance our understanding of EMS demand in order to help emergency service planners while satisfying the rigours of spatial analysis.

Chapter 2 Background and Literature Review

2.0 Introduction

Emergency Health/Medical Services (EHS/EMS) provide a critical link between a community's citizens and its health care facilities. The provision of emergency services involves the care and transportation of all types of ill and injured people. In Ontario, the provincial government has recently transferred responsibility for Emergency Health Services to upper tier municipalities. As regional governments tackle their new mandate, it is expected that they will attempt to deploy limited resources in such a way as to offer an efficient, yet effective, service that meets local needs (IBI Group, 2000: 27; Ontario Ministry of Health and Long Term Care, 2002: section 6).

Early analytical research in EMS centred on location-allocation models that used simple assumptions about demand (Achabal, 1978; ReVelle, 1991; ReVelle et al., 1977; Zaki et al., 1997; Larson, 1975; Daskin, 1983; Eaton et al., 1985). Concurrently, separate research on EMS demand modelling was being done, however, as will be shown in this literature review, many different models and methods were proposed. Certainly, many studies identified data collection (Siler, 1975; Williams & Shavlik, 1979; Schuman et al., 1977; Kvålseth & Deems, 1979; Rucker et al., 1997; Cadigan & Bugarin, 1989) and processing problems (Szplett, 1988; De Angelis, 1995). With the advent of Electronic Ambulance Call Reports and Computer Automated Dispatching (CAD) - Relational Database Management Systems (RDBMS) integrated with GIS and GPS - we should expect improved empirical data collection. And with dramatic improvements in computer processing capabilities, the development of much more sophisticated and realistic Decision Support Systems should be attainable.

Moreover, a review of the literature reveals that the spatial nature of geographic data has not been taken into consideration when modelling EMS demand. This thesis demonstrates that by using a GIS with a spatial statistics program, researchers can identify and manage spatial effects of increasingly geographically realistic datasets in order to properly specify a model of demand. This must be done before researchers can effectively integrate location-allocation solutions with demand modelling.

This chapter provides the reader with some background into Emergency Medical Services in the province of Ontario, and more specifically into EMS in the study site, which is the City of Sudbury. Afterwards, a section on the use of computer applications in EMS, which includes a concise review of location-allocation work in EMS is provided. Then, current use of computers for data collection and Computer Aided Dispatching in the Ontario EMS industry is briefly introduced.

Next, an extensive, if not exhaustive, review of studies of the demand for EMS that primarily used multivariate regression analyses is presented. The chapter is concluded with a summary of the basic demographic and socioeconomic dimensions that have been identified as factors in EMS demand in order to posit a model for this study.

2.1 EMS Background

2.1.1 EMS in Ontario[1]

Since 1967, the Ontario Ministry of Health and Long Term Care (MoHLTC; formerly the Ministry of Health - MoH) has been responsible for the provision of pre-hospital emergency care through legislation and regulation of a variety of service delivery models. At the time, the Ministry of Health passed legislation to co-ordinate and upgrade the delivery of ambulance services in the province because of public concerns over large disparities in Emergency Health Services (EHS[2]) between communities (Hanna, 1982). However, the take-over was seen as very expensive and, as a result, a plan was put into place that created five (Table 2.1) types of ambulance services (Hanna, 1982; Shapiro, 1988: 17). These have persisted until recently, with subsequent minor variations including the addition of Contract Service provision. This arrangement did not succeed in alleviating concerns of inequity, and in many cases was confusing and seen to compound the issue (Shapiro, 1988:18).

The state of affairs of EMS in Ontario remained basically unchanged until December 8, 1997, when the Services Improvement Act, Bill 152, received Royal Assent in the Ontario Legislature (Government of Ontario, 1997). Previously, the Ontario Ministry of Health was responsible for the funding of all ambulance services - with only rare exceptions. Part of the act proposed to devolve the responsibility of funding, administering and delivering Emergency Health/Medical Services to Upper Tier Municipalities (UTM), or to District Social Services

[1]The purpose of this section is not to give an authoritative historical or descriptive account, it is intended only to give the reader an overview of an institution that has had a dynamic and complex history.

[2]The province of Ontario has long used the terminology Emergency Health Services, while most other jurisdictions have used Emergency Medical Services. The term EMS is used here as it is more widely recognized and has also been accepted in this province more recently.

Administration Boards DSSABs in those areas where only unorganized townships existed or no upper tier (County, Regional or City) government existed. Municipalities were required to choose from three options given to them by the province to assure a complete transfer of service by Jan 1, 2000 (Ontario Ministry of Health and Long-Term Care, 1999a).

Table 2.1 Ontario EMS - Land ambulance service provider types until 2001.

Service Provider Type	Number of services
Private Operators	69
Hospitals	64
Municipalities	17
Volunteer Agencies	13
Ministry of Health	10
Total	173

Source: Ontario Ministry of Health and Long-Term Care, 1999b.

The legislation directed UTMs and DSSABs to continue with the current land ambulance operator, provide ambulance services themselves, or through a Request for Proposal (RFP) process select the "highest quality, best price operator" to deliver the service (Ontario Ministry of Health and Long-Term Care, 1999c). The Ministry of Municipal Affairs and Housing outlines the municipalities' responsibility and accountability as follows (Ontario Ministry of Municipal Affairs and Housing, 1999):

- full responsibility and accountability for funding and delivery
- ensure essential linkages with clients, customers and service providers
- public and client education.

The transition deadline was then extended to January 1, 2001 with an announcement that funding would be shared between the province and municipalities (Ontario Ministry of Health and Long-Term Care, 1999d).

Recommendations to update the Ambulance Act Regulation 501/97 were made in the report "Review of the Ambulance Regulation: Report of the Land Ambulance Transition Task Force to The Minister of Health and Red Tape Commission" (Ontario Ministry of Health and Long-Term Care, 1998). These recommendations included language that would require the Upper Tier Municipalities and DSSABs to be responsive to fluctuating health care and medical demands of a constantly changing demographic and socioeconomic environment (Ontario Ministry of Health and Long-Term Care, 1998: 3) within unique geographies.

To emphasize the point of local responsibility, the Land Ambulance Transition Practical Guide, August 1999 (Ontario Ministry of Health and Long-Term Care & MOH-AMO Land Ambulance Implementation Steering Committee, 1999: 6) discusses the statutory responsibility of "Emergency Response Needs Evaluation" that considers the social, economic, demographic, geographic, workload and health status characteristics of the communities, which drive current and future service needs." The Ministry of Health also promised service delivery evaluation tools which are discussed in a later section in this chapter.

Ball (1980) had long ago suggested that decisions should be made at the local level where EMS services are delivered to meet local needs. He posits that, "[t]he most easily managed and effective ambulance service is that which is locally organized and sensitive to local demands. It is also likely to be the most economical to operate (Ball, 1990: 387)."

In essence, the downloading of ambulance services is seen by many as offering the potential for communities to tailor EMS to the needs of their individual communities. Besides being mandated by the Ambulance Act, meeting local needs is more likely to be assured because the administration of EMS will be subject to local political forces and procedures including closer public scrutiny and debate.

2.1.1.1 EMS Priority codes in Ontario

Throughout Ontario, a common system of prioritizing EMS is used. Calls for request for emergency services, whether Ambulance, Police, or Fire, are received by a 911 dispatcher, who in turn routes the call to the appropriate agency. In cases where an ambulance has been requested, or is required, the call is transferred to a regional office of the Ministry of Health's Central Ambulance Communication Centre (CACC). There, a Communications Officer (dispatcher) asks the caller a strict routine of questions using a Dispatch Priority Card Index (the DPCI is designed to err on the side of caution) to assign a "dispatch" priority to each call. An ambulance is then assigned, by the Communications Officer, to the call on a priority based on the caller's (usually a lay person) information. Upon arrival at a scene, paramedics reassess the situation using clinical judgement and the call is then re-prioritized accordingly for the trip to the Hospital - if transportation is in order. This is known as the "return priority". The result is often a much lower incidence of returning emergency calls than those dispatched. Both dispatched and return priority codes are assigned using the same scale given in Table 2.2. It is important to note that a variety of priority systems are used throughout the world, and therefore comparisons between jurisdictions can be difficult.

Table 2.2. Ontario EMS ambulance call priority system. These include both "Dispatched" priorities assigned by CACC and "Returned" priorities assigned by paramedics en route to hospital or other destination.

Code	Priority	Definition	Example
1	Deferrable	Non emergency but requires ambulance	Back Pain, a Patient Returning after a Treatment or Diagnostic Transfer
2	Scheduled	Ambulance has been booked with dispatch centre	Diagnostic Transfer to MRI or CT, Airport Transfer
3	Prompt	Patient requires medical treatment but is stable	Minor Burns, Uncomplicated Limb Fractures
4	Urgent	Life or limb threatening injury or condition	Cardiac Arrest, Chest Pain, Difficulty Breathing, Motor Vehicle Collision with serious injuries, Other Serious Trauma, Critical Care Transfer
5†	Obviously Dead	As defined by the Ambulance Act of Ontario	Decapitation, Trans-section, Decomposition, Gross Rigor mortis, Grossly Charred
6†	Legally Dead	Patient pronounced dead at the scene by a physician	
7†	No Patient Carried		- No patient found at the scene - Patient refused transport - Patient taken to hospital by other means

† Additional Return priorities determined by paramedics after their arrival at a scene.

2.1.2 EMS in Sudbury: Ambulance Service Background since 1983[3]

At the time of this study, the city's Emergency Health Services were provided by a privately owned, provincially funded, ambulance service. The Sudbury and District Ambulance Service (SDAS) had a fleet of approximately ten ambulances of various

[3] It is beyond the scope of this thesis to present a complete and detailed historical and descriptive account of EMS in Ontario or Sudbury. This section is intended only to give the reader an overview of an institution that has had a dynamic and complex history.

configurations including a number of single and double stretcher emergency ambulances, one four-stretcher inter-hospital patient transfer unit (Multiple Patient Unit - MPU), an Emergency Support Unit for disaster response, a First Response Unit used during staff shortages, and an administration vehicle. The service's jurisdiction extended well beyond the Regional Municipality of Sudbury. The City of Sudbury proper was usually covered by two "twenty-four hour cars," staffed with twelve-hour shift crews, with the addition of an 8 a.m. to 8 p.m. "car" Monday to Saturday, and an 8 a.m. to 4 p.m. "car" on Sundays. Backup vehicles were found in four surrounding area municipalities ranging from 15 to 30 minutes away from the city of Sudbury, only two of which had twenty-four hour emergency coverage, one to the north, the second to the northwest. In 1999, the Sudbury and District Ambulance Service responded to approximately 22,000 calls in the area (Greater Sudbury EMS, 2001). Nearly 10,000 of these calls were requests for Code 3 and 4 service, the vast majority of which occur within the city boundaries (Greater Sudbury EMS, 2001).

By December 2000, the Regional Municipality of Sudbury, now the City of Greater Sudbury became the sole provider of pre-hospital emergency care through its Emergency Medical Services Division of the Department of Emergency Services. The community has seen dramatic changes in EMS resources, including a growth from 80 to 120 paramedics and staff and a corresponding increase in response vehicles, of various configurations, from 10 to 26. Not surprisingly, overall there have been significant reductions in emergency response times throughout the City of Greater Sudbury (the former Regional Municipality of Sudbury) and in the core (the former City of Sudbury). These improvements were partly due to strategic planning mandated under the Ambulance Act that requires communities' EMS to at least meet 1996 response times. More recently, communities have entered into Response Time Accountability Framework agreements with the province in order to secure further funding. These agreements require improvements and close monitoring of service performance. Unfortunately, by and large, improvements have been made by intuitive and anecdotal accounts of service demand with little technological assistance. Certainly, a decision support system based on empirical evidence would enhance strategic planning.

2.2 Computer Applications in Emergency Medical Services

To understand the potential use for Geographical Information Systems and Spatial Analysis in EMS planning it is useful to review past and current uses of Information Technology (IT) in the field. First, early research that focussed on location-allocation problems using linear programming and super computers is considered. Then, the most current use of IT in EMS, i.e., Computer Assisted/Aided Dispatching (CAD) is presented. Finally, the Ontario Ministry of Health's method of collection and distribution of EMS data is introduced.

2.2.1 Location-Allocation Models

Much of the early use of computers in EMS was combined with fire services location-allocation studies in the fields of Operations Research and Management Sciences. Location-allocation problems were dealt with using very sophisticated computer algorithms and linear programming intended for large super-computers (ReVelle, 1991). Works on optimal siting of Emergency Services had evolved from considering demand as discrete points in space

(ReVelle et al., 1976), to data/distance matrices representing demand along arcs (Toregas et al., 1971; ReVelle et al., 1976). Analysis zones have also been used to represent demand clusters (Daskin & Stern, 1981; Eaton et al., 1985) where "the expected travel time between zone centroids is a frequently used proxy for response time" (Daskin & Stern, 1981: 138). Because of resource limitations at the time, Symons' (1969) use of a location-allocation heuristic to propose optimal locations of EMS stations was limited to a small set of hypothetical subareas. These models were all considerably simplified constructs. Gough and McCarthy (1975) suggest that much of the early work was experimental and of little practical value. Furthermore, ReVelle (1991) notes that early optimising programs had little graphic representation abilities, but required considerable computer power. ReVelle et al. (1976) encountered some difficulties when dealing with more than ninety (90) arcs. Those using analysis zones were faced with progressive coding and computational constraints as the number of zones was increased so that they would be sufficiently small to reflect demand (Eaton et al., 1985).

Many assumptions were made in early modelling. Whereas Toregas et al. (1971) presumed that each service provider (ambulance) was capable of responding at all times, Volz (1971), considered that every time an ambulance went into service, the remaining vehicles were instantaneously optimally relocated. Another assumption was made by Eaton et al. (1985) and Eaton et al. (1986) whereby population data were used as surrogate for emergency demand because of data constraints. A similar assumption was made by Achabal (1978), however, he did acknowledge that socioeconomic variables would probably play an important part in calculating real demand for EMS. According to Gough and McCarthy

(1975), one assumption was ubiquitous in the eight models that they evaluated: mean response times were used as criteria for the value of each model's results. They remind us that there was evidence that reduction in the time taken to get to accident victims to hospital can reduce the fatality rate (Gough & McCarthy, 1975). This concept is crucial to the notion that EMS must be carefully sited.

Practical applications of EMS location-allocation models have met with considerable success in Denver (Plane & Hendrick, 1977), Austin (Eaton et al., 1985), and Santo Domingo (Eaton et al., 1986). However, it is not surprising that, with developments in computer software and hardware, some are calling for the integration of GIS and optimization models (Miller, 1996). Simply put: "Advances in geographical information systems (GIS) technologies provide a greatly enhanced ability to represent, store and manipulate spatial objects" (Miller, 1996: 792).

2.2.2 Current uses of GIS in EMS

Where Gough & McCarthy (1975) may have called for more practical models for EMS siting in the late 1970's, we continue to find a paucity of theoretical research in the field of EMS demand and a slow introduction of technological applications, such as GIS, especially in Ontario. Even abroad, the use of GIS as a planning tool in EMS is in its infancy. The Association for Geographic Information, based in London, UK, reports that implementation is recent but growing quickly in Europe. For combined Emergency Services, including Ambulance, Police, and Fire, use of GIS is devoted mostly to command and control and emergency planning. Only about 18% of the services polled use GIS for census/demographic analysis. A very low response rate to the poll by EMS specifically (Association for Geographic Information, 1997: 3) suggests that the usage is even lower in that industry.

The 1990's saw the emergence of Computer Automated Dispatching systems developed by several software companies including Siemens (1999), Tritech (1999), and Intergraph (1999). These systems are relatively widespread in the US, however less so in Canada, certainly in Ontario. Currently, the Toronto EMS is using a CAD along with Global Positioning System to track and deploy its ambulances to 265,000 requests for service per year (Toronto EMS, 1999). As for the rest of the province, which is still dispatched by the MoHLTC's CACCs, the progressive roll out of a CAD is expected to be fully implemented by 2005. The system, developed by Tritech Software Systems will upgrade the province's former Ambulance Response Information System (ARIS) with "ARIS II" a Windows-based CAD incorporating GIS. Many services have implemented, or are planning the implementation of, GPS / Automated Vehicle Location (AVL) which are expected to readily integrate with ARIS II. Full implementation of the provincial CAD promises the potential for management reporting and data extraction.

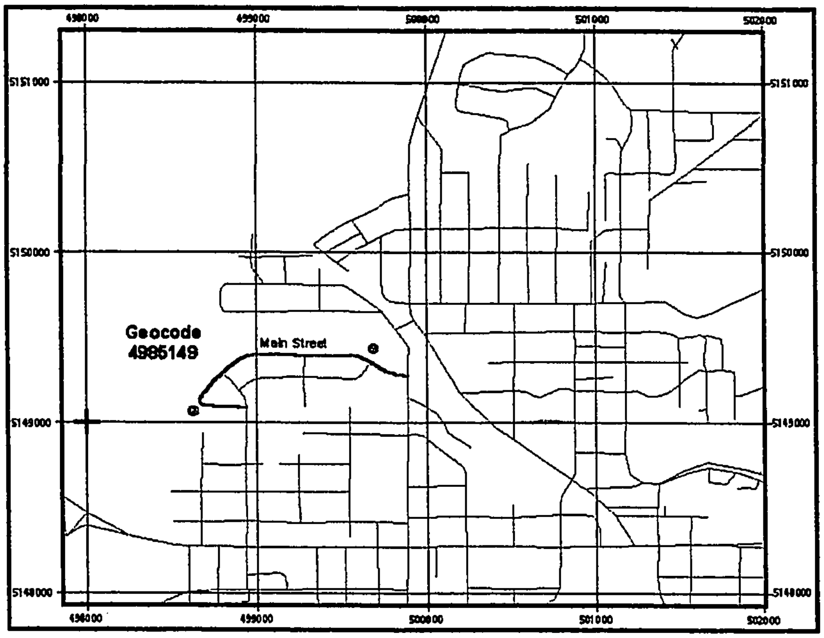
Currently, ARIS data is made available through a web-based download service known as ARIS Dispatch Data Access Service (ADDAS). ADDAS data are imported into a MicroSoft Access® database. The data consist of most of the CACC's dispatch data including call location (pick up address, municipality, Universal Transverse Mercator geocode) and response time (date, time of day, unit activity) information. Each municipality has access to its service activity data, and information on other services that have completed calls in its jurisdiction.

Associated with the distribution of ADDAS data is a MapInfo-based program developed by the Ministry of Health [based on Peters' Master's thesis (Peters, 1998)], known as Geoplot Standard Report. The software is intended to be a Spatial Decision Support System for EMS planners. It uses ADDAS data and reports call volumes, response times, and some basic statistics based on the one square kilometre Universal Transverse Mercator (UTM) grid. The results are, unfortunately, rather crude cartographically and of limited use when considering underlying social and economic structures (see De Angelis, 1995 below).

Although ARIS now assigns UTM geocodes to calls automatically during a dispatch session, the original reference database used for UTM assignment of streets was populated manually by Communications Officers from each CACC. Figure 2.1 illustrates the original manual geocoding assignment of streets using a UTM grid. Streets within a grid cell were assigned the cell's lower left UTM coordinates. However, streets that crossed more than one grid cell were often assigned only one geocode. In the example given below, EMS events at both the red dot and the blue dot could be assigned the geocode 4985149, depending on which grid cell "Main Street" was assigned to by the dispatcher (though usually the grid that contained the largest portion of a street was used) when the ARIS system was initially populated (O'Neil, 2004). The result of this inconsistent procedure is that automated geocoding of EMS events by CACC is prone to error, as De Angelis demonstrates (1995: 39-40). Concerns regarding this procedure have recently been reiterated by many municipalities, given that cross-border billing between jurisdictions is based on the geocode assignments.

Moreover, many EMS directors and managers lack the necessary training to use the MapInfo-based software; therefore, a number of Municipalities are considering the use of in-house GIS for more detailed scale EMS data representation and analysis.

Figure 2.1 UTM assignment of streets by CACCs



2.3 Studies in Determinants of EMS use

This section provides a review of previous studies that demonstrate the link between demographic, socioeconomic and other (usually land use) variables and the demand for pre-hospital emergency health care. Some studies on the demand for EMS have often been simply descriptive (Schuman et al., 1977; Butler, 1981; Gibson, 1971; Hisserich, 1971; Kvålseth & Deems, 1979; Pennycook et al., 1991; Fairhurst, 1992; Gardner, 1990; McConnel & Wilson, 1998; McConnel & Wilson, 2001), while others involved correlation analysis of either ecological variables or patient specific characteristics. However, this review focusses on works that have included regression analyses as these would, of course, offer more insight into the variables to include in this thesis. Notable throughout these studies is that, despite the use of geographical units of analysis such as counties and census tracts, none have included any spatial analysis techniques/diagnostics or discussed the potential effects of space on analytical results.

Probably the seminal, or at least the most frequently referenced, work in ecological analysis of EMS demand was performed by (Aldrich et al., 1971). In their study, the authors generated least squares multivariate linear regression models for a number of subcategories of ambulance calls as well as for all types combined. The subcategories included "Automobile Accidents," "Other Accidents," "Sickness: Cardiac and Poison," "Sickness: Other," and "Dry Runs." All had quite high coefficients of determination (R^2), from the lowest for "dry runs" at 0.598 to an incredibly high 0.927 for combined ambulance call types by census tract. They suggest that this high explanatory power may in part be due to reducing the sampling variation by pooling four years of data for all types of calls. When aggregating 632 census tracts into 157 contiguous areas, the R^2 was reduced, slightly, to 0.904 for pooled data. In light of what is known about the effects of aggregating spatial units -a tendency to increase R^2 - this decrease is a bit perplexing.

Aldrich et al. (1971) recognize as a significant problem, the gap between the census data collected six years prior to the EMS demand data. The magnitudes of the coefficients of determination seem rather unexpected considering the temporal mismatch of their data.

The three categories that most closely replicate the current study are "Other

Accidents," "Sickness Cardiac and Poison," (the combination of Cardiac and Poison is confusing but fortunately separated for interpretation) and "Sickness: Other." The interpretation of Aldrich et al.'s (1971) results are difficult as they postulate a hypothesis regarding each variable's effect on each individual model of demand posited. Sometimes, their variables have reverse effects depending on which type of call is examined. For example, "Percent Children" has a positive effect on "Other Accidents" and a negative effect on "Sickness: Other." This is a bit problematic if the results are to be used for general planning purposes.

When considering the overall demand model, most of the variables (20 of 31 -refer to Aldrich et al.: pp.1160-1161 for a detailed list and pg. 1164 for their statistical significance in the regression equation) had statistical importance. These included demographic (age, sex, race), socioeconomic (married status, employment, occupation), and urban structure/spatial variables (land use, population and housing density). It is clear that all priority of calls (urgent and non-urgent) have been included in this study. They conclude that areas with elderly and children generate high call volumes. Interestingly, this study found that commercial land use also generated frequent ambulance calls. The variable "Distance From a Hospital' contributed nothing to the results, and as a result, was dropped from the analysis.

Most importantly, they concluded that low income families and non-whites created a higher demand for public ambulance service in Los Angeles county. For them, "low socioeconomic status is defined in terms of income....race, unemployment, age (people over 65)....employment in low status jobs and crowded living conditions" (Aldrich et al., 1971:1158). They capture each of these components of low SES with approximately 30 variables; some of which seem a bit redundant. For example, they use both "Percent Males Employed" and "Percent Males Unemployed" in their models. Though they acknowledge collinearity in one of their models, they offer no means to re-specify the model. The strong correlation between these variables may have generated some spurious results. It is not surprising that at least one other author has criticized this model as over-specified (see below).

Gibson (1971) responds to Aldrich et al.'s (1971) analysis in a letter to the editor by providing his data for 75 communities in the Chicago area. Gibson (1971) claims to overcome some of Aldrich et al.'s (1971) shortcomings by including only cases where patients were transported to hospital, and adding users of private ambulance services. He used stepwise regression and compared two free public services to a private service which charges a fee-for-service. Ultimately, Gibson charges that Aldrich et al.'s (1971) models were over-specified. He acknowledges however, that Aldrich et al. (1971) modelled all types of calls, unlike his own work that examined only calls where patients were actually transported. Using only "Age," "Income," and "Number of Physicians per 1000 Population," Gibson (1971) generates an R^2 of 0.89 for all services combined in 75 "community areas" -which he does not clearly define- in Chicago. He gets a similar result (0.88) using "Age," "Education," and "Number of Physicians per 1000 Population" for public services. Unfortunately, Gibson's (1971) published tables do not reflect these results, and they also do not include the variable "Number of Physicians." The results table, including six variables from the 1960 census, shows significantly lower R^2 for each service type than those reported in the text; for total ambulance utilization the value is only 0.661. Interpretation is difficult because the text and table are incongruent, however it seems that for both public services and combined services, "Median Years of Education" has a positive regression coefficient and "Median Family Income," a negative coefficient. The author insists the work was ongoing but indicated that several social variables (race, income, and education in particular) were related to patients being transported to hospital. Gibson (1971) notes that, when compared to the model presented by Aldrich et al. (1971), the reverse impact of some variables in his model, may suggest a differentiation between models predicting overall usage - including patients not taken to hospital- versus models of emergency ambulance transport.

"Predicting private emergency ambulance utilization, however, seems to be more problematic" asserts Gibson (1971: 2158). Even with 18 predictor variables, a coefficient of determination of only 0.39 was obtained. Gibson (1971) offers no explanation; (Hisserich, 1971; see below) does. Gibson (1971) also notes that income has much greater negative effect on the use of free ambulance services than non-free services, though the effect is still negative for the latter.

A rebuttal to Gibson's (1971) comments is made, also in the form of a letter to the editor in the same journal, by Hisserich (1971), a co-author of the Aldrich et al. (1971) article. He claimed that, in a more detailed study of the Los Angeles data, an R^2 of 0.71 was achieved using only five predictor variables, namely: "Percent of Commercial Acreage," "Number Persons Working Within the Census Tract," "Number of Widowed Separated and Divorced Males," "Number of Unemployed Male Residents," and "Acres of Streets." These variables all had positive influences on demand for EMS, a relationship which he attributes to the high call volume in the downtown "Skid Row." Also, corrections for heteroscedasticity did nothing to change the pattern of variables. Curiously, removing "Skid Row" from the analysis made low income and minority variables more positively significant.

Hisserich (1971) proposes that Gibson's difficulty in predicting private ambulance utilization is a function of a random phenomenon that produces urgent conditions rather than "social conditions which inhibit access to regular medical care." That is, demand for public EMS, suggests Hisserich (1971), may be more strongly related to socioeconomic variates than is demand for private services. He suggests that private ambulances are typically dispatched at the request of a physician, supposing an urgent need, as opposed to public requests that may include non-urgent cases. The public system, posits Hisserich (1971), is filling a role of primary care to disadvantaged people who are immobile.

This distinction is important because the private model studied by Gibson (1971), with its low coefficient of determination (0.39), may more closely reflect the reality of the current analysis, whereby only actual transports to a hospital for emergency cases are included. Because of data limitations, non-emergency transports and dry runs cannot be considered in this thesis, cases that Hisserich (1971) submits are related to socioeconomic forces.

Siler (1975) used 1970 census data with 1973 EMS use data in Los Angeles County to develop a stepwise multiple regression model that resulted in a coefficient of determination of 0.921. Census tract data were aggregated for communities in the county. From an initial 136 independent variables, his methods first pared the data down to a set of 23 variables representing "demographic, employment, land use and emergency medical services characteristics of each community" (Siler, 1975:259). His final model included only four independent variables.

In his study of 81 communities in Los Angeles county, the most powerful determinant was "the employment in an area expressed as a proportion of the resident population" (Siler, 1975: 260). But the relationship was non-linear and in two forms, one the square, the other a reciprocal. Though he does not state it explicitly, by using two forms of the same variable he seems to suggest two different spatial regimes of EMS demand. Siler (1975) interprets the jobs to residents ratio as a surrogate for land use, i.e., the degree of 'residentialness' of an area. He explains that, up to a threshold "Employment to Resident ratio," there is an inverse relationship between the ratio and ambulance demand. Therefore, as residential areas are more "residential" there is an increase in call rates, posits Siler (1975). Beyond the threshold, his research suggests that there is a direct relationship with EMS demand. He then proposes that areas with high ratios are commercial/ industrial, and as an area becomes more so, so does demand for EMS increase. Next Siler (1975) finds a negative relationship with the "Log_{10} of Proportion White and Married People," suggesting demand is higher for "Unmarried Whites" and for "Nonwhites" regardless of marital status. Siler's (1975) third variable, the "Square of Housing Units per Residents," finds that demand is higher with lower occupants per household. He proposes that this represents areas of seniors or single youth. De Angelis' (1995:26) interpretation of Siler's (1975) findings is to suppose that "[s]ingle occupant households would appear to rely on ambulances in an emergency, whereas larger households would rely on their spouse, sibling, or house mate to provide transportation to emergency medical facilities." Finally, a positive regression coefficient for the reciprocal of "White to Blue Collar Ratio of Female Workers" (a similar value was found for males) implies that demand increases with a higher proportion of women holding blue collar jobs as compared to white collar.

Siler (1975:255) insists that "there are no underlying phenomena in ambulance demand that might lead one to reject nonlinear relationships." He recognizes that others (Aldrich et al., 1971) have had success with the simpler linear model, but ignores that others (Gibson, 1971; and Hisserich, 1971) supported that premise. Nevertheless, Siler (1975) has also not suggested good reasons for transforming the variables. The author does caution that, though his methods are applicable to other regions, the specific model developed for Los Angeles would not be. He adds that climate may have played a larger factor in an area with greater climatic extremes, causing some seasonal variation in demand. This is worth considering in the Canadian context.

Because of a study that demonstrated that nearly half of critical patients in emergency departments were not transported by ambulance, Schuman et al. (1977) proposed to estimate demand that was met, as well as unmet need. Both demand and need estimates were made by retrospectively examining patient records and considering the need for an ambulance on a case by case basis. They looked at all cases for 14 randomly sampled days in 1973 from three emergency departments in a rural county of western Pennsylvania. The county was divided into 26 districts that, we surmise, were relatively homogeneous in terms of their composite census tract population characteristics. Both demand and need equations had very high coefficients of determination, 0.932 and 0.903 respectively. For both equations, "Population" had a positive impact on call volume, while "Travel Time to the Emergency Department" and "Major Highway in District" had negative effects on demand. How the last two variables were operationalized is unclear as one table reports only a mean and standard deviation for these variables. For the Demand equation, "Years of Education" had a negative impact, while the "Presence of an Emergency Department" had a positive relationship with call volume. For the Need equation, "Total Employed" and "Percent Families under Poverty Level" had negative and positive impacts respectively. The "Population" variable seems rather obvious in that both that variable and the dependent variable, "Calls per District," seem to be measured in raw counts. The education, employment, and poverty variables seem to suggest that low socioeconomic status is negatively associated with ambulance call volume. "Travel Time to the ED" and "Presence of Major Highway" may be describing distance from an urban core. The positive relationship with the presence of an ER may also suggest a degree of urbanism. Unfortunately, the geography of the study site is not well described. Just as Gibson (1971) and Aldrich et al. (1971) noted, Schuman et al. (1977) remind us the methods they use may be applicable to other communities but the equations would likely take a different form. It is interesting to note that the authors assert that the residuals analysis shows that their variables were random and normally distributed. It is unclear if the residuals themselves behaved this way, aspatially or spatially.

Kvålseth and Deems (1979) use a stepwise ordinary least squares multiple linear regression model to examine demand for EMS in Atlanta, Georgia. Because of data problems, the study was restricted to 79 census tracts representing approximately 315,000 residents, or 63.4 percent of the city's population in 1970. Neither the spatial arrangement of these tracts nor their contiguity is clearly stated. Ambulance calls for one month were assigned to their respective census tracts. The severity, or priority, of these calls is also unclear, but given that "dry runs" are included, it is likely that all priority calls are included.

Kvålseth and Deems (1979) provide two types of models, first and second order models. First order models use non-transformed census variables. Second order models include some transformations but mostly interaction terms using census variables.

For the first order model, the stepwise method resulted in 18 variables entering the equation with a coefficient of multiple determination of 0.90. Kvålseth and Deems (1979) agree with Aldrich et al. (1971) that demographic, economic, housing and land use, and traffic characteristics of an area are associated with EMS demand. The authors summarize their first order model by stating "that the demand for EMS increases with decreasing family income, increasing unemployment among males, increasing acreage per capita, and increasing percentage of the black population (Kvålseth and Deems, 1979: 252)."

They note however, that "Acreage per Capita" and "Acreage" itself are the most important determinants, but that their signs are opposite. The former is calculated by dividing census tract area by population, resulting in an inverse measure of population density; the latter is simply the areal extent of a census tract. They further note that these two variables were highly collinear. "Smaller Acreage," suggest the authors, is seen as a surrogate for proximity to the centre of the city, so it is surprising that its effect on demand was negative, especially when considering that "Housing Age," which would also intuitively be related to proximity to the city core, was positively associated with demand. "Percent Commercial Land Use," and "Number of Workers Using an Auto to Work" also display a positive impact on EMS use. These last two variables may simply suggest high activity space. Percent Population over 65 and under 15 both had positive impacts on the regression, the former having a greater impact than the latter.

The authors then used the first six variables (responsible for an R^2 of 0.85) entered in the first order regression to compute a number of interaction variables. When considering their second order model, it doesn't seem terribly surprising that the coefficient of determination is so high (0.93), given that "Percent Commercial Land Use," responsible for more than 36 percent of variation in EMS demand, was used to compute three of the second order variables. These models are difficult to interpret, seem redundant, and only offered a 3% increase in explanatory power.

Szplett (1988), to the best of the current author's knowledge, was the first to undertake an ecological study of EMS demand in a Canadian context. As a researcher from Ryerson Polytechnical Institute, she prepared a report for the Municipality of Metropolitan Toronto Department of Ambulance Services. She examined the association between the number of emergency (Priority 4) calls per square mile and variables representing socioeconomic, demographic, workplace and transportation network characteristics.

Non-institutional ambulance call data from the Toronto Department of Ambulance Service was available for 1 square mile grid cells. Data for nine months in 1986 were gathered. Though the author does explicitly state that the calls were Priority 4 (Emergency) she does not distinguish whether this was the dispatch or return priority. Accident data from the Police Department were converted from address to grid cells. Employment and retail data from the Planning department were available in Minor Planning District areas. Socioeconomic and demographic data from the 1981 census were available at the census tract level. The author created a Fortran program to assign census tract and planning district data to the 1 square mile grid by "Areal Overlays." The author describes a complex system developed to resolve the problem of grid cells intersecting several census tracts. Clearly, since census tracts, and likely planning districts, would be much larger than the grid cells, a number of spatially autocorrelated grid cells would have been created. This is clearly evident when examining her maps of EMS call volumes and many other variables which are highly concentrated in the downtown area. Szplett (1988) does not acknowledge this, nor does she discuss its possible impact on the regression equations.

Separate regression equations were calculated for weekdays, weekends and four-hour time frames, as well as for the total calls. For the purposes of this review, only the "Total Calls" equation's results are discussed here. Szplett (1988) reports that the coefficient of determination was 0.87 for a model that included five variables, namely: "Retail*Apartment (interaction)," "Total Employment," "Single Parent Families," "Number of Major Accident Locations," and "Low Income Unattached Individuals." Szplett (1988: 23) concludes that her work "support[s] past studies which demonstrate that areas exhibiting disproportionately high numbers of low income and single parent families, single and divorced individuals and rental units and apartments also exhibit disproportionately high demand of [ambulance] services." Perhaps surprisingly, age did not enter any of the equations. If she suspects that age is excluded from the regression solution because of collinearity with other variables, she does not stipulate it.

Starting with 150 variables, Szplett (1988) used Principal Components Analysis to reduce the number of independent variables to 47, and mitigate the effects of multicollinearity in the regression equations. Ironically, Szplett (1988) recognises the importance of independence between variables but ignores dependence (latent and perhaps artificially created - see above) between the sample units that she created.

It is important to note that the two variables used to derive the most important variable in the equation, "Retail*Apartment," were assigned to the grid cells by the Fortran program from two different sources, the first from planning districts, the second from census tracts. The impact of these likely incompatible geographies is highly suspect. Otherwise, we would expect that the interaction variable indicates activity space.

Most important, in calculating the "Number of Major Accident Locations" variable, the author is being quite redundant, as this indicates locations where ambulances are required by definition. Therefore, that the grid cells containing a high number of accidents is correlated with a high ambulance call volume (the dependent variable) is obvious.

Spzlett (1988) does address the violation of the assumption of normality as it relates to her data and model. She notes that the dependent variable is positively skewed, as are most of the independent variables. She concedes that log transformations would alleviate the violation but argues that "regression is fairly robust (i.e. insensitive to distortions) for the normality assumption especially with large sample size." She advances that "this robustness is enhanced when the predictor variables are similarly skewed." She indicates that attempts at variable transformations did not yield significantly "better" results and entered similar variables. This begs the question as to whether a "better" model, in terms of explanatory power, is necessarily more appropriate.

Cadigan and Bugarin (1989) argued that guidelines for EMS transport (3.5% of population, or one transport per 10,000 persons per day) do not consider socio-demographic variables that could affect EMS demand. The authors proposed to model EMS transports and EMS demand (including "dry runs") using stepwise multivariate regression. The study was limited to 142 communities in Massachusetts, with populations less than 65,000, that could provide sufficiently complete EMS data. They concede that data were not drawn from a random sample. To avoid problems with multicollinearity and ambiguities in interpretation of results, only six dependent variables were used, including a dummy variable to represent non-resident populations.

"Median Income," "Percent Living Below Poverty Level," "Percent More than 65," and "Cape Cod" (dummy variable) were found to be significant in estimating both EMS response and transport demand. The variable "Number of Highway Miles" failed to enter the equation. Interestingly, they contend that attempts at using interaction or transformed variables did not improve the models. Ironically, when they reduce their equation for "Emergency Transports" for resident populations, it essentially replicates the rules of thumb with the addition of "Median Income." It is unclear whether the dependent variable was raw EMS transports or rates based on 1,000 persons because detailed regression results were not presented. If the dependent variable was number of transports, it is not surprising that the

"Population" variable would be important. In this regression analysis, communities with large populations would have caused a size bias.

De Angelis (1995) undertook the second Canadian investigation of the relationship between Ambulance call demand and underlying demographic and socioeconomic factors. The study included Code 3 and Code 4 calls from December 1993 to December 1994 in the Greater Kingston Area in Ontario. Census data were from the 1991 census. His study excludes, from the EMS database, Motor Vehicle Accidents (MVA) and institutional transfers to more closely reflect the relationship between census data and emergency calls. He does not, however, separate those calls where people were not at their own residence when they needed an ambulance. In the current study, these calls represent approximately 15 percent of cases. It is probably reasonable to have assumed that even when a person is not at home, they are likely at another residence that closely reflects their own social and economic status. However, calls can frequently occur while at work or in public spaces, places that can be significantly different than one's own social, demographic, or economic space. Braun (2000) suggests it is best to consider people's home addresses if ecological variables are used (see Krafft, 2000 below). This is with the view of developing education and prevention programs for those who use EMS regardless of where people are when the need arises.

De Angelis (1995) provides an interesting descriptive breakdown of the EMS calls database. He reports that 28.4% of calls were Code 3 while 71.6% were Code 4. He notes that 46% of patients were female and 44% male; 10% of cases were of unspecified sex. He also includes a breakdown of ambulance calls by patient condition. De Angelis (1995) does not specify whether call priorities refer to dispatch or return codes. However, given that the "Unknown" category is dominant - 38.7% - it seems likely the "Priority Code" discussed above is dispatched rather than return.

Unfortunately, Ministry of Health call volume data were disseminated for 1 square kilometre cells, while census data were available for census tracts. As De Angelis (1995) reports, this caused major concerns. We would expect census tracts to be typically larger than the Ministry's grid cells. Therefore, in rural areas of Greater Kingston, the tracts were very large, containing several grid cells. In this case, as the census tracts were considered homogenous, all 1 km^2 cells whose centroids were contained within a census tract were assigned that tract's census attributes. This dis-aggregation of a homogenous census tract into many identical grid cells would naturally lead to a very high degree of spatial autocorrelation.

Where census tracts were much smaller and closer in area to the grid cells, i.e., in the City of Kingston and its immediate surroundings, several census tracts' boundaries could be found within a single cell. De Angelis (1995) opted to assign, to the 1km^2 grid cell, the characteristics of the census tract that occupied most of the cell.

Because the study area was quite large, it included an extensive rural area where there were no calls. This resulted in many grid cells with no EMS calls, therefore, models for "Zero Cells" included and excluded were run. De Angelis (1995) also selected a smaller study area to analyse, which roughly represented the built up area of the city of Kingston and immediate surroundings.

Three similar dependent variables were examined for each of the large and small areas including and excluding zero cells. "Total Calls" "Demand per 1,000 Persons" and "Log 10 of Total Calls" were separately modelled by stepwise multiple regression. The last model was proposed because of a positively skewed distribution.

The independent variables first used by De Angelis (1995) included:

1. Average number of persons per household
2. Population Density
3. Percent Movers
4. Median Income
5. Percent Employment and
6. Percent Post-secondary Education

De Angelis (1995) used a guideline whereby an absolute bivariate correlation value of 0.7 between two independent variables would result in problematic multicollinearity. Therefore, because "Average Number of Persons per Household" was correlated to "Median Income," the latter was removed. However, why "Median Income" was removed rather than "Average Number of Persons" is not clear. Whether the income variable was removed from the pool by the researcher prior to regression analysis, or the exclusion was the result of the stepwise regression process itself is also unclear. Nevertheless, De Angelis (1995) proposes that, in a Canadian context, income may not play the same role as in previous American studies of EMS demand.

The coefficient of determination for his analyses of "Zero Cells" included and excluded were not remarkably different. For "Total Calls" and "Log 10" equations, the R^2 ranged from 0.58 to 0.65. However, the variables entered into each model were different. Persistent variables in these equations were "Average Number of Persons per Household" and

"Population Density."

However, results for the "Demand per 1,000" model were dramatically different both for the full and smaller areas. For the full study area, the R^2 was 0.012 for zero cells included and increased significantly for zero cells excluded to 0.117. For the small study area, slightly better results of 0.032 and 0.150 for zero cells included and excluded, respectively, were found.

