INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
A SPATIAL ANALYSIS OF THE RELATIONSHIP BETWEEN RESPIRATORY HEALTH AND MOTOR VEHICLE EMISSIONS AMONGST SMALL URBAN AREAS

by

David Llewellyn Buckeridge

A thesis submitted in conformity with the requirements for the degree of Master of Science in Epidemiology
Graduate Department of Community Health
University of Toronto
© Copyright by David Llewellyn Buckeridge 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

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

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

0-612-40706-3
ABSTRACT

Title A Spatial Analysis of the Relationship Between Respiratory Health and Motor Vehicle Emissions Amongst Small Urban Areas

Degree Master of Science in Epidemiology, 1998

Author David Buckeridge

Department Community Health, University of Toronto

This study assesses the relationship between exposure to particulate matter (PM) from motor vehicle emissions and hospital admission rates for respiratory illness among enumeration areas (EAs) in Southeast Toronto (SETO). Exposure is modelled from traffic count data using a geographical information system, and census data is used to control for the effect of socioeconomic status. Spatial relationships of variables are assessed, and Poisson regression is used to estimate the effect of exposure on hospitalization.

Exposure to PM from motor vehicles has a statistically significant effect on hospitalization rates for selected respiratory conditions, with a relative risk of 1.39 in the highest exposure quintile. The effect of exposure on non-respiratory hospitalization rates is weaker, but also significant. The non-specific effect of exposure suggests that control for confounders may have been incomplete. Some variables are positively spatially autocorrelated, but this does not appear to influence the results of the multivariate model.
ACKNOWLEDGEMENTS

The completion of this thesis was made possible through the contributions of a number of individuals.

To begin with, Dr Richard Glazier in his capacity as thesis Supervisor provided insightful criticism, support, and direction throughout the research and writing phases. In addition, the members of the thesis committee all provided exemplary support in their respective areas of expertise (Dr Bart Harvey - Community Medicine and Epidemiology; Professor Michael Escobar - Statistics and Biostatistics; Professor Carl Amrhein – Spatial Analysis and Geography; Dr John Frank – Community Medicine and Epidemiology), and demonstrated unflagging stamina in reading earlier drafts of the thesis.

Work with health data was greatly facilitated by the advice and assistance of Julie Arnold. On a number of occasions, Professor Deniz Karman provided expert environmental engineering advice that assisted in the refinement of a method to assess exposure to motor vehicle emissions.

In addition to these individuals, the agencies involved in the Southeast Toronto (SETO) Urban Health coalition laid the groundwork for this study by identifying respiratory health as a community concern and making relevant data available.

Despite the contributions of the aforementioned individuals, all errors and omissions in the thesis are clearly the sole responsibility of the author.

Finally, I would like to acknowledge the unquantifiable contribution made to this thesis by my wife Stéphanie. Without her support, patience and love this thesis could never have been written.
**TABLE OF CONTENTS**

Abstract .................................................................................................................. ii

Acknowledgements .............................................................................................. iii

Table of Contents .................................................................................................... iv

List of Tables ......................................................................................................... vi

List of Figures ....................................................................................................... vii

List of Maps .......................................................................................................... viii

Introduction ........................................................................................................... 1

  Problem Statement ............................................................................................ 2
  Aim ...................................................................................................................... 2
  Research Approach ............................................................................................ 3

Background ........................................................................................................... 4

  Current Evidence and Theory ........................................................................... 5
    Air Pollution and Respiratory Health .............................................................. 6
    Motor Vehicle Emissions and Respiratory Health ............................................. 9
    A Model for the Relationship between Motor Vehicle Emissions and Respiratory Health ......................................................... 25
  Preliminary Investigations ................................................................................. 31
  Study Area ........................................................................................................ 33

Original Work of Thesis ......................................................................................... 38

  Identification of Hypothesis .............................................................................. 39
    Areas Emerging from Background Research .................................................. 39
    Hypothesis ...................................................................................................... 40
  Selection of Method .......................................................................................... 42
    Overview of Method ....................................................................................... 42
    A Review of Methodologic Issues .................................................................... 43
  Method .............................................................................................................. 53
<table>
<thead>
<tr>
<th>Number</th>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Summary of Studies Examining the Relationship between Motor Vehicle Emissions and Respiratory Health</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>Table 2</td>
<td>Selected Educational, Employment, Income, Mobility and Social Indicators for Neighborhoods in Southeast Toronto</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>Table 3</td>
<td>Benefits and Limitations of Using Administrative Databases for Epidemiological Research</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Table 4</td>
<td>Diagnoses and Associated ICD-9 Codes Used to Abstract Records for Study</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>Table 5</td>
<td>Particulate Matter Emission Factors from PART5 Model and Distribution of Motor Vehicle Types</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>Table 6</td>
<td>Data Reduction for HMRI Data</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>Table 7</td>
<td>Observed and Poisson Distributions of Hospital Admissions by Individual</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>Table 8</td>
<td>Summary Statistics for Admission Counts, Crude Admission Rates and Standardized Admission Rates by Enumeration Area</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>Table 9</td>
<td>Summary Statistics for Exposure to Traffic Volume and PM 2.5 Emissions</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>Table 10</td>
<td>Census Variables Examined for SES Measurement</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>Table 11</td>
<td>Rank Correlation Matrix for Variables in Multivariate Analysis</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>Table 12</td>
<td>Results of Bivariate and Multivariate Poisson Regression Analyses to Address Hypothesis</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td>Table 13</td>
<td>Results of Poisson Regression Analyses to Address Ancillary Research Questions</td>
<td>140</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Overview of Method and Data Sources ........................................ 142</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Hospital Admissions Data Flow .................................................. 143</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Motor Vehicle Emissions Data Flow ............................................... 144</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Exposure Estimation Data Flow Diagram ........................................ 145</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Graphical Display of Point Count to Segment Volume Methods ............... 146</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Distribution of Hospital Admissions by Enumeration Area and Poisson Distribution of Hospital Admissions .................................. 147</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Scatterplot of PM 2.5 Exposure from Motor Vehicle Emissions and Standardized Hospitalization Ratio (SHR) for Respiratory Subset (First Admissions) in South East Toronto by Enumeration Area (N=229) ........................................ 148</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Relative Risk of Admission for Respiratory Subset (Rate of First Admissions) in South East Toronto by PM 2.5 Exposure Quintile 149</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Relative Risk of Admission for Respiratory Subset and GI Chapter (Rate of First Admissions) in South East Toronto by PM 2.5 Exposure Quintile ........................................ 150</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Relative Risk of Admission for Respiratory Subset (Rate of First Admissions) in South East Toronto by PM 2.5 and Traffic Volume Exposure Quintile ........................................ 151</td>
</tr>
<tr>
<td>Number</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Map 1</td>
<td>Neighbourhood and Enumeration Area Boundaries in South East Toronto</td>
</tr>
<tr>
<td>Map 2</td>
<td>Street Network in South East Toronto by Traffic Data Status with Location of Cordon Count Points</td>
</tr>
<tr>
<td>Map 3</td>
<td>Average Daily Emissions of PM 2.5 from Motor Vehicles in South East Toronto for Streets with Over 5000 Vehicles per Day</td>
</tr>
<tr>
<td>Map 4</td>
<td>Enumeration Area Boundaries and Suppressed Enumeration Areas in South East Toronto</td>
</tr>
<tr>
<td>Map 5</td>
<td>Counts of First Respiratory Subset Admissions in South East Toronto by Enumeration Area</td>
</tr>
<tr>
<td>Map 6</td>
<td>Indirectly Standardized First Respiratory Subset Admission Rates in South East Toronto by Enumeration Area</td>
</tr>
<tr>
<td>Map 7</td>
<td>Indirectly Standardized First Admission Rates for Respiratory Subset, Respiratory Chapter, Gastrointestinal Chapter, and Genitourinary Chapter in South East Toronto by Enumeration Area</td>
</tr>
<tr>
<td>Map 8</td>
<td>Average 24 Hour Exposure to PM 2.5 from Motor Vehicles in South East Toronto by Enumeration Area</td>
</tr>
<tr>
<td>Map 9</td>
<td>Proportion of Population with a University Degree in South East Toronto by Enumeration Area</td>
</tr>
<tr>
<td>Map 10</td>
<td>Standardized Likelihood Residuals from Poisson Regression Analysis Addressing Hypothesis</td>
</tr>
</tbody>
</table>
Chapter One

Introduction
Acute exposure to air pollution can cause adverse respiratory health outcomes in humans. A number of recent studies offer strong support for a similar relationship between chronic exposure to air pollution and adverse respiratory health outcomes. In particular, an association is consistently observed between chronic exposure to particulate matter (PM) air pollution and respiratory health. Measures of effect from large cohort studies in combination with data on numbers at risk indicate that the burden of illness from chronic exposure to PM may be considerable.

Motor vehicles are a major source of PM and other pollutants. The relative contribution from motor vehicles is greatest in urban areas, where large numbers of motor vehicles result in elevated PM concentrations near busy streets. Individuals residing near busy urban streets appear to be chronically exposed to these elevated concentrations of pollutants, and this exposure may adversely affect their health. A small number of studies have examined the existence of a relationship between exposure to motor vehicle emissions and respiratory health. Results from these studies tend to support the existence of a relationship, but are not conclusive.

The aim of this thesis is to examine the relationship between exposure to motor vehicle emissions and respiratory morbidity in an urban area.
Research Approach

This thesis is divided into four chapters. The rationale and aim for the thesis, as well as an overview of the thesis structure, are presented in the current chapter. Background information relevant to the thesis is discussed in the next chapter. This includes a review of current evidence and theory, a discussion of preliminary work, and a description of the study area. The method and results, preceded by brief literature reviews, are documented in the third chapter. Information from previous chapters is synthesized and discussed in the fourth chapter. Conclusions are drawn in the final chapter.
Chapter Two

BACKGROUND
The aim of this chapter is to provide a theoretical and community context for the thesis.

The chapter is divided into three sections. In the first section, current evidence and theory are reviewed and synthesized. Preliminary work performed in preparation for the thesis is presented in the second section. Finally, an overview of the socio-economic nature of the study area is presented in the third section.

CURRENT EVIDENCE AND THEORY

This section is divided into three parts. The first part briefly examines the evidence for a relationship between air pollution and respiratory health. This provides a background for the second part, where the evidence for a relationship between motor vehicle emissions and respiratory health is reviewed in detail. The last part of this section synthesizes information from the first two parts and other areas to develop a theoretical model of how motor vehicle emissions may affect respiratory health.

Throughout this section, particulate matter (PM) is used as an example of an individual pollutant found in the complex chemical mixture of motor vehicle emissions. An individual pollutant is identified in this manner to facilitate the discussion around exposure and risk assessment, as these topics are difficult to discuss for complex chemical mixtures. PM is used for this purpose as it appears to have an adverse effect on respiratory health at ambient concentrations, and it has physical properties that are similar to other pollutants found in motor vehicle emissions.
Air Pollution and Respiratory Health

Acute Effects

The ability of air pollution to adversely affect respiratory health has been clearly demonstrated through mortality time-series studies during episodes of heavy air pollution. One of the earliest episodes studied revealed a 10 times increase in mortality (60 deaths, as compared with approximately 6 deaths over the same period in adjacent years) during a pollution episode in Meuse Valley, Belgium (Firket, 1936). A similar pollution episode in Donora, Pennsylvania during 1948 was also associated with a ten times increase in mortality (20 deaths, as compared with approximately 2 deaths over the same period in adjacent years), and a considerable increase in morbidity (Ciocco and Thompson, 1961). The most widely known episode of heavy air pollution however, occurred in London during December 1952 (Ministry of Health, 1954). Despite lower relative increases in mortality (2.3 times) and hospital admissions (1.5 to 2.5 times) than seen in previous episodes, the absolute number affected was considerably larger — 3500 to 4000 deaths. The temporal relationship of exposure and outcomes, as well as the magnitude of the outcomes strongly suggest exposure to heavy air pollution is causally related to mortality and morbidity.