For the "Total Calls" model, De Angelis (1995) both mapped residuals and plotted them against the estimated dependent values. The maps showed no obvious patterns, however he does discuss some outliers and demonstrates the utility of the GIS in identifying flaws in the Ministry's data. Plotted residuals for the "Total Calls" model, for both the full and small study areas, shows an obvious spreading pattern that De Angelis (1995) does not discuss or even identify.

The small study area "Demand per 1,000" model, with zero cells excluded, most closely replicates the current study. Using "Average Number of Persons per Household," "Percent Movers," and "Percent Employed" only 15% of the variation in "Demand per 1,000" was explained. All three variables were negatively associated with demand. It is interesting to note that in the full study area, the first two of these variables had similar influences, however, "Percent Employed" was positively associated with demand, and "Percent Post-secondary Education" was also included as a negative influence.

De Angelis (1995) acknowledged that assigning data from one framework to another put into question the representativeness of the census characteristics of each cell. He adds in his conclusion that his results are inconclusive. It seems that the very large difference in the

"Demand per 1,000" and the "Total Calls" models, in terms of explanatory power, adds to this concern. De Angelis (1995) recommended that GPS be placed on board ambulances so that calls could be assigned to Enumeration Areas for improved analysis. Strangely, he ignores the use of address matching which is used in this thesis.

Rather than using an ecological approach and a classic linear regression model, Rucker et al. (1997) looked at patient specific predictors of EMS using a multivariate model and logistic regression. Socio-economic data were captured using questionnaires issued to patients and chart abstractions rather than census data. Five Emergency Departments (ED) in the northeastern United States participated. Specifically where the study was conducted is unclear. Data were gathered from each ED for one month between February to September, 1993. Only patients 18 years of age and older with a history of six selected chief complaints were interviewed. Exclusions included more severely injured patients (e.g. head injuries) and uninsured patients.

For patients who filled out the questionnaire, only "Insurance" and "Income" were statistically significant as predictors of EMS use, when Severity and Complaint Type were controlled for. Furthermore, "Patients 65 years of age and older," greater severity of complaint , and patients with certain chief complaints also contributed to EMS use. This additional assertion is probably without having controlled for "Severity" and "Complaint Type," but they are not explicit here. More important, they report that for patients as a whole, data from chart reviews showed "Age" and "Insurance Status" as the only predictors of EMS demand.

Rucker et al. (1997:490) conclude that "Ambulance use was significantly associated with age, clinical severity, insurance coverage and family income." They determined that "race, sex, education, Medicaid coverage, frequency of ED use, living arrangements, and primary physician availability were not predictive in multivariate analysis of surveyed patients" (Rucker et al.1997:484). Curiously, they suggest that "Employment Status" and "Living Arrangements," which they concede may be important in describing ambulance use on a geographic basis, appear to not be statistically significant in their model. However, for at least one of these variables, it is unclear how this was determined as they had previously indicated that "Employment Status" and "Housing Type" had not entered the model because of pre-screening using bivariate analysis.

Notable limitations of this investigation involve the exclusion of patients based on a number of clinical criteria, time of admission, and patients' option to participate. Furthermore, they identify a potential flaw in arbitrarily choosing socioeconomic variables for the analysis; however, their proposed variables seemed grounded in previous research. Like previous authors, Rucker et al. (1997) acknowledged that the characteristics of the urban communities involved were likely unique.

Braun (1998) examined EMS use in Santander Northern Spain as part of his requirements for a diploma at the Geographische Institute der Rheinischen Friedrich-Wilhelms-Universität Bonn[4]. Braun aims to investigate the relationship between EMS

[4] Mr. Braun provided a copy of his work in German with a summary of results in Spanish included. It is with the assistance of German and Spanish speaking colleagues that the current author tentatively report the results of this fine work.

demand and neighbourhood-level social structure in an attempt to identify inequities in health between social groups.

Census data from 1991 were compared to EMS usage during the month of March 1998. Therefore the results are subject to the same uncertainties as many studies. The results are from bivariate correlation analysis of EMS use and five social variables, and from subsequent cluster analysis of the 133 census areas in Santander.

The Spearman Rank Correlation Coefficient between "Unemployment Rate" and EMS emergency use was significantly positive, as was the value for "Unskilled Labourers." "University Education" and "Persons with Management/Directors Positions" were significantly negatively related to EMS emergency use. Braun (1998) concludes that low socioeconomic neighbourhoods generated approximately twice the demand for emergency calls.

At the 9th International Symposium in Medical Geography in Montreal, Thomas Braun - a co-author with Thomas Krafft and Barbara Raue - presented early work on the development of a GIS tool for EMS Demand Analysis and Spatial Decision Support. These researchers from the University of Bonn teamed up with the Richmond Virginia (USA) Ambulance Authority (RAA) to improve EMS system design in that community. Together, they recognized the lack of EMS demand knowledge and the long standing influence of "short-term, issue driven policies" on EMS planning.

At the time of the presentation, the framework for data collection had been proposed and was being implemented. A two pronged approach to EMS analysis was advanced. The first analysis would include the mapping of geocoded incidents to which RAA units responded (pick up locations). Braun (2000) suggests that only operational decision support and planning insights can be gleaned from this analysis. He insists that ecological analysis comparing socioeconomic variables with pick up locations is not sufficiently meaningful, given that many people who require an ambulance may not live in the area where they are picked up. The second part of the analysis involves geocoding patients' residential addresses then undertaking socioeconomic and demographic analysis with clear intent of "identifying high risk population/neighbourhoods for the implementation of prevention strategies." This endeavour is, to this author's knowledge, the first practical implementation of a GIS application in EMS demand analysis.

However, when asked in a personal interview, Mr. Braun (2000) was not aware of any efforts, or any plans, to consider the spatial characteristics of the data (e.g., spatial autocorrelation) in terms of their effects on the results of their analysis.

In a county level analysis of EMS use in the state of Kentucky, Svenson (2000) used multiple regression analysis with the following geographic and demographic independent variables: "Location Rural vs. Urban," "Level of Pre-hospital Care," "Access to 24 hour Emergency Room," "Availability of 911," "Poverty" (including Per Capita Income, % Households with No Wage Earner, % Residence below Poverty Level), "Education (% < Grade 9)," and "Availability of a Telephone."

All variables except "Education Less than Grade 9" were reportedly significant, but the effects of each were small. Unfortunately, specific regression parameters are not reported, nor is the coefficient of determination, therefore we cannot assess the overall explanatory power of this model. Given that the data are highly aggregated and no spatial effects were considered we might expect it to be rather high.

The author reports Relative Rate for each variable where he tells us that "Age" was the major determinant of demand. He reports that use of EMS by people over 65 increases exponentially. In Svenson's (2000) work, rates for those aged 1 to 14 were lowest, while for those aged 15 to 64, the rates were reported as constant. "Poverty Level" is also highlighted as an important determinant of EMS demand. Each of the three variables used to represent poverty were directly related to rate. Interestingly, as the number of homes without telephones increased, EMS usage decreased; but, usage increased in communities with no 911 service. The former may be reasonably intuitive, however the lack of 911 service associated with increased EMS use is confusing especially when the authors report that there is very little difference in rural and urban EMS rates. Perhaps communities without 911 service are also poorer communities. Many other authors have provided similar results of EMS use by the elderly (McConnel & Wilson, 1998; McConnel & Wilson, 2001; Wofford et al., 1995; Dickinson et al., 1996; Clark & Fitzgerald, 1999; Gerson & Skvarch, 1982; Béland et al., 1991) however these are not in the form of multivariate regression. They are typically simply descriptive of empirical findings in nature and in some cases the results of bivariate correlation analysis.

Reviewing the literature has highlighted a large number of variables that have been associated with EMS demand in a variety of geographic and clinical contexts. We can, nevertheless, broadly categorize these variables into four basic dimensions that have been successfully associated with EMS demand, namely, the demographic, social, economic, and geographic (urban structure/land use) dimensions of a community. What has clearly been demonstrated in past work is that people of disadvantaged social and economic status, and demographic groups related to disadvantage, e.g. advanced age and minority race, have been identified as higher users of Emergency Medical Services. Also, 'Geographic' variables such as population density and urban structure (proximity to the core of a city) have also been related to high EMS demand. The reader must keep in mind that the distinction between these dimensions is not always clear, i.e., a variable used to represent one dimension can often be associated with another. For example, unemployment status is as much a social concern as it is an economic indicator; dwelling type can reflect one's economic status as well as one's proximity to the centre of a community, transportation networks, or activity spaces.

The next chapter describes the study site and how the data were collected, the dataset's idiosyncrasies, and how it was spatially enabled so that, after first mimicking classical methods, the spatial characteristics of the data could be explored in order to potentially improve the model.

Chapter 3 Methods

3.0 Introduction

This chapter is divided into three parts. In part A, the study site is described in order to give the reader some context and demonstrate why the site offers unique research opportunities. In part B the data are introduced and methods used to prepare the data for analysis are explained. Part C outlines the analytical methods used to explore the data both aspatially and spatially. In this third part, two regression equations used to model EMS demand are proposed. The first is an exploratory tool used to prepare for the second regression equation that contends with the spatial effects of geographic units that past researchers have ignored. First however, the chapter is prefaced with a brief section in order to 1) remind the reader of the research context of this thesis, 2) introduce the conceptual framework under which this work is undertaken and 3) present the software environment in which this work was performed, and 4) offer a basic guide to the many steps used to complete this thesis.

3.1 Developing a Research Approach

3.1.1 Research Context

As previously stated, ambulances need to be located in strategic locations within a community if they are to be both effective and efficient. Currently, legislation requires that 90 percent of dispatched emergency calls (code 4) be reached within at least the same response time as had been the case in 1996. Furthermore, Response Time Accountability

Framework agreements with the Ministry of Health and Long Term Care (MoHLTC) require that these response times be reduced.

Two different approaches to setting response time goals in a municipality could be adopted. One can be thought of as a jurisdiction-based approach, the other a population-based approach. The former is where every citizen of a community is provided with the same predetermined response time by its Emergency Medical Service regardless of where she/he resides throughout an often vast area of responsibility or jurisdiction. The latter case is where EMS resources would be stationed to respond within a given time frame where the greatest concentration of people is seen. The former would be effective at providing service to everyone who needs it, but would likely be prohibitively expensive; the latter would be considered more efficient and seems reasonable in any environment where resources are limited. The emphasis is on "seems," because if all persons generated an equal degree of EMS demand, then the population-based approach would dictate that planning service areas for a set number of ambulance stations is a simple function of determining the population size of small subareas of the service's jurisdiction. This approach has certainly been used in early EMS siting models (Eaton et al., 1985; Eaton et al., 1986; Volz, 1971; ReVelle, 1991). However, this simple population-based framework ignores what other researchers (Aldrich et al., 1971; Gibson, 1971; Siler, 1975; *etc.*) have demonstrated as a link between demand and certain social and demographic forces. Although efficient, the population-based approach is not very effective at providing the service to people or areas that need it the most.

Therefore, a truly effective and efficient approach to siting EMS resources should consider where demand is generated and by whom. It seems a simple matter of tying operational research to demand modelling, which has been suggested (Eaton et al., 1985; Eaton et al., 1986; Achabal, 1978). As reviewed in the second chapter, modelling to predict areas or people that use EMS most frequently has been done with considerable success. That success, however, may have been inadvertently exaggerated, and before carrying on with the integration of operational and demand modelling, this issue needs to be addressed.

It is the current author's thesis that the explanatory power of these models has been inflated because past studies have ignored the spatial nature of the problem. It is agreed here that demand for EMS is a function of a number of socioeconomic and demographic forces, however, there are considerations about the geographic data which must be examined and which may possibly need to be addressed. What is often referred to as Tobler's "First law of geography" states that "everything is related to everything else, but near things are more related than distant things" (Cliff & Ord, 1981: 8 & 141; Anselin & Rey, 1991; Anselin, 1996: 112; Wong, 1996: 85). This first law, or heuristic, can more formally be described and measured as spatial autocorrelation. Unfortunately, spatial autocorrelation has not been examined, never mind addressed, in any ecological analysis of EMS demand. Moreover, it causes the violation of some of the assumptions of regression analysis used in previous work. GIS and loosely coupled spatial analysis software provide the opportunity to consider spatial autocorrelation in regression modelling without cumbersome programming or at least with relative ease.

3.1.2 Conceptual Framework

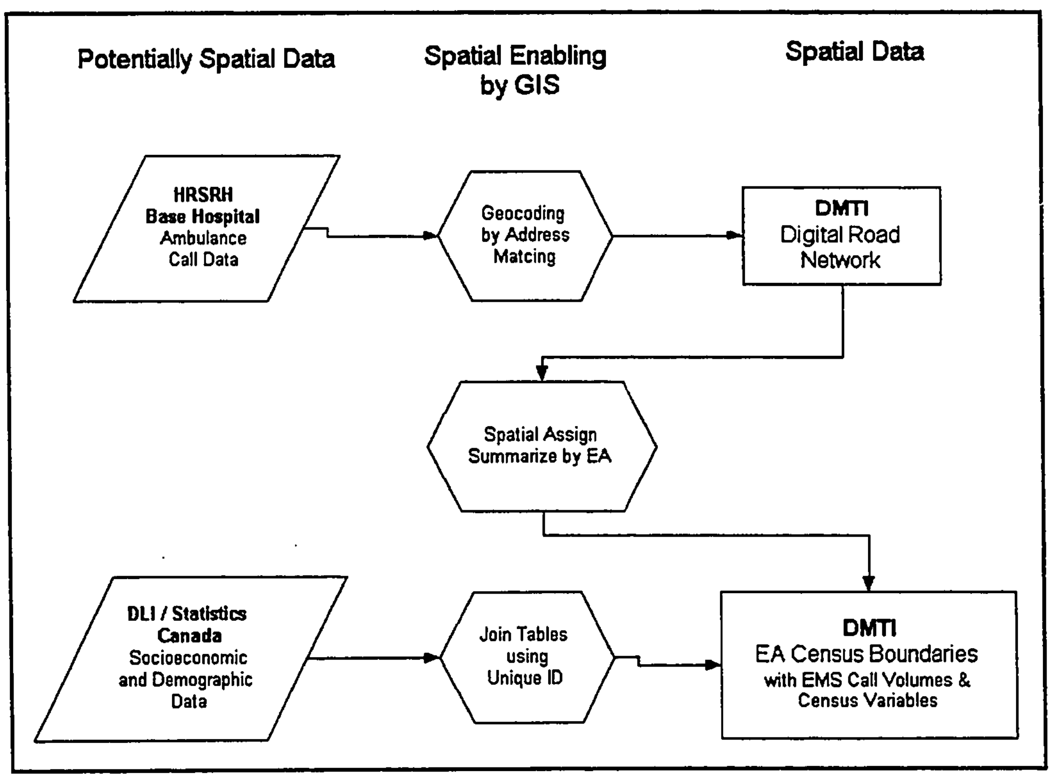
This research is concerned with ecological variates that govern the rate of demand for emergency medical services in a small city. Usually, demand is measured in number of EMS emergency calls per 1,000 persons per year. “Calls” or “runs” are common terms in the EMS industry that refer to an individual occurrence to which an EMS unit responded at a location on a particular date, and at a particular time. Calls or runs do not necessarily strictly include occurrences from which a patient was transported to hospital but, for the purpose of this research, a distinction is necessary due to the limitations of available data. Therefore, an “Event” in this work is an EMS call where a patient was taken to hospital. Several events can occur at one location and be generated by the same person at different times. The collection of all such records in a digital database is referred to as the “Events” file in this thesis.

The socio-economic data against which EMS events are compared are available from Statistics Canada at a variety of geographic levels. The smallest geographic unit at which census data are disseminated, the Enumeration Area, is used. Therefore, more appropriately, demand is measured by number of EMS emergency events per 1,000 persons per year per Enumeration Area.

In order to glean information from sizable EMS and census databases, a basic conceptual framework (Figure 3.1) was developed, whereby data are presumed “potentially spatial” until they are placed on their respective spatial frameworks. From there, a common framework is used so that spatial analysis and regression modelling can be performed.

For EMS events, a digital road network is used as a spatial reference or framework to locate each event’s address in space. For census data, the spatial framework is a digital file of Enumeration Area boundaries. Ecological variates or census data are joined to their respective spatial framework using a common unique identifier. Geolocated EMS events are then placed within the EA framework. "Placing" events and census values in their respective spatial frameworks is referred to as "spatially enabling" potentially spatial data. It is accomplished through GIS ArcView 3.2 so that spatial data can be analysed on a common spatial framework. Details of these procedures are described below in section titled "Spatially Enabling".

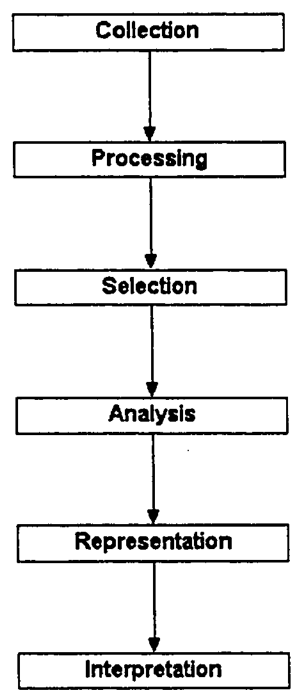
Figure 3.1 Conceptual Framework. The diagram illustrates how "potentially spatial" data from EMS events and census data are incorporated onto a common spatial framework for analysis and interpretation.



3.1.3 Research Approach

Within the conceptual framework presented above, a number of steps were undertaken to collect the data, process them, select candidate events and spatial units, analyse them, and represent the results in a variety of formats for final interpretation. Though Figure 3.2 is an over-simplification of the approach taken, it will hopefully serve to guide the reader through the process. Each step represented in this diagram will be expanded upon throughout the thesis, reflecting the complexities and intricacies of this approach.

Figure 3.2 Basic research approach to EMS data



It is hoped that EMS planners can apply the methods presented here to their own jurisdictions. Therefore, a reasonably simple environment composed of easily accessible software (Figure 3.3), based primarily on an ArcView 3.2® SPSS® and SpaceStat® is used.

The source data for EMS events were collected in MSAccess® and, because of personal familiarity with spreadsheets, are cleaned primarily in Excel®. The software environment could quite conceivably be simplified. Though SpaceStat has the functionality to perform regression and descriptive statistics, SPSS is preferred for these tasks because of significantly better graphics capabilities and personal familiarity. SpaceStat is used primarily for spatial analysis including lagging variables and measuring spatial statistics. An ArcView 3.2 extension acts as a loose coupling between programs. The extension allows preparation of datasets in ArcView for SpaceStat analysis and for representation of SpaceStat results in ArcView. Another SpaceStat product, known as Dynastat, is incorporated into ArcView as an extension. With it, Moran scatterplots and other graphs can easily be generated and dynamically linked to the geographic features in ArcView, providing a useful interactive exploratory tool. Movement of data from GIS to SPSS and database and spreadsheet programs is accomplished through use of common data formats.

Though the ultimate goal is to interpret collected data, the steps presented to that end are not as direct and linear as the basic research framework suggests. Figure 3.4 illustrates the research approach within the software environment used. This diagram is used to illustrate the complex and iterative data trajectory from collection to interpretation often encountered by the geographer. For example, before selecting relevant cases, data must first be cleaned. However, realistically, cleaning cannot occur without analysing and representing the data, both aspatially (manually checking for valid addresses), and spatially (visualizing the results of geocoding). Hence the bidirectional and circumventing arrows between the basic research steps in the diagram.

Figure 3.3 Software environment used for this thesis.

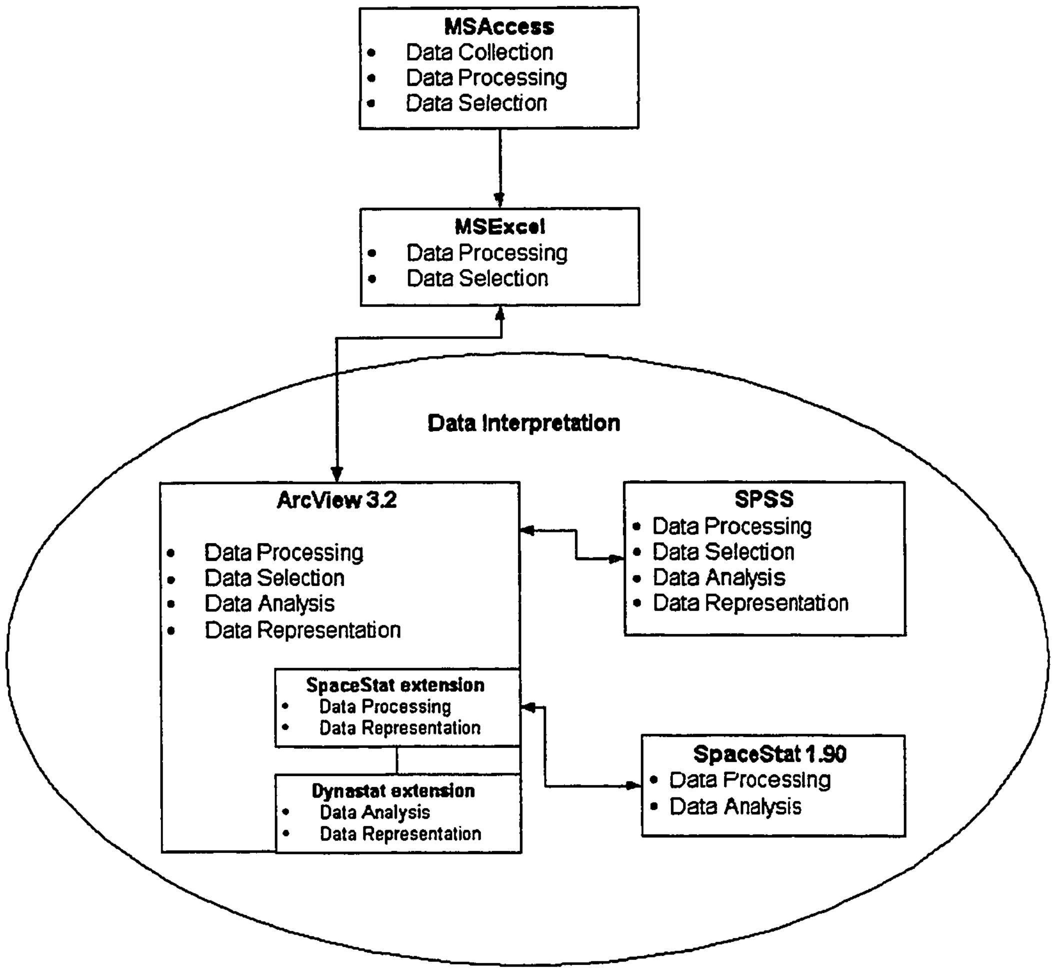
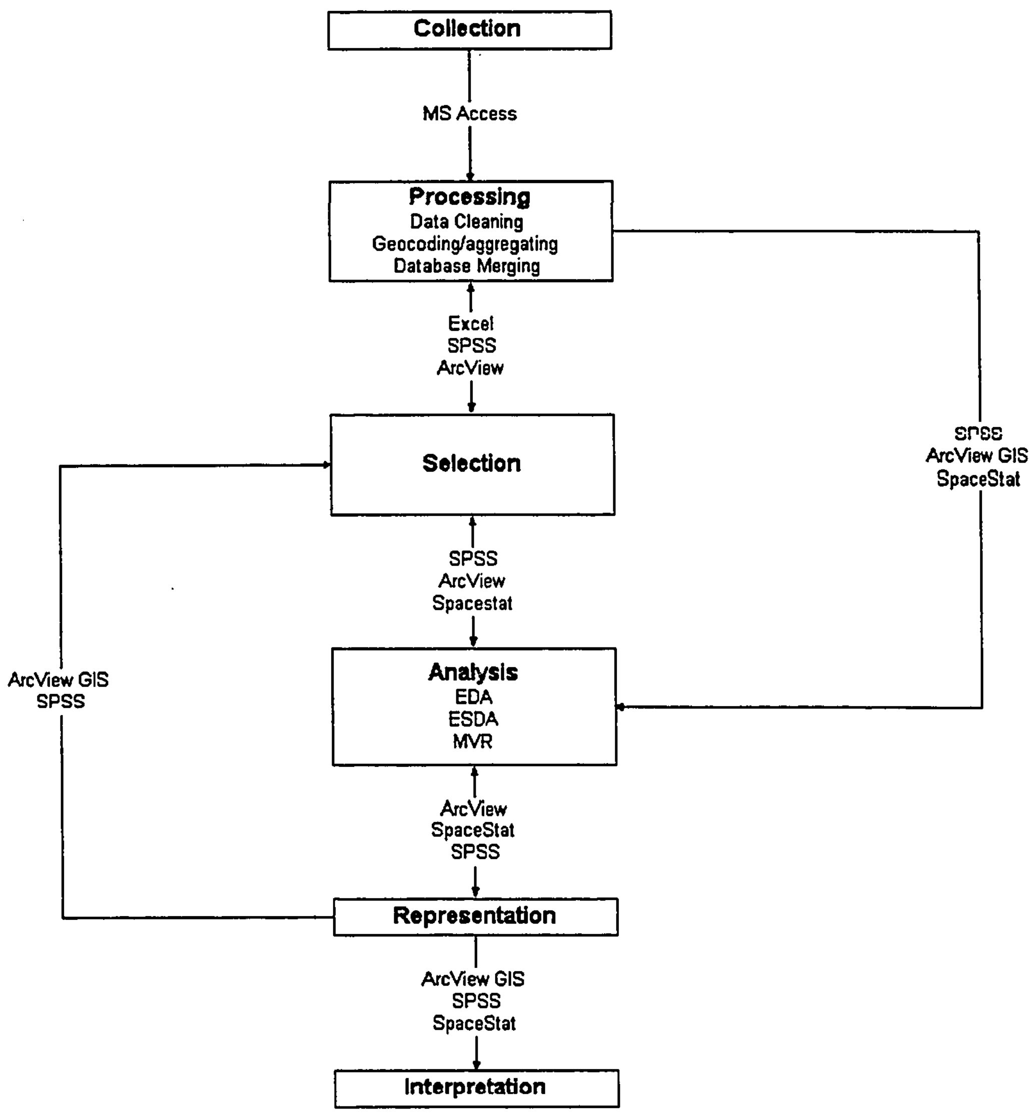


Figure 3.4 Detailed research framework. This diagram more precisely depicts the data trajectory from collection to interpretation.



Part A

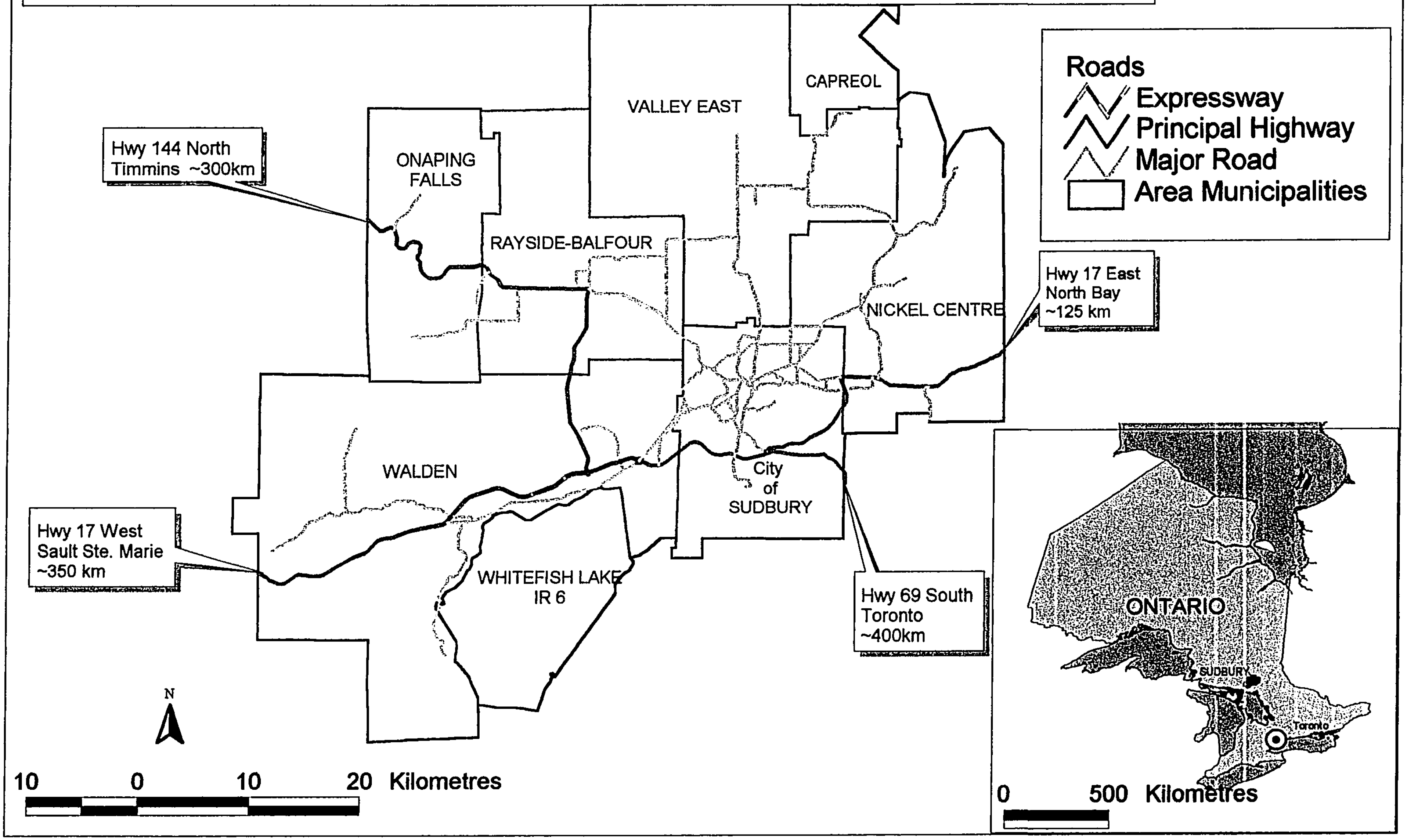
3.2 The Study Site

This part describes why the particular study site was selected. Prior to the Service Realignment Act, the now City of Greater Sudbury was nearly identical to the former Regional Municipality of Sudbury. The former Region was a both a Census Division and a Census Metropolitan Area[5] comprised of eight Census Subdivisions known as Area Municipalities (Table 3.1 and Figure 3.5). The most populated of these was the former City of Sudbury (hereinafter referred to as the City of Sudbury) which is the focus of this study. The 1996 census of Canada reported a population of 92,059 in an area of 263 square kilometres for the Census Subdivision of the City of Sudbury (Statistics Canada, 1998). The city was an important component of the Census Division of the Regional Municipality of Sudbury's more than 2,600 square kilometres and approximately 164,000 citizens (Statistics Canada, 1998). The City of Sudbury has long been the economic centre of Northeastern Ontario as well as its health care referral centre. Some of the health speciality institutions include a Lead Trauma Hospital, a Regional Cardiac Care Centre, and a Cancer Treatment Centre to name only a few (Sudbury Regional Development Corporation, 1998). Though there were four hospitals in the city at the time of the study, only the hospital with the Emergency Department is indicated on the maps in this thesis. The city is situated at N46° 30', W81° 00', approximately 400 kilometres north of Toronto, Ontario's provincial capital. Sudbury's built up area is elongated along two axes, one in an east-west direction and the other in a north-south direction.

---

[5] There is only a minor distinction between the CD and CMA in that the former includes one small town not yet incorporated into the CMA because of census administration reasons. The CMA does include a small First Nations territory that the CD does not.

Figure 3.5 The City of Sudbury within the Regional Municipality of Sudbury 1999



The axes of development stem from the Central Business District (CBD) extending towards two other main commercial districts, one in the northeast known as New Sudbury, and the other south of the downtown, known as the Four Corners area. The CBD, as well as these main commercial areas, are broadly outlined on Figure 3.6. They were abstracted from a land use map (Regional Municipality of Sudbury, 1985) and the Region's Official Plan (Regional Municipality of Sudbury, 1993). These generalized commercial nodes are used as reference points in the discussion of EMS demand, it is not the intention here to provide a detailed description and interpretation of land use and urban structure in Sudbury. The shape of the city is the result of many physical limitations to development typical of the Canadian Shield. These constraints include a very rocky and hilly topography and a number of lakes within the city boundary itself, the largest of which is Ramsey Lake, elongated in an easterly fashion from the centre of the city. These physical constraints can be considered when looking at interactions and autocorrelation between adjacent Enumeration Areas.

Sudbury's population has remained relatively stable over many years with only a slight decline between 1991 and 1996 (Table 3.1). This stability allows us to use census data collected at a slightly different time than our ambulance call data with less trepidation.

Another reason this city was selected as a study site is because of the author's familiarity with it. As a paramedic working in Sudbury for approximately 15 years, the current author acquired an intimate knowledge of the road network and considerable anecdotal experience with EMS demand there. Also, this author's undergraduate thesis involved the creation and use of a digital road network of the city of Sudbury to examine EMS station siting there.

Figure 3.6 Main Commercial Areas as Reference Points
City of Sudbury 1999

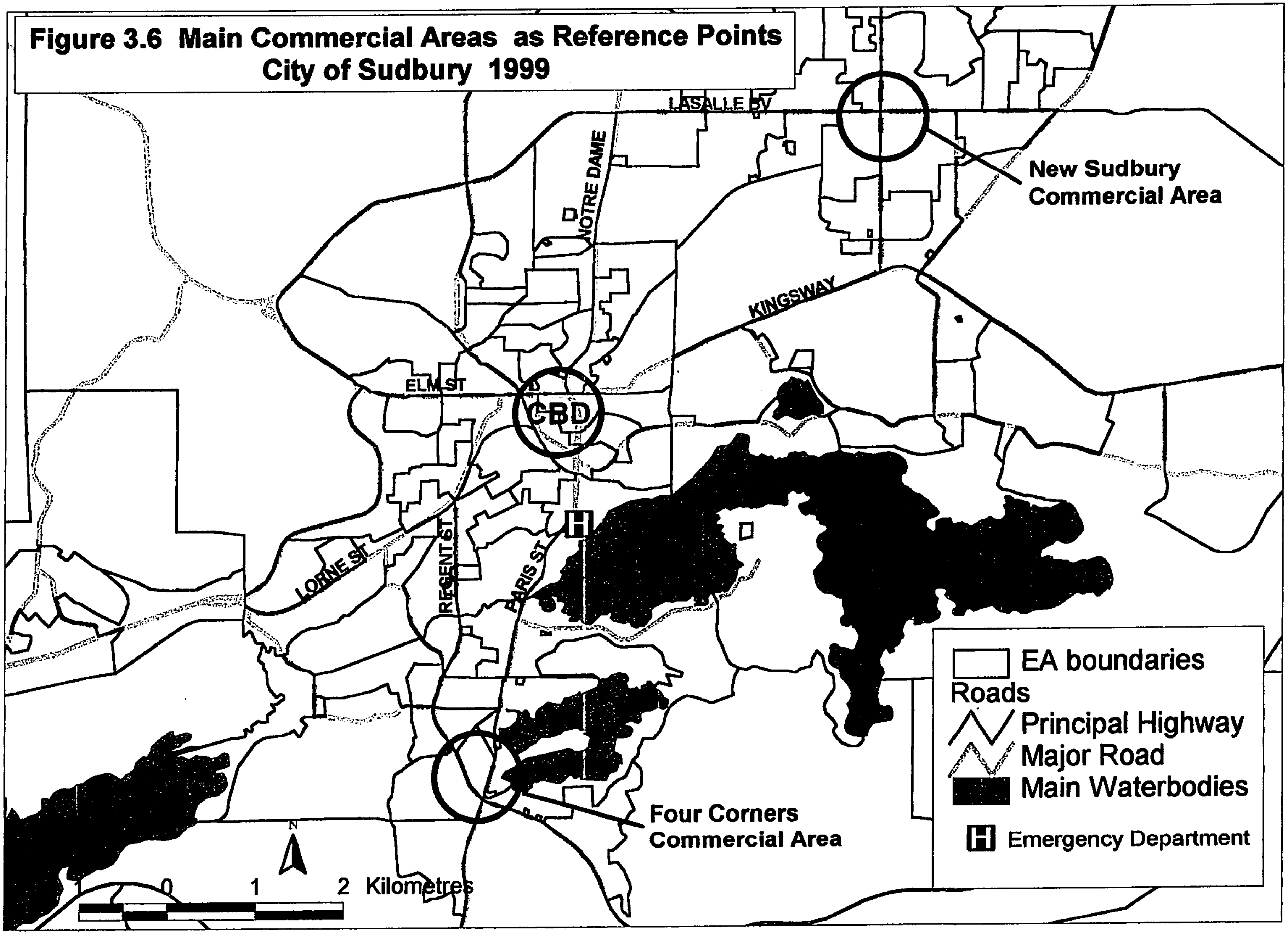


Table 3.1 Population profiles of the Sudbury CMA, CD, CSDs, in 1996.

Municipality Census Subdivision	Population (1996)	Population % Change (1991 -1996)	Area (Km^2)
Sudbury	92059	-0.9	262.73
Valley East	23537	7.3	518.03
Rayside-Balfour	16050	6.7	328.21
Nickel Centre	13017	5.6	378.36
Walden	10292	5	718.62
Onaping Falls	5277	-2.3	228.98
Whitefish Lake 6 (IR)	256	20.8	177.18
Total CMA	160488	1.8	2434.93
Capreol	3817	0.2	172.09
Total Census Division*	164049	1.8	2607.02

* does not include Whitefish Lake 6 IR
Source: Statistics Canada 1998.

Part B

3.3 The Data

This part of the chapter begins with a brief discussion of the issues that surround geographic data, i.e., Ecological Fallacy, the Modifiable Areal Unit Problem, Boundary Effects, and Spatial Autocorrelation.

Next, the section on data collection and acquisition introduces the datasets used in this thesis. Two datasets are used: the first a collection of EMS data, the second contains the ecological variables used in modelling EMS demand. Each dataset's attribute data and spatial framework are described.

The following section explains how data are processed. How data are spatially enabled using the address matching function in ArcView is described, as well as how EMS events are aggregated to the EA level. Also, issues of data accuracy and precision are identified, and the steps taken to clean the events file (ambulance calls) and the spatial reference file (road network) are explained. The placement of EMS events in the Enumeration Area framework are then validated.

The section on data selection outlines the criteria used to include EMS events and Enumeration Areas as candidates for analysis and interpretation.

The final section of part B is a return to data processing whereby the spatial model and the spatial weights used to operationalize that model are explained. These steps are necessary for spatial statistics to be calculated in SpaceStat.

3.3.1 Data Considerations in Geography

This section offers a brief discussion on a number of problems encountered in an ecological spatial analysis, i.e., Ecological Fallacy, the Modifiable Areal Unit Problem (MAUP), Boundary Effects, and Spatial Autocorrelation.

3.3.1.1 Ecological Fallacy

Ecological Fallacy or, as Fellers (2000:84) refers to it, "Aggregation Bias," defines the erroneous inferences from data collected at one scale to individuals or communities aggregated at other scales (Fellers, 2000: 96). Usually, the error is perpetuated by inferring group or aggregated characteristics down to the individual, however the reverse can be just as misleading. To avoid Ecological Fallacy, the discussion is simply limited to the smallest scale at which data are disseminated, i.e., the Enumeration Area. It is not the intent here to suggest that an *individual* person of a particular age group, social class, or educational level generates a specific need for EMS, but rather that geographic units, Enumeration Areas in this case, with identifiable social, economic, and demographic characteristics may be expected to create a certain degree of demand over the course of time. This scale of observation is also more practical in an EMS planning context; anticipating individual need is impractical.

3.3.1.2 Modifiable Areal Unit Problem (MAUP)

Any surface can be partitioned at any scale and in any number of ways. We know that this is problematic in geographical analysis because different partitioning of the same area can result in different analytical results (Griffith & Amrhein, 1997: 229; Wong, 1996; Cliff & Ord, 1981: 85; Fotheringham & Wong, 1991; Bailey & Gatrell, 1995: 37). The Modifiable Areal Unit Problem, as it is known, has two components, scale and aggregation.

The scale component of MAUP refers to the level of aggregation. For example, data collected at the Enumeration Area Level (approx. 250 households) could be aggregated to the Census Tract level (avg. 4,000 persons) that could in turn be aggregated to the Census Metropolitan Area level (population > 100,000). In a bivariate analysis, one clear relationship in terms of scale is that correlation coefficients increase as adjacent areas are aggregated to form larger areal units (Wong, 1996: 89; Haining, 1990: 48; Cliff & Ord, 1981:133). Sampling variance of regression parameters also increases with aggregation (Haining, 1990: 48; Fotheringham & Wong, 1991: 1026).

However, in the Multivariate environment the effects on statistical parameters do not follow any significant pattern (Fotheringham & Wong, 1991: 1042; Wong, 1996: 92), probably because the changes in variance and covariance cannot be anticipated (Fotheringham & Wong, 1991: 1030 and 1042). With non-zero spatial autocorrelation, multivariate parameters may or may not be significant at different levels of aggregation (Wong, 1996: 93; Anselin & Griffith, 1988; Cliff & Ord, 1981: 85), and the signs of those parameters, whether positive or negative, may also change (Wong, 1996: 93). Another significant problem encountered with scale is a dramatic increase in the multiple coefficient of determination (Fotheringham & Wong, 1991:1032, 1038, and 1041).

The aggregation, or zonation, component of the MAUP refers to the infinite number of ways an area can be partitioned into the same number of subareas (Wong, 1996: 84). The aggregation component may also produce a number of different results in multivariate analysis (Wong, 1996: 84), though the effects on parameters seem less sensitive to zoning than scale (Fotheringham & Wong, 1991). In Fotheringham and Wong (1991), there are notable cases where parameters change from not being statistically significant to being significant. Also problematic, the sign of the relationships switch from significantly positive to negative.

Clearly both the scale and zonation components of the MAUP can cause unpredictable fluctuations in Multivariate Regression results. To minimize the MAUP scale and zonation concerns, the analysis and discussion here is constrained to one scale and one set of zones, recognizing that the particular areal configuration used will in part determine statistical results of pattern or relationship analysis (Bailey & Gatrell, 1995: 256; Fotheringham & Wong, 1991). Researchers accept that they are presented with, and work with predefined sets of spatial units and their accompanying data (Bailey & Gatrell, 1995: 37 and 258; Fellers, 2000: 83).