While findings from these studies clearly identify the dangers of exposure to high levels of air pollution, the episodes studied bear little resemblance to the current situation in developed countries. Particle levels during the London episode ranged from 500 μg per m³ to 2 mg per m³ (Ministry of Health, 1954), compared to the current Canadian average of 46 μg per m³ (Brook et al., 1997). Despite the generally lower levels of air pollution seen today, concern still exists over the potential health implications of exposure at current levels. Specific attention has been directed towards discerning the health effects of exposure to fine particles smaller than 10 μm (PM 10) and 2.5 μm (PM 2.5) in diameter.
Time-series studies of daily mortality in relation to current PM 10 levels have been conducted in a number of North American, South American and European cities. A review of these studies suggests that in general, a 10 μg per m³ increase in PM 10 is associated with a 0.7% increase in mortality (Bates, 1996). The relationship appears to be monotonic, with no evidence of a threshold below which exposure has no effect. Furthermore, the association appears to be independent of other pollutants, and occur year round (Saldiva et al., 1995). The consistency of this relationship among different studies suggests that the association is unlikely to be due to chance (Lipfert and Wyzga, 1995).

For the relationship between particulate air pollution and respiratory health to be coherent, associations should also be seen between exposure and morbidity, as well as mortality (Bates, 1992). Time-series studies that demonstrate an association between daily exposure to PM 10 and hospital admission for asthma (Schwartz, 1994a), pneumonia (Schwartz, 1994b), COPD (Burnett et al., 1995), and all respiratory diagnoses (Burnett et al., 1994) add to the coherence of the relationship.

In summary, the evidence suggests that acute exposure to elevated levels of particle pollution is strongly, temporally, consistently and coherently associated with increased mortality. The association with respiratory morbidity at current levels of ambient exposure demonstrates similar qualities, with the exception of a weaker association. When taken together, the evidence suggests that ambient exposure at current levels has the potential to acutely affect the respiratory health of a large number of individuals.

**Chronic Effects**

Time-series studies can demonstrate the acute effects of exposure to particle air pollution, but are not well suited to assessing the chronic effects of ongoing
exposure to ambient levels of pollution. Consequently, a number of ecological and prospective cohort studies have been performed to assess the chronic effects of ongoing ambient exposures.

Ecological studies have consistently shown an association between ambient particle air pollution and mortality (Archer, 1990; Bobak and Leon, 1992), but the results were originally discounted by many due to the implausibly large magnitude of association and the potential for cross-level bias (Pope and Dockery, 1996). These results are being reevaluated following the completion of a number of prospective cohort studies. The cohort studies confirm that an association exists between ongoing exposure to ambient particle air pollution and mortality, after controlling for the effects of age, sex, race, smoking, body mass index, and other risk factors (Dockery et al., 1993; Pope et al., 1995). The adjusted mortality rate ratios for the most versus least polluted city were 1.26 and 1.17 in two of the largest studies (Dockery et al., 1993; Pope et al., 1995).

Prospective cohort studies have also shown coherent relationships between ambient particulate air pollution and morbidity outcome measures. Small negative associations have been found between particulate pollution and lung function (Chestnut et al., 1991; Dockery et al., 1989). Stronger positive associations have been consistently found for particulate pollution with COPD, bronchitis and asthma (Abbey et al., 1993; Dockery et al., 1993; Dockery et al., 1989). Based upon these studies, the approximate effect per 10 μg per m³ increase of PM 10 has been estimated to be a 0% to 2% decrease in lung function and a 10% to 25% increase in respiratory morbidity (Pope and Dockery, 1996).

The results of these studies indicate that ongoing exposure to increased levels of fine particles has the potential to significantly effect the long term health and survival of individuals (Bates, 1996).
Motor Vehicle Emissions and Respiratory Health

The evidence for a relationship between air pollution and respiratory health strongly suggests that this relationship is causal in nature. This ability for air pollution in general to affect respiratory health lends credence to the hypothesis that a relationship exists between motor vehicle emissions and respiratory health. With this background, the following section reviews the evidence for a relationship between motor vehicle emissions and respiratory health.

Only a few studies have examined the relationship between exposure to motor vehicle emissions and respiratory health. Evidence from toxicological, clinical and occupational studies is presented first to develop a picture of the relationship at the individual level and among those exposed to higher concentrations. Evidence from, and methods used in, population based studies are then reviewed. Greater attention is paid to the population level studies given their relevance to the thesis.

Toxicological, Clinical and Occupational Evidence

A small number of experimental studies have examined the toxicological effects of motor vehicle emissions on respiratory health. Findings from these studies are more easily generalized to urban human exposures if two conditions are met. These conditions are the use of subjects with a respiratory anatomy similar to humans, and exposure of subjects to diluted exhaust. One study that meets these conditions exposed female dogs to diluted gasoline engine exhaust for over 5 years (Lewis et al., 1974). Respiratory function did not appear to be affected during exposure, but a number of post-exposure effects were noted (Hyde et al., 1978). These included increases in lung volumes and deadspace ventilation and a decrease in alveolar-capillary gas exchange efficiency. Histopathological examination revealed distinct, but not severe, epithelial hyperplasia in small bronchioles and emphysema in proximal alveoli. In another study, cats examined
after 27 months of intermittent exposure to diluted diesel exhaust demonstrated respiratory function impairment and peribronchiolar fibrosis (Moorman et al., 1985; Plopper et al., 1983). It is not clear which component of motor vehicle emissions is responsible for the inflammatory respiratory changes seen in these studies, but similar changes have been seen in guinea pigs and rats exposed to particulate matter (Chen et al., 1992; Godleski et al., 1996).