3.3.1.3 Boundary Effects

Of course any geographic study has to have boundaries, but where does one literally draw the line? Neighbours that are excluded from the study area may have a spillover effect into our area of interest. However, the somewhat isolated nature of the northern city (Figure 3.5, Ch. 3) relieves some of this effect. If a similar study were conducted in southern Ontario, the effects of conurbations would likely be more pronounced. Furthermore, the unique geography (and historical settlement pattern) of the former Regional Municipality makes the City of Sudbury a nearly closed system because the area immediately surrounding the city is uninhabited mining property or very sparsely populated boreal shield.

Notwithstanding the mitigating geography described above, within the confines of the study area there is still a boundary consideration. Spatial units at the outer edges will naturally have fewer neighbours than those inside the study site (Bailey & Gatrell, 1995: 36-37; Boots & Tiefelsdorf, 2000). To this author's knowledge, SpaceStat has no function to deal with this, so we must simply be aware that local spatial statistics values on the edges are based on smaller samples.

3.3.1.4 Spatial Autocorrelation

Spatial autocorrelation if present, referred to as non-zero autocorrelation, can take on two forms, either positive or negative. The degree of autocorrelation, either positive or negative, is a function of the degree of similarity or dissimilarity between observations and their neighbours. Conditions where similar values are found in adjacent spatial units describes positive spatial autocorrelation. Negative spatial autocorrelation occurs when dissimilar areas are juxtaposed. No spatial autocorrelation, or zero autocorrelation, then refers to conditions where observations bear no significant similarity or dissimilarity with their neighbourhood. If Tobler's First Law of Geography holds true, then positive spatial autocorrelation would be detectable in most, if not all, geo-referenced datasets, as well as those that are not explicitly spatial, but were collected in space. The corollary would suggest we should expect to see negative autocorrelation less frequently in a geographic context.

Global measures of Spatial Autocorrelation summarize the spatial pattern for a whole area, whereas Local Indicators of Spatial Association (LISA) (Ord & Getis, 1995; Anselin, 1995; Getis & Ord, 1996; Boots & Tiefelsdorf, 2000; Zhang & Murayama, 2000) are used to measure the relationship of a particular subarea with its neighbours, the neighbourhood being defined by the researcher.

Unfortunately, it is well established that spatial autocorrelation in a dataset has significant effects on a variety of analytical parameter estimates (Griffith & Layne, 1999: 21; Cliff & Ord, 1981: 196). As Griffith and Amrhein (1997: 130) indicate "Non zero [significant positive or negative] autocorrelation tends to cause conventional confidence intervals for regression coefficients and the corresponding assessment of the significance of the covariates....to be invalid." Bailey and Gatrell (1995: 276) reaffirm this finding. Interestingly, Griffith and Amrhein (1997: 129) assert that positive spatial autocorrelation tends to make certain covariates appear to be more highly significant than they actually are and inflates the significance of the regression. While negative spatial autocorrelation will tend to make certain covariates less significant than they are and suppress the regression results (Griffith & Amrhein, 1997: 129).

Anselin and Rey (1991) recognize two forms of spatial autocorrelation, the first, an element of error terms, a nuisance which can simply be eliminated. Though Anselin and Rey (1991:113) use the term "nuisance" they agree with Cliff and Ord (1981: 197) that spatial autocorrelation alters the efficiency (variance) of OLS estimates, even though they remain unbiased. However, the interpretation of significance based on t-tests and measures of fit will be biased (Anselin and Rey 1991:113). Moreover, Anselin and Rey (1991) propose that "nuisance" spatial autocorrelation could stem from a mismatch of data and process. This may well be worth consideration in this ecological study. Since map patterns themselves generate a degree of latent spatial autocorrelation that increases with map complexity (Griffith, 2000: 155; Griffith & Layne, 1999:233), it seems possible that some nuisance autocorrelation would stem from the configuration of units in space.

The second form of spatial autocorrelation described by Anselin and Rey (1991:113) is termed "substantive," where the variable in question's values at one site are truly determined by values at other sites. In this case, they suggest that an autoregressive term should be included to create a *mixed regressive spatial autoregressive model* (Anselin & Rey, 1991:113).

Regardless of its form, spatial autocorrelation can also be a useful diagnostic in regression analysis in a geographic context. It can provide insight about the process under investigation and can suggest strategies for spatial modelling (Can, 1996). Spatial autocorrelation can of course be measured in all variables, however, most importantly, we should look for it in regression residuals (Cliff & Ord, 1981: 297).

Spatial autocorrelation in regression residuals can arise from, or imply:

- the presence of nonlinear relationships between dependent an independent variables (Cliff & Ord, 1981: 196) also (Griffith & Layne, 1999: 233),
- one or more missing regressor variables (Cliff & Ord, 1981: 196), possibly having conspicuous spatial autocorrelation itself (Griffith & Layne, 1999: 233),
- that the regressive model should have an autoregressive structure (Cliff & Ord, 1981: 196), a case of what Anselin and Rey (1991) refer to as *substantive* spatial autocorrelation.

To manage spatial autocorrelation, its presence and degree must first be determined. Moran's I and Geary's c are typically used to this end. These tests will be defined more precisely below.

Global and Local forces

Haining (1998: 30-31) and Bailey and Gatrell (1995: 32-35) illustrate spatial patterns as a function of broad, or global processes, and rough, or local processes. Bailey and Gatrell (1995: 32-35) define these effects more precisely as functions of mean, variance and covariance. Global or first order effects stem from a large scale or regional variation in the mean (Bailey & Gatrell, 1995: 33; Haining, 1998: 30); while local or second order effects stem from the spatial dependence between neighbouring sites (see Bailey and Gatrell, 1995 pp 32-35 for a useful explanation). Real life spatial patterns likely stem from a mix of first and second order processes (Bailey & Gatrell, 1995 :34).

Using both global and local statistics to examine the spatial pattern of EMS demand may imply the scale at which proposed explanatory variables are operating or suggest that there are spatial regimes that deserve special consideration. These tests for spatial autocorrelation are also useful in evaluating the independence of regression residuals (discussed further below), and therefore are useful regression diagnostic tools.

Measures of Spatial Autocorrelation

The most commonly used measures of spatial autocorrelation are Moran's I and Geary's c (Bailey & Gatrell, 1995: 280). Moran's I (Bailey & Gatrell, 1995: 270), for any variate (dependent, independent, residuals) $\boldsymbol{Z}$ can be represented as:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{\left(\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)}$$

and calculations for Moran's I at kth lag are represented by

$$I^{(k)} = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{(k)}\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{\left(\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2\right)\left(\sum\sum_{i\neq j} w_{ij}^{(k)}\right)}$$

while Geary's c is

$$c = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(z_i - z_j\right)^2}{2\left(\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)}$$

where z_i is the observed value of a random variable at a particular site, z_j the observed value at each of z_i's neighbours, and $\bar{z}$ is the mean of the random variable.

Moran's I is similar to the Pearson correlation coefficient in that it is a measure of covariance (Getis & Ord, 1996: 261), however it does not have a -1 to 1 range. Its expected value is $E(I) = -\frac{1}{n-1}$

and its range can extend beyond ±1 (Griffith & Layne, 1999: 12).

Geary's c has theoretical limits of 0 to 2, with 0 representing positive spatial autocorrelation, 1 equalling no spatial autocorrelation, and 2 representing negative autocorrelation. The test is based on squared differences between values (Getis & Ord, 1996: 261) and therefore would be less sensitive to skewed means.

Griffith & Layne (1999:15) refer to Moran's I as the more powerful measure of spatial autocorrelation. Nevertheless, analysis using Moran's I and Geary's c usually produces very similar results (Getis & Ord, 1996: 261). Naturally corroboration between the two statistics is desirable (Griffith & Layne, 1999: 15). Because of the more intuitive interpretation (negative values for negative autocorrelation and positive values for positive autocorrelation, approximately zero for no autocorrelation) of Moran's I, it is most frequently used and will be at the forefront of this study.

3.3.2 Data Collection and Acquisition

Two main datasets are used to explore the relationship between EMS demand and socioeconomic and demographic variables. These include: 1) the EMS dataset which includes EMS events located on a digital road network, and 2) the Ecological Variables dataset that is a mosaic of census enumeration areas for which a large number of socioeconomic and demographic variables are available. Both datasets are composed of potentially spatial attribute data and a spatial reference file. This section first describes potentially spatial data then the spatial frameworks of each dataset. The next section describes how these datasets were processed for further selection and analysis.

3.3.2.1 The EMS Dataset

Potentially spatial data: the EMS Events

The EMS Events file is a database of ambulance call records. Call information, including pick up location addresses were abstracted by the Sudbury Base Hospital Program from paper Ambulance Call Reports (ACR) into an MSAccess database.

Regulations made under the Ambulance Act (Ontario Ministry of Health and Long-Term Care, 2002) in Ontario require that each request for service or "Ambulance Call" for which a paramedic has made contact with a patient be recorded on an Ambulance Call Report (ACR- hard, or more recently, electronic copy). A copy of the ACR is distributed to the Base Hospital, which is responsible for medical oversight and shares the responsibility of quality control with EMS delivery agents. The Base Hospital in turn creates an electronic copy of the ACR for research and quality assurance purposes. The ACR captures scores of administrative and operational data, demographic, clinical history, care and outcome data. Most important for the geographer (but sadly least important for the clinician and health planners) some locational data are included. The "Pick Up Location" (the street address or intersection to which an ambulance responded) and the patient's mailing address are recorded on the ACR then entered into a database by the Base Hospital.

Access to these data was acquired through an agreement with the Hopital Régional Sudbury Regional Hospital after a presentation to their Research Ethics Committee to assure them that no patient identifiers would be used other than to aggregate the data, and that only aggregated data at the Enumeration Area level would be represented, discussed and disseminated in this thesis.

EMS Spatial Framework

To spatially enable the EMS data, individual events first need to be located using a spatial reference file. A digital road network with address ranges assigned to each road segment was used to match addresses in the events file. The spatial reference file is compatible with the census Enumeration Area framework. This file was purchased from DMTI™, a private vendor that distributes GIS products. It was anticipated that the private product would include a current road network. In the end, errors were found and required considerable efforts and field work to correct and verify; these will be discussed in the Results chapter.

3.3.2.2 The Ecological Variables dataset

Ecological variables against which EMS demand were modelled were primarily census data, however a few variables (population density, distance from hospital) were derived using GIS. This section describes the census variables and the Enumeration Area boundary file used as a spatial framework for analysis, representation and interpretation. The derived variables are described in the Data Processing section that follows.

Census Data

Statistics Canada's 1996 Census of Canada includes nearly 1,700 demographic and socioeconomic variables suitable for an ecological analysis. The data are available through the Data Liberation Initiative (DLI) at a variety of geographic scales including the Enumeration Area level, which is the finest scale at which census data are disseminated. The DLI is a subscription agreement between Statistics Canada and a consortium of Canadian universities whereby these variables are made available to academics at no cost. Data were downloaded from the TriUniversities Data Resources webpage (http://tdr.uoguelph.ca/cgi-bin/drc.cgi/other). These data were assigned to their respective census geographic boundary by simply using a unique identifier associated with each Enumeration Area making up the spatial framework.

Census Spatial Framework

A census boundary file was purchased from DMTI delineating Enumeration Areas in the Regional Municipality of Sudbury. Some minor geographic features were included in the files to provide context and facilitate mapping.

Enumeration Areas (EA) are the smallest geographic unit at which census socioeconomic and demographic data are disseminated (Statistics Canada, 1999: 210). They are designed to be as compact as possible and follow visible features. They encompass between a minimum of 125 dwellings in rural areas to 440 dwellings in large urban areas. Large apartment complexes can constitute an entire Enumeration Area (Statistics Canada, 1999: 210).

Because the Enumeration Area file and the road network used to locate EMS events are compatible, it is possible to overlay the two files and assign EMS events to their respective EAs. The details are described below in the "Aggregating EMS Events to EAs" subsection 3.3.3.1.

3.3.3 Data Processing

In order to perform spatial analysis, a dataset must reflect the relative positions in space of the events or objects in which we are interested. Unfortunately, EMS events, like many health data, are not typically conveniently placed within their respective geographic space (Albert et al., 2000b: 43). Nor are census data intrinsically part of digital boundary files. These data are referred to as "potentially spatial" because on their own, events data and census attributes are essentially aspatial until they are placed within a geographic context. Until then, the relative position and the space between events or objects cannot become meaningful.

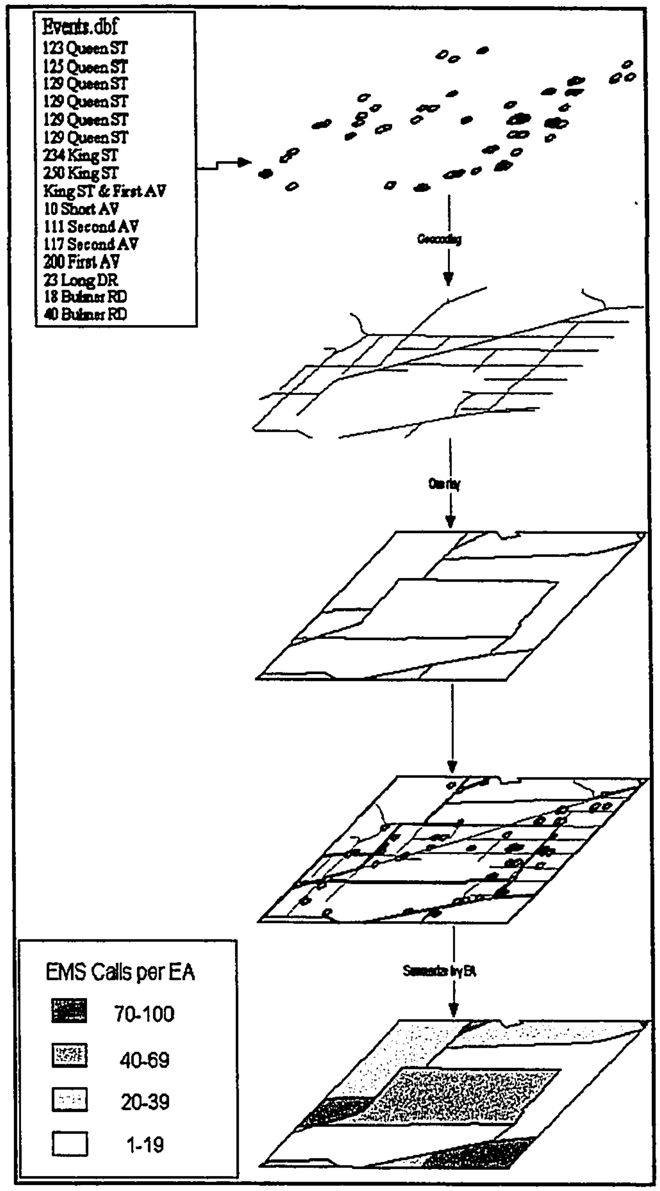
3.3.3.1 Spatially Enabling Potentially Spatial Data

To assign EMS call rates to individual EAs, two steps were necessary. First individual calls need to be geolocated, second those events needed to be "counted" in each EA. The first step involved using the GIS to address-match call locations to a road network. The second step spatially assigned the located calls to their respective EAs; then a tally of calls for each EA was computed. Both the address matching and aggregating processes are semi-automatic in the GIS environment. Figure 3.7 broadly illustrates this process. Each step is described below.

Address Matching

To geolocate addressed data, individual addresses are compared to street names and address ranges on a road network, then a point is "placed" along the appropriate road segment. Address matching requires an events data file with addresses and a spatial reference database. The pick-up addresses collected from the Base Hospital's database were used to create the events database. DMTI's roads file served as the reference file. It is a digital road network comprised of road segments, each of which is assigned the range of numbered addresses, as well as which side of the street, odd or even, numbers are found. Each road segment is populated with "left from" and "left to" and "right from" and "right to" address ranges.

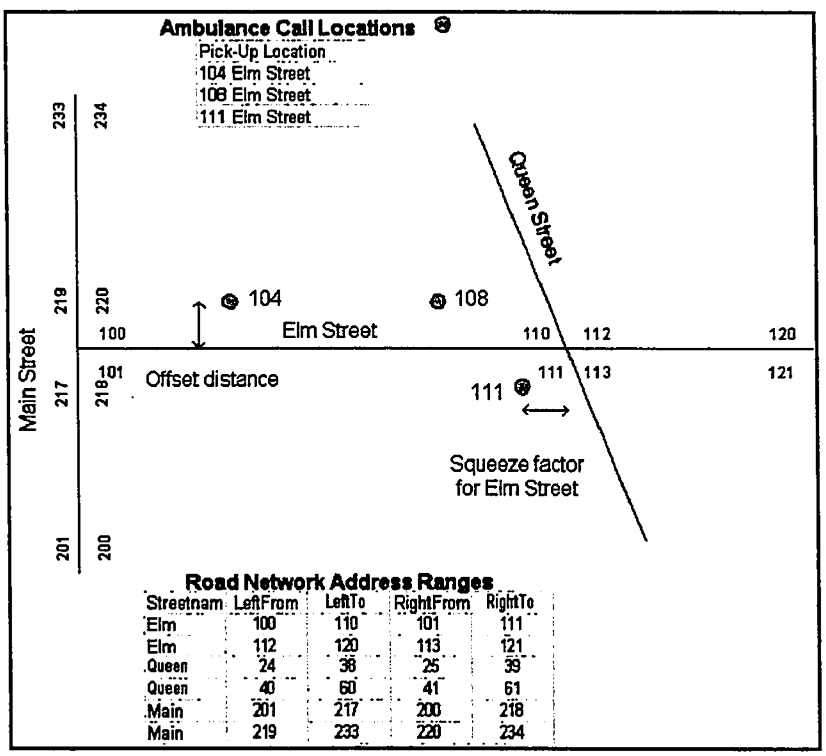
Figure 3.7 Address matching, geolocating and aggregating EMS events to Enumeration Areas



Once an address in an events database is found in a road segment's address range, the GIS places a point along the road segment, on the appropriate side of the road, by estimating the approximate location of the address number given the "from" and "to" range.

The user selects an offset distance to place the point away from the digitized road segment. This step is crucial to ensure that an address on a road that coincidently delineated an EA boundary is placed in its appropriate EA. To prevent address numbers that are at the extreme range of a road segment to be placed at its very end, ArcView applies a 2.5% "squeeze factor," therefore the distance from the terminus is a function of the segment's length. Otherwise, an address at the very end of a segment might be placed on another road segment that is perpendicular the first one's end (Figure 3.8).

Figure 3.8 Geolocating events by address matching using offset distance and squeeze factor.



Streetnam	LeftFrom	LeftTo	RightFrom	RightTo
Elm	100	110	101	111
Elm	112	120	113	121
Queen	24	38	25	39
Queen	40	60	41	61
Main	201	217	200	218
Main	219	233	220	234

ArcView's address matching function necessitates that certain requirements be met. First, a road network's address format must be clearly defined for it to be used as a spatial reference file. Second, the events database should have an address format that is compatible with the reference network. In this case "US Streets" format (number, name, type, direction) was used. Geocoding addresses is then done through a simple series of drop down menus and windows.

Matches are rated as "good match," "partial match" or "no match" depending on the preferences set by the user. Initially, the default spelling sensitivity and minimum match scores are used. These are used and adjusted iteratively to help identify problems with the address and reference databases. In order to reduce possible error, the final address matching process preferences are set to 100% essentially leaving us with matched or unmatched addresses.

Aggregating EMS events by EA

Once EMS events are individually located using a road network, they are aggregated within census enumeration areas using ArcView's Geoprocessing wizard. To do this, the Geoprocessing Wizard function "Spatial Assign" joins an EA's unique identifier (PRFEDEA) to each EMS event within which it is geolocated. Then using ArcView's Table module, the Events table is summarized by the PRFEDEA variable. The result is a small table consisting of 134 EAs as observations with a tally of events in each EA.

This summary table is then joined to a shapefile to assign to each geographic unit the sum of all calls within its boundaries. Counts are then converted to rates per 1,000 residents (RTSP1000) using the population values (POP96) associated with each EA.

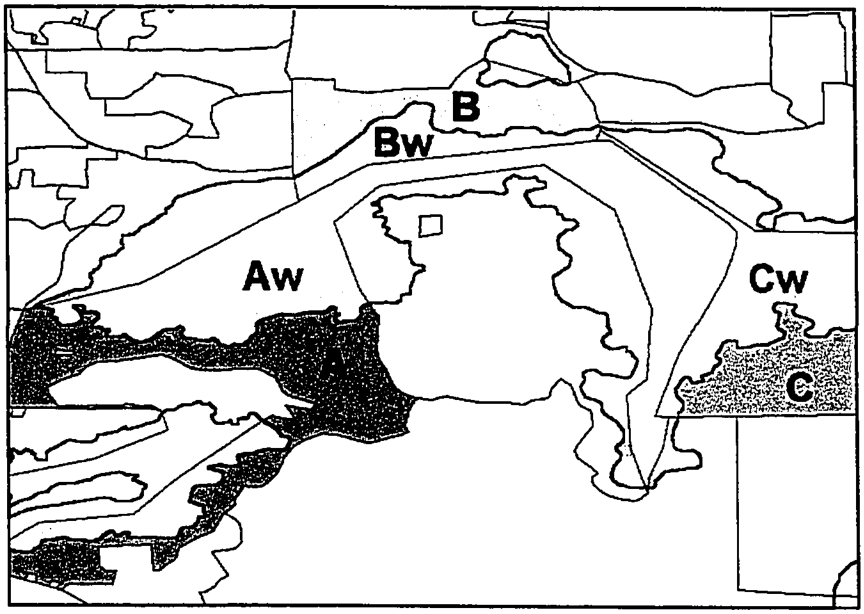
3.3.3.2 Derived Spatial Variables

Two variables to be used in the analysis needed to be derived using GIS. "Population Density" was simply calculated using the population (POP96) variable described below, and the square area in kilometres (SC_AREAR) of each EA. A second variable, "Distance," representing the approximate straight line distance from each EA centroid to the Emergency Department of the Sudbury Regional Hospital was calculated using ArcView's Geoprocessing Wizard. This last calculation is automatically generated by selecting the "Assign Data by Location (Spatial Join)" function.

3.3.3.3 Modifications to the EA spatial framework

Enumeration Areas are delineated by Statistics Canada such that they form a space exhaustive surface within the boundaries of Census Subdivisions. Boundaries between EAs are sometimes drawn in waterbodies, however, their extension into lakes and rivers creates false connectivity between some EAs. This is problematic given that the adjacency of EAs is used to define a spatial model (see below). Figure 3.9 illustrates how a degree of false connectivity between EAs can be introduced in the City of Sudbury. EA "B" is only adjacent to "A and "C" if these EAs' parts that are in Ramsey Lake are considered (Aw, Bw, and Cw). To mitigate the false connections between EAs, a "Waterbody" layer is used to "cookie cut" EAs that extend into lakes. In ArcView 3.2, this is easily accomplished using the Geoprocessing Wizard.

Figure 3.9 False adjacency between Enumeration Areas caused by borders extending into a waterbody.



3.3.3.4 Data Cleaning and Validation

As mentioned above, the geographer must contend not only with typical data errors but also errors that cause cases to be misrepresented within space.

Corrections to the EMS events address database are undertaken after a number of iterations of geocoding attempts. Obvious typographic errors in the "Pick Up Location" field of the events data are corrected and changes to suspect addresses are carefully verified using Canada Post's website, and maps with address ranges. Efforts are made to ensure that the same spelling was used both in the ambulance call database and the road network reference files. Apartment numbers are removed from the address (after determining whether or not they had been transposed with the street number) to improve the address matching scores.

The road network is also scrutinized as some outdated address ranges are found.

Because of recent changes in the road network, some streets in the core of the city were renamed to eliminate the direction designation North, South, East, or West. To update the digital road network, a Postal Code Conversion File is used to map Postal Code representative points; then address ranges for each postal code are collected from Canada Post's website and compared interactively against the address ranges of suspect digital road segments. Figure 3.10 illustrates the process of cleaning both the EMS events database and the road network reference file.

Before the final selection of candidate EMS events can be undertaken, it is necessary to ensure that both the events database and the road network file used to place the events into their respective EAs are reasonably error free. As described above, this is achieved by careful examination of the ambulance call data and the digital road network. A variety of sources are used to ensure that EMS calls are located in their proper EAs.

Additionally, a random sample of 100 successfully geocoded events is selected using SPSS. Then, using the Canada Post website, each sample event's postal code is recorded. Statistics Canada's Postal Code Conversion File (PCCF) includes the EA designation for each Postal Code. Therefore, each sample event's EA ascribed by the PCCF is also recorded manually. The EA assigned by the GIS aggregation procedure described above is then compared to the EA determined using the PCCF. Figure 3.11 details the procedure used to verify Enumeration Area assignments.

Figure 3.10 Events file and reference file cleaning and validation

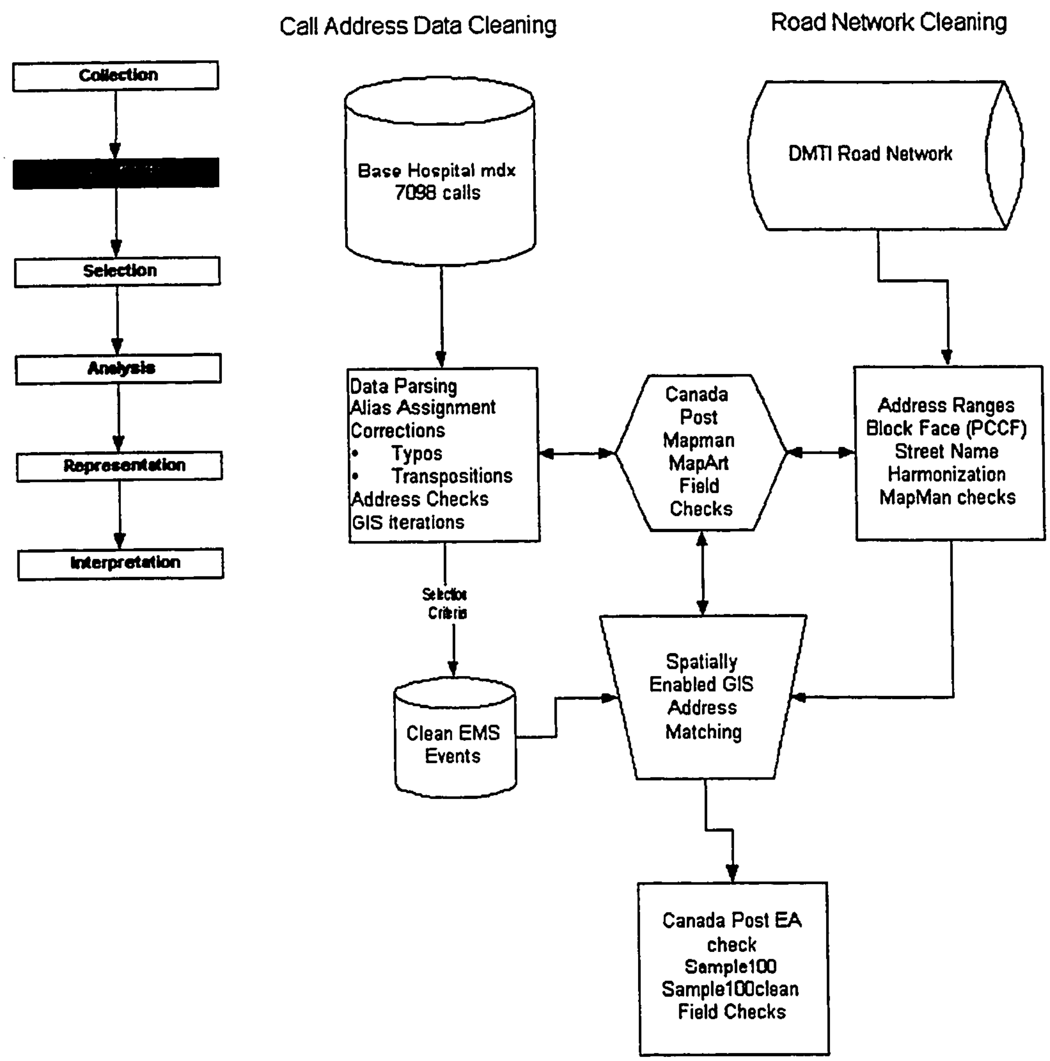
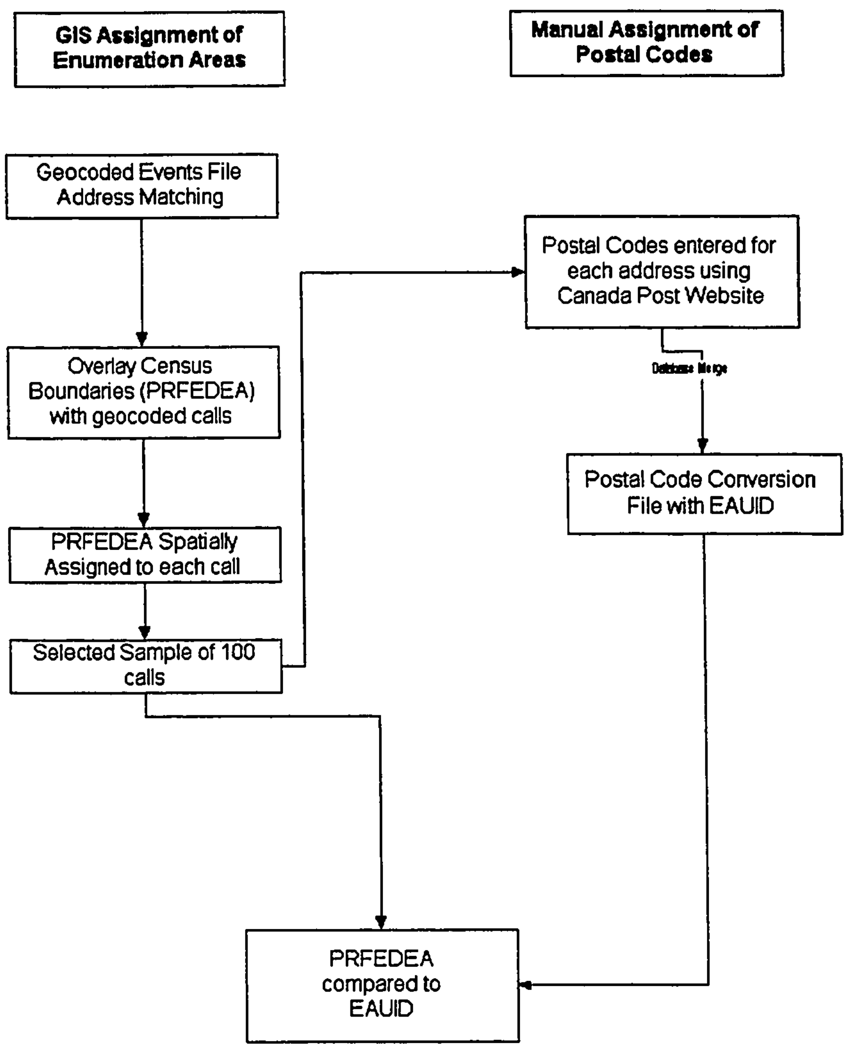


Figure 3.11 EA assignment verification by postal code.



3.3.4 Data Selection

This section describes how individual EMS events are selected, which census variables are used and how Enumeration Areas are chosen as candidates for analysis in this thesis. Ambulance calls are first chosen based on date, priority etc. and on whether they could be successfully geolocated. Census ecological variables are selected based on past research and the availability of variables. Enumeration Areas with suppressed census data are excluded from the study. Details of each selection process are described below. Figure 3.12 outlines the process of selecting both the EMS call data and the candidate Enumeration Areas for final analysis.

3.3.4.1 Selecting EMS events

Unfortunately, the Base Hospital does not enter all ambulance calls into their database; it only includes cases where paramedics determined the call warranted a Code 3,4,5, or 6[6] as a "Return Priority." Therefore, excluded from the Base Hospital records are a very large number of low priority code 1 and 2 returns, and "no patient carried" code 7's. This last category is sometimes referred to in the literature as "dry runs." These can include cases where no patient is found, or where the patient is not transported to hospital by ambulance. In the latter case, a patient may be treated by paramedics but refuse transport or chose another suitable means of transport to hospital. No such distinction is made in the Ontario EMS priority system (see Table 2.2).

---

[6]There are rare cases of codes 1 and 7 where a delegated medical act was performed either on site, or, in the case of code 1,possibly en-route, by the paramedics. However these calls were excluded.

The database provided by the Base Hospital included calls for the entire Regional Municipality of Sudbury for the years 1993 to1999. Initially, all available calls from 1999 were selected, as this was the first year the data entry program was in vigour. Furthermore, Base Hospital staff assured that this year was most complete (Gregoris, 2000). However, by exploring the data, it became evident that the database was limited. Ultimately, all calls in 1999 where paramedics indicated that the patient's mailing address was the same as the pick up location, that were dispatched code 4 and returned code 3,4,5, or 6, within the boundaries of the City of Sudbury were used. Figure 3.12 illustrates the selection of EMS events, however, details of the selection results are presented in chapter 4.

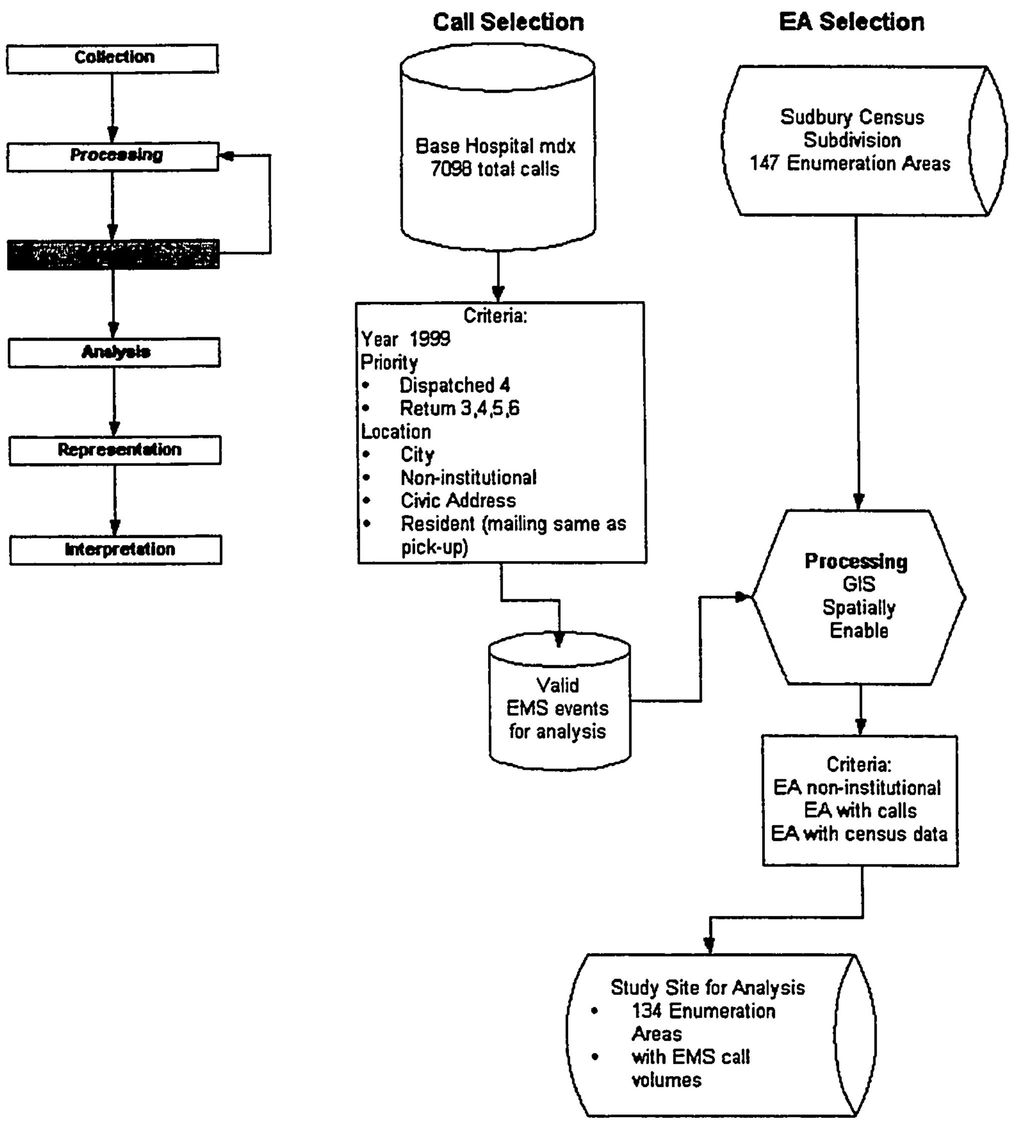
3.3.4.2 Selecting Ecological variables

Certainly, the literature review demonstrates that a number of sets of explanatory variables have been used to model EMS demand but naturally, these sets are governed by the availability of the variables. The 1996 quinquennial census of Canada offers nearly 1,700 demographic and socioeconomic variables to chose from. These data are categorized by population, family, household, and dwelling "universes." Depending on the variable, the data are collected from 100% of the population, or a 20% sample. To ensure confidentiality, Statistics Canada uses a number of strategies, two of which deserve particular attention. First, raw data are "randomly rounded" up or down to multiples of 5 for each EA. Statistics Canada assures that does not add significant error to the data (Statistics Canada 1999:358), however, percentage calculations can be slightly distorted, and for areas with very small counts, distortions may be greater.

At the onset of this study, only 1996 Census data were available. Though it is less than ideal to "compar[e] dataset that do not correspond in scale and time" (Gesler, 1986: 968), research is constrained by available data. As previously stated, census data used for this study were made available through Data Liberation Initiative via the Tri-University Group Data Resources web site (http://tdr.uoguelph.ca/cgi-bin/drc.cgi/other). Figure 3.12 outlines the EA selection process.

Of course an analysis of all available census variables would be unwieldy and unnecessary. A very large number of variates would lead to considerable risk of multicollinearity and difficulty in interpretation of results. Furthermore, there is sufficient understanding of EMS demand to suggest pertinent explanatory variables, making data mining techniques, such as Factor or Principal Components Analysis, seem redundant. Also, it is hoped that this research will provide a framework that Emergency Service Planners could use in analysing their own communities; therefore using a cumbersome method, such as PCA, with its abstract interpretations of factor loadings *etc.* is undesirable. To assure some parsimony in the model, 32 variables representing 4 major forces that have, in the past, been related to EMS demand are used. Where possible, variables similar to the two Canadian studies of EMS demand (Szplett, 1988; De Angelis, 1995) are included.

Figure 3.12 EMS Events and census data selection



The dimensions associated with EMS demand can be broadly listed as Demographic, Social, Economic, and Geographic-Urban Structure as follows:

- the Demographic dimension includes age and ethnicity variables,
- the Social dimension includes education and occupation, living arrangement and mobility,
- the Economic dimension includes employment, income and dwelling value, and,
- the Geographic-Urban Structure dimension includes dwelling type, population density, and distance from the hospital. Dwelling type is derived directly from census data, while distance from hospital and population density are calculated using ArcView GIS.

Some variables first need to be aggregated to create desired variates. For example, the census 5 year age clusters are summed to create total senior, adult, teen, and child counts for each EA. The totals are used to create a "percent" age category for each EA. Appendix 1 provides a detailed list of census data used to derive the variates in Table 3.2.

Table 3.2 List and brief definition of census and derived variables

Variable Name*	Definition**
DEMOGRAPHIC	
psnr	senior, person > 64 years of age
pchild	child < 15
pteen	teen 15 to 19
padlt	adult 20 to 64
pmdage	middle Age 50 to 64
pabor	aboriginal
pcitocan	citizenship other than Canadian
pnoholan	non-official home language
pvismin	visible minority
SOCIAL	
Education	
pedlo9	less than grade 9
pseced	completed secondary education
Occupation	
pcwtcol	white collar professions
Living arrangement and mobility	
plivealo	persons living alone
pmove1	movers 1 year
pmove5	movers 5 years
psepdiv	separated / divorced
plnprnt	lone parent families
pcsnr	unpaid care to seniors
ECONOMIC	
emppopr	employment to population ratio
uemrt	unemployment rate
a95hsinc	average 1995 household income
stdeinc	standard error of 1995 household income
agrrent	average gross rent
avaldwl	average value of dwelling
GEOGRAPHIC-URBAN STRUCTURE	
distance	distance from emergency hospital
pcar	people who use of car/truck to work
popdens	population density per square kilometre
anumphld	average number of persons per household
powned	owned dwelling
psuburb	detached and semi-detached dwellings
papt	apartment, detached duplex, row houses, apartment bldgs < 6 stories
ptow	apartment buildings > 5 stories

* "p" preceding a variable name indicates a derived percent calculation

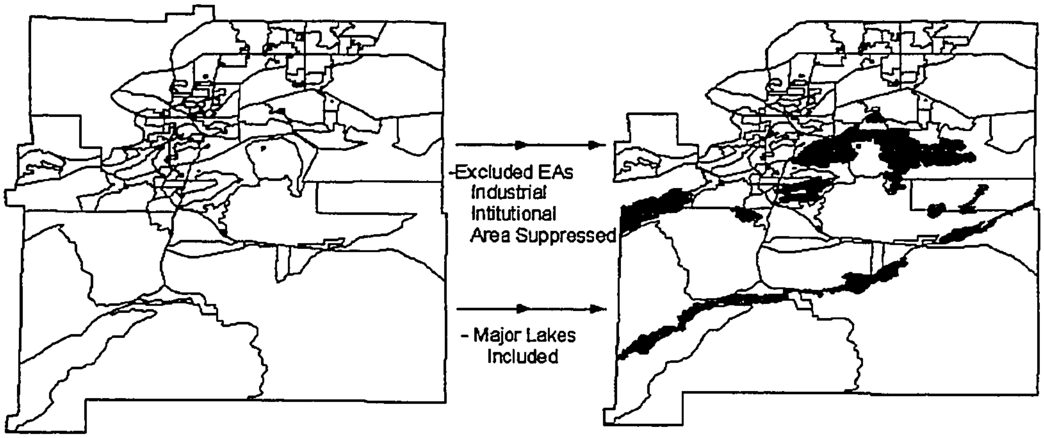
** see Appendix 1 for complete details of calculation of derived variables from raw census data

3.3.4.3 Selecting EAs

For a variety of reasons given by Statistics Canada, not every Enumeration Area includes all census data (Statistics Canada 1999: 358). Three main strategies used in disseminating census data are pertinent to this thesis. They account for some EAs' exclusion from the analysis because of insufficient census data. First, to ensure confidentiality, Statistics Canada adopts a strategy known as Area Suppression, whereby most data tabulated from populations less than 40 (in an EA) are suppressed (Statistics Canada 1999: 357). Second, Institutional EAs only register basic statistics including age, sex, marital status, and mother tongue. Institutional EAs include hospitals, nursing homes and seniors' residences. Third, some EAs, referred to as "Empty Places," can represent industrial or remote areas (Statistics Canada 1999: 255) with no resident population. These Empty Places are constructed by Statistics Canada to avoid gaps in the continuous EA surface and to allow for future population growth.

Area Suppressed EAs, Industrial EAs, and Institutional EAs are removed from the spatial framework using ArcView. ArcView uses "Shapefiles" with associated "Attributes Tables" to represent spatial objects, such as EA polygons, and their respective aspatial data. These shapefiles are easily edited to exclude EAs with missing data from further analysis and from playing part in operationalizing the spatial model presented below as Figure 3.13. The result is a study site composed of 134 EAs, reduced from the original 147.

Figure 3.13 Excluding EAs from the study site



3.3.4.4 Joining EMS and census datasets

In order to prepare a dataset for regression analysis in SPSS, the EA shapefile attributes data were exported to dBase format. These were then joined to the DLI data in SPSS by the "Merge" function using the common key variable PRFEDEA. The result is a dataset of EAs with their respective EMS events rates (from aggregating events by EA), and socioeconomic and demographic data from Statistics Canada.

3.3.5 Operationalizing the Spatial Model

As indicated from the beginning, the task of preparing the data is an iterative one. Though the basic research design suggests a linear trajectory from selection to analysis,

Figure 3.11 reminds that backtracking is necessary. This section describes a return to data processing that occurs after candidate EAs are selected for analysis. It also identifies how the remaining EAs are included in the operationalizing of a spatial model to be used in later spatial analysis.

How data are represented can govern how it is treated and analysed. To re-iterate, the rate of EMS emergency calls within Enumeration Areas is the variable of interest. Bailey and Gatrell (1995: 248) provide a good example of operationalizing a spatial model for such areas in space. For the most part, the spatial model in this study is operationalized using their approach.

Suppose our study site or region, the Census Subdivision of the City of Sudbury, composed of 134 Enumeration Areas (reduced from original 147 as explained above), is $\mathcal{R}_s$. If all sub-areas (EAs) are defined by $\mathcal{A}$, then a individual data site is $\mathcal{A}_i$. $\boldsymbol{Z}(\mathcal{A}_i)$ is a random variable for a particular data site. As Bailey and Gatrell (1995: 248) suggest, we can simplify the notation of $\boldsymbol{Z}(\mathcal{A}_i)$ to $\boldsymbol{Z}_i$. $\boldsymbol{Z}_i$ then is a set of all possible z, and z_i is an observed value.

Thus

$\{\boldsymbol{Z}(\mathcal{A}): \mathcal{A} \in \boldsymbol{R}\}$ represents a *spatial stochastic process*, or data set where a random variable is defined at each sub-area.

If we look at a snapshot in time

$z(\mathcal{A}): \mathcal{A} \in \boldsymbol{R}$

is a single realization at one point in time, which can be represented graphically as a map; essentially, it is a sample of 1.

3.3.5.1 Spatial weights

We must also operationalize neighbourhoods if we want to measure the relationship between z_i and its neighbours z_j. A spatial weights matrix $\boldsymbol{W}$ is generated in SpaceStat based on this relationship. A number of parameters including spatial autocorrelation statistics are computed in SpaceStat based on the spatial weights.

The simplest form of spatial weights is used here, a binary first order neighbourhood based on simple queen's case adjacency where:

$$w_{ij} = \begin{cases} 1 & \text{if } A_i \text{ and } A_j \text{ are contiguous} \\ 0 & \text{otherwise} \end{cases}$$

Weighting schemes can be much more complex. For example, they can include length of common border, distance between polygon centres *etc.*(Cliff & Ord, 1981: 17-19; Bailey & Gatrell, 1995: 262). But, as Cliff and Ord (1981:17) say, "[c]are must be used in the choice of weights if spurious correlations are to be avoided."

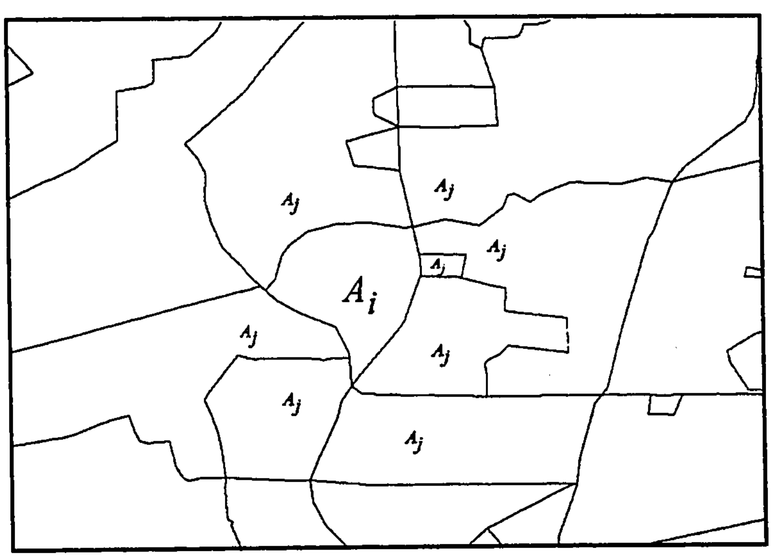
In the case of distance between polygon centres, Shen (1994: 170) highlights that irregular sized and shaped polygons can cause bias. He explains that if a large enough distance is not selected, large polygons could be excluded from calculations or found to have no neighbours (Shen, 1994).

Moreover, using a complex weight scheme does not necessarily provide dramatically different results than a simple one. Can (1996) used a simple binary contiguity scheme which she compared to a much more sophisticated generalized weights definition based on both distance and length of common boundary. Measures for a simple scheme were 0.430854 and 0.558308 for Moran's I and Geary's c respectively; while for the complex scheme ,Moran's I was 0.507094 and Geary's c 0.484108. Both proximity schemes exposed considerable spatial autocorrelation with no impressive differences between them for either statistic.

One final reason for using a simple first order adjacency is that Boots and Tiefelsdorf (2000: 344) have demonstrated that "the standardized form of I can be used safely for evaluating the null hypothesis of no spatial autocorrelation" for large enough samples using a first order lag. However, for $k > 1$ the normal approximation of I no longer holds true (Boots and Tiefelsdorf 2000:344). Therefore, the significance of the test would be in question.

Figure 3.14 is a diagram of all A_j first order, or lag 1, neighbours to A_i as defined above. Collectively, these will be referred to as an EA's first order neighbourhood throughout this thesis.

Figure 3.14 First order, lag 1, neighbourhood used to calculate spatial weights.



3.3.5.2 Using ArcView and SpaceStat to operationalize the data

With the spatial model defined above, it is now possible to use SpaceStat to operationalize the spatial model. The developers of SpaceStat provide an extension for ArcView that generates datasets for use in SpaceStat. The extension's drop-down menu permits a queen's case sparse contiguity matrix to be generated from the EA shapefile. The sparse matrix is saved as a *.GAL file that includes a list of each polygon (identified here by their PRFEDEA unique identifier) followed by its number of neighbours, then a list of each of those neighbours (Figure 3.15). This file is used in SpaceStat to generate a binary contiguity matrix used in SpaceStat to compute spatial statistics and spatially weighted (lagged) variables.

Figure 3.15. The .GAL file produced by the SpaceStat extension in ArcView. The file has 134 cases, the first EA #35082301 has 4 neighbours, 35082305, 35082306, 35082270, and 35082304.

```
134
 35082301 4
 35082305 35082306 35082270 35082304
 35082302 4
 35082303 35082266 35082270 35082306
 35082270 3
 35082306 35082301 35082302
 35082263 8
 35082167 35082265 ....
```

Data tables are also exported for analysis in SpaceStat. Outputs from SpaceStat can in turn be imported into the GIS for visualization and representation of spatial analysis results. This software environment was introduced in section 3.0.3 and illustrated using Figure 3.3.

Part C

3.4 Analytical Methods

This part of the chapter describes the analytical methods used including EDA, ESDA, and regression modelling. The first section describes how the data are subjected to Exploratory Data Analysis (EDA) and Exploratory Spatial Data Analysis (ESDA) using a classical statistical software SPSS and the spatial analytical software program SpaceStat.

The second section posits two regression models. The first is an aspatial statistical model, specifically an Ordinary Least Squares (OLS) regression. A brief discussion includes how geographic data often violate one or more of the data assumptions compulsory to OLS, and the significance of not dealing with these transgressions. Therefore, a second regression model is proposed that considers the spatial characteristics of the data.

3.4.1 Exploratory Data Analysis (EDA)

Some degree of informal data exploration is naturally applied when data are first acquired, both before and during the processing and selection steps described above. Examining and sorting cases is used extensively to determine which cases might be candidates for analysis, where obvious errors exist and what data massaging should be considered. This section on EDA, however, deals more specifically with methods used to determine basic statistical properties of both the dependent variable RTSP1000 and the ecological variables used to model EMS demand with a regression equation. RTSP1000 is herein referred to as the response variable, the ecological variables as the explanatory variables.

Histograms are used to examine the frequency distribution of each variable. Descriptive statistics are used to measure centrality (mean, median, and mode) and dispersion (standard deviation) of the response and explanatory variables.

Correlations matrices are generated for EMS call rates and each explanatory variable. (Norušis, 1990) notes that the behaviour and relationships of variables can be quite different in a multivariate versus bivariate environment, therefore the bivariate matrix is not used deterministically, but rather as an exploratory tool and to provide some clues as to possible problems with multicollinearity between explanatory variables.

Scatterplots are also generated to examine the shape (linearity) of the relationships between the response variables and each explanatory variable. These can also give a sense of the strength of relationships between variables.

3.4.2 Exploratory Spatial Data Analysis (ESDA)

Gatrell and Bailey (1996:843-44) describe three levels of spatial data analysis (SDA): visualization, exploratory methods, and modelling. They suggest that the first two levels of SDA may be sufficient, but if the researcher wants to estimate the relationship between events and ecological variates, it is not sufficient to use classical methods that assume independence (Gatrell & Bailey, 1996:844). They also suggest use of exploratory techniques to identify the spatial dependent nature of a phenomenon then incorporate it into spatial models (Gatrell & Bailey, 1996: 844). These techniques are referred to as

Exploratory Spatial Data Analysis (ESDA). This then is the purpose of undertaking the following spatial analyses.

Bailey and Gatrell (1995: 23) admit that the distinction between visualization and EDA is blurry, artificial and really just a matter of sophistication. Maps can be viewed by some as a rather unsophisticated form of ESDA. Nevertheless, they are indispensable for the spatial analyst or the geographer. Maps were generated to visualize the spatial distribution of the response variable. They were created using ArcView GIS. Extreme values of RTSP1000, as identified by SPSS, are highlighted and discussed.

More sophisticated ESDA are available using SpaceStat. Using the ArcView extension described above, Moran's I and Geary's c statistics were generated for each explanatory variable.

These measures of spatial autocorrelation can be calculated over several lags. SpaceStat is used to create up to 5th order lag values of Moran's I for each variable. These results are imported into MSExcel to generate graphic representation of the spatial correlogram. Gesler states that "correlograms provide an indication of the scale of at which spatial patterning is operating"(Gesler, 1986: 967). Local Moran I statistics are also generated for the response variable using SpaceStat. The results are mapped in ArcView to identify spatial outliers.

The Moran Scatterplot is a useful tool found in SpaceStat that compares the relationship between a value of a random variable at a specific location to the values of the same variable in its neighbourhood, as defined by the researcher. With it, the association between any given value z_i and weighted average wz_i of its neighbourhood (defined above) can be plotted (Anselin, 1996). The Moran Scatterplot has four quadrants. The upper right and lower left represent cases that are surrounded by similar values, therefore, positive spatial association. The upper right represents high values surrounded by high values, while the lower left has low values surrounded by low values (Anselin, 1996). The upper left and lower right quadrants contain observations surrounded by dissimilar values (Anselin, 1996). The upper left has low values surrounded by high values, and lower right holds high values surrounded by low values.

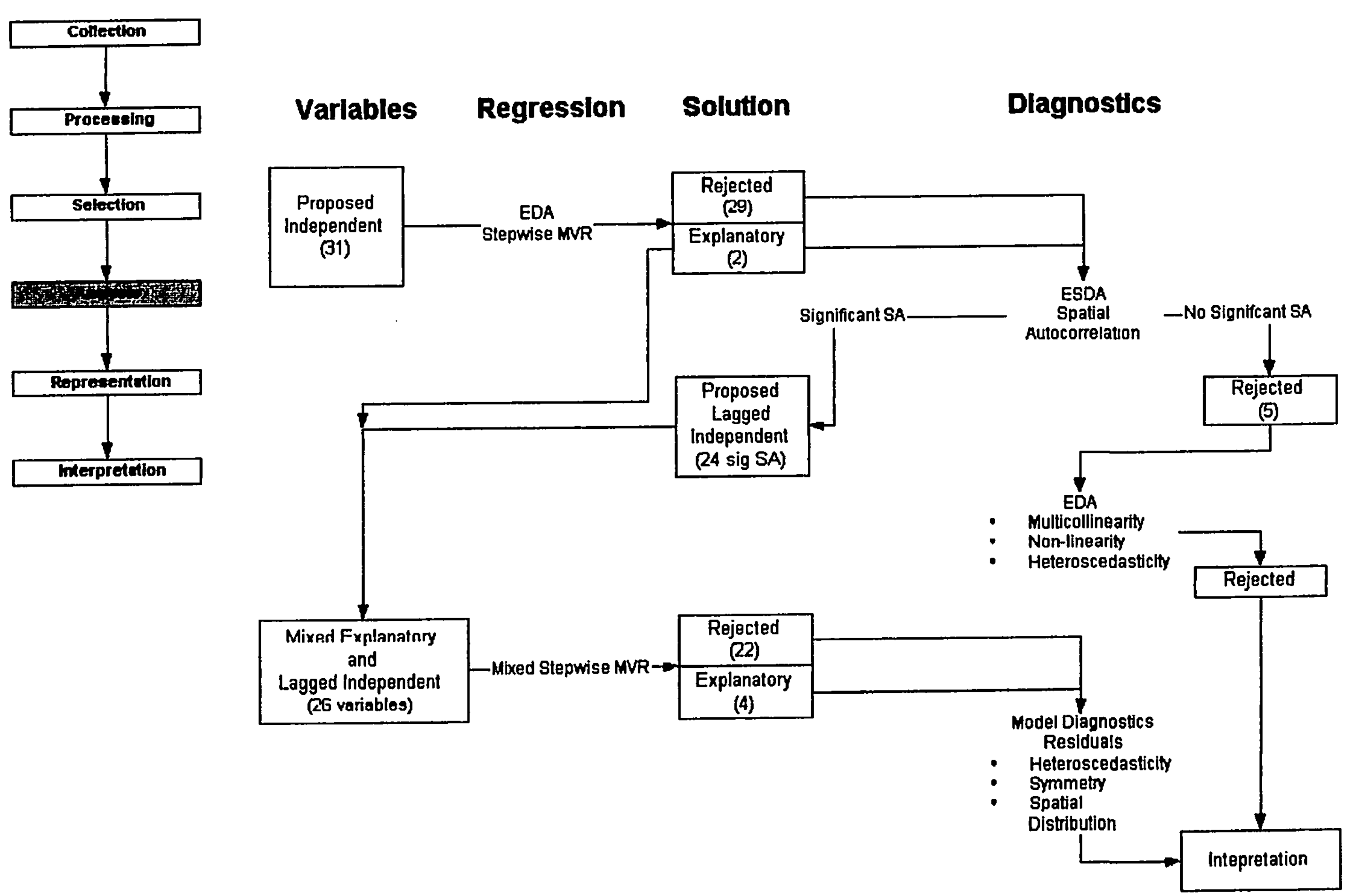
The quadrants assigned to each observation are automatically coded 1 through 4 by SpaceStat and saved to a dataset that is imported into ArcView. In ArcView, a map showing each EA's position in the Moran Scatterplot, is easily generated.

3.4.3 Regression Modelling

Two regression models are proposed in this thesis. The first is essentially used as a form of EDA and is referred to as the Exploratory Multivariate Regression (EMVR). The second is referred to as the Mixed Multivariate Regression (MMVR). Both use the stepwise method of ordinary least squares regression in SPSS, and both include RTSP1000 as the response or dependent variable. They differ in terms of which variables are proposed as explanatory variables. The first model uses only the variables listed in Table 3.2. The second model includes spatially lagged variables created with SpaceStat. This section explains both models. Figure 3.16 illustrates the modelling approach used to carry the research from classical aspatial regression modelling to regression modelling that contends with spatial effects in the explanatory variables.

The Coefficient of Determination (R^2) is used as a measure of the explanatory power of each model. Partial Beta (regression) coefficients are used to examine the relative impact of each explanatory variable. Analysis of regression residuals is performed to identify any violations of assumptions of regression modelling.

Figure 3.16 Exploratory and Mixed Multivariate Regression approach.



3.4.2.1 The Exploratory Multivariate Regression - An Aspatial Statistical Model

This section deals with the non-spatial statistical modelling of the relationship between demographic, socioeconomic, urban geographic forces and the demand for EMS in the selected study area. The statistical model needs to be acceptable in three respects 1) parsimony, 2) explanatory power and 3) specification. Therefore, the aim is to have a model that 1) is not too cumbersome, 2) has reasonable explanatory power, and 3) mitigates violations of assumptions while missing no critical explanatory variables.

As the literature review demonstrated, much work has examined the relationship between demand for EMS and a number of socioeconomic forces. Most of the studies have used simple linear regression (Aldrich et al., 1971; Kvålseth & Deems, 1979; Gibson, 1971; Szplett, 1988; De Angelis, 1995; Schuman et al., 1977) and in many cases have had significant and compelling explanatory power. Ordinary Least Squares (OLS) regression is utilized in this thesis as it is relatively simple to use and its limitations reasonably well understood. The OLS method calculates or draws a line between all data points such that the sum of the squared vertical distances from the data points to the regression line is minimized (Norušis, 1990: 188). More specifically, a stepwise regression method is used here; a method that enters and eliminates independent variables iteratively or in steps (Norušis, 1990: 273). Each variable's contribution to the model is reassessed as a new variable is entered into the equation (Norušis, 1990: 273). Yeates (1974:120-121) states that a stepwise method can be used as a "search" procedure, and is only acceptable where a "variety of models are theoretically tenable for a particular problem". Given that the literature review provided a number of potential models using a variety of variables, the stepwise approach seems appropriate. A model using specific dependent variables (a rather long list is used) is not posited beforehand which results in the regression (used iteratively) as form of Multivariate Exploratory Data Analysis.

First, the Multiple Linear Regression equation is defined as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots . \beta_n X_{ni} + e_i$$

where:

- Y_i is the Dependent or Response[7] variable, in this case the Number of Emergency Calls/1000 persons in each Enumeration Area (RTSP1000)
- β_0 is a constant
- β_{ni} are the partial regression coefficients for each X_{ni}
- X is a random independent, or explanatory variable from a pool of demographic and socioeconomic data
- n is the number of Independent variables
- e is an error term, that part of the variation in Y that cannot be explained by the model. This residual error is not trivial as its analysis in terms of distributional behaviour and pattern are critical in evaluating the correctness of the specification of the model presented here. They will at least hint at what needs to be addressed both in the aspatial and spatial model.

Informally then, it is proposed that:

Number of Emergency Calls/1000 persons in each EA = Constant + Demographic and Socioeconomic Forces ± Error

[7] The reader is reminded that the term "Response" in the EMS industry and literature often refers to "Response Time" The present research does not include any response time analysis or time interval analysis and for the nature of this work the term Response is used in the statistical context where it refers to a variable that is affected by one or a number of Explanatory variables.

Assumptions in Multivariate Linear Regression and Consequences of Violations

Now that the variates to be used in the regression have been defined, we must consider the properties of each variate and the relationships between the dependent and the independent variables. Though OLS may be considered robust by some (Szplett, 1988), it is based on a number of important assumptions (Norušis, 1990:247) namely:

1. linearity, whereby a linear relationship between the response and explanatory variables must exist;
2. normality, whereby each variable is normally distributed;
3. independence between observations; and
4. homoscedasticity, meaning equal variance of any one explanatory variable at any value of the dependent variable.

Typically one deals with violations of these assumptions by transformations. Fortunately, these transformations often have the serendipitous effect of mitigating more than one violation at a time; for example a logarithm transformation could contend with heteroscedasticity, linearity, and normality at same time (Griffith & Layne, 1999). But Griffith & Amrhein (1997: 19) and Griffith & Layne (1999: 75) also caution that transforming a variable just for the sake of fitting a model is not always a meaningful strategy. Non-linearity can be corrected by a power transformation of the dependent or independent variable or both if required. However this can lead to difficult interpretation as it is less intuitive to consider a value in its transformed, e.g., logarithm, form.

As was discussed in the section on spatial autocorrelation (section 3.3.1.4) the very nature of geographic data assures us that at least the assumptions of independence and homoscedasticity will be violated. Therefore, before considering any transformations each variable is first examined in terms of its spatial properties. The relative impact of violating these assumptions are clearly outlined in Griffith and Layne (1999:71). In order of lesser to greater impact on the regression parameters, they are: normality, independence and homoscedasticity.

3.4.2.2 The Mixed Multivariate Regression - An Aspatial/Spatial Model

To incorporate the spatial effects of explanatory variables into a regression model, the degree and significance of spatial autocorrelation for each variable is determined by ESDA described above. Those variates that demonstrate significant spatial autocorrelation are transformed using the spatial weights matrix described. For each EA, a new value for each variable is calculated using the weighted average of the value for that variable in its neighbourhood. This new variable is referred to as the lagged variable. It is identified by prefixing a variable name with "W_".

The Mixed Multivariate Regression (MMVR) simply consists of a stepwise OLS regression analysis that incorporates spatially lagged variables, but only those that demonstrated significant spatial autocorrelation in ESDA. However, it is important to note that any aspatial variables that provide a solution to the first regression are also included in the MMVR, hence the term "mixed." The same parameters, R^2, partial beta coefficients, etc., as in the case of the aspatial model described above, are compared and evaluated.

3.5 Summary

In the first part of this chapter the study area is described to give the reader some insight into the geographic context of this work. The second part of this chapter was dedicated to describing the datasets used in this study. These include EMS calls data acquired from the Sudbury Regional Hospital's Paramedic Base Hospital Program and census data acquired from the Data Liberation Initiative. The steps necessary to incorporate these "potentially spatial data" into a common spatial framework, i.e., a mosaic of Enumeration Areas representing the City of Sudbury were also outlined. Though these steps have been described in the order presented in the basic framework, the reader should be cognizant of the many iterations required at each step of preparing the data for analysis and ultimately, interpretation. The third part of this chapter outlined the analytical methods used to glean meaningful information from the data. These methods included EDA, ESDA and regression modelling. What is unique in this work is that the spatial behaviour of the data are considered, particularly during the modelling phase of the analysis.

The following chapter reports the highlights of these analyses and discusses their implications. It is hoped that some insights into the forces that generate demand for EMS can be gathered from these explorations.

Chapter 4 Results

4.0 Introduction

This chapter provides the results and interpretation of both the data processing and analysis undertaken. A number of figures, tables, and graphs illustrating results from both aspatial and spatial analysis are presented and interpreted. Though ultimately the goal is to interpret collected data, the steps to that end, presented in the previous chapter (Figure 3.2), are not as direct and linear as the basic research framework suggests. Figure 3.4 perhaps represents a more realistic depiction of the research process used. It shows that the flow from raw data to interpreted results is not necessarily unidirectional; the analysis of geographic data requires a complex and iterative framework. This is true because the accuracy of the data needs to be validated, and the data need to be spatially enabled. For example, before selecting relevant cases, the data must first be cleaned. However, realistically, cleaning cannot occur without exploring the data, both aspatially and spatially.

This highlights another problem that may be more relevant to geographers but should probably be a concern to any researchers whose work involves anything that has a locational or spatial component, (hence, just about anything physical or social). Exploratory Data Analysis and Exploratory Spatial Data Analysis are often inextricable. For example, a classical statistical package, SPSS, is used to identify outliers in EMS call rates. Strictly speaking, this an aspatial exploration. But it would be impractical and confusing to develop a whole new section in this paper simply to present and discuss the location of those outlying values. Therefore, many EDA and ESDA results are presented together. Perhaps for the geographer, it would be best to ignore the term EDA and rather incorporate it into his/her ESDA. Maps, a spatial representation for visual interpretation, are presented alongside aspatial tables, scatterplots, and histograms. However, particular attention is given to methods that are specifically measuring spatial characteristics such as spatial association.

This chapter is subsequently divided into three main parts. The first part reports on the data processing required to prepare a dataset for analysis, including cleaning of the EMS events file (ambulance call location address file) and spatial reference file (digital road network). The second part describes the selection criteria and process used to prepare a dataset for further processing and analysis. The last part focuses on the results of EDA and ESDA beginning with a close look at the variable RTPS1000 (EMS calls per 1,000 residents -picked up at their own residence- per EA per year). Finally, the results of the Exploratory Multivariate Regression and the Mixed Multivariate Regression models are presented and compared. This includes an analysis of residuals to diagnose violations of assumptions associated with regression modelling.

Part A

4.1 Data Processing

4.1.1 Data Cleaning

Considerable effort and a number of procedures were required to provide a sufficiently "clean" dataset for final analysis. Harries (1999: 99) proposes that a 60 percent hit rate (addresses successfully geocoded) is unacceptable, but admits that there are no standards as to what is considered acceptable. The approach here was to correct as many addresses as possible without inadvertently introducing error into the dataset. Ultimately, 96 percent of 2,051 events were successfully geocoded. To achieve this, both the events file (ambulance call addresses) and the geographic reference file (digital road network) needed attention. This section describes how, and what type of problem addresses were detected. It also presents the corrective actions taken and the possible consequences of those actions.

4.1.2 GIS as an exploratory tool

Errors in both the events file and the reference file were largely detected by simple visual examination of the dataset, but GIS proved invaluable in highlighting problems. A number of iterations of geocoding attempts were used to detect problem addresses whether the errors were in the address database or the digital road network. The GIS's address matching function served as an invaluable exploratory spatial data tool. Recall that an address in an events file is assigned a match score based on the degree to which it matches an address range in a reference file. The score for a successful match, however, is set by the user in ArcView. It might be tempting to use default settings or to lower settings of minimum scores for a candidate address to be considered a match. However, by setting a minimum score of 100 (all elements of an address must match exactly) subtle errors in both the events file and the reference file were detectable.

To maximize the hit rate, two procedures were found to be useful. First, by harmonizing the events and reference files and second, by simplifying the geocoding process, it was possible to increase the hit rate without inadvertently introducing error (moving an existing event to a different location by manipulating its address elements) into the final geocoded product.

4.1.3 Harmonizing Addresses in the Events and Reference Files

Lack of data standards allows for a variety of address formats to be used, both in the events and the reference files, none of which are necessarily wrong. For example, "number" streets such as First, Second, Third, can be correctly entered in a variety of ways: First, 1st, 1 street. Best results were achieved when both the addresses in the events file and the street names in the reference file were converted to a common format. Also, street names with apostrophes or spaces, e.g., "O'Connor," "St Paul," behaved much better when non-alphabetic characters were eliminated or replaced with an underscore. In the case of "St", the GIS address matching function is probably confusing "Saint" with "Street." Finally, though alias names for apartment building, nursing homes, etc. can be used in a GIS address matching environment, it was easier to simply use street addresses. Figure 4.1a represents some of the inconsistencies in address data entry found in the original EMS Events file.

4.1.4 Simplifying the Address Matching Process

Another means of reaching a maximum hit rate is to simplify the parameters of the address matching process. Each component of a street address - street number, street name, street type, direction- can be used in address matching. In some trials with this data, simply not having the correct street type, e.g., DR (drive) rather than BV (boulevard), decreased a match score from 100 to 75. Fortunately, in Sudbury only one street name is duplicated with a different "street type," one of which is a non residential street. Because of this, it was possible not to use "street type" as a geocoding parameter. This eliminated the need to examine, alter and validate any corrections in the street type data field. First however, address elements were parsed into newly created data fields (Figure 4.1b).

In larger centres where different street types are used commonly, e.g., Carlyle Drive and Carlyle Place both being valid addresses, this strategy is not recommended. In this case, it would be best to ensure that an address database of events includes a field for every element of an address, including street type, and that the entries in these fields is mandatory and correct. Figure 4.2 provides a simplistic example of a database structure that could be used. In this case, the street element fields could easily be concatenated to accommodate the requirements of a GIS's address matching format.

Another means of simplifying the geocoding process is to not use apartment or suite numbers. Though GIS manufacturers claim that the geocoding process can manage such addresses, it is felt that the plethora of apartment delimiters: -, /, ,(comma) , Apt. after the street name, Apt., before the street number, sometimes confounded the GIS. Therefore, clearly recorded apartment address elements were removed from the "pick up location" field and moved to a newly created data field. Figure 4.1a and b illustrate the results of parsing the apartment numbers into a new field.

Figure 4.1a. Example of typical address entries[8] in the Events database provided by the Sudbury Base Hospital Paramedic Program.

CALL ID	PICK UP LOCATION	MAILING SAME AS PICK UP
12345	123 King street north, apt 4	T
12346	2002-1247 Ramsey View CT	T
12347	89 Birch	F
12348	400 First	T
12349	King ST & Main AV	F
12400	332 St. John ST	

Figure 4.1b. Data parsed into newly created fields to simplify geocoding and improve hit rates. Only the Pick Up Location field is used for address matching.

CALL ID	PICK UP LOCATION	STREET TYPE	STREET DIRECTION	UNIT / APT	MAILING SAME AS PICK UP
12345	123 King	ST	N	4	T
12346	1247 Ramsey View	CT		2002	T
12347	89 Birch				F
12348	400 First	AV			T
12349	King ST & Main AV				F
12400	332 St_John	ST			T

Figure 4.2 Example events database using separate data fields for each address element.

CALL ID	STREET NUMBER	STREET NAME	STREET TYPE	STREET DIRECTION	UNIT / APT	MAILING SAME AS PICK UP
12345	123	King	ST	N	4	T
12346	1247	Ramsey View	CT		2002	T
12347	98	Birch	ST			F
12348	400	First	AV			T
12349		King ST & Main AV				F
12400	332	St_John	ST			T

[8]For reasons of confidentiality, all addresses used in this paper bear no relation whatsoever to the address database used. Any similarity to real addresses is strictly coincidental.

The geocoding process also allows for the use of "Alias" tables, i.e. matching commonly used building names, e.g. "Lakeshore Towers" with their respective street addresses through a Look Up Table. Rather than using Look Up Tables however, building alias name entries were converted to their respective civic addresses. This method was preferred because it allowed for easier sorting of the events files in order to visually inspect and validate addresses.

By simplifying the geocoding process and ensuring that the events and reference files were harmonized, it was possible to achieve a respectable hit rate and, more importantly, one based on exact matches (match score of 100). The next section describes errors in the events and reference files and how they were managed in order to achieve these scores and hit rates without introducing undue error.

4.1.5 Errors Detected during Address Matching

4.1.5.1 Errors in the EMS Events File

This section deals with the source and types of errors detected in the events file (ambulance call addresses). How these errors were managed and how they can be avoided in future work is also discussed.

The events file contained address data for individual requests for EMS service. Data fields used in this study included:

- Call Date (self explanatory);
- Pick Up Location (street address or intersection to which an ambulance responded). This field included address data elements street number, street name, apartment number, street type, street direction, and sometimes alias names for apartment or institutional buildings;

- Pick Up Town (municipality in which call originated);
- Same as Pick up (True /False) (whether or not the patient's mailing address was the same as the pick up address); and
- Mailing Address (patient's mailing address including municipality).

Error sources

These data were transcribed from Ambulance Call Reports by a data clerk at the Base Hospital. Unfortunately, this process is subject to a number of potential errors, many of which were realized. The source of these errors is difficult to detect. It is impossible to assess, without direct access to the original ACRs, whether errors in a patient's street addresses were a function of:

1. Paramedics collecting an error (patient tells the paramedic that they reside at an address that they do not);
2. Paramedics entering errors on the ACR themselves (a paramedic enters a different address/municipality than he/she responded to, or misspells the address). These errors should not have been corrected at the data processing level by a data clerk for fear of modifying a medical record. Therefore, the error would in turn be transcribed during data entry into the Base Hospital database;
3. Data clerk entering a typographical error into the database. The presence of this type of error became obvious as some addresses included non-alphanumeric symbols, e.g., the pick up town "Sudbury" entered as Sudbur6y [*sic.*], or Suysudbury [*sic.*].

Error types

Error types detected in address data fields and their elements i.e., street number, street name, street direction, street type, unit or suite number included:

1. Spelling or typographical errors in street names, type or direction, for example:
   a. Misspelled street names: Michelle vs. Michel
   b. Wrong street type: Carlyle CT vs. Carlyle CR
   c. Wrong street direction: 100 Elm W vs. 100 Elm E
2. Use of old addresses. Some apartment complexes underwent address changes in order to meet requirements of Enhanced 911 whereby every property required a number referenced on a municipal street
3. Spelling or typographical errors in municipality name
4. Wrong municipality
5. Typographical errors in street numbers, eg. L29 vs. 729
6. Missing address elements
7. Transposition of address elements. For example, apartment 2 at 61 Regent Street should have read 2-61 Regent, but could be found as 61-2 Regent.

Error Corrections

To minimize introducing error into the dataset, a number of reference sources were consulted for each suspect address. The first approach was to check whether a valid address had been entered in the "mailing address" field for an observation. This often solved the problem of missing data elements, spelling, and typographical errors, assuming that the mailing address was correct.

To correct spelling and typographical errors in the street names, type or direction, several sources were consulted collectively. Paper maps by Rand McNally© and the Ministry of Health's map book, as well as a MapArt© CD were cross referenced. Also, a very useful locally published map book, MapMan2© (Solonynka, 1999) was used. This clever book (based on Regional Municipality data) was designed by a local cabdriver and includes landmarks and address ranges for streets sections and apartment complex configurations. The book is widely used by paramedics and, in fact, is preferred to the Ministry of Health's map book.

These same reference materials were used to ensure errors in municipality names were detected. Given that the original dataset included EMS calls performed throughout the region, considerable care was taken to ensure that an address did not belong to another neighbouring municipality. The Municipality to which a call belonged was confirmed by first consulting the Ministry of Health UTM (Universal Transverse Mercator) geocode in the database. The MoH's dispatch system assigns a seven digit UTM code to each call (Figure 2.1, Ch. 2). Paramedics are obligated to enter that code into the ACR which in turn is entered into the Base Hospital database. The "Mailing City" field was also compared to the "Pick Up Location." Next, the Canada Post website was used to confirm that the address was valid for the address range of that street in that municipality.

Staff at the Sudbury Base Hospital Paramedic Program were consulted to confirm address changes of apartment buildings as a result of E-911 (Enhanced 911) requirements. These discrepancies were detected by simple exploration of the events database.

Finally, suspected transposed address data elements were carefully checked before any changes were affected. Transpositions between apartment and street number are particularly problematic. Personal knowledge of the city played a particularly important role in detecting this sort of error. If an address, for example 519-1 Adelaine, street did not successfully address match, it was quite clear that the apartment and street numbers were transposed and should have read 1-519 Adelaine. However, before any change was affected, the address range for that street was confirmed using the Canada Post website and the MapMan2 map book (Solonynka, 1999). Finally, where addresses could not be resolved in this manner, field checks were undertaken.

Extreme care was taken not to make a correction that was not validated. If an address could not be fully reconciled, it was left unchanged and would result in a unmatched event address.

4.1.5.2 Suggested Corrective Actions for Future EMS Databases

Clearly, there needs to be some mechanism to ensure the quality of the data that is collected and entered into a health database. Regrettably, little attention has been given to locational data. Quality assurance programs typically address the accuracy of clinical skill or the completeness of ACRs. But, because so little spatial planning of EMS services has occurred in Ontario, no one has seen right to ensure that locational data is collected and entered accurately.

Many of the problems encountered with this data set could easily be avoided. Opportunity for improvement is presenting itself as more and more Emergency Medical Services providers consider the use of automated data entry by paramedics in the field. These systems use tablet or laptop computers to enter operational and clinical data.

Spelling and typographical errors could be eliminated by using drop down lists of street names, while address numbers could easily be validated as they are entered. It would also be quite useful to have address data elements entered in separate fields (Figure 4.2), i.e. individual data fields for apartment number, street number, street name, direction (suff and prefix), street type, community. Strictly speaking, this structure is not necessary for geocoding, but it would facilitate error detection and data validation, reduce error, and simplify analysis. With addressed data parsed into its components it is easier to check for errors. However, when needed, it is quite easy to concatenate fields to meet the needs of address matching. This data structure (particularly a separate field for apartment or suite numbers) would ensure consistency and minimize the possibility of transpositions, typographic, and spelling errors, resulting in a high quality data set ready for geographic analysis.

4.1.5.3 Errors in the Spatial Reference File -Digital Road Network

Several iterations of "exploratory" address matching also revealed that the reference file needed some modifications. These modifications were not necessarily performed as a result of errors in the road network itself. Harmonizing street names, as explained above, was done simply to improve the hit rate of the address matching process. And in some cases, address ranges assigned to a street segment were correct, but because of the digital precision of the GIS, addresses might be inadvertently placed in the wrong enumeration area. But, there were some modifications that needed to be done as a result of the road network not being current. This section describes how the modifications were undertaken.

Events and Reference File Harmonizing

As explained above (section 4.1.3), changes in both the events and reference files were made to harmonize the street names. A new data field was created with the modified street names (apostrophes and spaces removed) and the geocoding parameters in ArcView were changed accordingly (Figure 4.1b).

Network Modifications

The digital environment of a GIS offers a degree of precision that may be incompatible with modelling the real world. When geocoding, events are placed where the computer interpolates the street number to be, based on the address range of a road segment and a user defined offset distance from the roadway (described in section 3.3.3.1).

However, in reality, some buildings, especially apartment complexes, are often significantly offset from the road network. In the case where a building complex is also an Enumeration Area, this can be particularly problematic depending on how the census boundaries are drawn. The result is that, despite the road network being properly populated in terms of address ranges, events with valid addresses, can be placed outside of their respective EA. Figure 4.3a shows how an irregularly shaped EA, and Figure 4.4a a small EA, could have resulted in misplaced call locations.

To determine the extent of this problem and mitigate it, all EMS calls at apartment complexes were "address matched." Then the results were visually inspected to ascertain whether or not events had been successfully geocoded to their proper EAs. Twenty five EA/apartment complexes were examined and another 25 non-EA apartment complexes were also examined. In all, 19 event addresses missed their target EA using a 10 metre offset distance. Only two of the non-EA apartment complexes were misplaced (during a field check these were found to be very near an EA boundary). It is noteworthy that some of these addresses were institutional (hospital, nursing homes and retirement residences) and were eventually removed from the study altogether. Because each apartment complex could represent several events, not managing this potential source of error could have had significant impact on the rate of EMS calls per EAs.

Therefore, one of two strategies is adopted where appropriate. If the EA is significantly offset from the street, a new road segment is digitized in the EA with a limited address range that "forces" an event at that address to be placed in the EA. The original road network is then modified to reflect this change. Figures 4.3b and 4.4b demonstrate this modification process. In Figure 4.3a the red marker indicates the placement of an event based on the correct address range of the unmodified road network. However, because of the irregular shape of the Enumeration Area containing the apartment complex, the event has been misplaced in a neighbouring EA. The interpolation mechanism is not problematic in this case, but the offset distance is. Figure 4.3b shows that by adding a road segment with a limited address range the geocoded event (green marker) is "forced" into its proper Enumeration Area. The old road has been split and each new segment of "Main Crescent" has been populated with a new address range.

If, on the other hand, a small EA abuts a road segment, or the road segment transects the EA, a slightly different approach is taken. The offset distance in this case is not problematic, but the interpolation mechanism is because, in the real world, address numbers are not necessarily evenly distributed along a road. To rectify this, the original road segment is split at the EA boundary. That part of the road within the EA is modified by changing the address range to force the placement of an event there. The road segment's address ranges on either side of the EA are adjusted accordingly.

Figure 4.3a Placement of EMS events in wrong EA -irregular shaped EA.

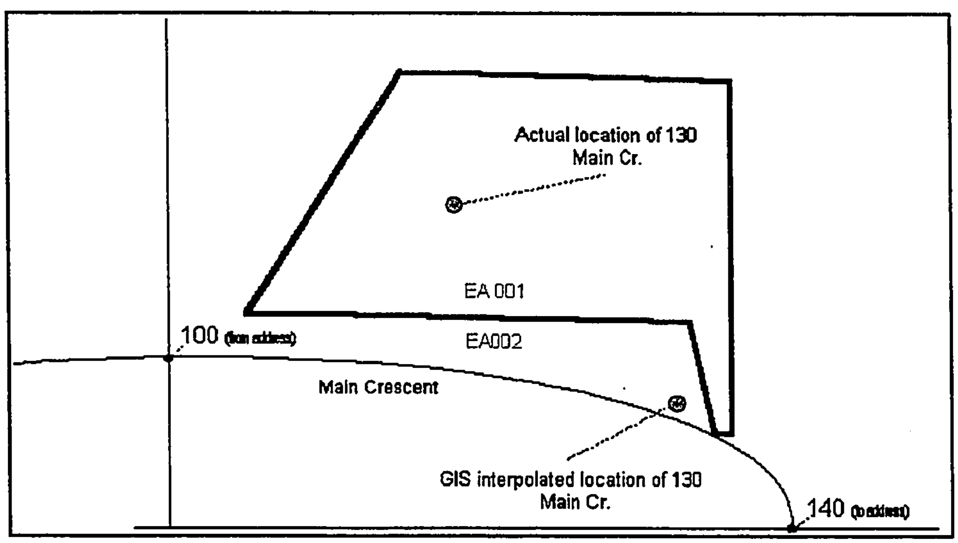


Figure 4.3b Placement of EMS events in correct EA -irregular shaped EA.