The few clinical studies to examine the effect of motor vehicle emission exposure on humans have involved small numbers of subjects and short exposure periods. These factors make it difficult to generalize findings to humans who are chronically exposed to emissions. Nevertheless, two clinical studies have demonstrated that exposure to emissions can affect the respiratory system. A study that exposed eight subjects to diluted diesel exhaust for one hour found changes consistent with decreased intrapulmonary immune function in cell samples obtained by lavage (Rudell et al., 1990). Another study found an approximately 20% increase in airway resistance after exposing twelve normal subjects to diluted diesel exhaust for one hour (Rudell et al., 1996). Interestingly, clinically significant respiratory changes have not been found after similar exposures to respirable carbon particles (Anderson and Avol, 1988).

Short-term occupational studies of respiratory function in workers with heavy diesel emission exposures have reported mixed results. Significant workshift decrements in forced expiratory function have been found for ship workers (Ulfvarson et al., 1987), but not miners (Ames et al., 1982) or garage workers (Ulfvarson et al., 1987). In contrast, studies of longer-term effects of occupational exposure to diesel emissions have more consistently demonstrated adverse respiratory outcomes associated with chronic exposure. Reductions in expiratory flow rate have been found for miners (Ames et al., 1984) and stevedores (Purdham et al., 1987), while increased wheezing and cough have been found in bus garage workers (Gamble et al., 1987). The most compelling evidence comes
from a prospective study of nearly 500,000 US males which estimated significantly elevated relative risks for subjects exposed to diesel exhaust of 1.2 for emphysema and COPD, and 1.7 for pneumonia (Boffeta et al., 1988).

To summarize, the small number of toxicological and clinical studies to date have consistently found exposure to vehicle emissions to be related to mild yet significant inflammatory and obstructive pulmonary changes. These findings have been inconsistently supported by short-term occupational studies, but longer-term occupational studies offer more consistent support.

**Population Level Evidence**

Nine published studies and one unpublished study have assessed the nature of the relationship between motor vehicle emissions and respiratory health at the population level. As both the methods and results of these studies are relevant to the thesis, they are reviewed in detail. The pertinent parts of the studies are reviewed according to the following topics: study design and population; measurement of respiratory health; assessment of exposure; measurement of confounding variables; and, analysis and results. The studies are summarized by these topics in Table 1.

**Study Design and Population**

Two of the ten studies employ a case-control design (Livingstone et al., 1996; Waldron et al., 1995), five use a cross-sectional design (Brunekreef et al., 1997; Nitta et al., 1993; Oosterlee et al., 1996; Weiland et al., 1994; Wjst et al., 1993), and two an ecologic design (Buckeridge et al., 1998; English, 1997).

Of the three types of study design, case-control is traditionally viewed as the strongest design from an etiological perspective. This relative strength is a result of the ability of a case-control design to discern, albeit retrospectively, that
exposure preceded the onset of disease (Rothman and Greenland, 1998a). Unfortunately, the studies that employ a case-control design measure exposure in a cross-sectional manner, and do not account for duration of exposure at the ascertained level (Livingstone et al., 1996; Waldron et al., 1995). Consequently, these studies do not benefit from this design strength. Interestingly, this shortcoming also limits the potential effect of recall bias with respect to exposure.

The two major limitations of a cross-sectional design are the inability to determine the temporal relationship of exposure and disease, and the propensity to identify cases with longer disease duration (Kelsey, 1996). One method of addressing the temporal ambiguity of exposure and outcome is to ascertain previous exposure rather than current exposure. If this is done, a cross-sectional study can be viewed as a case-control study with an excessively large control group, and biased case ascertainment due to the inclusion of prevalent as opposed to incident cases (Rothman and Greenland, 1998d). Three of the cross-sectional studies have assessed previous exposure (Nitta et al., 1993; Oosterlee et al., 1996; Wjst et al., 1993), and a fourth reports that 85% of subjects lived at the same address for at least three years (Brunekreef et al., 1997).

A number of practical reasons exist to support the use of an ecologic study design to assess the relationship between motor vehicle emissions and respiratory health. These include low cost, convenience, and measurement limitations of individual-level studies (Morgenstern, 1998). Measurement limitations of individual-level studies relate to the difficulty of assessing individual level exposure. This difficulty leads all the individual-level studies to assign an ecological exposure value to individuals. Such an approach may result in a biased estimation of relationships in individual-level studies due to correlated error terms amongst individuals (Duncan et al., 1998). Despite these advantages, ecologic studies suffer from methodological limitations that include ecological bias, problems of confounder control, and within-group classification (Morgenstern, 1998). The use of smaller
units in an ecologic study may help to address ecologic bias and within-group misclassification by making the groups more homogeneous with respect to exposure (Morgenstern, 1998). The ecologic study that uses the smaller Canadian enumeration area as the unit of analysis (Buckeridge et al., 1998) accomplishes this to a greater degree than the study that employs the larger US zip code (English, 1997). The problem of confounder control is more difficult to address, and will be discussed below in the section on measurement of confounding variables.

With the exception of one study (Nitta et al., 1993), all of the studies include children in the study population. Five of the remaining study populations are comprised solely of children (Brunekreef et al., 1997; Edwards et al., 1994; Oosterlee et al., 1996; Waldron et al., 1995; Weiland et al., 1994), and the other three include subjects of all ages (Buckeridge et al., 1998; English, 1997; Livingstone et al., 1996). Unfortunately, neither the upper nor lower age bounds are uniformly defined for study populations composed of children.

Measurement of Respiratory Health

Measurement of respiratory health can be compared amongst the studies with regard to what is measured and how it is measured. Four studies measure only asthma (Edwards et al., 1994; English, 1997; Livingstone et al., 1996; Waldron et al., 1995), while the remaining studies measure an assortment of respiratory conditions or symptoms.

The method used to measure respiratory health data is closely related to the choice of study design. Studies with case-control and ecological design all rely upon administrative databases for measurement of respiratory health. One of the case-control studies uses a general practitioner (GP) database to identify cases by diagnostic and medication criteria (Livingstone et al., 1996). The other case-
control study and both ecologic studies rely upon hospital discharge databases to measure respiratory health. A GP database clearly identifies cases of lesser severity than a hospital discharge database, and measures of association from these different approaches may not be directly comparable. Methodological issues around the use of administrative databases are discussed in the methods section of the thesis.

In contrast to the case-control and ecologic studies, the cross-sectional studies all rely upon self-reported symptoms of respiratory health status or physical measurements. The measurement of respiratory health status through a survey offers a number of practical advantages such as low cost and the ability to gather information on other variables of interest simultaneously. However, underreporting and overreporting of conditions on surveys may result in inaccurate measurement of health status when compared against physician reports. Estimates of underreporting on surveys range from 35% to 68%, and estimates for overreporting range from 8% to 60% (Jabine, 1987). While generally viewed as more objective than surveys, physical measurements such as pulmonary function tests (PFTs) may also result in inaccurate or unreliable measurements.

Assessment of Exposure

A standard method for defining or assessing exposure to motor vehicle emissions does not exist, and consequently, a variety of approaches have been employed. Studies differ in what exposures are assessed, where they are assessed, and how they are assessed.

Four of the studies assess exposure to specific pollutants, while the other studies assess exposure to vehicle emissions in general. The studies that examine individual pollutants assess exposure to either particles (Buckeridge et al., 1998), NO$_2$ (Nitta et al., 1993; Oosterlee et al., 1996), or both (Brunekreef et al., 1997). Given that particles and NO$_2$ have similar diffusion characteristics (Thatcher and
Layton, 1995), one would not expect the choice of pollutant to drastically alter the study results.

All of the studies assess location-specific exposure, rather than individual-specific exposure. The majority of the studies assess exposure by location of residence, while one assesses exposure by location of school (Wjst et al., 1993), and another at both locations (Brunekreef et al., 1997). Subjects in these two studies are restricted to children. Notably, the other three studies that examine effects of exposure on children do not assess exposure at school (Edwards et al., 1994; Oosterlee et al., 1996; Waldron et al., 1995), and none of the studies with adult subjects assess exposure at work. The decision to assess exposure at only one location reduces the resources required to complete a study. However, as individuals are exposed to air pollutants at differing concentrations in differing locations, this decision inevitably introduces error into the results of a study. This subject is discussed further in the model of the relationship between motor vehicle emissions and respiratory health.

All of the studies model exposure from traffic data, and one study supplements modelled exposure with measurements of pollutant concentration (Brunekreef et al., 1997). The rationale for modelling exposure is discussed in the methods section of the thesis and is essentially a matter of convenience. One study (Oosterlee et al., 1996) employs the validated CAR (calculation of air pollution from road traffic) model to achieve the most rigorous measurement of exposure (Eerens and Sliggers, 1993). Details of this model are discussed in the methods section of the thesis. The other individual-level studies combine measured data on vehicle traffic or emissions with distance from the street to estimate exposure.

Individual-level studies select highways or major streets in the study area, then assess traffic on the streets, usually from counts taken by traffic authorities. Distance to the closest street is most often measured for each subject from a
paper or digital map, using postal code or street address. Finally, exposure is assessed by including traffic and/or distance to the street as predictor variables in the analyses. Despite the similar method employed by these individual-level studies, it should be pointed out that some studies simplify this approach to the extent that a considerable amount of error is likely introduced into the assessment of exposure. Some of the studies do not account for traffic volume at all (Livingstone et al., 1996; Nitta et al., 1993; Waldron et al., 1995), while another relies upon subject-reported frequency of truck traffic (Weiland et al., 1994). When taking into account distance to the nearest major street, one study determines only if it is in the same school district (Wjst et al., 1993), while another determines only if it is in the same electoral ward (Waldron et al., 1995).

Ecologic studies take a slightly different approach. In these studies, traffic is measured for all major streets in the study area and the traffic counts or emissions for each street are transferred to the surrounding study unit using a Geographic Information System (GIS). Distance is taken into account by dividing traffic counts or emissions for each study unit by the area around streets (Buckeridge et al., 1998; English, 1997), or the total area of the study unit (English, 1997). Finally, the distance-weighted traffic count or emission value is used as a measure of exposure. This approach provides a benefit over the methods used by individual-level studies in that it allows the effect of all major streets in the area to be assessed, as opposed to only the one closest street. This is particularly important for residences and schools near a number of streets or intersections.

**Measurement of Confounding Variables**

The measurement and subsequent control for potential confounding variables is performed to differing degrees amongst the studies. All of the cross-sectional studies collect extensive data on confounding variables using the same questionnaire employed to assess respiratory health status. Of the two case-
control studies, one does not measure confounding variables (Edwards et al., 1994), and the other measures some variables for individuals using administrative data, and other variables in an ecological manner from census data (Livingstone et al., 1996). As discussed above, the attribution of area measures to individuals may bias a study's results. Finally, both ecologic studies employ age standardized rates, and measure socio-economic status with census data (Buckeridge et al., 1998; English, 1997). The limited measurement of confounding variables in the case-control and ecologic studies is a direct result of their use of administrative databases to measure health. This limitation of using administrative databases in epidemiological research is discussed further in the methods section of the thesis.

Analysis and Results
Analytic approaches taken in the studies include analysis of tabular data, multiple logistic regression, and multiple Poisson regression. Despite the similarity in overall analytic approaches, a number of important differences exist in the application of these methods. For the most part, these differences relate to the manner in which exposure is included in the analysis.

The two case-control studies employ either chi-square or multiple logistic regression analytic methods. In the study using a chi-square analysis (715 cases and 736 controls), the association of asthma admissions with both traffic volume and distance to street are assessed independently through stratification (Edwards et al., 1994). As data on potential confounding variables are not collected in this study, the analysis does not control for variation in confounders among subjects. The authors report a significant association with traffic volume (greater or less than 24 thousand per day) using community (OR 1.40; 1.13-1.74) and hospital controls (OR 1.29; 1.04-1.50). This relationship is analyzed further by calculating odds ratios for six levels of traffic volume. With hospital controls, a significant trend of increasing strength is seen among subjects residing within 500 meters
(m) of a major street. A significant association with distance to the street (greater or less than 200 m) is also seen with community controls (OR 1.52; 1.22-1.90), but this effect is not significant with hospital controls (OR 1.16; 0.94-1.44). Interestingly, the authors also report that children admitted to hospital for any cause are significantly more likely to live within 200 m of a major street. While the findings of this study support a relationship between respiratory health (asthma admissions) and exposure (both traffic volume and distance to street), this support must be interpreted in light of potential bias from two sources. To begin with, there appears to be an element of selection bias as evidenced by the different magnitude of association seen with the two control groups. Bias from confounding is also likely to exist given the lack of control for confounding variables in the analysis. This source of bias may contribute to the observed relationship with all admissions.