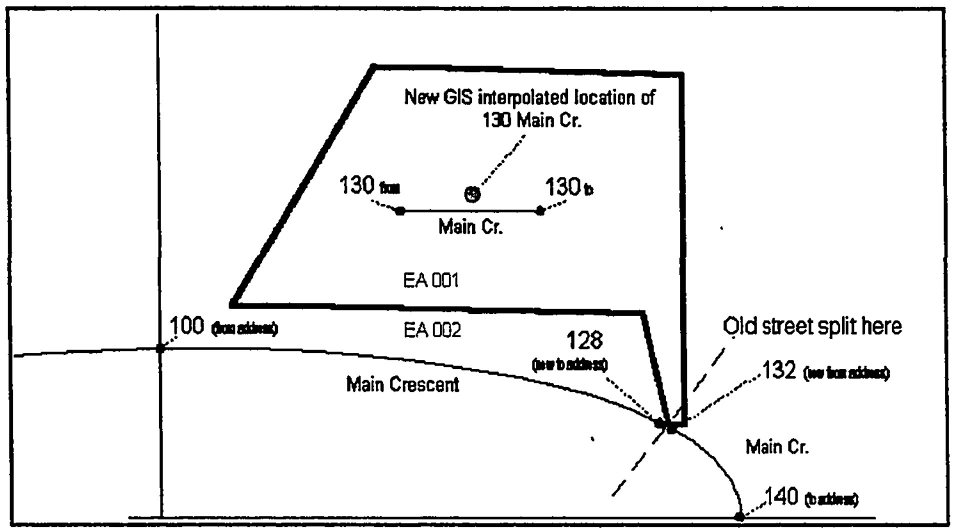


Figure 4.4a Placement of EMS events in wrong EA -small EA

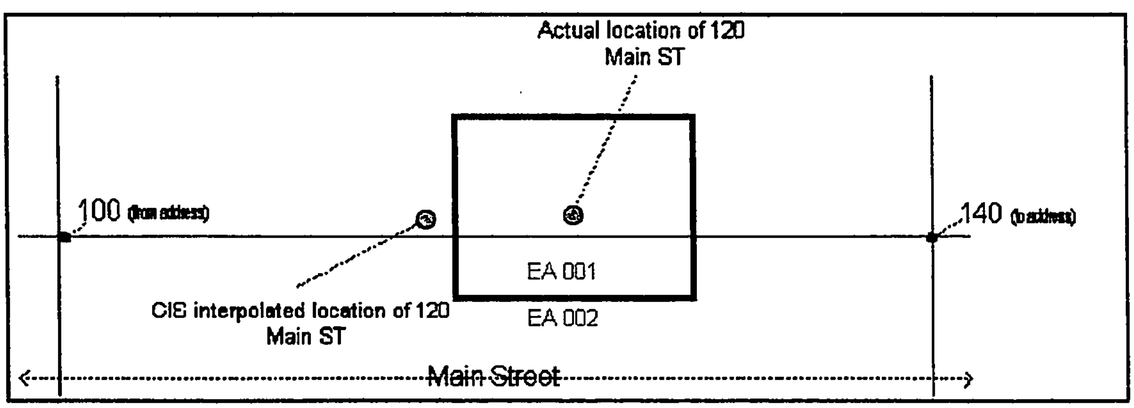
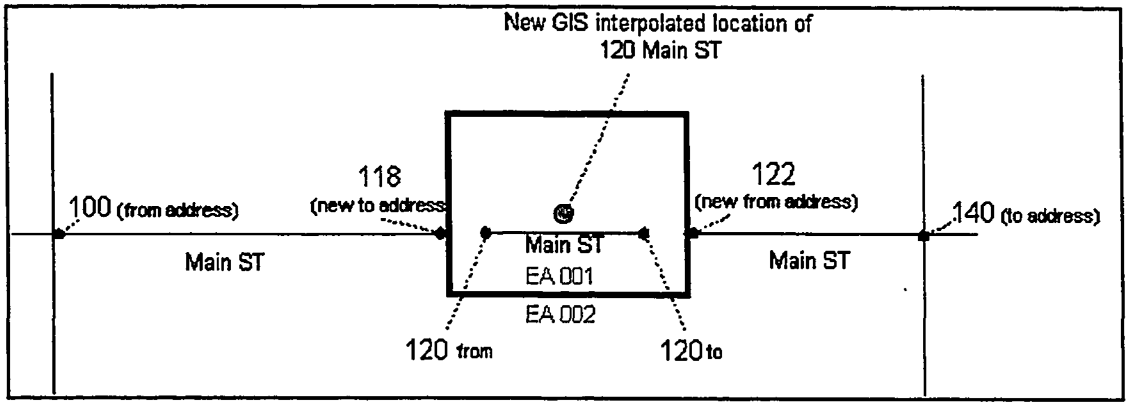


Figure 4.4b Placement of EMS events in correct EA -small EA.



Finally, addresses are re-geocoded using the modified road network, and all modifications are found to be effective at properly locating events in their respective EAs. These results are confirmed with field checks.

Updating Address Ranges

Despite claims by the digital road network manufacturer to the contrary, a few address ranges were found to be outdated or incomplete. Most of the errors detected were the result of the city changing the road direction designation (North, South, East, West) in order to simplify their infrastructure. A strategy to reduce any other missing values in address ranges was developed. To see which addresses needed updating, the following steps were taken:

1. The MapMan website (http://www.mapmanonline.com) was consulted to see which streets were newly built. And, with the advice of the Regional Municipality of Sudbury's Engineering and Planning departments, particular attention was given to streets with East/West or North/South designations. The Region had undergone a program to eliminate most of these road designations to simplify the city's road network.

2. A query that identified road segments with unpopulated address ranges on either or both sides of the street segment was developed.
3. Events that had failed to be geocoded were used as a guideline as to which streets would need attention.

After problem streets were identified, a reference file was needed to properly populate the address ranges. Statistics Canada's 1999 Postal Code Conversion File (PCCF) was incorporated into ArcView by geolocating each postal code using its assigned latitude and longitude. Then each Postal Code's address range was determined individually by using the Canada Post website. A road segment's address range was then compared to the Postal Code address range.

The network modifications described above, including splitting existing roads, digitizing new segments to force address placement, and updating address ranges resulted in 169 road segment changes. As will be seen in the next section, when used with clean addresses, these modifications resulted in significant improvement in successful address-matches.

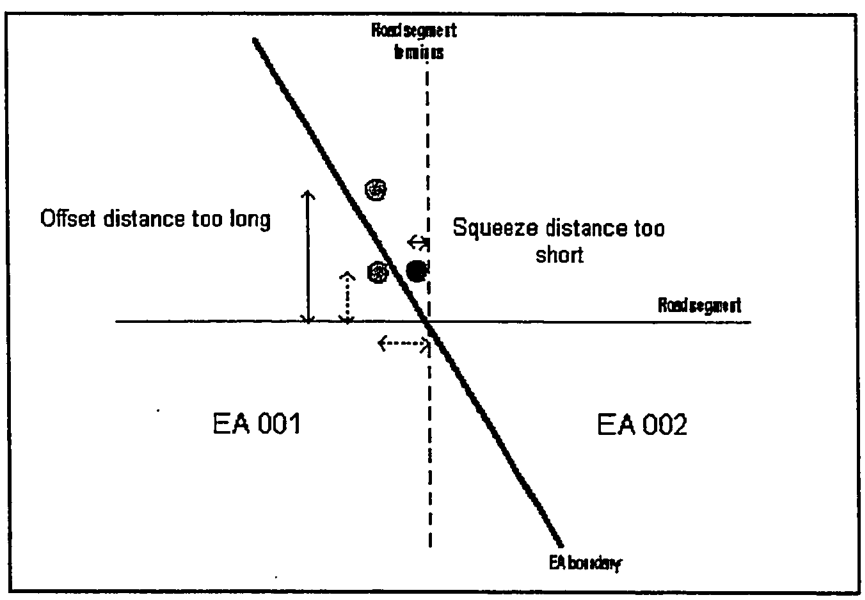
4.1.6 Verifying EA Assignment

As explained above, particular attention was given to the proper placement of multi-events addresses, i.e., apartment or building complexes. However, single events near EA borders can also be "sent" into wrong EA, especially when the shape of the EA is irregular. To identify these problem areas, EA boundaries and geocoded events were visualized using the GIS. Of course, particular attention was given to those areas where events geocoded near the EA boundaries.

As described in the methods section, two user defined distances can be selected to place an event along a road segment. An offset distance of 10 metres (user selected) was used to avoid placing an event on an EA border that coincides with a street The default squeeze factor was sufficient to prevent an extreme address (a value at either extreme of the "to" or "from" address range of a road segment) from being placed at the terminus of a road segment that might also coincide with an EA boundary. The 10 metre offset distance and the 2.5% squeeze factor were found to be reliable at keeping events within their respective EAs.

Figure 4.5 shows the potential misplacement of a valid address geocoded near an EA boundary that is not perpendicular to a street segment. The red symbol is placed in its proper Enumeration Area by using an appropriate offset and squeeze distance.

Figure 4.5 Possible misplacement of EMS event by offset and squeeze distance.



4.1.6.1 EA assignment validation by Postal Code Check

To verify that addresses were placed in their appropriate EAs, one final validation strategy was developed. A random sample of 100 "cleaned" addresses that were successfully geocoded was selected using SPSS. Postal Codes for each address were determined using the Canada Post website. Then, the postal codes were assigned their corresponding EA's using the Postal Code Conversion File (the PCCF lists each postal code's Enumeration Area using the field EAUID). By using ArcView Geoprocessing Wizard's Spatially Join function, each event was assigned its respective EA (census boundary files use field name PRFEDEA). Finally, the PCCF's EAUID was compared to the GIS's PRFEDEA.

Of the 100 addresses checked, 96 EA assignments matched initially. The four apparently mismatched EA were subject to a field check. Two of the addresses were in fact correctly placed in their respective EAs. They had failed the comparison check because they were both in the town of Copper Cliff that uses a rural postal code. The two final addresses would have been properly located had they existed. Upon examining the original address file it was discovered that these addresses were not modified during the cleaning process because, though they were data entry errors, they were "valid" addresses. Using this sample, we can then surmise that the cleaned address file has a 2 percent error.

4.1.7 Address Matching Results

Table 4.1 shows the results of attempting to address match 2,051 emergency calls for the study. Clearly, the most significant improvement in address matching came from cleaning the events file, whereby the hit rate rose from 18% to 84% using the unmodified street network. Then by making modifications and corrections to the road network, an additional 12% of emergency calls were successfully geocoded to a total of 96%. The reader should keep in mind that the criteria for address matching was a match score of 100 which requires an exact match in all address data elements. These improvements were the result of scores of "exploratory" address-matching attempts, close examination of partial matches and unmatched address, and careful modifications and validation checks.

Table 4.1 Address matching results

Events File	Network	
	Original (%)	Modified (%)
Original	362 matched (18) 1689 unmatched (82)	389 matched (19) 1662 unmatched (81)
Cleaned	1733 matched (84) 318 unmatched (16)	1975 matched (96) 76 unmatched (4)

As Albert et al. (2000a:57) exclaim, "GIS is no Panacea!" in medical geography and the efforts to provide a "clean" dataset for this project bears that out well. Fellers (2000: 82) reminds us that if the address-matching process is not carefully monitored, error can be introduced into a dataset. These same authors report that Tobias et al.(1996) reached a 100 % final match rate after carefully reviewing match failures. A 28% initial match score was achieved by Vine et al. (1997) using rural addresses in North Carolina

(The match score criteria and the number of events are not given). Vine et al. (1997) recount their personal communication with staff at the Carolina Population Center (CPC) who maintain that 20 percent address matching hit rates are not unusual in rural centres. According to Vine et al. (1997), the CPC staff also claim that urban area matches as high as 98 % can be reached with automated address matching.

Nevertheless, the result of cleaning the events dataset (ambulance calls) and updating and modifying the geographic reference file (street network) is a dot map[9] of EMS calls and a choropleth map of EMS call rates per 1,000 persons that reasonably represent, in this author's opinion as a paramedic who worked in this city for 15 years, the call locations and volume in the City of Sudbury.

4.2 Data Selection

4.2.1 Selecting EMS Events

1. The hospital database held 7,098 individual entries. Of these 5,691 represented EMS calls responded to in 1999 throughout the Regional Municipality of Sudbury and beyond. After the date entries were verified, addresses were cleaned and validated. Fifty six (56) entries had no address or had incomplete addresses that could not be extrapolated from the mailing address or by any other means.

2. Initially, calls dispatched code 3 or 4 were considered for analysis which resulted in 5,566 cases.

3. By selecting from these cases only those found in the former City of Sudbury (including the amalgamated Town of Copper Cliff), this number was reduced to

[9] This dot map is not presented in this thesis because of an agreement not to graphically identify individual EMS events as a means of assuring patient confidentiality requirements of the Ethics Review Committee of the Sudbury Regional Hospital Corporation condition of releasing the Base Hospital data.

3753 calls, representing approximately 67% of emergency calls responded to in the Sudbury Area.

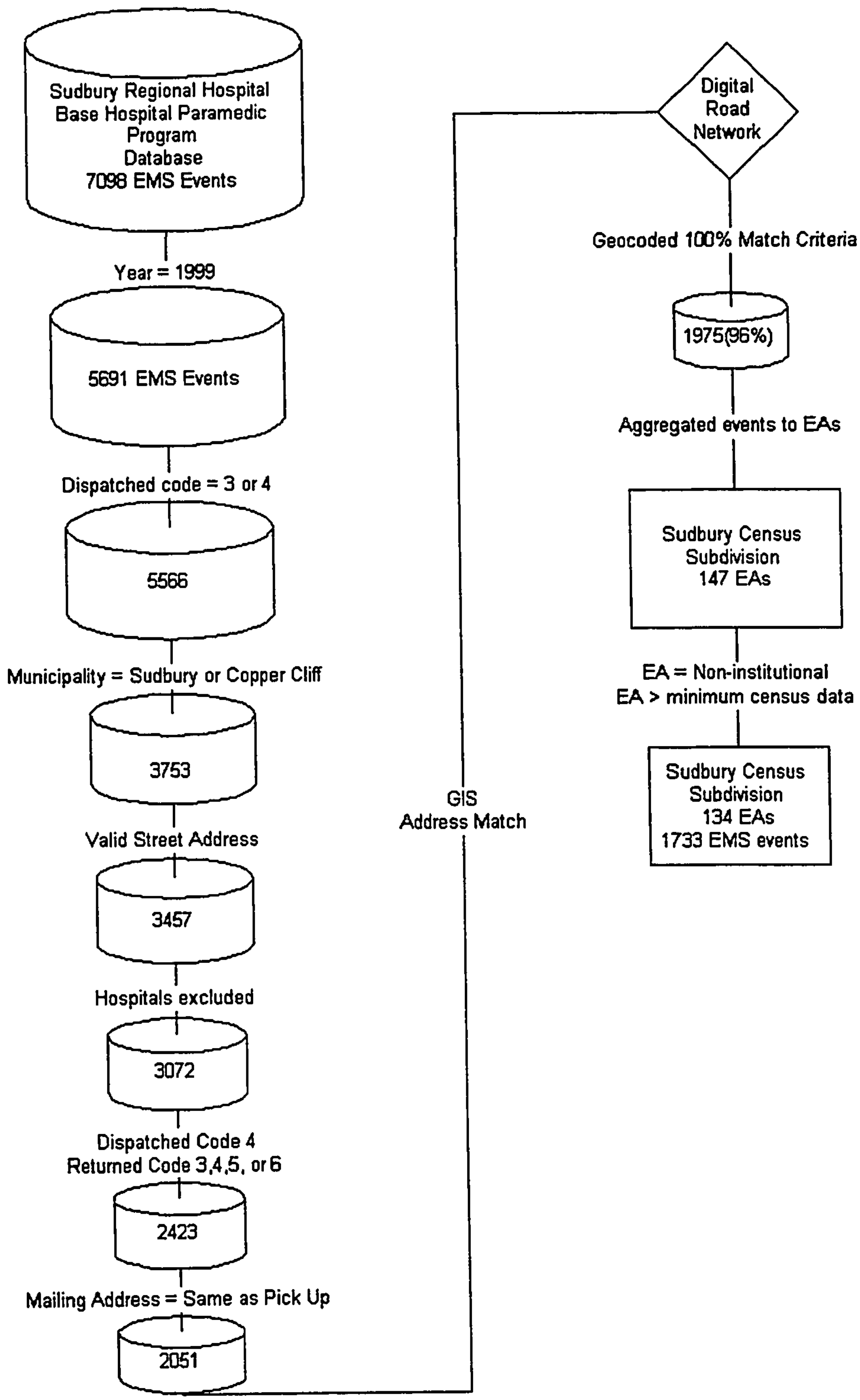
4. Addresses were examined again and only calls with valid civic addresses were selected. 296 cases had intersections or incomplete addresses entered as "Pick Up Location[s]." Though 70 of these had "Same as Pick-up" checked in the dataset, it was impossible to infer the address from the mailing address field. Clearly, the "Same as Pick-up" field had often been inadvertently selected during the data entry process. This was obvious because the corresponding mailing addresses were public spaces (e.g., a park) or an intersection.

5. Next, calls to all four city hospitals were removed from the dataset because, despite the fact that this represented a considerable call volume (385), it did not relate to the demand created by the population at these institutional EAs. In fact, these emergencies reflect high priority, inter-facility transfers originating from various hospital departments including critical admission transfers from the Emergency Department. The result from culling the inter-facility calls was 3,072 emergency calls responded to within the city of Sudbury.

6. After exploring the remaining calls, an additional 649 cases were removed from the dataset. These included calls dispatched code 4 but returned priority code 1, 2 or 7 (see Table 2.2, Ch. 2), and calls dispatched code 3 regardless of their return priority. This selection was undertaken after consulting with the Base hospital and establishing that not all code 3 calls (240) had been entered in the database. Furthermore, only return codes 1, 2, or 7 where a delegated medical act (paramedic skill) had been performed were included in the database, but these were not included in this analysis. Ultimately, this provided a cleaner database as the Base Hospital staff confirmed that every effort was made to enter all calls dispatched code 4 that returned code 3, 4, 5, or 6. Essentially, the remaining 2,423 calls could reflect true emergency calls within the City of Sudbury.

7. Of the 2,423 true emergency calls, 2,051 (84.6%) had a mailing address that was "Same As Pick Up." By selecting only those EMS events where a patient resides where she/he was picked up, an ecological analysis of underlying socioeconomic and demographic variables is more meaningful. Figure 4.6 illustrates the selection results described above as well as the EA selection results described below.

Figure 4.6 Results of EMS events and EA selection.



4.2.2 Selecting Enumeration Areas

The digital boundary file of the Census Subdivision of the City of Sudbury is comprised of 147 Enumeration Areas that form a space exhaustive non-overlapping irregular tessellation. However, a number of EAs were excluded from the study area resulting in a non-space exhaustive mosaic of 134 Enumeration Areas (Figure 3.13, Ch. 3). These Enumeration Areas were digitally removed from the dataset so as not to be considered when determining neighbours in the creation of the spatial weights matrix and lagged variables. EAs considered for exclusion were identified by exploring the data, both with ArcView GIS and SPSS. Thirteen EAs were excluded from the study for various reasons described below.

Two of the excluded EAs were industrial areas with no residents, or "Empty Places" (Statistics Canada, 1999:255). Also, only basic statistics (age, sex, marital status, and mother tongue) are collected for institutional residents (Statistics Canada, 1999: 42). This includes residents in hospitals, nursing homes and seniors' residences. Therefore, four hospital EAs, three nursing home EAs, and 3 senior care residence EAs were excluded from further analysis. Furthermore, because of Statistics Canada's rules on data suppression (Statistics Canada, 1999: 357) for EAs with populations less than 40 persons, one EA that is a religious residence was excluded.

Though hundreds of calls originated from the hospital EAs, their respective rate of EMS calls do not reflect the demand created by permanent residents there. The extremely high rates of EMS use (3,564 per 1000 - all hospitals combined) clearly stems from inter-facility transfers from a variety of departments, particularly the Emergency Room.

Nursing homes and seniors' residences do reflect the demand for EMS by their residents, however. With a mean of 283 calls per 1,000 residents for the 3 nursing homes and senior care EAs combined, these institutions clearly have an impact on EMS use. Unfortunately, the paucity of census data relating to these residences precludes them from being used in an ecological analysis based on census variables. However, not all senior residences constitute an entire EA; which means that some senior care facilities may be engulfed by a surrounding EA. We might expect these EAs to generate significant demand.

With the institutional EAs effectively removed from the analysis attempted in this work, we can only allege to model demand for EMS generated mostly by non-institution population. Nevertheless, the emergency planner would be wise to think of EAs such as hospitals, nursing homes and seniors' residences as hot spots by default. They must be considered when siting EMS stations or developing strategic deployment plans.

Part B

4.3 Analytical Results

In this section the analytical results are divided into two parts. The first presents the empirical results of having spatially enabled the emergency calls datset within the census boundary spatial framework. It examines the rate of emergency calls per 1,000 residents (RTSP1000) at the Enumeration Area scale through EDA and ESDA. The second part focusses on two regression models. The first model is aspatial, much like those undertaken in past research. The second expands on the first by accounting for spatial effects.

4.3.1 Empirical Results of Emergency calls in the City of Sudbury

As described above, our study site is now a mosaic of 134 non-overlapping, non-space exhaustive units. Figure 4.7 shows the rate of Emergency call per 1,000 residents and highlights the five lowest and seven highest values (a discussion on these extreme values follows). A general trend of high values near the core of the city with rates decreasing outwardly is apparent. This trend is interrupted by a slight increase in rates near the well established commercial nodes at the north east, i.e., the "New Sudbury Shopping Centre" area, and in the south near the "Four Corners" area. The mean annual rate of emergency calls per 1,000 residents (RTSP1000) in 1999 was 23.5; the median and mode were 16 and 9 calls per 1,000 residents respectively. This relationship between the measures of central tendency reflects the positive skewness of this variable as demonstrated in the histogram of values (Figure 4.8). SPSS calculated a skewness index of 3.078 and a kurtosis index of 12.31, which further illustrates the departure from a normal distribution. The form of this distribution becomes important in the choice of assumptions when calculating some spatial statistics below.

The rate of emergency calls per 1,000 residents ranges from zero to 179. The minimum of zero does not include enumeration areas that were excluded from the analysis. The standard deviation of the variable RTSP1000 was 26.16. For a normal distribution, rates beyond two standard deviations from the mean (75.82) will be referred to as outliers, while those beyond three standard deviations from the mean (101.98) as extreme. Due to the distribution of the variable, there are no outliers or extremes in terms of negative deviations from the mean. Because of the shape of the distribution, the five lowest and seven highest (mean + 2 std. dev.) values in the dataset are identified.

Figure 4.7 EMS Rate per 1000 population per EA City of Sudbury 1999

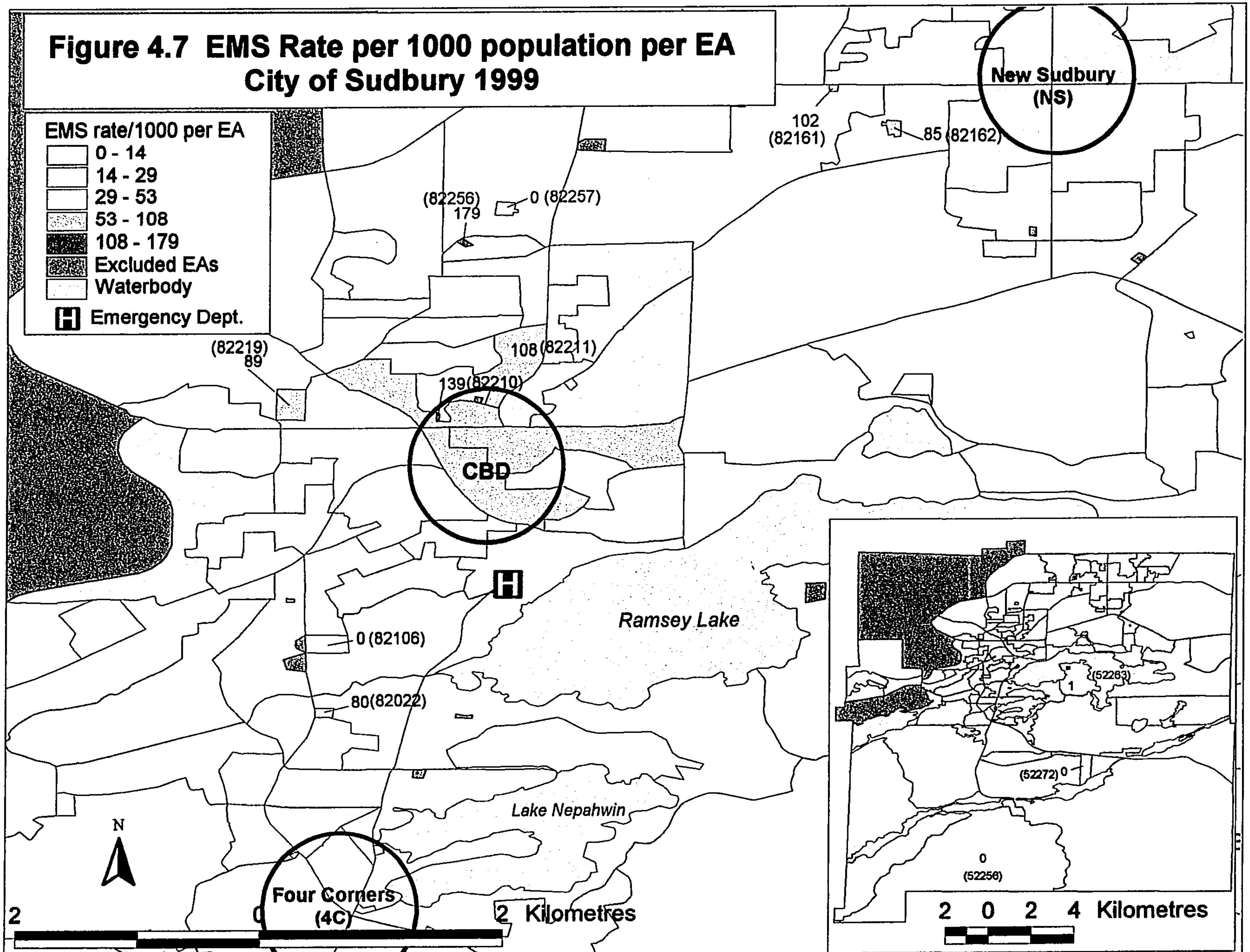
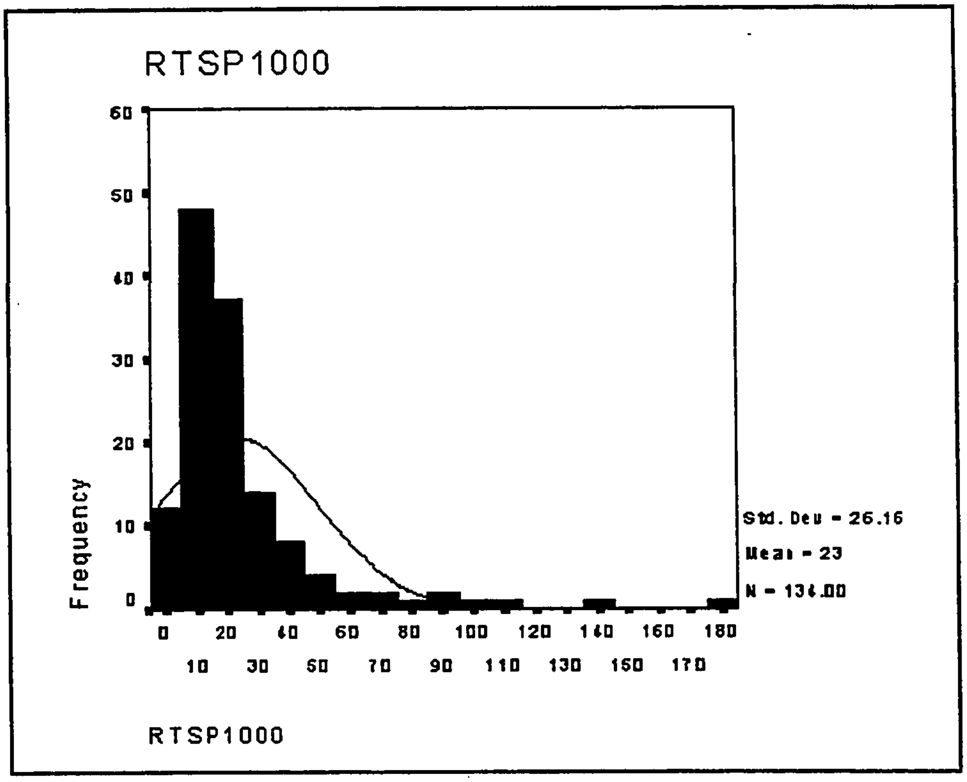


Figure 4.8 Histogram of Rate of Emergency calls by enumeration area.



4.3.1.1 High EMS Rates

The highest rate, 179 emergency calls per 1,000 residents, is slightly north of the city core (EA # 82256) and represents an apartment tower well known to paramedics for its high percent of low income seniors. The next two highest rates are in the core of the city. One is a known seniors' building with a rate of 139 calls per 1,000 (82210). The other EA with a rate of 108, contains - aside from a variety of dwelling types - two apartment towers inhabited mostly by seniors (82211).

The fourth highest rate, 102 (82161), is found in another apartment tower/EA, north and east of the core in an area known as New Sudbury. The fifth highest value at

89 is an EA (82219) just west of the core of the city. This EA includes a community care health centre that has non-permanent residents. The high rate found in this EA is likely a result of clients reporting, or paramedics recording, this centre as the patients' mailing address. The sixth highest value is found in New Sudbury at another apartment tower, its usage rate is 85 calls per 1,000 residents (82162). The final outlier is the seventh highest value at 80 calls per 1,000 residents (82022). This EA is comprised of two apartment towers, which are defined as buildings with more than 5 stories.

Though most of the discussion about these outlying and extreme values will follow the results of the regression models, it is noteworthy at this point to indicate that five of the seven highest generators of EMS calls per residents are EAs made up almost exclusively of apartment towers. Also, each of these five EAs has senior populations of greater than 30%, which is more than double the median (14.2%) senior population found throughout the city.

4.3.1.2 Low EMS Rates

Four EA report no calls per 1,000 resident and one reports 1 call per 1,000 residents (Figure 4.7). Two EAs with no EMS calls are found at the south end of the city (EAs 52256 and 52272). One other is south west of the city core in an area known as the hospital area. This EA (82106) has a small population of 85 residents. The last EA with no emergency calls reported is slightly north of the core of the city (EA 82257), quite near the highest call generating EA. The two EAs at the South End are generally considered affluent as is the EA (52263) near Lake Ramsey with an EMS demand rate of 1 call per 1000 residents.

4.3.2 Discussion on Preliminary Results

If we use the guideline presented by Cadigan and Bugarin (1989) to calculate the expected EMS transport rate for the population (90,923 persons) of the 134 EAs remaining in the analysis, the result would be approximately 35 calls per 1,000 persons. This would be consistent with other studies such as those conducted by Aldrich et al. (1971), Gibson (1971), and Siler (1975). Then, the mean rate of 23.5 calls per 1,000 residents discovered here would seem rather low. However, a comparison with communities studied in previous works is difficult, as most studies do not clearly define what selection criteria were used to include candidate calls. This is particularly problematic in terms of dispatched versus returned priorities. Likely, other studies have included a broader range of call priorities than this study where only calls dispatched at the highest priority, i.e., code 4 that return as at least a "prompt" call - code 3 or higher can be considered. This limitation was imposed as a function of the available data.

Also contributing to the lower utilization rate in this study is the fact that nursing homes and hospital calls were excluded because of lack of corresponding census data. Furthermore, by using only cases where a patient's mailing address was "Same as Pick-up", approximately 15% of emergency calls were effectively eliminated from the analysis. As Braun (2000) indicated, an ecological analysis using underlying socioeconomic variables is more meaningful this way. It identifies the users of the system - or more correctly to avoid ecological fallacy - the characteristics of the Enumeration Areas that generate EMS demand. Having such an empirical profile of a community can certainly assist health planners in targeting areas for education and prevention programs.

Clearly, the wide range of demand between EAs suggests that there are important factors contributing to EMS use. The results of the regression analysis discussed later in this chapter will identify some of those factors. Before that, however, the exploration of the dependent variable is completed by reporting on the methods used to identify its spatial characteristics.

4.3.3 A Focus on Space

The discussion of classical descriptive statistics provided above could not be done by a geographer without at least some reference to a map. If visualizing general trends and highlighting outliers on a map are ESDA, then EDA and ESDA are, as already suggested, inextricable. However, the results presented in this section are from more deliberate measurements of spatial association, both globally and locally, of the variable RTSP1000 (Emergency calls per 1,000 residents per year). Global statistics include Moran's *I* and Geary's *c* at first order contiguity. The spatial correlogram is used to attempt to identify whether or not spatial effects persisted beyond the first lag. Local effects, or second order effects, are determined using local Moran's *I* and are presented here with a map of significant LISA values.

4.3.3.1 Global Measure of Spatial Dependence

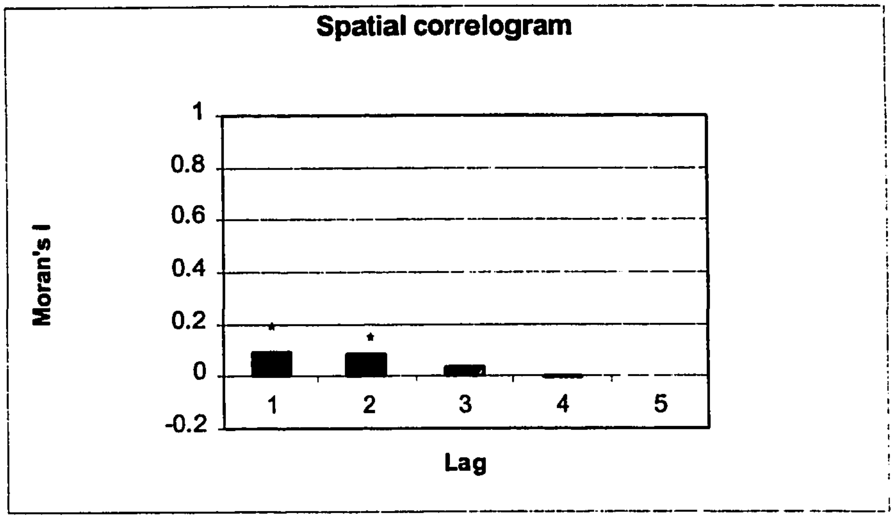
This section begins with an exploration of the global measures of spatial association. The histogram of RTSP1000 (Figure 4.8) presented earlier clearly indicates that the normal assumption for the statistics would be inappropriate. Therefore, a randomization assumption to test the significance of the values was chosen. Recall that the spatial weights matrix is based on first order contiguity using the queen's case. The weights matrix was row standardized.

With the criteria established above, SpaceStat calculated a global Moran's *I* of 0.095106 (p= 0.051430). This probability value is on the border of rejection / acceptance using a 95% confidence interval. To support a decision as to the statistics's significance, global Geary's *c* was also calculated. Geary's *c* was 0.7095004 (p = 0.005841) which confirmed that there was significant spatial autocorrelation in the variable RTSP1000. The expected values of no spatial autocorrelation are -0.00752 and 1 for Moran's *I* and Geary's *c* respectively. The observed values are indicative of very weak positive spatial autocorrelation. The decision to accept Moran's *I* with a *p* value so close to 0.05 is based on the supportive results of Geary's *c*, which is clearly statistically significant. These quantitative results confirm the visual interpretation of decreasing rates away from the core of the city.

4.3.3.2 The Spatial Correlogram

To further examine the extent of spatial autocorrelation, the spatial correlogram was used. The correlogram can also, in some cases, give an indication as to whether global or local effects are at play in a landscape. Figure 4.9 shows the results of creating a spatial correlogram to five lags. There is significant (indicated by an *) spatial correlation at the first and second lags. The first lag value is reported above as the global Moran's *I*. The value at the second lag is slightly less at 0.08312189 (p = 0.005180). Moran's *I* is not significant past the second lag.

Figure 4.9 Spatial correlogram of RTSP1000 to lag 5.



This would initially suggest that the rate of EMS use is under local influences in the study area, i.e. there appear to be no marked global (city wide) patterns of variation in EMS use in Sudbury. Bailey and Gatrell (1995: 270) caution that the spatial correlogram has limitations in that the correlation at successive lags is in part a function of the value at previous lags. This makes it difficult to determine then whether significant spatial dependence persists beyond first order neighbours. Certainly, if this is the case, it is very weak and not significant past the second lag. Large scale trends that would be evidenced by persistent autocorrelation through several lags are not apparent here. The general trend noted when simply visualizing the data may, more likely, be the result of some local effects in what is otherwise a random process. By random process we are not referring to spatial randomness but rather the equal chance of any person requiring EMS for an emergency. The following section will discuss the results of testing specifically for these local effects.

4.3.3.3 Local Indicators of Spatial Association (LISA)

The Local Moran's *I* is used as a measure of local spatial dependence or second order effects. Though it is not the only LISA available in SpaceStat, it is the most practical to use because of the loose coupling with ArcView that maps statistically significant measures of Local Moran's *I*. Also, with the Dynastat extension, a Moran scatterplot can easily be generated and dynamically linked to the geographic features in the ArcView. This makes it possible to interactively identify individual cases selected from the scatterplot or the View[10]. However, the Moran scatterplots presented in this thesis were replicated using SPSS because of its superior graphic capabilities.

The Moran Scatterplot

The Moran Scatterplot is a useful tool in assessing the relationship of an observation z_i (in this case RTSP1000 in an EA) and the weighted average of its neighbouring values (Anselin, SpaceStat User's guide:38) Wz_j. The scatterplot is divided into four quadrants. The upper right and lower left represent positive association. The upper right holds high values of z_i surrounded by high values of Wz_j, and the lower left, low values surrounded by low values. Negative autocorrelation is found in the upper left and lower right quadrants. High values of z_i are surrounded by low Wz_j values in the lower right, and low z_i are surrounded by high Wz_j in the upper left quadrant.

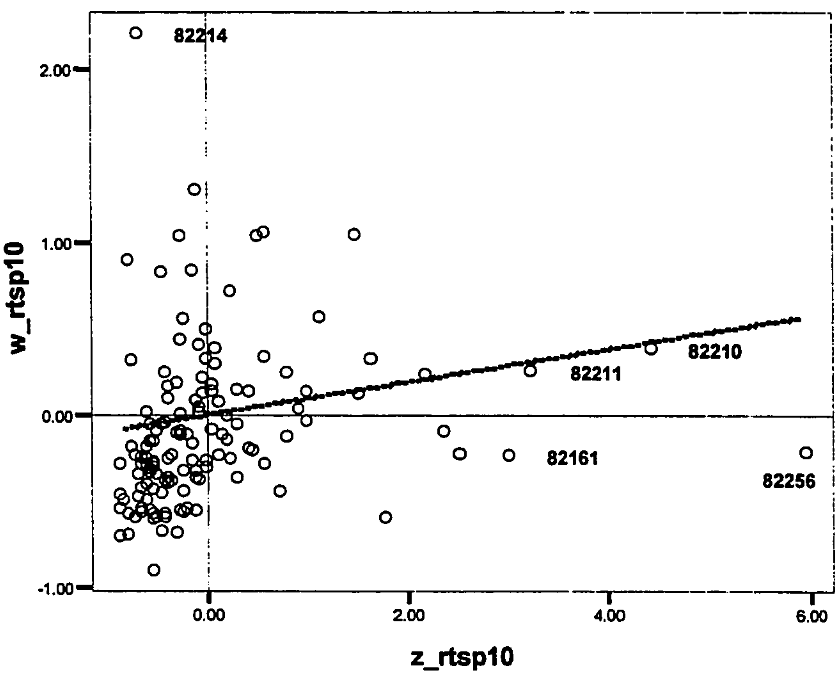
Like a typical bivariate correlation scatterplot, a general trend or regression line can be drawn from all values to summarize the overall relationship between points. In the

---

[10] A "View" in ArcView is an interactive on-screen map.

case of plotting z_i and its neighbours Wz_j, the resulting slope of the regression line is also the global Moran statistic. The dotted line in Figure 4.10 indicates a very shallow slope, reflecting weak spatial autocorrelation, or a small Moran's I.