In the other case-control study (1066 cases and 6233 controls), multiple logistic regression is used to estimate the effect of distance to the nearest major street (traffic volume greater than 1000 vehicles per hour at peak) on asthma prevalence while controlling for age and sex (Livingstone et al., 1996). The effect of traffic volume on the nearest street is not assessed. The authors measure, but do not adjust for smoking or socioeconomic status as they do not appear to be related to distance to the street in their data. No significant effect is found for distance to the street (greater or less than 150 m) among children (OR 0.96; 0.78-1.22) or adults (OR 1.00; 0.84-1.19). Similar findings are reported for an analysis of distance as a continuous variable using spatial point process methods. These results do not support a relationship between respiratory health (asthma prevalence) and exposure (distance to street only). However, the results must be interpreted in light of potential sources of bias. Misclassification of exposure from the lack of consideration given to traffic volume is likely the most important source of bias. It is noteworthy that neither socio-economic status nor smoking
status is related to the measure of exposure (i.e. distance to nearest major street) in this study.

Cross-sectional studies also employ either chi-square or multiple logistic regression analytic methods. The study with the most rigorous approach to exposure estimation (779 subjects) uses the CAR (calculation of air pollution from road traffic) exposure model to identify exposed (116 to 150 μg/m³ hourly NO₂ or approximately 10000 to 30000 vehicles per day on facing street) and unexposed residences (those in same area facing a street with little traffic) (Oosterlee et al., 1996). Multiple logistic regression is used to assess the association of respiratory symptoms with exposure while controlling for potential confounding variables. Significant associations are reported for respiratory medication use in children (OR 2.2; 1.1-4.6) and dyspnoea in adults (OR 1.8; 1.1-3.0). Crude measures of association are also elevated for chronic cough (OR 2.5) and physician diagnosed asthma (OR 2.5) in adults, but model instability precludes the estimation of adjusted measures. A sub-analysis by gender for children reveals associations in females with wheeze (OR 4.4; 1.4-13.6), dyspnoea (OR 4.8; 1.3-17.7), and respiratory medication use (OR 2.9; 1.1-7.9), but no significant associations in males. The findings are suggestive of an association between respiratory health and exposure for children (with measures of association above unity for seven of nine conditions examined) but not for adults. The response rate of approximately 65% does not represent a serious threat to validity but the inherent limitations of the study design must be noted. As discussed above, these include the possibility of bias in symptom reporting and selective migration of subjects.

In another cross-sectional study (1068 subjects), multivariate logistic regression is used to assess the association between measures of respiratory health in children (symptoms and PFTs) with both traffic volume (continuous) and distance to
nearest motorway (home greater or less than 100m) while controlling for confounding variables (Brunekreef et al., 1997). Analyses including all children living within 1000m of a highway reveal consistent but non-significant associations between distance from motorway and chronic cough (OR 1.64; 0.98-2.74) and wheeze (OR 2.00; 0.99-4.03). Similar results are seen for volume of truck traffic with cough (OR 1.61; 0.91-2.84) and wheeze (OR 1.71; 0.72-4.08). A companion analysis of pulmonary function test (PFT) data shows consistent, but non-significant reductions in expiratory flow (Peak Expiratory Flow or PEF, and Forced Expiratory Flow in midexpiratory phase or $\text{FEF}_{25-75\%}$) associated with distance to street, and significant reductions in PEF associated with traffic volume (van Vliet et al., 1997). When the analysis is restricted to children living within 300m of a highway, truck traffic volume is more strongly associated with reductions in expiratory flow, but not respiratory symptoms. Further analysis by gender demonstrates a stronger relationship between traffic volume with both symptoms and PFTs in females. Significant associations are also seen in boys between traffic volume and PEF, but not symptoms. Taken together, the results of this study support a relationship between respiratory health in children (symptoms and PFTs) and exposure (traffic volume). Limited support is provided for the association between respiratory health and distance to street. This finding may be due in part to misclassification introduced by using a binary measure of distance in the analysis. As with another study, stronger associations are seen for females than males. It has been hypothesized that this may be due to the higher prevalence of respiratory conditions in males, and the resulting difficulty of separating the 'signal' of the effect from the greater 'noise' of respiratory morbidity in males (Pershagen et al., 1995). Potential sources of bias in this study are similar to other cross-sectional studies, and the response rate of 71.3% somewhat limits the potential for selection bias. The influence of symptom reporting bias also appears to be limited given the complimentary results seen with symptoms and PFTs.
A series of three cross sectional studies (1148/1758/1916 subjects) uses multiple logistic regression to assess the association between respiratory symptoms and distance of residence to the nearest busy street (20m/20-50m/50-150m) while controlling for confounding variables (Nitta et al., 1993). The traffic volume on the nearest street is not measured. In all three studies, living closer to a street is consistently associated with an elevated prevalence of respiratory symptoms. Living closer to a street is also associated across all three studies with chronic cough (OR 1.62/1.78/1.45) and chronic cough with phlegm (OR 1.47/1.85/1.26). The odds ratios for both symptoms are significantly different from unity for the first two studies, but not the third. Measures of association for chronic wheeze, dyspnoea and cold with phlegm are also consistently elevated, although generally not significantly so. These results support a relationship between respiratory symptoms and exposure (distance to busy street). The studies achieved response rates ranging from 77.3% to 84.7% and only included subjects living at their current address for greater than three years. Limitation by duration of residence helps to limit the possibility of bias from mobility, but may limit the external validity of the study. Other sources of potential bias in this study include the reporting of symptoms, and the classification of exposure (due to the lack of measurement of traffic volume).

The largest cross sectional study (4320 subjects), uses multiple logistic regression to assess the relationship between measures of respiratory health (symptoms and PFTs) and traffic volume (continuous, on busiest street in school district) while controlling for confounding variables (Wjst et al., 1993). Distance to the busiest street is not measured. The analysis estimates small but statistically significant reductions in a number of PFT measures with increasing traffic volume. These include PEF, FEF_{25-75%}, and forced expiratory volume in one second over forced vital capacity (FEV_{1}/FVC). Measures of association between a number of symptoms and traffic volume are also significantly elevated. Lifetime croup (OR
1.09; 1.00-1.18), recurrent wheezing (OR 1.08; 1.01-1.16), and recurrent dyspnoea (OR 1.10; 1.00-1.20) show small increases, while current upper respiratory tract infection (URTI) shows a stronger association (OR 1.20; 1.08-1.34). Measures of association reflect an increase in exposure of 25000 vehicles per day. For comparison, the range of traffic volume in Toronto is from nearly 0 to approximately 80000 vehicles per day. The results of this study support a weak relationship between respiratory health (PFTs and symptoms) and exposure (traffic volume). The authors go to great lengths to limit potential sources of bias in this study as evidenced by the 90.9% response rate. A number of confounding variables are controlled for and the analysis is restricted to subjects residing at their current address for greater than 5 years. Restriction of subjects by duration of residence may decrease the external validity of the study. The greatest potential threat to internal validity is exposure misclassification, as exposure is assessed only for traffic volume and not distance to the street.

One of the most straightforward cross sectional studies (1936 subjects) uses a chi-squared analysis to examine the association between asthma symptoms and traffic volume (presence of highway in electoral ward) (Waldron et al., 1995). Neither distance to the highway, nor potential confounding variables are measured. The prevalence of cough, wheeze and asthma are not elevated in subjects residing in an electoral ward containing a highway. These results do not support an association between respiratory symptoms and exposure (traffic volume), but numerous potential sources of bias are present. To begin with, no potential confounding variables are measured. More importantly, the measurement of exposure is simplified to the point that considerable misclassification is inevitable. Traffic volume is measured in a binary fashion (highway present or absent in large geographical area), and distance to the highway is not measured. Other potential sources of bias include mobility
amongst subjects and symptom reporting bias. The 81.1% response rate suggests that selection bias may not be a problem.

Another relatively straightforward study (2050 subjects) uses multiple logistic regression to assess the relationship between asthma symptoms and self-reported exposure (traffic volume and street type) while controlling for some potential confounding variables (Weiland et al., 1994). Wheezing is associated with both 'frequent' truck traffic (OR 1.53; 1.06-2.20) and 'constant' truck traffic (OR 1.67; 1.05-2.66) when compared to 'no' truck traffic. In addition, wheezing is associated with living on a main street (OR 1.28; 1.01-1.64) as opposed to a side street. These results support a relationship between wheezing and exposure (traffic volume and street type), but they may be influenced by several sources of bias. The most notable potential source is exposure misclassification because of self-reported traffic volume. Other potential sources of bias include confounding from lack of control for socioeconomic status, and bias introduced from subject mobility.

The two ecologic studies use multiple Poisson regression to assess the association between respiratory hospital admissions and exposure (geographic information system model based upon traffic counts and area around streets, or area of study unit), while controlling for socioeconomic status. Specifically, one of the studies looks at age-standardized asthma admissions for children, and calculates exposure by zip code (English, 1997). The other study examines age-standardized respiratory admissions for all ages, and calculates exposure by census enumeration area (Buckeridge et al., 1998). Both of the studies report a dose-response gradient with increasing risk of hospitalization associated with increasing exposure. The magnitude of association is not reported for one of the studies (English, 1997) and appears to be implausibly large in the other study (Buckeridge et al., 1998). Nevertheless, the complimentary findings of the two
studies suggest that an ecologic approach to this problem is reasonable, but that methodological refinements are necessary.

In summary, a number of points from the studies reviewed above are noteworthy. The most apparent is that eight of the ten studies report a significant association between respiratory health and some measure of exposure to motor vehicle emissions. The consistency of the observed association among these studies with different methods and populations supports the interpretation that the association may be causal in nature. From a respiratory health perspective, the studies support an association with respiratory hospital admissions, symptoms of lower airway inflammation (e.g. cough and phlegm production), and PFT changes suggestive of lower airway obstruction (Weinberger, 1992).

Closer examination of the methods reveals a consistent association between exposure and respiratory health in all three studies that use an exposure model to estimate the effect of traffic volume and distance to street in an integrated fashion. In contrast, an association with respiratory health is inconsistently observed with distance to street or traffic volume in isolation. Respiratory health is associated with distance to street in three of four studies, and with traffic volume in four of five studies. Although it is difficult to generalize from a small number of studies, it appears that the complex procedure of exposure assessment is best accomplished by an exposure model that integrates the effects of traffic volume and distance to street. This is supported by the observation that both studies reporting no association assess exposure through only traffic volume or distance to street.

It is also noteworthy that the strength of association between respiratory health and exposure to motor vehicle emissions appears to be stronger in children, and especially in girls. This finding is plausible in light of the high volume of air inspired by children (per kilogram of body weight) and gender differences in the
prevalence of respiratory disease as discussed above (Pershagen et al., 1995). In addition, the only study that looks at the association between non-respiratory conditions and exposure reports a relationship between all hospital admissions in children and distance to the street (Waldron et al., 1995). This finding may be a result of confounding by socioeconomic status, but other studies have found little or no relationship between distance to street and socioeconomic status (Brunekreef et al., 1997; Livingstone et al., 1996). The lack of such an association is counter-intuitive, and may be due to imperfect measures of socioeconomic status and/or distance to the street.