Figure 4.10 Moran Scatterplot showing visually extreme values of RTSP1000



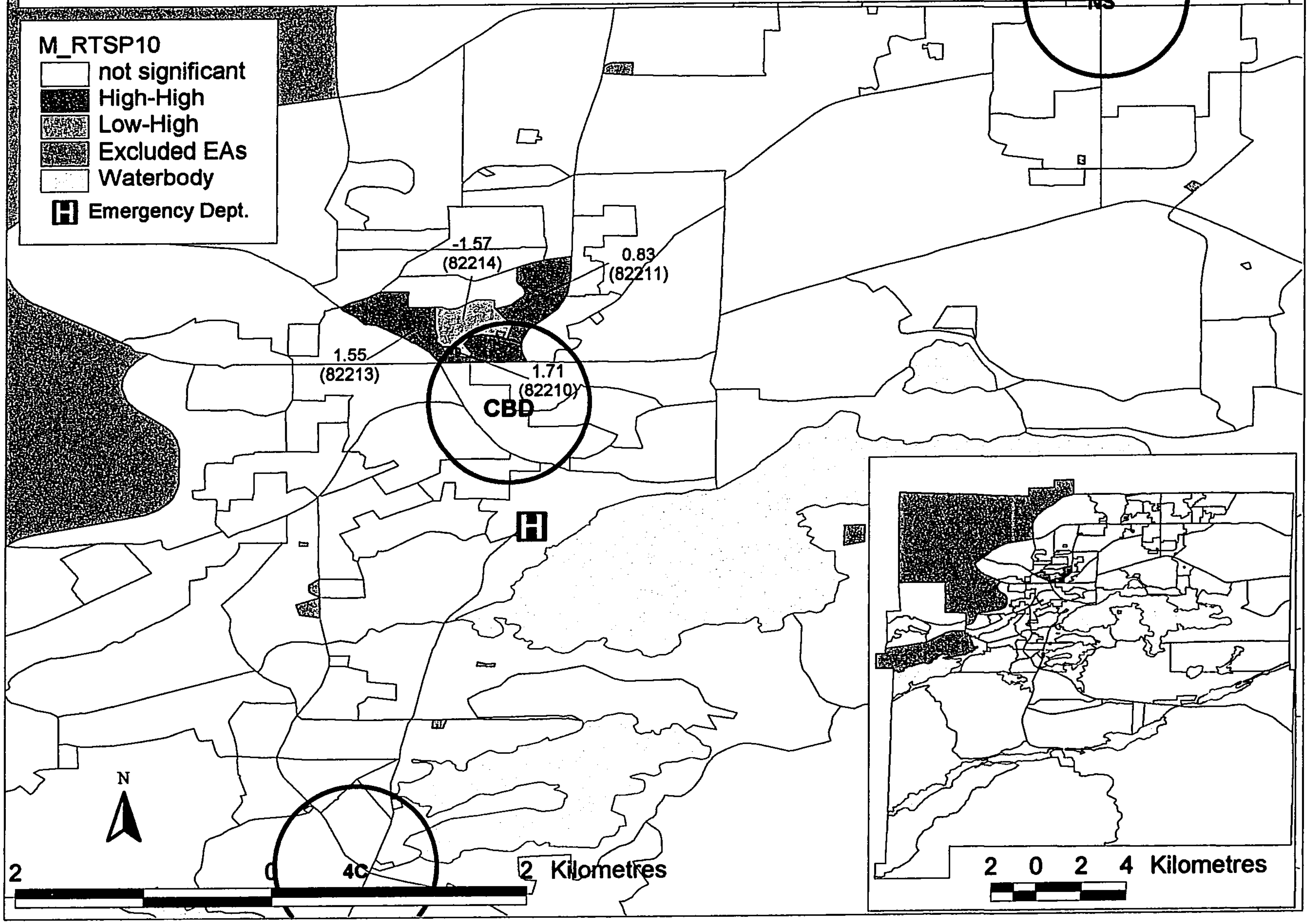
Anselin (1996) suggests that the Moran Scatterplot should be looked at for outliers that influence the slope of the Moran regression line. If we do this visually using Figure 4.10, the most obvious outlier is in the upper left quadrant (EA 82214), therefore it is a low value surrounded by high values. This represents an EA in the core of the city with an EMS utilization rate of 5 calls per 1,000 residents (Figure 4.7). Two of its four adjacent neighbours were identified as extreme cases with EMS rates of 108 (EA 82211) and 139 (EA 82210) calls per 1,000 persons. The other two neighbours have rates of 16 (EA 82212) and 62 (EA 82213) EMS calls per 1,000 residents. Not surprisingly then, it was quickly identified in the upper left corner of the Moran scatterplot.

There are no obvious outliers in the lower left quadrant. The lower right quadrant has two cases that have high z_i values (EAs 82161 and 82256) but their Wz_i are not particularly remarkable. The same is true with two high z_i values in the upper right quadrant (EA 82211 and 82210). What these four extreme z_i values indicate are EAs with extreme EMS rates that are surrounded by approximately average rates. When identified using the GIS and SpaceStat extension, we find that these, not surprisingly, are the four most extreme values that were identified earlier, i.e., EAs with rates greater than 100 calls per 1,000 residents.

Local Moran's *I*

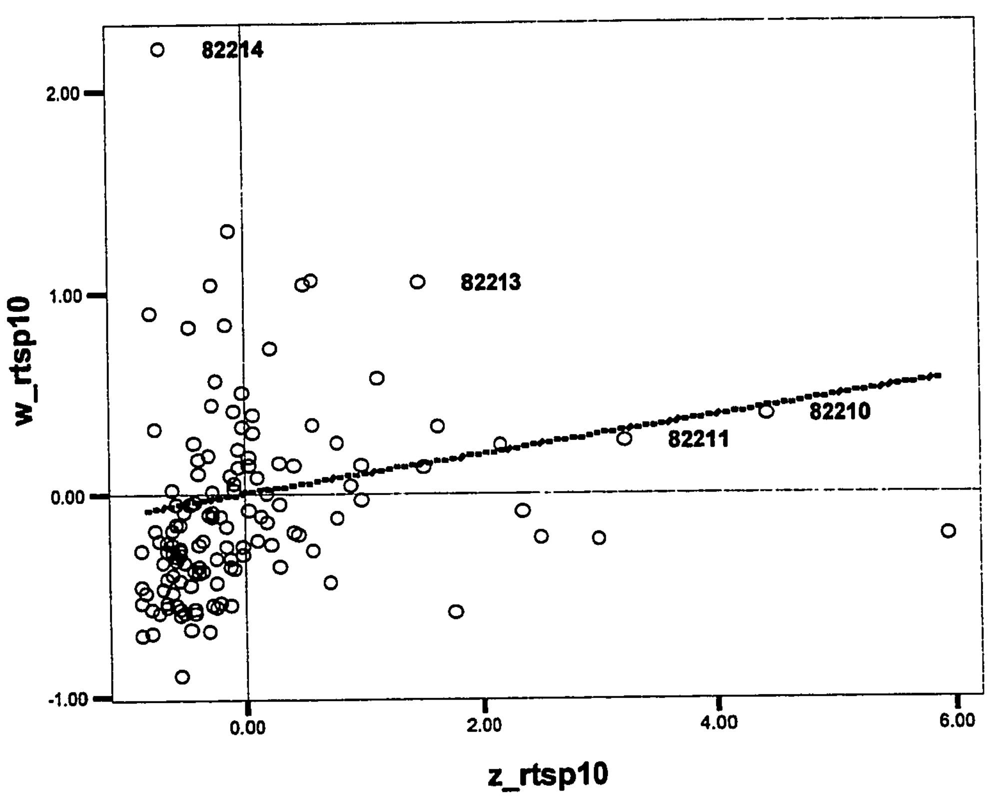
Inspecting the scatterplot visually, like visualizing a map pattern, can be deceiving. Fortunately, Spacestat computes the significance of the local Moran statistic for each spatial unit. Figure 4.11 maps only those values that were statistically significant at $p < 0.05$ and includes the respective Local Moran's *I* values for these EAs. However, Sokal et al. (1998: 348) warn that, in the presence of global spatial autocorrelation, "statistical tests of individual local SA [spatial autocorrelation] will be far too liberal." Therefore more LISA values would be found significant than there really are. However, Boots (2002) does suggest that LISAs may still be used for Exploratory (ESDA) purposes. Nevertheless, Boots (2002) still cautions that global spatial association should be measured before computing LISAs.

Figure 4.11 Local Moran Significance - EMS Rate per 1000 population



Thanks to a dynamic link between Dynastat and ArcView, EAs selected interactively in the GIS are highlighted in the Moran scatterplot. Figure 4.12 is similar to Figure 4.10 above, except that rather than examining cases that visually appear to be outliers, cases that have statistically significant Local Moran's *I* are indicated. The value in the upper left quadrant (EA 82214) is confirmed as statistically significant with a Local Moran's *I* (LMI) of -1.570185. This negative association is due to this EA's small EMS rate of 5 calls per 1,000 surrounded by 4 neighbours with rates of 139, 108, 62, and 16 calls per 1,000.

Figure 4.12 Moran Scatterplot showing EAs with significant Local Moran's *I*



The two values in the upper right quadrant that were seen as extreme are also confirmed; they had LMIs of 1.708573 (EA 82210) and 0.832342 (EA 82211). The two cases in the lower right quadrant that were identified above with visual inspection, do not have significant LMIs. More interesting, one case in the upper right quadrant does show as having a significant LMI, but was not identified visually - at least initially. This EA (82213), with a LMI of 1.547474, has a rate of 62 EMS calls per 1000 residents; nearly three times the mean EMS rate. It has 11 neighbours, with EMS rates ranging from 5 to 139, seven of which are greater than 38 calls per 1000.

Given the pattern of high values towards the core of the city, it is not surprising that these EAs show as significant local statistics. Certainly, this furthers the suspicion of a separate spatial regime withing the core of the city. However, although the degree of global spatial autocorrelation is very small, it is nonetheless statistically significant; therefore, we must consider the test for significant local statistics as liberal.

4.3.3.4 Section Summary

In this section, EDA and a map were used to determine basic descriptive statistics. The Rate of EMS calls per 1,000 residents was mapped and a general trend of high values in the core of the city, gradually decreasing outward with interruptions of moderate values near commercial areas at the northeast and towards the south end of the city, was initially proposed. Then, extreme values were identified and mapped. Low values were found mostly outside the core of the city, while high values were in the core and in the northeast. The high EMS rate EAs typically included apartment towers (buildings with greater than 5 stories) with relatively large concentrations of seniors.

Next, the focus was on the global spatial statistics used to quantify the patterns of the variable RTSP1000 apparent on the map. In the case of Moran's *I*, the p value is extremely close to the 0.05 level of significance, therefore a decision as to the significance of the weak positive spatial autocorrelation detected is supported by the clear significance ($p<0.05$) of Geary's *c* statistic. A spatial correlogram was generated that confirmed the weak spatial dependence at the first lag, possibly persisting into the second lag but not beyond. This suggested that some local or second order factors may be at play in determining the landscape of demand in the city of Sudbury.

To further explore these possible local effects, the Moran Scatterplot and Local Moran's *I* were used. Using the Moran Scatterplot, many of the extreme values identified using EDA were also apparent. However, when considering only significant local Moran's *I* values, only four EAs in the core of the city were found to be significantly associated with their respective neighbours.

There is a degree of weak spatial dependence when considering the landscape of demand for EMS in the city of Sudbury. This may simply be nuisance autocorrelation as a result of the pattern of spatial units. Local statistics suggest, albeit tenuously, a hotspot in the core of the city, which may be a separate spatial regime, where a different regression model may be necessary to more effectively explain the variation in EMS demand. What follows is a report and discussion of the results of the regression equations used to propose factors affecting demand for EMS. The results may shed more light on what appears to be two spatial regimes of EMS demand, one in the core and the other throughout the rest of the city.

4.3.4 Regression Results

In this section, the results of two regression equations are presented and discussed. The first regression is similar to those used in previous studies in that it ignores any spatial characteristics of the study site. It is used as an exploratory tool to determine which variables might be included in the subsequent regression model that will consider spatial effects. The reader is referred to Figure 3.16 for a reminder of the approach. These two models are compared in terms of explanatory power and the behaviour of their respective residuals. This should demonstrate that the first law of geography, i.e., spatial autocorrelation, cannot be ignored when proposing a model of EMS demand using socioeconomic and demographic variates.

4.3.4.1 Exploratory Multivariate Regression (EMVR)

The stepwise regression approach is used here primarily as an exploratory tool. Explanatory variables that are not excluded during the stepwise process will be retained for use in the subsequent model along with excluded variables that have demonstrated significant spatial autocorrelation. This section discusses the results of the stepwise exploratory multivariate regression (EMVR).

The EMVR uses 32 proposed explanatory or independent variables (Table 3.2 Chapter 3). The response or dependent variable is of course the number of EMS calls per 1,000 residents per year (RTSP1000) in each Enumeration Area. The adjusted coefficient of determination - adjusted R^2 is 0.394 (Table 4.2). Using the stepwise method of regression, only two independent variables remain as significant explanatory variables in this model (Table 4.3). They are PLIVEALO (percent of people living alone), and PSEPDIV (percent separated or divorced). Twenty nine other variables were excluded during the stepwise iterative process. Therefore, approximately 40 percent of the variation in EMS demand per 1,000 per EA can be explained by changes in the two variables PLIVEALO and PSEPDIV.

The R^2 is significantly lower than the models used in American contexts and Szplett's (1988) Toronto study. However those models that used fairly large geographic units, such as counties in the American studies, probably ignored the aggregation effects that are unpredictable in a multivariate environment (see section 3.3.1.2 MAUP above). More important, spatial autocorrelation, whose effects in multiple regression analysis are complex (Anselin & Griffith, 1988) and poorly understood (Griffith & Layne, 1999: 21), was not considered in these models. As indicated earlier, the Toronto study suffered from incompatible spatial frameworks (i.e., a regular grid of ambulance call volumes overlain onto an irregular network of census tracts). Certainly, the results of that study should be considered with some skepticism. De Angelis' (1995) Kingston study was also plagued with incompatible geographies, but the R^2 was much lower at 0.15 (the most comparable of his twelve regression attempts).

The method used in previous Canadian studies, that of extrapolating ecological variates from irregular census boundaries to a regular grid containing EMS calls, is quite simply the worst kind of MAUP. If we imagine several grid cells all being assigned exactly the same value from a large underlying census tract we could expect to generate extremely high values of artificial spatial autocorrelation between grid cells. The inconsistent results of the two Canadian studies may well be in part a due to this unfortunate limitation of previously available data. Surely, the method provided here, whereby events are allocated to their respective census units - which was one of De Angelis' (1995) recommendations - should provide more credible results in terms of an ecological analysis using census variables. It would seem imperative that good locational information be gathered if good science is to be performed, and in turn if good decisions are to be made.

The partial correlation coefficient indicates the relationship between an individual independent variate and the dependent variable when the effects of other independent variables have been controlled (Norušis, 1990: 266). Table 4.3 shows that PLIVEALO and PSEPDIV both had positive partial correlation coefficients. This tells us that these variables had a positive relationship with the demand for EMS. As the percent of persons living alone or those separated/divorced increases in an EA, so does the demand for EMS. Also, the relative importance of each variable can be assumed by the relative absolute magnitude of the partial correlation coefficient. In this case the partial correlation coefficient for PSEPDIV is minutely higher than that of PLIVEALO, suggesting that both variables have very similar importance in this model.

The percent of people living alone has been suggested as a determinant of EMS demand in past literature. Siler (1975) claims that the positive effect of his variable "Square of Housing Units per Area Resident" -which he translates to low number of occupants per household- on EMS demand "is suggestive of areas with relatively large numbers of senior citizens and/or youthful single persons"(Siler, 1975: 261).

Table 4.2 EMVR and MMVR Model Summaries

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
EMVR	.637	.405	.394	12.745
MMVR	.730	.533	.518	18.160

Table 4.3 EMVR and MMVR Regression Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
EMVR	(Constant)	-4.396	3.375		-1.302	.196					
	PSEPDIV	1.143	.387	.345	2.955	.004	.597	.283	.228	.436	2.293
	PLIVEALO	.346	.120	.336	2.873	.005	.595	.276	.222	.436	2.293
MMVR	(Constant)	108.801	34.533		3.151	.002					
	PLIVEALO	1.001	.106	.697	9.481	.000	.649	.641	.571	.671	1.491
	W_PADLT	-1.731	.542	-.200	-3.193	.002	-.268	-.271	-.192	.923	1.083
	W_PAPT	.501	.123	.319	4.065	.000	.361	.337	.245	.589	1.698
	W_PLIVEA	-.843	.180	-.401	-4.690	.000	.220	-.382	-.282	.495	2.020

a Dependent Variable: RTSP1000

De Angelis'(1995) interpretation of Siler's (1975) findings is that "[s]ingle occupant households would appear to rely on ambulances in an emergency, whereas larger households would rely on their spouse, sibling, or house mate to provide transportation to emergency medical facilities"(De Angelis, 1995:26). This theory is not implausible given the significant volume of emergency department patients that are brought in by private vehicle or cab[11]. This suggests however, that EMS may be under-utilized by people who need immediate care at home but who, through lack of awareness, do not take advantage of life saving techniques can be performed at their home and en route to the hospital by trained paramedics. De Angelis (1995) also finds (for all but one of his many regression equations) that as the number of persons per household decreases, EMS rates increase. It is important to note that "Average Number of Persons per Household," did not present itself as an explanatory variable in the stepwise regressions attempted here, but it is very strongly correlated with PLIVEALO (-0.931 Appendix 2)

Variables relating to marital status have also been clearly associated with EMS demands; with single, divorced, or separated individuals usually associated with increased demand (Szplett, 1988; Siler, 1975). Aldrich et al. (1971), however, finds that demand increases for single males but decreases for single females..

[11] This is simply anecdotal evidence from the author's personal experience as a paramedic.

Analysis of EMVR Residuals

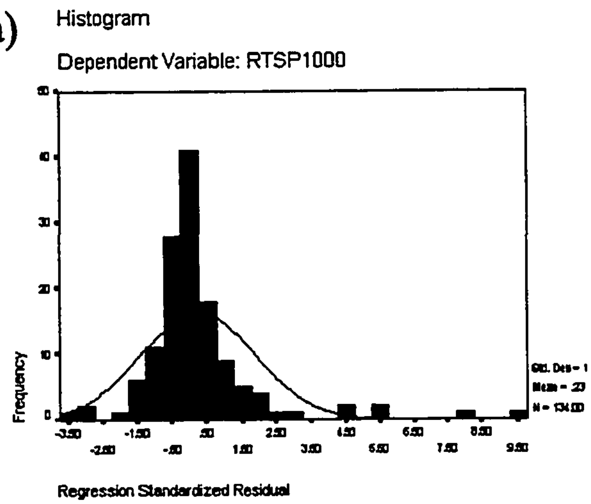
An ideally specified regression model would result in residuals that are normally distributed, independent of each other, and have no relationship with either the dependent or independent variables. These assumptions are often violated to some extent. For this reason an exploration into residual behaviour is important in order to determine the extent of violations, and to suggest some mitigating strategies, i.e., hints as to how to better specify the model.

As a residual diagnostic, Figure 4.13a, a histogram of regression standardized residuals for the EMVR, clearly indicates that the assumption of normality has been violated. Figure 4.13b is a scatterplot of standardized residuals and predicted values of RTSP1000. There should not be a discernable pattern to the relationship between these values on the scatterplot. But clearly, the spread of residuals increases with the magnitude of predicted values, suggesting that the equality of variance assumption may also have been violated.

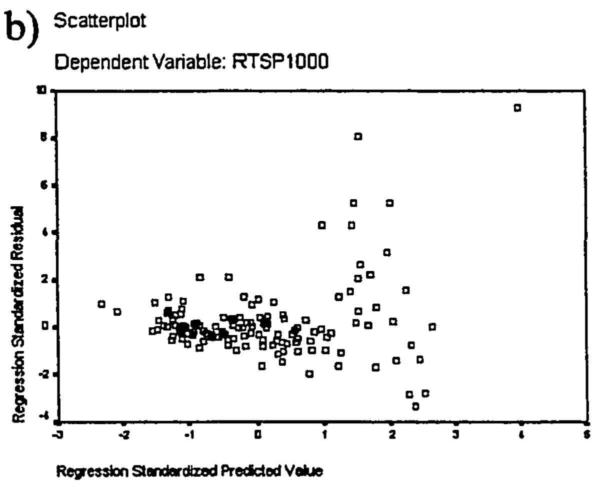
Fortunately, for the geographer, determining the degree of independence in residuals is relatively simple. A simple map of residuals may be quite revealing in a landscape where residuals are highly correlated. Figure 4.15 is a map of the residuals of the EMVR. There appears to be concentrations of large positive residuals near the core of the city and to the northeast in the New Sudbury commercial area. Areas with large positive residuals (observed - predicted) would be those where the model has tended to under-predict, and areas with large negative residuals where it has tended to over-predict EMS demand.

Figures 4.13 a and b, EMVR residual diagnostics

a)

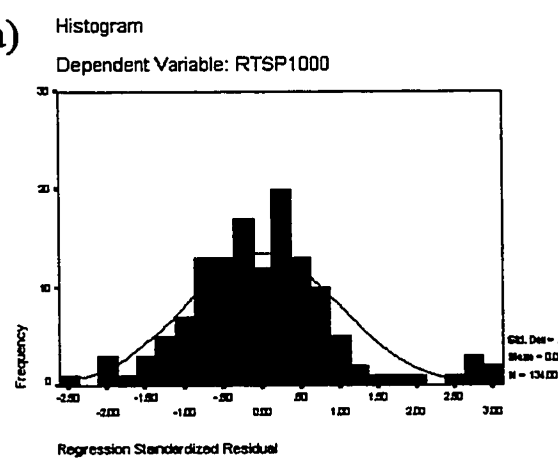


b)



Figures 4.14 a and b MMVR residual diagnostics

a)



b)

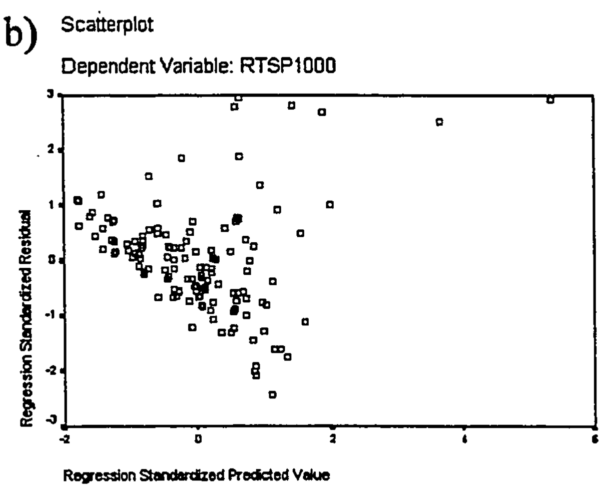
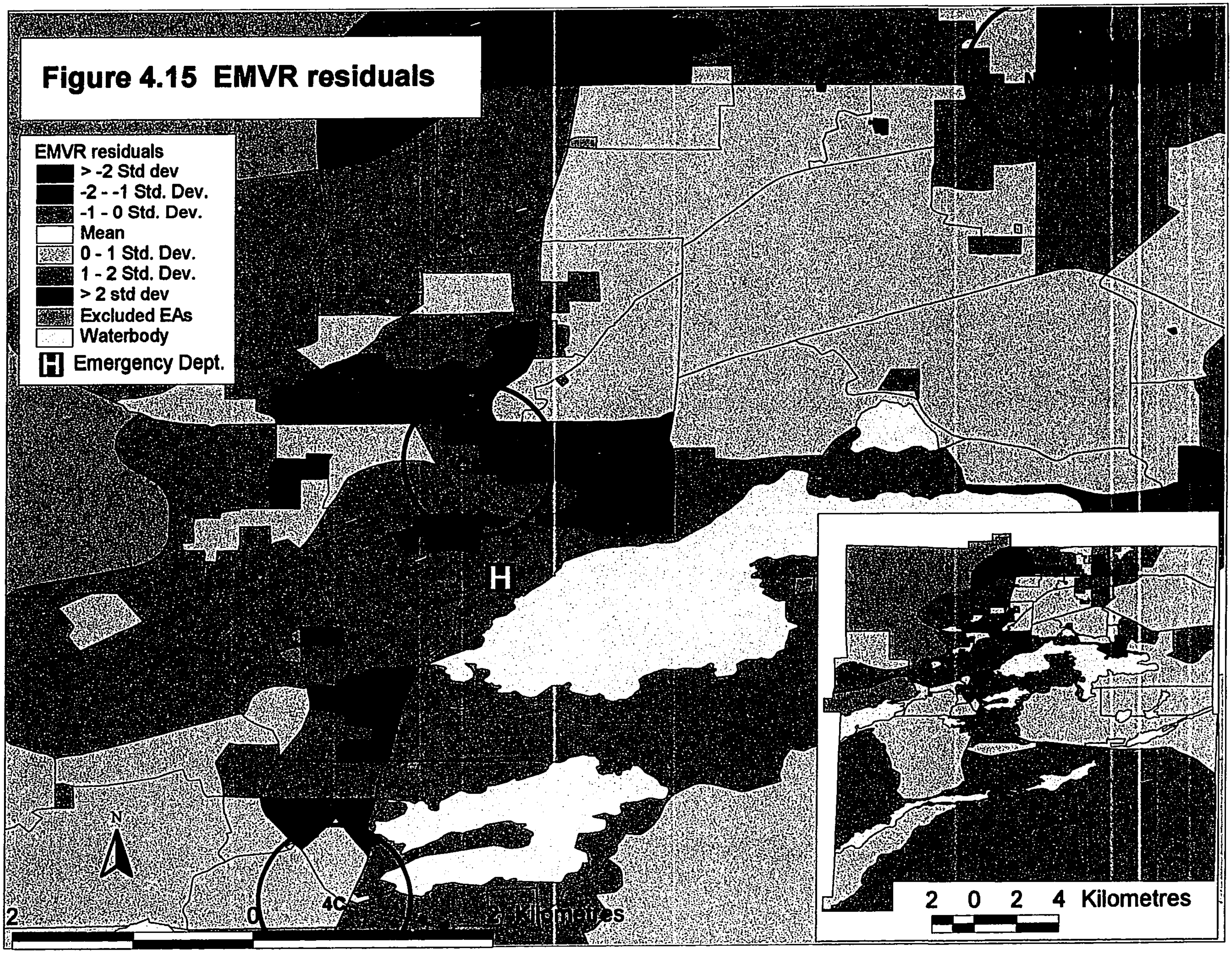


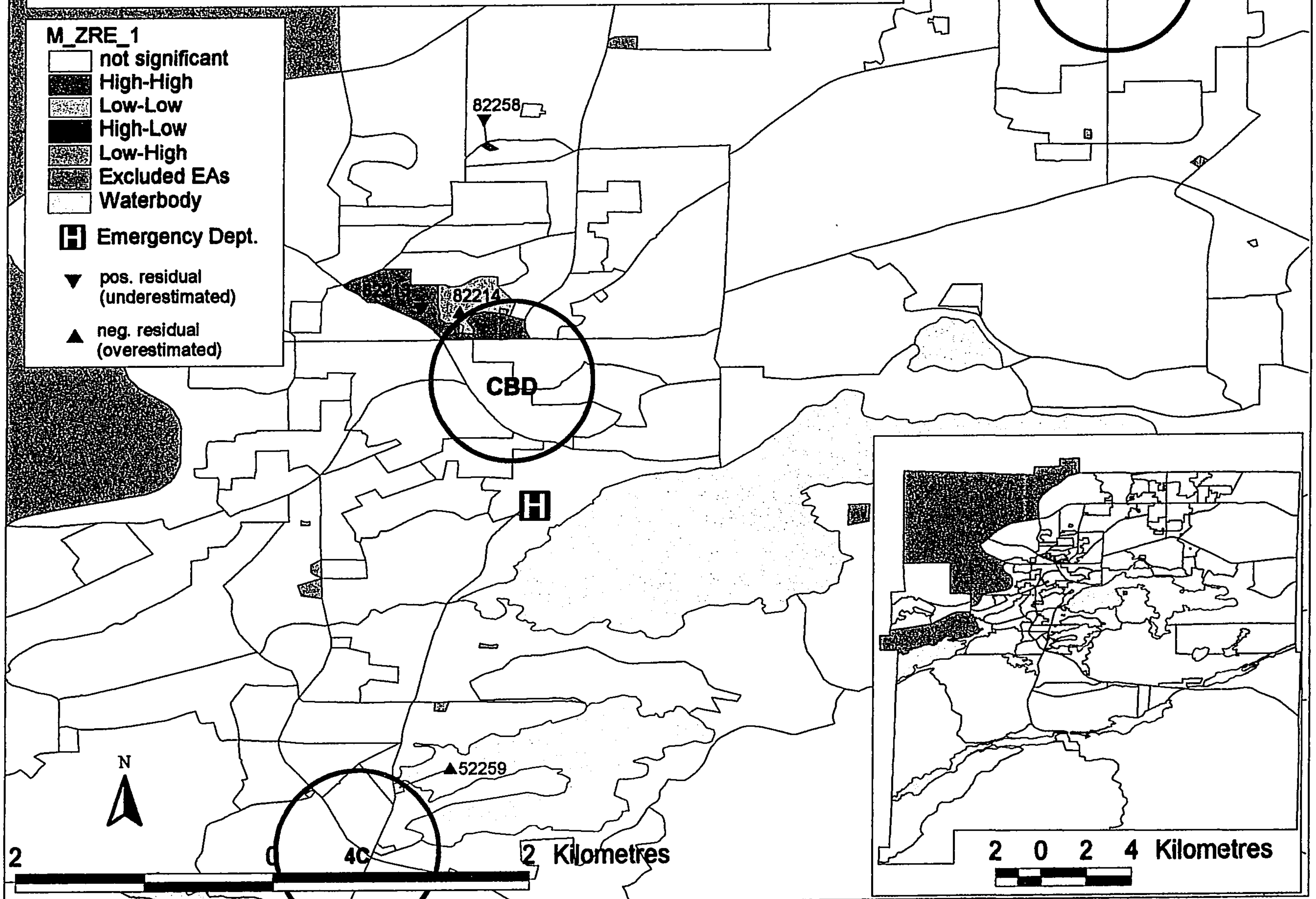
Figure 4.15 EMVR residuals



Maps can be deceiving in the face of subtle patterns, so the spatial analysis can turn to global and local measures of spatial dependence to examine this important assumption in regression analysis. A global Moran's *I* of -0.007664136 (p=0.99781) was calculated for the residuals of this first regression. It appears that there is no spatial dependence in the residuals. However, a global Geary's *c* of 0.7730116 (p=0.02627) refutes this assumption of independence, and, given the non-normal distribution of the residuals, is probably a more appropriate test of spatial autocorrelation. The observed value of Geary's *c* is quite close to the expected value of 1 for the case of no dependence, therefore, though it is statistically significant, it is certainly only weakly positive. The map pattern may again possibly be responsible for some nuisance spatial autocorrelation. Figure 4.16 highlights those EAs with statistically significant Local Moran's *I*. This regression is preliminary, so an in-depth discussion on the significance of these tests is withheld at this point. Suffice it to say that once again, the core of the city is highlighted.

With the residuals showing violations of normality, independence, and equality of variance, there is cause to re-specify the regression model. As Bailey and Gatrell (1995: 277) warn, transformations will not generally relieve the problem of spatially correlated residuals or second order effects. It makes sense then to identify spatial effects before attempting any transformations of the data. However, unlike other studies in EMS demand, this work attempts to re-specify by first identifying spatial effects in the explanatory variable and subsequently incorporating them in a mixed aspatial/spatial regression equation. This strategy will prove to be beneficial in mitigating two of three violated assumptions.

Figure 4.16 Local Moran Significance - EMVR residuals



4.3.4.2 Testing for Spatial Effects in the Original Variables

Twenty four of the thirty one variables[12] were found to have significant Moran's *I*. All 24 demonstrated significant positive spatial autocorrelation ($p<0.05$), no significant negative autocorrelation was detected (Table 4.4). The degree of positive spatial autocorrelation ranged from quite weak [0.099 for "Average Gross Rent" (AGRRENT)] to considerably strong [approximately 0.46 for "Average Number of Persons per Household,"(ANUMPHLD) and "Percent of People Living Alone" (PLIVEALO)]. The two "explanatory" variables from EMVR, PLIVEALO and PSEPDIV, are included in the list of original variables showing significant spatial autocorrelation. The statistics were calculated in SpaceStat using a weights matrix of row standardized first order neighbours (queen's case). A randomization assumption was used to test significance because of the non-normal distribution of many of the variables.

Subsequently, SpaceStat was used to create a lagged version of those variables with identified spatial dependence. For each variable z_i in each EA, SpaceStat computes a new variable, prefixed *W_z* in the output, based on the weighted mean of its first order (lag 1) neighbours as defined in section 3.3.5.1 above. SpaceStat does not include cell z_i value in this calculation (Anselin, 1992: §18.2.1). The formula for this calculation can be written as $\sum_j w_{ij} z_j$.

[12] Distance from hospital by definition naturally had significant spatial autocorrelation and is therefore not included as a candidate for lagging.

Table 4.4 Moran's *I* test for Global spatial autocorrelation of explanatory variables (randomization assumption). Weights matrix is row standardized

VARIABLE	*I*	PROB
RTSP1000	0.095106	0.051430
DISTANCE	0.885866	0.000000
UNEMRT	0.176902	0.00031
A95HSINC	0.143955	0.005952
STDEINC	0.063321	0.191765
AVALDWL	0.085349	0.089664
AGRRENT	0.099026	0.051751
PWHTCOL	0.254913	0.000002
PSNR	0.197227	0.000096
PCHILD	0.155705	0.003080
PTEEN	0.209003	0.000064
PADLT	0.121153	0.015798
PMDAGE	0.100455	0.046467
PABOR	0.160715	0.002059
PCITOCAN	-0.054252	0.362599
PNOHOLAN	0.163195	0.001950
POPDENS	-0.009263	0.972357
PEDLO9	0.171953	0.000806
PLIVEALO	0.458234	0.000000
POWNED	0.392734	0.000000
PSEPDIV	0.303313	0.000000
PCAR	0.334285	0.000000
PMOVE1	0.170020	0.001274
PMOVE5	0.166556	0.001612
PLNPRNT	0.161194	0.002083
PCSNR	0.073909	0.139639
PCVISMIN	-0.007752	0.996307
PTOW	0.252076	0.000002
PSUBURB	0.408742	0.000000
PAPT	0.272930	0.000000
ANUMPHLD	0.460153	0.000000
EMPPOPR	0.308438	0.000000
PSECED	-0.019549	0.82394

4.3.4.3 Mixed Multivariate Regression (MMVR)

A mixed (aspatial/spatial) multivariate regression was then performed using the two explanatory variables PLIVEALO and PSEPDIV from the EMVR above and the 24 spatially lagged variables. Of course the response variable is still RTSP1000.

The stepwise method excluded 22 of the 26 proposed variables resulting in a solution with four explanatory variables. The adjusted coefficient of determination R^2 of 0.518 (Table 4.2) offered significant improvement over the first regression. The solution included PLIVEALO, W_PLIVEA (lagged percent persons living alone), W_PAPT (lagged percent row housing, and apartment dwelling <5 stories), and W_ADLT (lagged percent adult-aged 20 to 64) (Table 4.3). Nearly 52 percent in the variation in Rate of EMS calls per 1,000 residents per enumeration area is now explained by the model using these four variates.

PLIVEALO and W_PAPT have a positive association with the rate of ambulance calls, as indicated by their respective partial correlation coefficients presented Table 4.3; while W_PADLT and W_PLIVEA have negative partial correlation coefficients.

Interestingly, PLIVEALO persists as an explanatory variable from the first attempt. This suggests that the percent of persons living alone is an important determinant of EMS demand. SEPDIV has now been dropped from the equation to be replaced by three lagged variables, suggesting that these slightly larger scale variates are more closely associated with demand.

When looking at the bivariate correlation matrix (Appendix 2) it seems quite clear why PLIVEALO has persisted as an explanatory variable. It has an absolute correlation coefficient greater than 0.625 with nine other independent variables which include:

- average income
- percent seniors
- percent child
- percent low education
- percent separated/divorced
- employment to population ratio
- percent suburbs
- percent owned
- average number of persons per household

What is significant here is that many of these variables have in past studies been associated with EMS including average income, percent seniors, percent low education, percent separated/divorced, and employment to population ratio. The strong relationship with the variable PLIVEALO may explain why these expected variables did not appear as significant in either regression - with the exception of PSEPDIV. Furthermore, PLIVEALO also has a moderate (>0.4) correlation coefficient with unemployment rate and population density, which have shown in the past to be related to EMS demand.

Moreover, the lagged version of this variable, W_PLIVEALO has the second largest partial correlation coefficient. However, in this case, it is negative. This reinforces the power of the variable "Percent Live Alone," but emphasizes the localized scale at which this variable operates to affect EMS demand. If we consider the EA as the scale at which EMS demand is generated, high percentages of people living alone in the EA generate high EMS demand. But EMS rates in an EA also increase as the percent of people living alone in its first order neighbourhood decreases, i.e., when an EA is distinctively different than its immediate neighbourhood.

W_PAPT, has a positive relationship with RTSP1000. The variable is a lagged version of PAPT, i.e., the percent of small apartment buildings (<5 stories), townhouse complexes, and apartments in detached dwellings in an EA. The source variable, "Percent Apartments," has moderate positive bivariate correlation with "Percent Lone Parent" (0.406), and a larger negative (-0.560) correlation with "Average Household Income." Also, the variables "Percent Movers" (one year and 5 years) have bivariate correlations with "Percent Apartment" of 0.540 and 0.518 respectively. Certainly others (Szplett, 1988; Kvålseth & Deems, 1979; Siler, 1975; Hisserich, 1971; Aldrich et al., 1971) have identified commercial land use (or some derivation of it - see the literature review) as an important component of their regression models. More important than capturing the urban structure of the city, the lagged "Percent Apartment" variable, given its correlations with "Lone Parent" and "Movers" variables, probably represents areas of high mobility. High mobility is typical of low income residents. The fact that the lagged version of the variable "Percent Apartments" entered this equation, rather than in its original form, speaks to the slightly broader scale (albeit only a first order neighbourhood in this case) at which the variable has a relationship with EMS demand in an EA. EMS demand in an EA is a function of the dwelling characteristics of its immediate neighbouring EAs.

Finally, the variable W_PADLT, is a lagged version of the age group from 20 to 64. The partial correlation is negative. Not surprisingly, this variable has a significant negative correlation (-0.656) with "Percent Seniors." With a moderate positive correlation with "Employment to Population Ratio" (0.434), and a negative (-0.446) correlation with "Education Below Grade 9," this variable suggests a dimension of education and work. Again, the variable is lagged and therefore, the scale at which it operates is broader than just the EA. This suggests that EAs with high EMS rates are associated with first order neighbourhoods that do not have a large proportion of people aged 20 to 64 (a working age group with some education).

Analysis of MMVR Residuals

If we again use the residuals to evaluate the model assumptions, we find significant improvements with this second model. The histogram (Figure 4.14a) shows much more normally distributed residuals. The handful of errors at the positive tail of the distribution may support the concept of a separate spatial regime where the model has tended to under-predict y. Figure 4.17 is a map of standardized residuals of the MMVR. Concentrations of high or low residual values are not very apparent visually.

If we consider the degree of spatial association in the residuals of this model, Global Moran's I and Geary's c for the residual are -0.0484038 (p=0.4569) and 0.8725342 (p=0.07239), both not significant. This supports the visual impression suggested above. So far, the Mixed MVR seems to be reasonably specified in that the assumptions of normality and independence are not violated.

Finally, the scatterplot of regression standardized residuals versus predicted values (Figure 4.14b) clearly shows a pattern of spreading as predicted values increase. This is indicative of non-constant variance, a violation of an important assumption in regression analysis as indicated earlier.

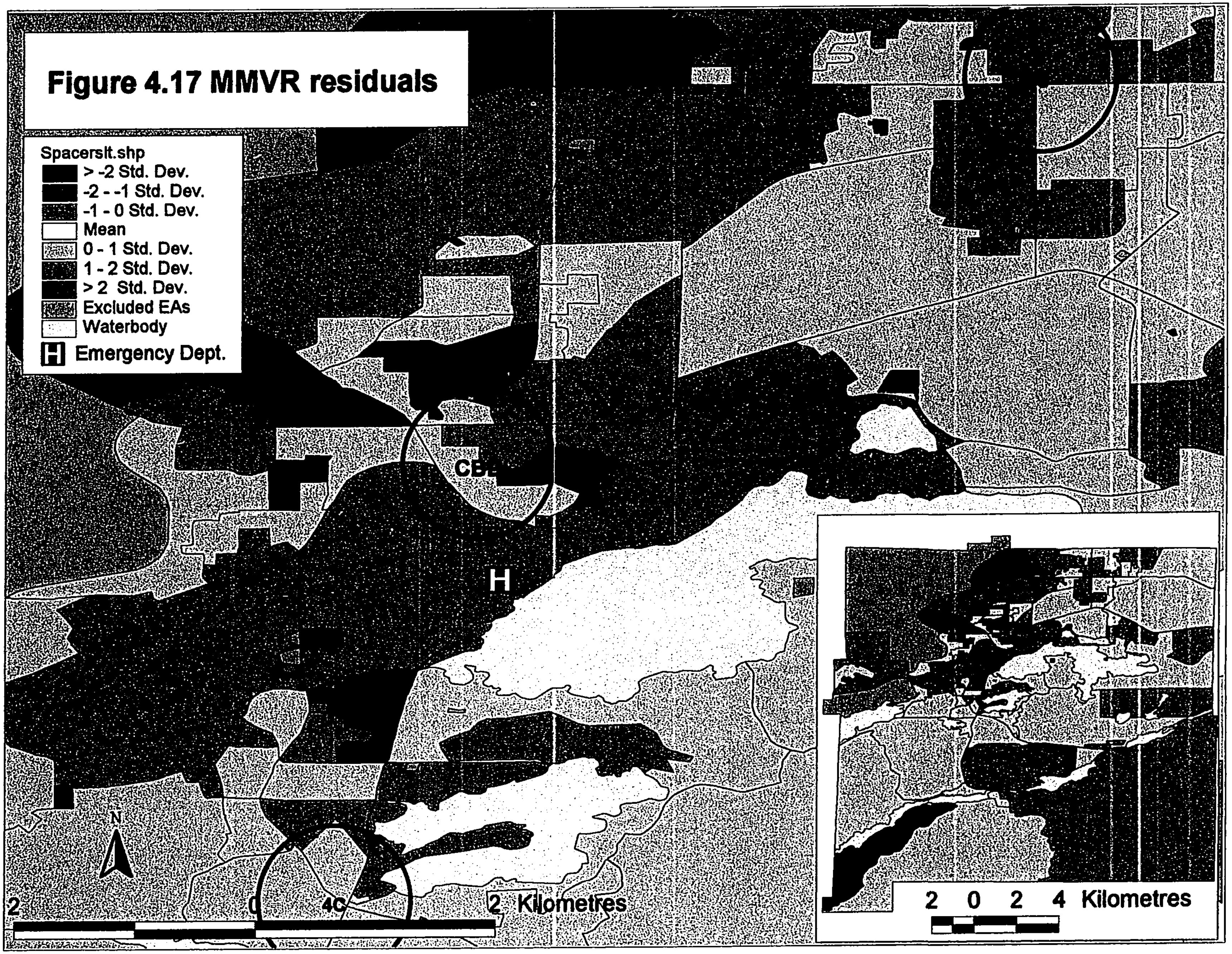


Figure 4.17 MMVR residuals
Spacerslt.shp
> -2 Std. Dev.
-2 - -1 Std. Dev.
-1 - 0 Std. Dev.
Mean
0 - 1 Std. Dev.
1 - 2 Std. Dev.
> 2 Std. Dev.
Excluded EAs
Waterbody
H Emergency Dept.
CB
H
4C
N
2 0 2 Kilometres
2 0 2 4 Kilometres

Nevertheless, the behaviour of the residuals of the mixed aspatial/spatial regression (MMVR) has improved considerably over the previous aspatial regression (EMVR) both in terms of normality and independence. The MMVR seems to be a better model, and one that has taken into account the spatial characteristics of the explanatory variables proposed for the equation. Unfortunately, it fails the test of constant variance, but so did the first regression.

Local Statistics Tests of MMVR Residuals

There is further opportunity to investigate the presumption of separate spatial regimes of EMS demand in Sudbury. If the regression model fails in distinct areas, it could be assumed that a different model needs to be considered there. Though the global statistic indicates that overall there is no dependence in the error terms, using Local Moran's *I* may help to identify areas where the model has failed or tended to significantly over or under predict the dependent variable.

Figure 4.18 reveals 6 EAs with significant Local Moran's *I*. This map indicates three characteristics of the residuals. First, only EAs with residuals that have significant LMIs are mapped. Second, as the legend indicates, the Moran scatterplot quadrant (high-high, low-low etc.) of an EA residual is indicated by the colour in the legend. And third, the triangles indicate whether or not the residual over or under-predicted the dependent variable in that EA. The inverted triangle ▼ indicates a positive residual where values for Y were under-predicted, and the triangle ▲ a negative residual where Y was over-predicted. Figure 4.19 is also included to show the position of these EAs in terms of their Moran scatterplot quadrants.

Figure 4.18 Local Moran Significance - Mixed Regression

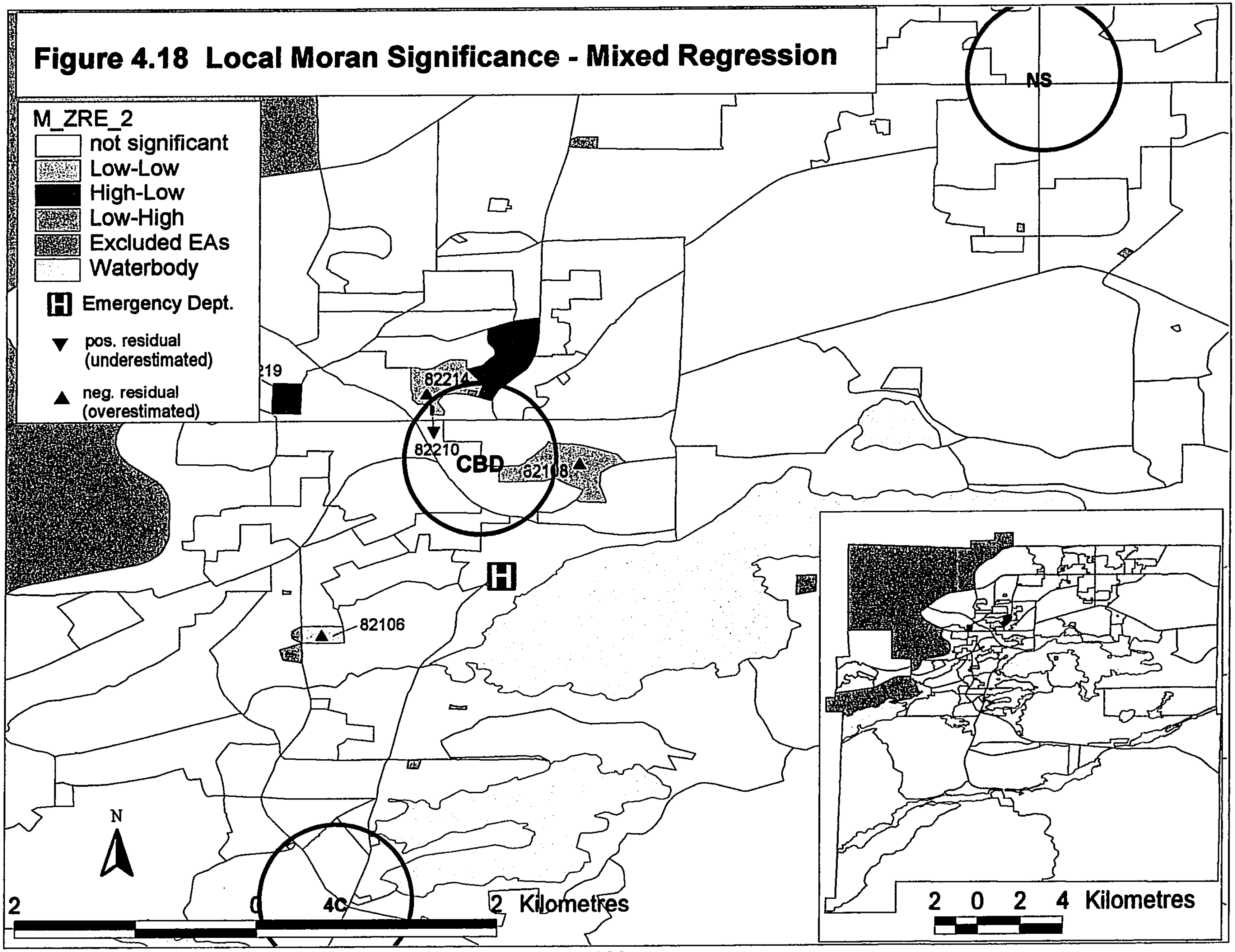
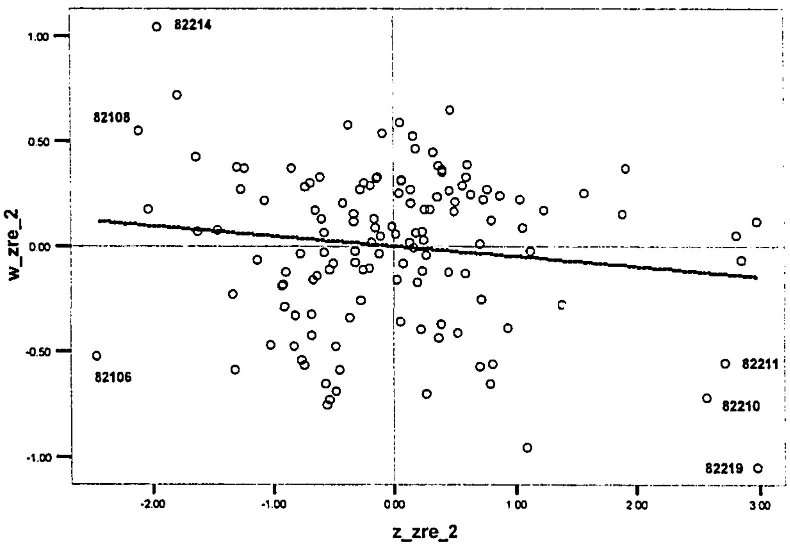


Figure 4.19 Moran Scatterplot of residuals from the Mixed Spatial Multivariate Regression



Three EAs have high residuals surrounded by low residuals, i.e. negative association, therefore, are found in the lower right quadrant of the scatterplot. The positive residuals ▼in these three EAs indicate that the model has under-predicted the rate of EMS here. It is not surprising, therefore, that these EAs were originally identified as having extreme high values for the variable RTSP1000. They include 1) a very small EA (82210) with a rate of 139 calls per 1,000, that is an apartment tower very near the core of the city, 2) an EA (82211)with a rate of 108 calls per 1,000, that encompasses 2 seniors buildings, and 3) to the west of the core, an EA (82219) that had 89 calls per 1,000. This last EA as discussed above, includes a community health care centre. The two EAs in the core of the city (82210 and 82211, Figure 4.18) also had significant positive LMIs in terms of RTSP1000 (high values surrounded by high valued neighbours; Figure 4.11). The EA to the west of the core did not.

The two EAs indicated (82214 and 82108) in the upper left quadrant of the scatterplot (again, negative autocorrelation) represent EAs with small residuals surrounded by relatively large residuals (Figure 4.17). The negative ▲ values of the residuals in these locations indicate that the model has over-predicted the dependent variable in that EA. One of these two EAs (82214) was identified earlier as having a significant negative Local Moran's *I* for its EMS rate. This EA has a low call rate and is surrounded by EAs with high call rates.

One EA in the south-central part of the city (EA 82106) also has a negative residual, but it is found in the lower left corner of the scatterplot (positive association) because it is surrounded by other relatively small residuals (Fig 4.17). This then is another area where the model has over-predicted the rate of EMS use. This EA has a very small population and produced no recorded emergency calls in 1999. The general area in which this EA is situated is known as the hospital area and is generally considered reasonably affluent. However, this affluence cannot be determined because no income data is available because of the small population there.

The purpose of using local statistics in regression residuals analysis is to attempt to identify hotspots or coldspots of the predictive power of the regression equation. The assumption is that if the model fails in specific areas, this may again reinforce the suggestion of a separate spatial regime.

In fact, significant local statistics of the residuals do highlight EAs in the core of the city that have first been identified as extreme values in terms of the dependent variable, and in some cases have shown significant local positive spatial autocorrelation.

If significant clusters of high residuals exist there, then the regression model, even when including lagged variables, is failing. It seems that the model fails in both over and underpredicting EMS demand where there are many extremely high values, i.e., near the core of the city. Certainly, this would suggest that there is in fact a separate spatial regime within the core of the city and that there are very localized effects. Notably, unlike the previous regression there is no significant global spatial autocorrelation in the residuals, therefore, our confidence in the significance of the local statistics is deepened.

4.3.5 Part B Summary

This part of the chapter has reported first on the results of the standard regression equation and then a mixed aspatial/spatial multivariate regression equation. This first model (classically aspatial) had modest explanatory power, but diagnostics of the residuals revealed that the assumptions of normality, independence and constant variance had been violated.

In an attempt to improve the specification of the regression, spatial autocorrelation inherent in many of the explanatory variables was first considered. Those that had significant spatial autocorrelation were transformed to lagged form based on their immediate neighbourhoods.

These lagged variables were then incorporated into a regression along with the two explanatory variables from the first equation. The result was a considerable improvement in the explanatory power of the model but more important, the residuals behaved much better. This new regression's residuals were approximately normally distributed and tests showed no significant global dependence.

However, a plot of the residuals clearly indicated that the assumption of constant variance was violated. Violation of this assumption, however, may support the concept of separate spatial regimes of EMS demand in the city of Sudbury. Coupled with the obvious outliers of EMS rates in the core of the city and the presence of local spatial association there, it would seem that at least two spatial regimes might exist.

Additionally, though the residuals showed no global spatial dependence, local statistics showed that there were some localized areas, including the core of the city, where the regression equation over-predicted and under-predicted the rate of EMS calls per 1000 residents. This too would support the idea of a separate spatial regime in the city core.

Interestingly, the pattern of demand seems reminiscent of Burgess' Concentric model of urban land-use as presented in Carter (1995). However, the city of Sudbury is relatively small and young, therefore, it is difficult to presume that other models of urban structure such as Hoyt's Sector, or Harris-Ullman, Multi-nuclei models (Carter, 1995) would not be just as apparent in time. It is beyond the scope of this thesis to place EMS demand within urban models, however, future work in larger centres may reveal some interesting results. Certainly, the inclusion of land-use variables should be considered if a comprehensive EMS database were available. Not limiting the analysis to emergency calls only and including calls where patients are picked up while at work or in public places may more closely replicate results from previous work.

4.4 Chapter Summary

In the first part of this chapter, problems with both the EMS events file (ambulance call location addresses) and spatial reference file (digital road network) were identified as a result of poor address matching hit rates. After data were cleaned, significantly better hit rates were achieved, and a sample demonstrated that these addresses were geolocated in their proper Enumeration Areas. The second part described the selection criteria for EMS calls and Enumeration Areas in order to prepare a dataset for final analysis. The third part focussed on the EDA and ESDA and regression modelling results. A number of proposed explanatory variables were found to have significant spatial autocorrelation. In the end, a parsimonious (four explanatory variables), reasonably specified (spatial effects taken into account, and well behaved residuals) regression model was developed that offers reasonable explanatory power. The rate of EMS calls per 1000 persons in an Enumeration Area can be modelled by considering "Percent People Living Alone" at a very localized scale, i.e., the EA level, the "Percent of Apartment Dwellings" at a slightly broader scale, i.e., the first order neighbouring EAs, and the "Percent of People Aged 20 to 64" again in the first order neighbourhood.

The next chapter will highlight and summarize the findings of this work. The limitations of the current work are identified and recommendations are made to address those limitations. Suggestions for future research in the field of EMS demand are also proposed.

Chapter 5 Summary and Conclusion

5.0 Introduction

The purpose of this chapter is to summarize the work that was done, to highlight its findings and to identify its limitations. Recommendations are made to mitigate each limitation identified. Finally, some concluding remarks are offered to place this thesis and its contributions within the context of EMS demand research.

5.1 Purpose and Summary of Findings

The goal of this research was to examine the relationship between EMS demand and demographic, socioeconomic and geographic variates in the city of Sudbury, Ontario. However, unlike previous work examining EMS demand in other jurisdictions that used classic regression methods, the spatial nature of the problem was addressed here. To do so, EMS data were used in a Geographical Information System (ArcView 3.2) to geolocate ambulance call events onto a spatial framework for which 1996 census data were available. This was achieved by address matching data collected by the Sudbury Base Hospital Program against a commercially available digital road network. A spatial analysis program (SpaceStat 1.90) was used in conjunction with the GIS to operationalize a spatial model, and perform a battery of spatial diagnostics to mitigate any misspecification of the regression model used to explore EMS demand.

Initially, only 18% of EMS events were successfully address-matched using ArcView GIS. However, after careful data cleaning, a hit rate of 96% was achieved.

These results were consistent with those submitted by Vine et al. (1997) and Tobias et al. (1996) as reported in Fellers (2000: 82). Furthermore, a sample of 100 observations was used to demonstrate that 98% of cleaned addresses are geolocated in their proper Enumeration Areas.

The EMS demand rate per 1,000 residents in the 134 EAs ranged from 0 to 179 with a mean of 23.5. This mean is slightly lower than previous work which typically report between 30 and 40 calls per 1,000 persons. It is important to note that only calls for which a patient's "Pick Up Location" was the same as their "Mailing Address" were used in this thesis. As Braun (2000) insists, this assures a more meaningful ecological analysis when using census data. The distribution of the EMS demand rate in Sudbury was quite positively skewed, with a median of 16 and mode of 9 EMS calls per 1,000 residents. Spatially, the distribution of higher call rates was clearly centred near the core of the city but also included EAs that contain apartment buildings that are known to house significant numbers of senior residents outside the core.

A global Moran's *I* test of the EMS rate variable revealed weak positive, but questionably significant, spatial autocorrelation. However, a global Geary's *c* test confirmed the statistical significance of the Moran's *I* test. A spatial correlogram revealed weak spatial autocorrelation only to a second order lag, suggesting that there are no marked city wide patterns of variation in EMS use in Sudbury. Local Moran's *I* statistics indicated statistically significant concentrations of high EMS rates in the core of the city. This supports the visual impression of mapped EMS rates that showed a concentration of high call rates in the core of the city interrupted by pockets of high values to the northeast and the south.

Two stepwise regressions were used to propose explanatory variates of EMS demand in Sudbury. The first model attempted was quite similar to previous models, particularly in that it was aspatial. It was used here as an exploratory tool. The aspatial regression equation explained only 32% of variation in EMS demand using two variates. This result is significantly lower than previous American (and one Canadian) studies that reported R^2 's in the order of 0.8 and greater (Cadigan & Bugarin, 1989; Szplett, 1988; Kvålseth & Deems, 1979; Schuman et al., 1977; Siler, 1975; Gibson, 1971; Aldrich et al., 1971). However, an analysis of residuals from this aspatial model showed evidence of violations of regression assumptions. Residuals were found to be non-normally distributed, and a significant degree of spatial dependence in the residuals was suggested using global Geary's *c*, and less definitively using global Moran's *I*. Furthermore, twenty four of the 31 originally proposed explanatory variables were found to have significant global spatial autocorrelation (the distance from the hospital is not included here as it naturally would have significant spatial autocorrelation).

Therefore, a second regression was proposed to address the issue of spatial dependence. Those proposed explanatory variables that showed significant spatial autocorrelation, were then converted to lagged versions using a simple weights scheme in SpaceStat. These lagged variables were then incorporated into a regression equation which included the explanatory variables from the first regression.

The final model explained 52% of the variation in EMS calls per 1,000 persons with the variates: "Percent People Living Alone" and its lagged version, a lagged version of "Percent of Apartment Dwellings" and lagged "Percent of People Aged 20 to 64." Though not identical to previous studies, the results are somewhat consistent in that age and variables suggesting deprivation are evident.

However, neither model was perfectly specified; there was a noticeable spread of plotted residuals as predicted values of the dependent variable increased in both regression attempts. This suggests non-constant variance in the dependent variable. There were also EAs near the core of the city where regression residuals showed statistically significant positive and negative local spatial autocorrelation using local Moran's *I*. These represent clusters of EAs where this second regression model under and over-predicted the response variable near the core of the city. Though the significance of local statistics is suspect in the presence of global spatial autocorrelation, we should consider that the global spatial dependence detected may simply be the result of a complex map pattern. Together with high EMS rates and significant local Moran's *I* in these areas, these results suggest that further work is needed to explore a possibly separate spatial regime near the core of the city that requires a different regression model.

5.2 Limitations and Recommendations

There are a number of limitations of note in this work that are presented in this section along with specific recommendations to address these limitations.

5.2.1 EMS Data Availability

Due to limited availability of data, only EMS calls dispatched as "Urgent" and returned to the hospital (and, in some cases left at a residence for body removal services) as "Prompt," "Urgent," "Obviously dead," and "Legally Dead" were analysed. Essentially, this excluded a large portion of calls attended to by paramedics both dispatched and returned as lower priorities. Therefore, comparisons to previous studies in other jurisdictions is difficult because they do not typically identify both the priority to which an ambulance was dispatched and returned to the hospital. The use of a subset of total EMS calls would clearly govern the results discovered here. We should expect that, as we consider all calls together, very different explanatory variables would survive a stepwise regression analysis, possibly including some that have been proposed in previous work.

Also, the exclusion of lower priority calls in this thesis probably accounts for the lower average EMS demand rate found when compared to those discovered by Svenson (2000), Cadigan and Bugarin (1989), Siler (1975), Gibson (1971), and Aldrich et al. (1971).

Moreover, it is probable that much of the strength of previous regression models is a function of non-urgent EMS demand significantly related to demographic and socioeconomic variables. However, it may be impractical, and inconsistent with their mandate, for Base Hospitals (the only agency from which data were readily accessible at the onset of this study) to collect such data. Nevertheless, the non-urgent component of EMS demand, whether dispatched or returned, needs to be clearly identified and used in future attempts at modelling.

Also, the analysis undertaken here encompassed data from a full year. Siler

(1975) suggested that areas of climatic extremes may find seasonal variations in EMS demand. With multi-year data from ADDAS, it may be possible to acquire sufficiently large seasonal samples to explore the possibility that different variables may play different roles at different times of the year in determining EMS demand.

Recommendations:

- It is recommended that a more comprehensive database of all call types be used for future work. The provincial ADDAS data, currently downloaded regularly by municipal EMS services, could be used to this end, given the comprehensive nature of these data.
- Any concerns regarding the quality of ADDAS data could be addressed by data quality checks undertaken collaboratively between the province, Base Hospitals, and municipal EMS providers.
- Separate analyses involving urgent and non-urgent calls should be undertaken. The results may provide considerable insight into mis-use or abuse of precious emergency resources. Public education and injury prevention programs could be developed to address any discovered trends.
- Seasonal variations in EMS demand and its determinants could be undertaken using multi-year ADDAS data.

5.2.2 EMS Data Quality

The strategic placement of EMS resources is clearly linked to a sound representation of the location of events to which the service responds. Currently MoH's Geoplot program developed by Peters (1998) uses a one square kilometre grid as a spatial framework. This is very limiting when undertaking ecological analysis; a fact which Peters (1998) himself and De Angelis (1995) recognize. It is possible to more precisely locate EMS calls so that ecological analysis can be done at the finest scale possible, which Gesler (1986) recommends. However to do so accurately requires a high quality locational dataset.

The low address matching hit rate achieved initially in this work is indicative of a data quality problem. A number of error sources were suspected and error types were discovered. Many errors likely stemmed from data being abstracted from paper ACR to a digital database. However, it is impossible to assess, without direct access to the original ACRs, whether errors in patients' street addresses in the database were a function of:

- Data error collection by paramedics
- Data error transcription by data entry personnel
- Data error creation in the data entry process

Error types included:

- Spelling or typographical errors in street names, type or direction
- Use of outdated addresses
- Spelling or typographical errors in municipality names
- Wrong municipality names used
- Typographical errors in street numbers
- Missing address elements
- Transposition of address elements, particularly street number and apartment/suite number

Recommendations:

To mitigate the number and types of errors, and reduce any possible sources of errors found in EMS databases, it is recommended:

- that the importance of accurate locational data be recognized as a useful tool for the EMS administrator and planner,
- that paramedics be made aware of the importance of accurate data collection through policy and educational materials,
- that Electronic Ambulance Call Reports (E-ACRs) be generated in the field by paramedics using currently available technology to eliminate possible errors generated during the abstraction of paper ACRs to electronic databases,
- that E-ACRs s include a validation mechanism for patient's "Pick Up Location" and "Mailing Address" data entries,
- that "Pick Up Location" and "Mailing Address" entries in the E-ACR include separate fields for each address element (street number, street name, type, direction, apartment). Should this be impractical, at least the apartment/suite element be clearly separated from the street number to facilitate geocoding of addressed events,
- that municipal EMS compare E-ACR data to corresponding provincial ADDAS data as a means of data quality assurance. This mechanism could prove invaluable as a data quality check for both Ministry of Health and Long Term Care and the municipal EMS, and
- if E-ACR are not readily available, that frequent exploratory analyses (including address matching attempts) using ADDAS and Base Hospital data be made by municipal EMS, and errors be reported to the MoHLTC and Base Hospitals in order to ensure consistency and improve the quality of the EMS databases.

5.2.3 Spatial Reference Files

For EMS to effectively and efficiently respond to community needs, it is essential that demand be clearly understood. As is clearly demonstrated by this thesis, GIS offers unparalleled opportunities to represent the EMS demand landscape in a community. However, for accurate address data to be correctly geo-located, it is essential that a reliable spatial framework be used. The commercial product used here was found to be somewhat outdated and wanting. As municipal GIS becomes more commonly used, it is anticipated that the spatial frameworks -particularly digital road networks also become more comprehensive and accurate.

Recommendations:

- That complete and accurate, municipally generated, digital road networks be used by EMS planners as part of a GIS-based spatial decision support system.
- That, whatever the source of locational data (ADDAS, E-ACR, GPS, Base Hospital databases), EMS events should be geo-located using GIS in order to map a community's EMS demand landscape with particular attention to identifying areas of very high demand.
- That results of GIS analysis be considered in order to strategically deploy EMS resources to meet community needs.

5.2.4 Census Data

While the EMS events data used here represent calls for the year 1999, the ecological variates are from the 1996 census. Many researchers have faced this problem. (Aldrich et al., 1971; Siler, 1975; Schuman et al., 1977; De Angelis, 1995; Braun, 1998)

Recommendation:

- A similar analysis to the one presented in this thesis should be attempted using EMS and census data from the same year. Statistics Canada has changed the geography at which data are disseminated in the 2001 census. New boundaries, referred to as Dissemination Areas (DA), are slightly larger than the previous EAs, but still smaller than Census Tracts. Unfortunately, this may cause some difficulty in comparing results to EAs longitudinally. DA's may relieve the problem of spatial autocorrelation if they are appropriately delineated, but at the same time they introduce a degree of aggregation bias.

Another limitation concerning census data is its representation of residents in collective dwellings and the homeless. Previously referred to as "institutional populations," Statistics Canada only collects basic data for people residing in group homes, prisons, nursing homes and seniors' residences. As a result, EAs that represent collective dwellings were excluded in this thesis to make ecological analysis more meaningful. However, their exclusion may also account for the moderate explanatory power of the regression models compared to past studies. Moreover, not all collective dwellings comprise an entire EA. Smaller collective dwellings may well be engulfed by an EA, produce a high EMS call volume, but not include representative census data.

Of course, homeless people are not counted in a census, and therefore, not likely represented by ecological variables. In this study, it is likely that some homeless clients were included in the events file, as evidenced by the high EMS rate and a significantly high regression residual in an EA that has a community care facility which serves homeless populations.

Recommendations:

Two approaches are possible to mitigate the problem of representation of residents of collective dwellings and the homeless.

- One would exclude, from the events database, all EMS calls originating from such collective dwellings and community care facilities that homeless people might indicate as a mailing address. This would include not only institutional EAs but events at institutions within larger EAs. Again, this is not with the intent of ignoring the importance of demand generated by their residents, but rather to improve a regression analysis that uses available ecological variates.

- Perhaps a better approach is to include the homeless and collective dwelling residents by introducing a dummy variable. This approach is preferred as it would possibly highlight the importance of age and deprivation in the demand for EMS, which have, in this work, shown themselves only indirectly.

5.3 Suggestions for Future Research

In light of the limitations and recommendations listed above, there are opportunities for future research that may provide further insights into Emergency Services Demand. They include:

- A more complete picture of EMS demand should include calls performed at locations other than a patient's residence, for example at work or in public spaces. A land use variable would have to be assigned to census boundaries to make this type of analysis meaningful.
- Separate analysis by call priority (both dispatched and returned) should be undertaken to identify areas that generate high volumes of non-essential or low priority calls. Strategies could be developed to assure better usage of EMS in these areas to free up important emergency resources.
- Analysis should be performed in other communities to determine the generalizability of the model presented here. It is likely, that the uniqueness of communities will provide different explanatory variables, nevertheless commonalities may also be discovered. Using the methods presented here in larger, more developed cities may also further the concept of separate spatial regimes of EMS demand within communities.
- To further test for the existence of distinct spatial regimes in the City of Sudbury, separate stepwise regression analyses could be performed on EAs that closely represent the community's downtown and those that represent the rest of the city. We should expect different results than the ones discovered here, both in terms of the power of the models and the explanatory variables associated with EMS demand. We should also see different results between proposed spatial regimes. If the models showed some homoskedasticity, which the present models failed to demonstrate, we might gain further insight into the factors that drive EMS demand in differing parts of a city.
- Given the parallels with other emergency agencies, similar methods to the one presented in this thesis for EMS demand analysis could be undertaken by police and fire services to identify ecological determinants of their demand landscape.

Although it was convenient to place the EMS events onto a spatial framework for which census data was available, that framework and how we chose to operationalize a spatial model based on it determines to some extent the results of our analysis.

As we have discussed, the partitioning of an area of interest can determine the degree of spatial autocorrelation detected in a map. Griffith (2000: 155) and Griffith and Layne (1999:233) remind us that this "nuisance" spatial association increases with map complexity. Undoubtably, the EA configuration in the City of Sudbury is a complex one which may account for the weak global spatial autocorrelation found in the response variables, the explanatory variables and some of the residuals.

Also, the definition of spatial weights will affect which spatial units are neighbours and therefore the results of association between them and the variables within them. Though a very simple weighting scheme was used in this thesis, it is recognized that weighting schemes can be much more complex. For example, they can include length of common border, distance between polygon centres, distance decay functions *etc.*(Cliff & Ord, 1981: 17-19; Bailey & Gatrell, 1995: 262). Though Can (1996) found little difference in the results of spatial autocorrelation tests between simple and complex weighting schemes, it would be worthwhile to explore the results of various spatial structural models.

It is also important how we choose to define certain variables. For example, distance in this thesis was measured using straight line distance between each EA centroid and the Emergency Department. Other distances could well have been used such as network distance along the road network, distance from the CBD, or in perceived distance, travel time as it varies throughout the day. Though in this case, where the study site is relatively small, it is possible that people's perception of distance (travel time under varying traffic conditions) to the hospital may affect their choice to call EMS in case of an emergency. For some the time spent waiting for an ambulance may seem too long; they might choose to drive themselves or a patient to the hospital themselves. For others, the care provided by paramedics may be worth the wait. Studies into the perception of distance may provide some insight into EMS use.

Furthermore, research into EMS demand/use should not be restricted to regression analysis. Other epidemiological approaches such as probability mapping using relative risk, Poisson probabilities and Empirical Bayes estimates are some examples given in Bailey and Gatrell (1995) that lend themselves well to lattice data such as Enumeration Areas.

However, given that EMS events are initially represented as points in space, points pattern analyses including Nearest Neighbour and K function analysis could easily be undertaken. Bailey and Gatrell (1995) demonstrate that multivariate point pattern analysis is possible. Cross K functions are certainly useful to compare the distribution of events and populations at risk in an epidemiological context.

Bailey and Gatrell (1995) also introduce the work of Bracken and Martin (1989) as a means of analysing areal and point data. Given that areal data can be represented by a centroid (or weighted centroid based on population location) Bracken and Martin developed an adaptive Kernel estimation method that deals with what they claim is choropleth maps' tendency to "misrepresent the underlying distributions precisely in areas possessing extreme socioeconomic characteristics (Bracken and Martin,1989: 540)." An EMS demand surface generated from the geocoded locations could conceivably be compared to an underlying (or overlain) census data surface.

5.4 Conclusion

In the end, a parsimonious (four explanatory variables), reasonably specified (independent and normally distributed residuals) regression model is developed that offers adequate explanatory power. Small areas (EAs) with high a percent people living alone, surrounded by areas with a low percent living alone, high percent apartments, and low percent of persons aged 20 to 64 tend to generate high rates of EMS demand in the city of Sudbury, Ontario.

The success of this ecological analysis has been in great part due to GIS functionality. GIS proved invaluable as a tool to circumvent previous problems of using disparate spatial frameworks used in past Canadian studies. Also, by geolocating events in space, smaller geographic units (EAs) than those used in previous research help identify more localized hotspots of EMS demand. Furthermore, GIS allows researchers to use much larger samples of events than was feasible when location of EMS incidents were assigned to their census tracts or counties manually.

Moreover, the geographic reality of spatial dependence naturally violates the regression assumption of independence. However, though this violation has not been addressed in past EMS research, this thesis demonstrates that spatial autocorrelation can be addressed using a loosely coupled spatial analysis software. All this leads to more confident results.

Clearly, demand for Emergency Medical Services varies greatly from place to place within a community. And this variety is, in part at least, related to underlying demographic and socioeconomic realities. Strategic deployment of resources based on these realities could assure provision of more effective and efficient Emergency Medical Services. Also, many of the incidents to which paramedics are called are preventable through healthy lifestyle choices and safety awareness. Injury prevention and health promotion programs, usually administered through Public Health Units - with increasing involvement by paramedics-, could be targeted more precisely to groups and areas in need through the help of EMS demand analysis.

Appendix 1 Variables computations from raw data

Variable Name*	Long Name	Sample	Definition	Variables used to compute
DEMOGRAPHIC				
pop96c	Population 1996	100	Total Population, 1996	v0001
v0002	Total Population	100	Total population by sex and age groups (100% data)	v0002 used to calculate percentages because of random rounding
ttlchild	Total Child	100	m0-4, m5-9, m10-14 f0-4, f5-9, m10-14	v0004+v0005+v0006+v0028+ v0029+v0030
ttlteen	Total Teens	100	m15-19, f15-19	v0012+v0036
ttladlt	Total Adult	100	m20-24 to m60-64 f20-24 to f60-64	v0013 to v0021 + v0037 to v0045
ttlsnr	Total Senior	100	mf65 to 85+	v0022 to v0026 + v0046 to v0050
mdage	Total Middle Age	100	m/f 50 to 64	v0019 + v0020 + v0021 + v0043 + v0044 + v0045
Percent Calculation	using v0002 to contend with random rounding			
psnr	Senior	100	>64	ttlsnr / v0002
pchild	Child	100	<15	ttlchild / v0002
pteen	Teen	100	15 to 19	pteen = ttlteen / v0002
padlt	Adult	100	20 to 64 (adult working age)	padlt = ttladlt / v0002
pmdage	Middle Age	100	50 to 64	pmdage = mdage / v0002
pabor	Aboriginal	20	Total Aboriginal population / Total population by Aboriginal groups and non-Aboriginal population	v0470/v0469
pcitocan	Citizenship	20	Citizenship other than Canadian	v0121 / v0119
pnoholan	Non-official Home Language	20	Non-official languages / Total population by home language	v0323 / v0319
pvismin	Visible Minority	20	Persons, other than Aboriginal peoples, non-Caucasian / Total population by visible minority	v0781/ v0780

Variable Name*	Long Name	Sample	Definition	Variables used to compute
SOCIAL				
Education by 15 years and over by highest level of schooling v1338				
pedlo9	Below grade 9	20	Less than grade 9	v1339 / v1338
pseced	Completed Secondary Education	20	With Secondary School Certificate	v1342 / v1338
Occupation				
pcwtcol	white collar	20	(A-1034) Management Occupations (B0-1040) Professional Occupations in Business and Finance (C0-1047)Professional Occupations in Science and Natural Resources (D0-1050) Professional Occupations in Health (D1-1051) Nurse Supervisors and Registered Nurses (E0-1055) Judges, lawyers, psychologists, social workers, ministers of religion, and policy and program officers (E1-1056) Teachers and professors (F0-1059) Professional Occupations in Arts and Culture	(v1034 + v1040 + v1047 + v1050 +v1051 + v1055 + v1056 + v1059) / v1033
Living arrangement and mobility				
plivealo	persons living alone	20	1 person / Total number of private households by household size	v0114 / v0113
pmove1	Movers 1 year	20	Movers / Total by mobility status 1 year ago	v1378/v1376
pmove5	Movers 5 years	20	Movers / Total by mobility status 5 years ago	v1387/v1376
psepdiv	Separated / Divorced	100	Separated, but still legally married + Divorced / Total population 15 years and over by legal marital status	(v0054+0055)/v0051
plnprnt	Lone parent	20	Number of lone-parent families / Total number of census families in private households by number and status of family members in the labour force	v1574/v1567
pcsnr	Hours of care to seniors	20	<5 hrs, 5 to 9, 10+ hours of care to seniors / Population 15 years and over by hours of unpaid care to seniors	(v1291 + v1292 + v1293) / v1289

Variable Name*	Long Name	Sample	Definition	Variables used to compute
ECONOMIC				
emppopr	Employment: Population ratio	20	employment : pop'n ratio	v0800
uemrt	Unemployment rate	20	unemployment rate	v0801
a95hsinc	Avg Household Income	20	average 1995 household income	v1551
stdeinc	Std Error of Income	20	standard error of average income	v1552
agrent	Avg Gross Rent	20	average gross rent	v1601
avaldwl	Avg Value of Dwelling	20	average value of dwelling	v1581
GEOGRAPHIC-URBAN STRUCTURE				
distance	Distance	-	distance from centroid of EA to Emergency hospital	calculated using ArcView
pcar	Car to work	20	M/F Car, truck, van as driver/ M/F with usual place of work or no fixed workplace	(v1317+v1326) / (v1316+v1325)
popdens	Population Density	-	population / square kilometres	v0001 / sc_area
anumphld	Avg No. Persons per Household	100	Average number of person per household	v1559
powned	Owned Dwelling	20	Owned / Total number of occupied private dwellings	v1582 / v1578
psuburb	detached and semi-detached	20	detached and semi-detatched dwelling	(v0105 + v0106) / v0104
papt	Apartment	20	row house, apt in a detached dwelling, apt <5 stories	(v0107+v0108+v0110) / v0104
ptow	Tower	20	≥ 5 stories	v0109(>5 sto) /104

* "p" preceding a variable name indicates that a percentage was calculated

Appendix 2 Correlation Matrix - Response (RTSP1000) and Explanatory Variables

		RTSP1000	Unemployment rate	Average 1995 household income $	Standard error of average household income $	Average value of dwelling $	Average gross rent $
RTSP1000	Pearson Correlation	1.000	.378**	-.479**	-.304**	-.208*	-.151
	Sig. (2-tailed)	.	.000	.000	.001	.022	.085
	N	134	134	111	111	121	124
Unemployment rate	Pearson Correlation	.378**	1.000	-.530**	-.422**	-.261**	-.420**
	Sig. (2-tailed)	.000	.	.000	.000	.004	.000
	N	134	134	111	111	121	124
Average 1995 household income $	Pearson Correlation	-.479**	-.530**	1.000	.807**	.732**	.457**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000
	N	111	111	111	111	108	107
Standard error of average household income $	Pearson Correlation	-.304**	-.422**	.807**	1.000	.750**	.305**
	Sig. (2-tailed)	.001	.000	.000	.	.000	.001
	N	111	111	111	111	108	107
Average value of dwelling $	Pearson Correlation	-.208*	-.261**	.732**	.750**	1.000	.120
	Sig. (2-tailed)	.022	.004	.000	.000	.	.208
	N	121	121	108	108	121	111
Average gross rent $	Pearson Correlation	-.151	-.420**	.457**	.305**	.120	1.000
	Sig. (2-tailed)	.085	.000	.000	.001	.208	.
	N	124	124	107	107	111	124
PWHTCOL	Pearson Correlation	-.310**	-.361**	.751**	.634**	.636**	.340**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	134	134	111	111	121	124
PSNR	Pearson Correlation	.647**	.206*	-.372**	-.221*	-.100	-.083
	Sig. (2-tailed)	.000	.017	.000	.020	.277	.359
	N	134	134	111	111	121	124
PCHILD	Pearson Correlation	-.494**	-.084	.309**	.139	.128	-.151
	Sig. (2-tailed)	.000	.333	.001	.146	.163	.095
	N	134	134	111	111	121	124
PTEEN	Pearson Correlation	-.418**	-.162	.575**	.373**	.188*	.090
	Sig. (2-tailed)	.000	.062	.000	.000	.039	.321
	N	134	134	111	111	121	124
PADLT	Pearson Correlation	-.354**	-.108	.083	.091	-.029	.164
	Sig. (2-tailed)	.000	.215	.389	.340	.755	.068
	N	134	134	111	111	121	124
PMDAGE	Pearson Correlation	.107	-.026	.313**	.374**	.297**	.330**
	Sig. (2-tailed)	.219	.769	.001	.000	.001	.000
	N	134	134	111	111	121	124
PABOR	Pearson Correlation	.263**	.279**	-.392**	-.243*	-.165	-.293**
	Sig. (2-tailed)	.002	.001	.000	.010	.071	.001
	N	134	134	111	111	121	124
PCITOCAN	Pearson Correlation	-.057	.088	.013	.048	-.050	-.137
	Sig. (2-tailed)	.515	.313	.889	.620	.589	.129
	N	134	134	111	111	121	124
PHOHOLAN	Pearson Correlation	-.075	.084	-.059	.108	.042	-.188*
	Sig. (2-tailed)	.387	.334	.541	.260	.651	.037
	N	134	134	111	111	121	124
POPDENS	Pearson Correlation	.570**	.452**	-.261**	-.221*	-.066	-.058
	Sig. (2-tailed)	.000	.000	.006	.019	.469	.521
	N	134	134	111	111	121	124
PEDLO9	Pearson Correlation	.650**	.325**	-.596**	-.356**	-.296**	-.226*
	Sig. (2-tailed)	.000	.000	.000	.000	.001	.011
	N	134	134	111	111	121	124
PLIVEALO	Pearson Correlation	.649**	.407**	-.710**	-.440**	-.240**	-.268**
	Sig. (2-tailed)	.000	.000	.000	.000	.008	.003
	N	134	134	111	111	121	124
POWNED	Pearson Correlation	-.521**	-.359**	.760**	.488**	.324**	.113
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.210
	N	134	134	111	111	121	124
PBRNHM	Pearson Correlation	.505**	.514**	-.750**	-.540**	-.399**	-.230*
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.010
	N	134	134	111	111	121	124
PCAR	Pearson Correlation	-.329**	-.418**	.646**	.441**	.314**	.236**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.009
	N	132	132	110	110	120	122
PMOVE1	Pearson Correlation	.243**	.326**	-.627**	-.442**	-.293**	-.041
	Sig. (2-tailed)	.005	.000	.000	.000	.001	.653
	N	134	134	111	111	121	124
PMOVE5	Pearson Correlation	.145	.204*	-.564**	-.399**	-.192*	.085
	Sig. (2-tailed)	.095	.018	.000	.000	.035	.349
	N	134	134	111	111	121	124
PLNPRNT	Pearson Correlation	.172*	.357**	-.625**	-.507**	-.479**	-.327**
	Sig. (2-tailed)	.047	.000	.000	.000	.000	.000
	N	134	134	111	111	121	124
PCSNR	Pearson Correlation	-.235**	-.092	.318**	.249**	.368**	.099
	Sig. (2-tailed)	.006	.289	.001	.008	.000	.276
	N	134	134	111	111	121	124
PCVISMIN	Pearson Correlation	-.078	.020	.190*	.239*	.075	-.029
	Sig. (2-tailed)	.371	.819	.046	.011	.413	.748
	N	134	134	111	111	121	124
PTOW	Pearson Correlation	.495**	.195*	-.309**	-.150	-.022	.079
	Sig. (2-tailed)	.000	.024	.001	.116	.807	.386
	N	134	134	111	111	121	124
PSUBURB	Pearson Correlation	-.528**	-.351**	.710**	.441**	.287**	.085
	Sig. (2-tailed)	.000	.000	.000	.000	.001	.347
	N	134	134	111	111	121	124
PAPT	Pearson Correlation	.073	.204*	-.560**	-.365**	-.334**	-.202*
	Sig. (2-tailed)	.401	.018	.000	.000	.000	.024
	N	134	134	111	111	121	124
Average number of persons in private households	Pearson Correlation	-.592**	-.355**	.772**	.501**	.343**	.170
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.060
	N	134	134	111	111	121	124
Employment-population ratio	Pearson Correlation	-.568**	-.576**	.716**	.486**	.334**	.309**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	134	134	111	111	121	124
PSECED	Pearson Correlation	-.421**	-.049	.511**	.194*	.259**	-.052
	Sig. (2-tailed)	.000	.571	.000	.041	.004	.567
	N	134	134	111	111	121	124
X_LONG	Pearson Correlation	-.063	-.085	.169	-.040	-.026	.136
	Sig. (2-tailed)	.466	.330	.076	.674	.780	.132
	N	134	134	111	111	121	124
Y_LAT	Pearson Correlation	.161	.115	-.258**	-.348**	-.346**	-.033
	Sig. (2-tailed)	.064	.184	.006	.000	.000	.713
	N	134	134	111	111	121	124
DISTANCE	Pearson Correlation	-.251**	-.147	.364**	.192*	.143	.204*
	Sig. (2-tailed)	.003	.090	.000	.044	.118	.023
	N	134	134	111	111	121	124
QUAD	Pearson Correlation	-.098	-.079	.218*	.331**	.336**	-.058
	Sig. (2-tailed)	.260	.364	.022	.000	.000	.525
	N	134	134	111	111	121	124

		PWHTCOL	PSNR	PCHILD	PTEEN	PADLT	PMDAGE	PABOR	PCITOCAN	PHOHOLAN
RTSP1000	Pearson Correlation	-.310**	.647**	-.494**	-.418**	-.354**	.107	.263**	-.057	-.075
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.219	.002	.515	.387
	N	134	134	134	134	134	134	134	134	134
Unemployment rate	Pearson Correlation	-.361**	.206*	-.084	-.162	-.108	-.026	.279**	.088	.084
	Sig. (2-tailed)	.000	.017	.333	.062	.215	.769	.001	.313	.334
	N	134	134	134	134	134	134	134	134	134
Average 1995 household income $	Pearson Correlation	.751**	-.372**	.309**	.575**	.083	.313**	-.392**	.013	-.059
	Sig. (2-tailed)	.000	.000	.001	.000	.389	.001	.000	.889	.541
	N	111	111	111	111	111	111	111	111	111
Standard error of average household income $	Pearson Correlation	.634**	-.221*	.139	.373**	.091	.374**	-.243*	.048	.108
	Sig. (2-tailed)	.000	.020	.146	.000	.340	.000	.010	.620	.260
	N	111	111	111	111	111	111	111	111	111
Average value of dwelling $	Pearson Correlation	.636**	-.100	.128	.188*	-.029	.297**	-.165	-.050	.042
	Sig. (2-tailed)	.000	.277	.163	.039	.755	.001	.071	.589	.651
	N	121	121	121	121	121	121	121	121	121
Average gross rent $	Pearson Correlation	.340**	-.083	-.151	.090	.164	.330**	-.293**	-.137	-.188*
	Sig. (2-tailed)	.000	.359	.095	.321	.068	.000	.001	.129	.037
	N	124	124	124	124	124	124	124	124	124
PWHTCOL	Pearson Correlation	1.000	-.216*	.050	.153	.091	.157	-.271**	-.078	-.088
	Sig. (2-tailed)	.	.012	.569	.078	.295	.070	.002	.371	.312
	N	134	134	134	134	134	134	134	134	134
PSNR	Pearson Correlation	-.216*	1.000	-.663**	-.409**	-.656**	.216*	-.023	-.104	-.051
	Sig. (2-tailed)	.012	.	.000	.000	.000	.012	.790	.231	.560
	N	134	134	134	134	134	134	134	134	134
PCHILD	Pearson Correlation	.050	-.663**	1.000	.391**	.118	-.396**	.172*	-.048	-.063
	Sig. (2-tailed)	.569	.000	.	.000	.176	.000	.047	.580	.471
	N	134	134	134	134	134	134	134	134	134
PTEEN	Pearson Correlation	.153	-.409**	.391**	1.000	-.018	-.153	-.114	.091	-.064
	Sig. (2-tailed)	.078	.000	.000	.	.835	.077	.190	.295	.465
	N	134	134	134	134	134	134	134	134	134
PADLT	Pearson Correlation	.091	-.656**	.118	-.018	1.000	.230**	-.026	.087	.197*
	Sig. (2-tailed)	.295	.000	.176	.835	.	.008	.765	.317	.023
	N	134	134	134	134	134	134	134	134	134
PMDAGE	Pearson Correlation	.157	.216*	-.396**	-.153	.230**	1.000	-.312**	-.246**	.006
	Sig. (2-tailed)	.070	.012	.000	.077	.008	.	.000	.004	.948
	N	134	134	134	134	134	134	134	134	134
PABOR	Pearson Correlation	-.271**	-.023	.172*	-.114	-.026	-.312**	1.000	-.096	-.100
	Sig. (2-tailed)	.002	.790	.047	.190	.765	.000	.	.268	.251
	N	134	134	134	134	134	134	134	134	134
PCITOCAN	Pearson Correlation	-.078	-.104	-.048	.091	.087	-.246**	-.096	1.000	.360**
	Sig. (2-tailed)	.371	.231	.580	.295	.317	.004	.268	.	.000
	N	134	134	134	134	134	134	134	134	134
PHOHOLAN	Pearson Correlation	-.088	-.051	-.063	-.064	.197*	.006	-.100	.360**	1.000
	Sig. (2-tailed)	.312	.560	.471	.465	.023	.948	.251	.000	.
	N	134	134	134	134	134	134	134	134	134
POPDENS	Pearson Correlation	-.123	.509**	-.354**	-.295**	-.268**	.362**	.020	-.163	-.135
	Sig. (2-tailed)	.156	.000	.000	.001	.002	.000	.822	.059	.120
	N	134	134	134	134	134	134	134	134	134
PEDLO9	Pearson Correlation	-.395**	.718**	-.465**	-.460**	-.446**	.074	.163	-.145	.119
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.393	.061	.094	.170
	N	134	134	134	134	134	134	134	134	134
PLIVEALO	Pearson Correlation	-.333**	.676**	-.625**	-.540**	-.187*	.077	.222**	-.038	.066
	Sig. (2-tailed)	.000	.000	.000	.000	.030	.376	.010	.665	.450
	N	134	134	134	134	134	134	134	134	134
POWNED	Pearson Correlation	.366**	-.439**	.377**	.538**	.102	.032	-.261**	-.015	.053
	Sig. (2-tailed)	.000	.000	.000	.000	.242	.718	.002	.860	.545
	N	134	134	134	134	134	134	134	134	134
PBRNHM	Pearson Correlation	-.443**	.265**	-.261**	-.448**	.123	.072	.331**	-.026	-.013
	Sig. (2-tailed)	.000	.002	.002	.000	.158	.406	.000	.764	.878
	N	134	134	134	134	134	134	134	134	134
PCAR	Pearson Correlation	.393**	-.328**	.090	.236**	.050	.042	-.214*	-.022	-.049
	Sig. (2-tailed)	.000	.000	.305	.006	.571	.480	.014	.803	.580
	N	132	132	132	132	132	132	132	132	132
PMOVE1	Pearson Correlation	-.317**	.002	-.161	-.337**	.292**	-.140	.451**	.128	.124
	Sig. (2-tailed)	.000	.986	.063	.000	.001	.106	.000	.141	.156
	N	134	134	134	134	134	134	134	134	134
PMOVE5	Pearson Correlation	-.136	-.044	-.093	-.361**	.206*	-.131	.331**	.029	.012
	Sig. (2-tailed)	.116	.612	.285	.000	.017	.133	.000	.738	.891
	N	134	134	134	134	134	134	134	134	134
PLNPRNT	Pearson Correlation	-.341**	-.106	.172*	-.190*	.028	-.346**	.271**	.067	-.100
	Sig. (2-tailed)	.000	.222	.047	.028	.747	.000	.002	.441	.249
	N	134	134	134	134	134	134	134	134	134
PCSNR	Pearson Correlation	.298**	-.218*	.227**	.231**	-.089	.020	-.098	-.052	-.073
	Sig. (2-tailed)	.000	.011	.008	.007	.306	.823	.260	.549	.400
	N	134	134	134	134	134	134	134	134	134
PCVISMIN	Pearson Correlation	.003	-.108	-.060	.248**	.064	.126	-.089	.653**	.324**
	Sig. (2-tailed)	.972	.216	.493	.004	.463	.147	.306	.000	.000
	N	134	134	134	134	134	134	134	134	134
PTOW	Pearson Correlation	-.105	.685**	-.522**	-.442**	-.346**	.351**	-.088	-.144	-.158
	Sig. (2-tailed)	.226	.000	.000	.000	.000	.000	.310	.007	.068
	N	134	134	134	134	134	134	134	134	134
PSUBURB	Pearson Correlation	.300**	-.466**	.431**	.568**	.112	.025	-.239**	-.052	.032
	Sig. (2-tailed)	.000	.000	.000	.000	.196	.778	.005	.549	.711
	N	134	134	134	134	134	134	134	134	134
PAPT	Pearson Correlation	-.255**	-.206*	.069	-.172*	.243**	-.443**	.381**	.236**	.128
	Sig. (2-tailed)	.003	.017	.430	.047	.005	.000	.000	.006	.139
	N	134	134	134	134	134	134	134	134	134
Average number of persons in private households	Pearson Correlation	.344**	-.649**	.705**	.612**	.095	-.096	-.170*	-.017	-.068
	Sig. (2-tailed)	.000	.000	.000	.000	.276	.271	.049	.847	.433
	N	134	134	134	134	134	134	134	134	134
Employment-population ratio	Pearson Correlation	.356**	-.634**	.416**	.467**	.434**	.038	-.242**	-.028	.027
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.662	.005	.746	.761
	N	134	134	134	134	134	134	134	134	134
PSECED	Pearson Correlation	.269**	-.496**	.456**	.189*	.314**	-.127	.110	-.122	.124
	Sig. (2-tailed)	.002	.000	.000	.029	.000	.144	.205	.160	.154
	N	134	134	134	134	134	134	134	134	134
X_LONG	Pearson Correlation	.066	-.168	.150	.305**	-.016	.078	.048	-.087	-.296**
	Sig. (2-tailed)	.446	.052	.083	.000	.857	.369	.583	.317	.001
	N	134	134	134	134	134	134	134	134	134
Y_LAT	Pearson Correlation	-.263**	.004	.004	-.007	-.060	-.207*	.255**	.090	-.092
	Sig. (2-tailed)	.002	.968	.962	.938	.487	.016	.003	.302	.292
	N	134	134	134	134	134	134	134	134	134
DISTANCE	Pearson Correlation	.176*	-.280**	.317**	.386**	.018	.149	-.150	-.095	-.148
	Sig. (2-tailed)	.041	.001	.000	.000	.837	.086	.067	.277	.087
	N	134	134	134	134	134	134	134	134	134
QUAD	Pearson Correlation	.174*	.079	-.045	-.011	-.019	.117	-.174*	-.024	.161
	Sig. (2-tailed)	.044	.362	.602	.900	.832	.177	.044	.779	.063
	N	134	134	134	134	134	134	134	134	134

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

POPDENS	PEDLOG	PLIVEALO	POWNED	PBRKHM	PCAR	PMOVE1	PMOVES	PLNPRNT	PCSNR	PCVISMIN	PTOW	PSUBURB	PAPT	Average number of persons in private households	Employment-population ratio	PSECED	X_LONG	Y_LAT	DISTANCE	QUAD
.370	.650**	.649**	-.521**	.505**	-.329**	.243**	.145	.172*	-.235**	-.078	.495**	-.526**	.073	-.502**	-.505**	-.421**	-.063	.161	-.251**	-.098
.000	.000	.000	.000	.000	.000	.005	.095	.047	.006	.371	.000	.000	.401	.000	.000	.000	.466	.064	.003	.260
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.452**	.325**	.407**	-.359**	.514**	-.418**	.326**	.204*	.357**	-.092	.020	.195*	-.351**	.204*	-.355**	-.576**	-.049	-.085	.115	-.147	-.079
.000	.000	.000	.000	.000	.000	.000	.018	.000	.289	.819	.024	.000	.018	.000	.000	.571	.330	.184	.090	.364
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.261**	-.566**	-.710**	.780**	-.750**	.646**	-.627**	-.564**	-.625**	.318**	.190*	-.309**	.710**	-.560**	.772**	.716**	.511**	.169	-.258**	.384**	.218*
.006	.000	.000	.000	.000	.000	.000	.000	.000	.001	.046	.001	.000	.000	.000	.000	.000	.076	.006	.000	.022
111	111	111	111	111	110	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
-.221*	-.356**	-.440**	.488**	-.540**	.441**	-.442**	-.399**	-.507**	.249**	.239*	-.150	.441**	-.385**	.501**	.486**	.194*	-.040	-.348**	.192*	.331**
.019	.000	.000	.000	.000	.000	.000	.000	.000	.008	.011	.116	.000	.000	.000	.000	.041	.674	.000	.044	.000
111	111	111	111	111	110	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
-.066	-.298**	-.240**	.324**	-.369**	.314**	-.293**	-.192*	-.479**	.369**	.075	-.022	.287**	-.334**	.343**	.334**	.259**	-.026	-.346**	.143	.336**
.469	.001	.008	.000	.000	.000	.001	.035	.000	.000	.413	.807	.001	.000	.000	.000	.004	.780	.000	.118	.000
121	121	121	121	121	120	121	121	121	121	121	121	121	121	121	121	121	121	121	121	121
-.058	-.226*	-.268**	.113	-.230*	.236**	-.041	.085	-.327**	.099	-.029	.079	.085	-.202*	.170	.309**	-.052	.136	-.033	.204*	-.058
.521	.011	.003	.210	.010	.008	.653	.349	.000	.276	.748	.386	.347	.024	.060	.000	.567	.132	.713	.023	.525
124	124	124	124	124	122	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124
-.123	-.395**	-.333**	.386**	-.443**	.380**	-.317**	-.136	-.341**	.298**	.003	-.105	.300**	-.255**	.344**	.380**	.269**	.066	-.263**	.176*	.174*
.156	.000	.000	.000	.000	.000	.000	.116	.000	.000	.972	.226	.000	.003	.000	.000	.002	.448	.002	.041	.044
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.509**	.716**	.676**	-.439**	.265**	-.326**	.002	-.044	-.106	-.218*	-.108	.685**	-.466**	-.206*	-.649**	-.634**	-.496**	-.168	.004	-.293**	.079
.000	.000	.000	.000	.002	.000	.986	.612	.222	.011	.216	.000	.000	.017	.000	.000	.000	.052	.968	.001	.362
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.354**	-.465**	-.625**	.377**	-.261**	.090	-.161	-.093	.172*	.227**	-.060	-.522**	.431**	.069	.705**	.416**	.450**	.150	.004	.317**	-.045
.000	.000	.000	.000	.002	.305	.063	.285	.047	.008	.493	.000	.000	.430	.000	.000	.000	.083	.962	.000	.602
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.295**	-.460**	-.540**	.538**	-.448**	.239**	-.337**	-.361**	-.190*	.231**	.246**	-.442**	.568**	-.172*	.612**	.467**	.189*	.305**	-.007	.388**	-.011
.001	.000	.000	.000	.000	.006	.000	.000	.028	.007	.004	.000	.000	.047	.000	.000	.029	.000	.938	.000	.900
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.266**	-.446**	-.187*	.102	.123	.050	.292**	.206*	.028	-.089	.064	-.346**	.112	.243**	.095	.434**	.314**	-.016	-.060	.018	-.019
.002	.000	.030	.242	.158	.571	.001	.017	.747	.306	.463	.000	.196	.005	.276	.000	.000	.857	.487	.837	.832
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.362**	.074	.077	.032	.072	.062	-.140	-.131	-.346**	.020	-.126	.351**	.025	-.443**	-.096	.038	-.127	.078	-.207*	.149	.117
.000	.393	.376	.718	.408	.480	.106	.133	.000	.823	.147	.000	.778	.000	.271	.662	.144	.369	.016	.086	.177
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.020	.163	.222**	-.261**	.331**	-.214*	.451**	.331**	.271**	-.098	-.089	-.088	-.239**	.381**	-.170*	-.242**	.110	.048	.255**	-.159	-.174*
.822	.061	.010	.002	.000	.014	.000	.000	.002	.260	.306	.310	.005	.000	.049	.005	.205	.563	.003	.067	.044
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.163	-.145	-.038	-.015	-.026	-.022	.128	.029	.067	-.052	.653**	-.144	-.052	.236**	-.017	-.028	-.122	-.087	.090	-.095	-.024
.059	.094	.665	.860	.764	.803	.141	.738	.441	.549	.000	.097	.549	.006	.847	.746	.160	.317	.302	.277	.779
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.135	.119	.066	.053	-.013	-.049	.124	.012	-.100	-.073	.324**	-.158	.032	.128	-.068	.027	.124	-.296**	-.092	-.148	.161
.120	.170	.450	.545	.878	.580	.155	.891	.249	.400	.000	.068	.711	.139	.433	.761	.154	.001	.292	.087	.063
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
1.000	.524**	.440**	-.440**	.447**	-.198*	.106	.114	.038	-.179*	-.149	.694**	-.452**	-.240**	-.415**	-.525**	-.535**	.012	.027	-.070	-.043
	.000	.000	.000	.000	.023	.225	.190	.660	.039	.086	.000	.000	.005	.000	.000	.000	.894	.758	.421	.623
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.524**	1.000	.631**	-.441**	.381**	-.360**	.091	.060	.074	-.212*	-.196*	.538**	-.464**	-.061	-.599**	-.687**	-.417**	-.201*	.094	-.315**	-.030
.000		.000	.000	.000	.000	.294	.302	.394	.014	.023	.000	.000	.484	.000	.000	.000	.020	.281	.000	.732
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.440**	.631**	1.000	-.758**	.696**	-.476**	.399**	.321**	.260**	-.342**	-.117	.566**	-.735**	.262**	-.931**	-.657**	-.318**	-.312**	.003	-.560**	.093
.000	.000		.000	.000	.000	.000	.000	.002	.000	.180	.000	.000	.002	.000	.000	.000	.000	.976	.000	.287
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.440**	-.441**	-.758**	1.000	-.806**	.525**	-.622**	-.661**	-.475**	.290**	.069	-.624**	.954**	-.456**	.792**	.686**	.417**	.207*	-.152	.524**	.053
.000	.000	.000		.000	.000	.000	.000	.000	.001	.429	.000	.000	.000	.000	.000	.000	.016	.080	.000	.545
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.447**	.381**	.696**	-.806**	1.000	-.566**	.660**	.559**	.548**	-.310**	-.086	.506**	-.782**	.369**	-.713**	-.564**	-.231**	-.132	.068	-.399**	-.039
.000	.000	.000	.000		.000	.000	.000	.000	.000	.323	.000	.000	.000	.000	.000	.007	.128	.436	.000	.657
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.198*	-.360**	-.476**	.525**	-.566**	1.000	-.358**	-.334**	-.421**	.138	.029	-.293**	.496**	-.283**	.456**	.494**	.026	.098	-.240**	.460**	.137
.023	.000	.000	.000	.000		.000	.000	.000	.115	.739	.001	.000	.001	.000	.000	.771	.265	.006	.000	.116
132	132	132	132	132	132	132	132	132	132	132	132	132	132	132	132	132	132	132	132	132
.106	.091	.399**	-.622**	.660**	-.358**	1.000	.728**	.474**	-.260**	.125	.183*	-.629**	.540**	-.490**	-.362**	-.084	-.090	.179*	-.320**	-.101
.225	.294	.000	.000	.000	.000		.000	.000	.002	.149	.034	.000	.000	.000	.000	.333	.299	.039	.000	.246
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.114	.090	.321**	-.661**	.559**	-.334**	.728**	1.000	.430**	-.165	.020	.236**	-.649**	.518**	-.420**	-.350**	-.167	-.134	.087	-.289**	-.075
.190	.302	.000	.000	.000	.000	.000		.000	.057	.818	.006	.000	.000	.000	.000	.053	.122	.319	.001	.392
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.038	.074	.260**	-.475**	.548**	-.421**	.474**	.430**	1.000	-.102	-.091	.121	-.454**	.406**	-.274**	-.403**	-.052	-.013	.131	-.237**	-.065
.660	.394	.002	.000	.000	.000	.000	.000		.241	.294	.162	.000	.000	.001	.000	.550	.883	.132	.006	.458
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.179*	-.212*	-.342**	.290**	-.310**	.138	-.260**	-.165	-.102	1.000	.046	-.248**	.282**	-.043	.386**	.177*	.193*	.060	-.024	.208*	-.069
.039	.014	.000	.001	.000	.115	.002	.057	.241		.601	.004	.001	.625	.000	.041	.026	.493	.784	.016	.426
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.149	-.196*	-.117	.069	-.086	.029	.125	.020	-.091	.046	1.000	-.192*	.024	.191*	.079	.061	-.101	.106	.122	.077	-.100
.086	.023	.180	.429	.323	.739	.149	.818	.294	.601		.026	.784	.027	.363	.487	.247	.221	.160	.378	.248
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.694**	.538**	.566**	-.624**	.506**	-.293**	.183*	.236**	.121	-.248**	-.192*	1.000	-.649**	-.356**	-.577**	-.616**	-.574**	-.098	-.096	-.230**	.110
.000	.000	.000	.000	.000	.001	.034	.006	.162	.004	.026		.000	.000	.000	.000	.000	.259	.268	.008	.205
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.452**	-.464**	-.735**	.954**	-.782**	.496**	-.629**	-.649**	-.454**	.282**	.024	-.649**	1.000	-.469**	.782**	.714**	.430**	.205*	-.140	.513**	.036
.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.784	.000		.000	.000	.000	.000	.018	.107	.000	.682
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.240**	-.061	.262**	-.456**	.369**	-.283**	.540**	.518**	.406**	-.043	.191*	-.356**	-.469**	1.000	-.306**	-.154	.109	-.143	.307**	-.362**	-.198*
.005	.484	.002	.000	.000	.001	.000	.000	.000	.625	.027	.000	.000		.000	.075	.209	.100	.000	.000	.022
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.415**	-.599**	-.931**	.792**	-.713**	.456**	-.490**	-.420**	-.274**	.386**	.079	-.577**	.782**	-.306**	1.000	.658**	.377**	.336**	-.012	.551**	-.040
.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.363	.000	.000	.000		.000	.000	.000	.886	.000	.650
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.525**	-.687**	-.657**	.686**	-.564**	.494**	-.362**	-.350**	-.403**	.177*	.061	-.616**	.714**	-.154	.658**	1.000	.474**	.150	-.155	.347**	.061
.000	.000	.000	.000	.000	.000	.000	.000	.000	.041	.487	.000	.000	.075	.000		.000	.084	.074	.000	.480
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.535**	-.417**	-.318**	.417**	-.231**	.026	-.084	-.167	-.052	.193*	-.101	-.574**	.430**	.109	.377**	.474**	1.000	.060	.011	.072	-.010
.000	.000	.000	.000	.007	.771	.333	.053	.550	.026	.247	.000	.000	.209	.000	.000		.486	.900	.411	.910
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.012	-.201*	-.312**	.207*	-.132	.098	-.090	-.134	-.013	.060	.106	-.098	.205*	-.143	.336**	.150	.060	1.000	.471**	.362**	-.510**
.894	.020	.000	.016	.128	.255	.299	.122	.883	.483	.221	.259	.018	.100	.000	.084	.488		.000	.000	.000
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
.027	.094	.003	-.152	.068	-.240**	.179*	.087	.131	-.024	.122	-.096	-.140	.307**	-.012	-.155	.011	.471**	1.000	-.097	-.784**
.758	.281	.976	.080	.436	.006	.039	.319	.132	.784	.160	.268	.107	.000	.886	.074	.900	.000		.263	.000
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.070	-.315**	-.560**	.524**	-.399**	.460**	-.320**	-.289**	-.237**	.208*	.077	-.230**	.513**	-.362**	.551**	.347**	.072	.362**	-.097	1.000	-.189*
.421	.000	.000	.000	.000	.000	.000	.001	.006	.016	.378	.008	.000	.000	.000	.000	.411	.000	.263		.029
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134
-.043	-.030	.093	.053	-.039	.137	-.101	-.075	-.065	-.069	-.100	.110	.036	-.198*	-.040	.061	-.010	-.510**	-.784**	-.189*	1.000
.623	.732	.287	.545	.657	.116	.246	.392	.458	.426	.248	.205	.682	.022	.650	.480	.910	.000	.000	.029	
134	134	134	134	134	132	134	134	134	134	134	134	134	134	134	134	134	134	134	134	134

Reference List

Achabal, D. D. (1978). The Development of a Spatial Delivery System for Emergency Medical Services. Geographical Analysis, 10, 47-64.

Albert, D. P., Gesler, W. M., & Horner, R. D. (2000a). Geographic Information Systems in Health Services Research. In D.P.Albert, W. M. Gesler, & B. Levergood (Eds.), Spatial Analysis, GIS and Remote Sensing Applications in the Health Sciences (pp. 55-76). Chelsea, MI: Ann Arbor Press.

Albert, D. P., Gesler, W. M., & Wittie, P. S. (2000b). Geographic Information Systems: Medical Geography. In D.P.Albert, W. M. Gesler, & B. Levergood (Eds.), Spatial Analysis, GIS and Remote Sensing Applications in the Health Sciences (pp. 39-54). Chelsea, MI: Ann Arbor Press.

Aldrich, C. A., Hisserich, J. C., & Lave, L. B. (1971). An Analysis of the Demand for Emergency Ambulance Service in an Urban Area. American Journal of Public Health, 61, 1156-1169.

Anselin, L. (1992). SpaceStat Tutorial A Workbook for Using SpaceStat in the Analysis of Spatial Data. Morgantown, WV: Regional Research Institute, West Virginia University.

Anselin, L. (1995). Local Indicators of Spatial Association - LISA. Geographical Analysis, 27, 93-115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M.Fischer, H. J. Scholten, & D. Unwin (Eds.), Spatial Analytical Perspectives on GIS (pp. 111-125). London: Taylor & Francis Ltd.

Anselin, L. & Griffith, D. A. (1988). Do Spatial Effects Really Matter in Reggression Analysis? Papers of the Regional Science Association, 65, 11-34.

Anselin, L. & Rey, S. (1991). Properties of Tests for Spatial Dependence in Linear Regression Models. Geographical Analysis, 23, 112-131.

Association for Geographic Information (1997). Emergency services GIS survey - a report commissioned by the emergency services special interest group of the AGI London, UK: The Association for Geographic Information.

Bailey, T. C. & Gatrell, A. C. (1995). Interactive Spatial Data Analysis. Harlow, UK: Addison Wesley Longman Limited.

Ball, T. (1980). Ambulance Services: A logical, effective and economic solution. Health and Social Service Journal, 90, 386-387.

Béland, F., Lemay, A., Philibert, L., Maheux, B., & Gravel, G. (1991). Elderly Patients' Use of Hospital-Based Emergency Services. Medical Care, 29, 408-418.

Boots, B. & Tiefelsdorf, M. (2000). Global and Local Spatial Autocorrelation in Bounded Regular Tessellations. Journal of Geographical Systems, 2, 319-348.

Boots, B. (2002). Local measures of spatial association. Ecoscience, 9, 168-176.

Bracken, I. & Martin, D. (1989). The generation of spatial population distribution from census centroid data. Environment and Planning A, 21, 537-543.

Braun, T. (1998). Sozialstruktur un Notfallmedizin zusammenhänge am Beispiel von Santander. Diplomarbeit Bonn.

Braun, T. 2000. 7-4-2000.
Ref Type: Personal Communication

Butler, J. R. G. (1981). Factors Affecting the Demand for Ambulance Services. International Journal of Transport Economics, 8, 225-238.

Cadigan, R. T. & Bugarin, C. E. (1989). Predicting Demand for Emergency Ambulance Service. Annals of Emergency Medicine, 18, 618-621.

Can, A. (1996). Weight matrices and spatial autocorrelation statistics using a topological vector data model. International Journal of Geographical Information Systems, 10, 1009-1017.

Carter, H. (1995). The Study of Urban Geography. (4 ed.) London, UK: Arnold.

Clark, M. J. & Fitzgerald, G. (1999). Older People's Use of Ambulance Services: a Population Based Analysis. Journal of Accident & Emergency Medicine, 16, 108-111.

Cliff, A. D. & Ord, J. K. (1981). Spatial Processes: Models and Applications. London: Pion Limited.

Daskin, M. S. (1983). A Maximum Expected Covering Location Model: Formulation, Properties and Heuristic Solution. Transportation Science, 17, 48-69.

Daskin, M. S. & Stern, E. H. (1981). A Hierarchical Objective Set Covering Model for Emergency Medical Service Vehicle Deployment. Transportation Science, 15, 137-151.

De Angelis, A. (1995). Factors affecting the demand for emergency services in the Greater Kingston Area. M. Pl Report, School of Urban and Regional Planning, Queen's University.

Dickinson, E. T., Verdile, V. P., Kostyun, C. T., & Salluzo, R. F. (1996). Geriatric Use of Emergency Medical Services. Annals of Emergency Medicine, 27, 199-203.

Eaton, D. J., Daskin, M. S., Simmons, D., Bulloch, B., & Jansma, G. (1985). Determining Emergency Medical Service Vehicle Deployment in Austin,Texas. Interfaces, 15, 96-108.

Eaton, D. J., Sánchez U., H. ML., Lantiqua, R. R., & Morgan (1986). Determining Ambulance Deployment in Santo Domingo, Dominican Republic. Journal of the Operational Research Society, 37, 113-126.

Fairhurst, R. J. (1992). Use of the emergency ambulance service. Journal of the Royal Society of Medicine, 85, 306.

Fellers, R. A. (2000). GIS-Aided Environmental Research: Prospects and Pitfalls. In D.P.Albert, W. M. Gesler, & B. Levergood (Eds.), Spatial Analysis, GIS and Remote Sensing Applications in the Health Sciences (pp. 77-110). Chelsea, MI: Ann Arbor Press.

Fotheringham, A. S. & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A, 23, 1025-1044.

Fotheringham, S. A. (1997). Trends in quantitative methods I: stressing the local. Progress in Human Geography, 21, 88-96.

Gardner, G. J. (1990). Use and abuse of the emergency ambulance service: some of the factors affecting the decision whether to call an emergency ambulance. Archives of Emergency Medicine, 7, 81-89.

Gatrell, A. & Senior, M. (1999). Health and health care applications. In P.Longley, M. F. Goodchild, D. J. Maguire, & D. Rhind (Eds.), Geographical information systems (pp. 925-938). New York: John Wiley & Sons Inc.

Gatrell, A. C. & Bailey, T. C. (1996). Interactive Spatial Data Analysis in Medical Geography. Social Science and Medicine, 42, 843-855.

Gerson, L. W. & Skvarch, L. (1982). Emergency Medical Service Utilization by the Elderly. Annals of Emergency Medicine, 11, 610-612.

Gesler, W. (1986). The Uses of Spatial Analylsis in Medical Geography: A Review. Social Science and Medicine, 23, 963-973.

Getis, A. & Ord, J. K. (1996). Local spatial statistics: an overview. In P.Longley & M. Batty (Eds.), Spatial Analysis: Modelling in a GIS Environment (pp. 261-277). Cambridge: GeoInformation International.

Gibson, G. (1971). Letters to the Editor: An Analysis of the Demand for Emergency Ambulance Service in an Urban Area. American Journal of Public Health, 61, 2158-2160.

Gough, J. & McCarthy, O. (1975). The Ambulance Facility Location Problem - A Survey of Methods and a Simple Solution (Rep. No. 73). Cantebury, NZ: The Agricultural Economics Research Unit, Lincoln College, University of Cantebury.

Government of Ontario. An Act to improve Services, increase Efficiency and benefit Taxpayers by eliminating Duplication and reallocating Responsibilities between Provincial and Municipal Governments in various areas and to implement other aspects of the Government's "Who Does What" Agenda. 152. 1997. 2nd session, 36th Parliament.
Ref Type: Bill/Resolution

Greater Sudbury EMS. Upper Tier Report Frequency Distribution of Call Volume -From 1/1/99 to 12/31/99. Ontario Ministry of Health and Long Term Care. 2001. Sudbury, On, Greater Sudbury EMS. Aris Dispatch Data Access Service. 7-13-2001.
Ref Type: Data File

Gregoris, D. Sudbury Base Hospital EMS Database Completeness. 2000. 2000.
Ref Type: Personal Communication

Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation problem. Journal of Geographical Systems, 2, 141-156.

Griffith, D. A. & Amrhein, C. G. (1997). Multivariate Statistical Analysis for Geographers. Upper Saddle River, NJ: Prentice Hall.

Griffith, D. A. & Layne, L. J. (1999). A Casebook for Spatial Statistical Data Analysis. Oxford: Oxford University Press.

Haining, R. (1990). Spatial data analysis in the social and environmental sciences. Cambridge, UK: Cambridge University Press.

Haining, R. (1996). Designing a Health Needs GIS with Spatial Analysis Capability. In M.Fischer, H. J. Scholten, & D. Unwin (Eds.), Spatial Analytical Perspectives on GIS (pp. 53-66). London UK: Taylor & Frances Ltd.

Haining, R. (1998). Spatial Statistics and the Analysis of Health Data. In A.C.Gatrell & M. Löytönen (Eds.), GIS and Health (pp. 29-47). London: Taylor & Francis Ltd.

Hanna, J. (1982). A Century of Red Blankets A history of Ambulance Service in Ontario. Erin, ON: The Boston Mills Press.

Harries, K. (1999). Mapping Crime Principle and Practice. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

Hisserich, J. C. (1971). Reply to Dr. Gibson: An Analysis of the Demand for Emergency Ambulance Service in an Urban Area. American Journal of Public Health, 61, 2160-2161.

IBI Group (2000). Summary Report Sudbury ON: Regional Municipality of Sudbury.

Intergraph (1999). Intergraph Public Safety, Products and Services. http://www.intergraph.com/pubsafety/ProdSer.htm [On-line].

Kotzé, J. M. (1990). The role of the ambulance service as part of the health profession. South African Medical Journal, 78, 320-322.

Kvålseth, T. O. & Deems, J. M. (1979). Statistical Models of the Demand for Emergency Medical Services in an Urban Area. American Journal of Public Health, 69, 250-255.

Larson, R. C. (1975). Approximating the Performance of Urban Emergency Service Systems. Operations Research, 23, 845-868.

McConnel, C. & Wilson, R. W. (2001). Patterns of use of EMS transport. American Journal of Emergency Medicine, 19, 84.

McConnel, C. E. & Wilson, R. W. (1998). The Demand for Prehospital Emergency Services in an Aging Society. Social Science & Medicine, 46, 1027-1031.

Messmer, J. (1973). Emergency Medical Services in Metropolitan Areas A bibliography (Rep. No. 493). Monticello, IL: Council of Planning Librarians.

Miller, H. J. (1996). GIS and geometric representation in facility location problems. International Journal of Geographical Information Systems, 10, 791-816.

Norušis, M. J. (1990). SPSS Introductory Statistics Students Guide. Chicago: SPSS Inc.

O'Neil, S. 2004. 4-15-2004.
Ref Type: Personal Communication

Ontario Ministry of Health and Long-Term Care (1998). Review of the Ambulance Regulation: Report of the Land Ambulance Transition Task Force to the Minister of Health and Red Tape Commission Ontario Ministry of Health and Long-Term Care.

Ontario Ministry of Health and Long-Term Care (1999a). EMERGENCY HEALTH SERVICES Land Ambulance Transition Frequently Asked Questions. http://www.gov.on.ca:80/MOH/english/program/ambul/transfaq_dt.html [On-line].

Ontario Ministry of Health and Long-Term Care (1999b). PROGRAMS & SERVICES Emergency Health Services Land and Air Ambulance. http://www.gov.on.ca/MOH/english/program/ambul/ehs_mn.html [On-line].

Ontario Ministry of Health and Long-Term Care (1999c). Land Ambulance Services. Backgrounder August,1997. http://www.gov.on.ca:80/MOH/english/pub/ministry/amb.html. [On-line].

Ontario Ministry of Health and Long-Term Care (1999d). Land Ambulance/Public Health Announcement Questions and Answers. http://www.gov.on.ca:80/MOH/english/pub/ministry/qa_032399.html [On-line].

Ontario Ministry of Health and Long-Term Care. General Regulations. 257/00. 2002. Government of Ontario. (As amended by O. Reg. 386/01 and O. Reg. 229/02).
Ref Type: Bill/Resolution

Ontario Ministry of Health and Long-Term Care & MOH-AMO Land Ambulance Implementation Steering Committee (1999). Land Ambulance Transition Practical Guide Ontario Ministry of Health and Long-Term Care.

Ontario Ministry of Municipal Affairs and Housing (1999). "Who does what?" January 1998. Section 3. Social and Community Health Services. www.mmah.gov.on.ca/business/wdw/section3_6_e.html [On-line].

Ord, J. K. & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geographical Analysis, 27, 286-306.

Parker, E. B. & Campbell, J. L. (1998). Measuring access to primary medical care: some examples of the use of geographical information systems. Health and Place, 4, 183-193.

Pennycook, A. G., Makower, R. M., & Morrison, W. G. (1991). Use of the emergency ambulance service to an inner Accident and Emergency Department - a comparison of general practioner and '999' calls. Journal of the Royal Society of Medicine, 84, 726-727.

Peters, J. (1998). An evaluation of ambulance service performance using a geographic information system.

Plane, D. R. & Hendrick, T. E. (1977). Mathematical Programming and the Location of Fire Companies for the Denver Fire Department. Operations Research, 25, 563-578.

Regional Municipality of Sudbury. Map A Land Use Plan Draft. 1985. Sudbury, ON, Regional Municipality of Sudbury. 1985.
Ref Type: Map

Regional Municipality of Sudbury (1993). Official Plan for the Sudbury Planning Area Sudbury, ON.

ReVelle, C. (1991). Siting Ambulances and Fire Companies. Journal of the American Planning Association, 57, 471-484.

ReVelle, C., Bigman, D., Schilling, D., Cohon, J., & Church, R. (1977). Facility Location: A Review of Context-free and EMS Models. Health Services Research, Summer, 129-146.

ReVelle, C., Toregas, C., & Falkson, L. (1976). Applications of the Location Set-covering Problem. Geographical Analysis, 8, 66-76.

Rucker, D. W., Edwards, R. A., Burstin, H. R., O'Neil, A. C., & Brennan, T. A. (1997). Patient-specific predictors of ambulance use. Ann Emerg Med, 29, 484-491.

Schuman, L. J., Wolfe, H., & Sepulveda, J. (1977). Estimating Demand for Emergency Transportation. Medical Care, 15, 738-749.

Shapiro, J. R. (1988). Final Report of a commission of public inquiry into ambulance and Emergency Health Services delivery in the province of Ontario Commissioned by the Ontario Public Service Employees Union.

Shen, Q. (1994). An Application of GIS to the Measurement of Spatial Autocorrelation. Computers, Environment and Urban Systems, 18, 167-191.

Siemens (1999). PROFESSIONAL SERVICES Computer Aided Dispatch Systems, Solution-related Services. http://www.sni.be/en/serviceware/s1_12_nl.htm [On-line].

Siler, K. F. (1975). Predicting demand for publicly dispatched ambulances in a metropolitan area. Health Serv Res, 10, 254-263.

Sokal, R. R., Oden, N. L., & Thomson, B. A. (1998). Local Spatial Autocorrelation in a Biological Model. Geographical Analysis, 30, 331-354.

Solonynka, O. MapMan 2 The Sudbury Region. 1999. Sudbury, ON, Mapman Publications.
Ref Type: Map

Statistics Canada. 1998 E-STAT. Statistics Canada. [CD ROM, 10F0174XCBWWIN]. 1998. Ottawa, ON, Government of Canada, Minister of Industry.
Ref Type: Data File

Statistics Canada (1999). Statistics Canada 1996 Census Dictionary. (Final ed.) (Vols. Catalogue No. 92-351-UIE) Ottawa, ON: Government of Canada, Ministry of Industry.

Sudbury Regional Development Corporation (1998). Sudbury Region - Land of Discovery - Health Care Services. http://srdc.on.ca/medical/index.html [On-line].

Svenson, J. E. (2000). Patterns of Use of Emergency Medical Transport: a Population-Based Study. American Journal of Emergency Medicine, 18, 130-134.

Symons, J. G. Jr. (1969). Locating Emergency Medical Service Vehicle Dispatch Centres in a Metropolitan Area. Master of Science University of Washington.

Szplett, E. (1988). The underlying socioeconomic, workplace, and accident correlates of demand for ALS services in Metropolitan Toronto Toronto: The Municipality of Metropolitan Toronto, Department of Ambulance Services.

Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The Location of Emergency Service Facilities. Operations Research, 19, 1363-1373.

Toronto EMS (1999). Automatic Vehicle Location. http://www.city.toronto.on.ca/ems/communications/vr_test_area/avl.htm [On-line].

Tritech (1999). *Safety Solutions Made Easy.* VisCAD for Ambulance. http://www.tritech.com/products/va.htm [On-line].

Vine, M. F., Degnan, D., & Hanchette, C. (1997). Geographic Information Systems: Their Use in Environmental Epidemiologic Research. Environmental Health Perspectives, 105, 598-605.

Volz, R. (1971). Optimal Ambulance Location in Semi-Rural Areas. Transportation Science, 5, 193-203.

Williams, P. M. & Shavlik, G. (1979). Geographic patterns and demographic correlates of paramedic runs in San Bernadino, 1977. Social Science and Medicine, 13, 273-279.

Wofford, J. L., Moran, W. P., Heuser, M. D., Schwartz, E., Velez, R., & Mittelmark, M. B. (1995). Emergency Medical Transport of the Elderly: A Population-Based Study. American Journal of Emergency Medicine, 13, 297-300.

Wong, D. (1996). Aggregation Effects in Geo-Referenced Data. In S.L.Arlinghaus, D. A. Griffith, W. C. Arlinghaus, W. D. Drake, & J. D. Nystuen (Eds.), Practical Handbook of Spatial Statistics (pp. 83-106). New York: CRC Press.

Yeates, M. (1974). An Introduction to Quantitative Analysis in Human Geography. (2 ed.) Toronto, ON: McGraw-Hill, Inc.

Zaki, A., Cheng, H. K., & Parker, B. R. (1997). A Simulation Model for the Analysis and Management of An Emergency Service System. Socio-Economic Planning Science, 31, 173-189.

Zhang, C. & Murayama, Y. (2000). Testing local spatial autocorrelation using k-order neighbours. International Journal of Geographical Information Systems, 14, 681-692.