A Model for the Relationship between Motor Vehicle Emissions and Respiratory Health

An underlying goal of nearly all epidemiological studies is to explore aspects of what might be a causal relationship. No study in isolation, and certainly not an ecologic study, can prove a causal relationship of this nature. Nevertheless, it is worthwhile to explore the plausibility of such a relationship using a theoretical model. This serves to identify assumptions, and establish the existence of a plausible basis for a causal relationship independent of the study results.

A simple theoretical model that draws on published results from a number of areas is presented. This model follows emissions of PM 2.5 from motor vehicles and uses data on dispersion, exposure and respiratory physiology to demonstrate a plausible link between motor vehicle emissions and adverse respiratory health.

Nature of Emissions from Motor Vehicles

Emissions from internal combustion engines found in motor vehicles contain a number of substances. The exact composition varies in relation to a number of factors including fuel, speed, load, and temperature (Westerholm and Karl-Erik,
1994), but some substances are found in nearly all emissions. These 'core' substances include CO, NO, hydrocarbons and particulate matter (Sawyer, 1993). All of these substances have been shown to have adverse effects on human health (Bascom et al., 1996).

A great deal of attention has recently been given to particulate matter (PM), and much of this attention has been on respirable PM 2.5. Evidence for the respiratory effects of PM 10 and PM 2.5 has been reviewed above, and PM 2.5 is used as an example of emissions in this model. Two reasons suggest that PM 2.5 is the most appropriate choice. The first is that PM 2.5 appears to have greater respiratory toxicity than PM 10 (Samet, 1996). Another reason is that PM 2.5 is a subset of PM 10, and due to the preponderance of small particles PM 2.5 accounts for approximately 90% of the particles found in PM 10 (Brook et al., 1997; Koutrakis and Sioutas, 1996).

**Contribution of Motor Vehicle Emissions to Ambient Levels**

The adverse health effects of PM emissions from motor vehicles are a function of the inherent toxicity of PM and the ambient concentration of PM from motor vehicles. Inherent toxicity has been discussed previously, and this section examines the contribution of motor vehicles to the ambient concentration.

The ambient concentration is related to the emissions and dispersion of PM from motor vehicles, as well as the contribution of PM from other sources. At a national level, transportation is estimated to account for 10% of direct PM 2.5 emissions (Environment Canada, 1997). However, in urban areas, source apportionment studies have estimated this contribution to range from 25% (Japar, 1995) to 35% (Chow et al., 1988). This increased contribution is likely due to the large number of vehicles in urban areas.
Based on this evidence, it appears that the contribution of motor vehicle emissions to ambient levels of PM 2.5 is considerable, especially in urban areas. Moreover, it also appears that levels of PM 2.5 vary within urban areas. A recent study found that the average ambient concentration of PM 2.5 near a major street was 31% higher than the background level in a Canadian urban center (Brook et al., 1997). This implies that the contribution of motor vehicle PM 2.5 emissions to ambient levels is likely to be considerable, and most pronounced near major streets in urban areas. Models (Eerens and Sliggers, 1993) and measurements (Nitta et al., 1993) of dispersion of motor vehicle emissions support this conclusion.

**Relationship between Ambient Levels and Exposure**

The argument to this point suggests that the ambient concentration of PM 2.5 (and likely other pollutants) near streets in urban areas is considerably higher than background levels. However, the existence of elevated concentrations in these locations does not necessarily imply that individuals are exposed to these concentrations. Individual exposure is a function of both ambient concentration and time spent in an environment with that concentration.

Varied activity patterns of individuals make it difficult to account for exposure in all environments when estimating a person's total daily exposure. To avoid this problem, many studies estimate exposure for only the one environment where individual's spend the majority of their time. As mentioned previously, this environment is clearly inside an individual's residence. Detailed time budget studies suggest that on average individuals spend 72.8% of their time inside their home (Moschandreas, 1981). Stratified analyses suggest that this value ranges from approximately 60% for those in the labor force to nearly 90% for those not in the labor force (Moschandreas, 1981). Despite this variation, it is clear that the
average individual spends the majority of their time inside their place of residence.

Up to this point it has been argued that PM 2.5 concentrations are higher near busy streets, and that individuals spend the majority of their time inside their homes. These two elements are necessary but not sufficient to determine that individuals residing near busy streets are exposed to elevated levels of PM 2.5 from motor vehicle emissions. The third and final element that must be considered with respect to determining exposure is the relationship between outdoor air quality and indoor air quality.

Indoor concentration of PM 2.5 is influenced by particles generated indoors and particles generated outdoors. The relative influence of particles generated outdoors on indoor concentration is a function of a number of factors. These factors include the ambient particle concentration outdoors, the ability of outdoor particles to penetrate into a building, and the concentration of particles in a building attributable to indoor particles. The evidence presented above suggests that the ambient particle concentration outside buildings near busy streets is elevated relative to other parts of urban areas.

Studies of residential air exchange rates have shown that on average between 50% and 150% of the air within a residence is completely exchanged each hour (Murray and Burmaster, 1995). The lower value represents the median value for US homes in the winter and the upper value the median value in summer. In addition to seasonal variation, exchange rates also vary by building type and age, but in general air within a residence is completely replaced by outdoor air a number of times each day. Due to their small diameter, PM 2.5 particles have similar physical dispersion characteristics to non-reactive gases and penetrate building envelopes at the same rate (Thatcher and Layton, 1995).
Using the data presented above and semi-empirical physical models, it has been predicted that between 60% and 80% of the outdoor PM 2.5 concentration is found indoors at equilibrium (Ozkaynak and Spengler, 1996). These predictions are supported by field studies that have found indoor to outdoor PM 2.5 ratios of 60% to 70% in homes with no smokers (Wallace, 1996). The relative contribution of outdoor particles to the indoor particle concentration is reduced when indoor particles are generated from activities such as smoking and cooking with gas stoves. Unfortunately, accurate estimates of the magnitude of the contribution from these indoor sources are not currently available (Ozkaynak and Spengler, 1996). However, studies employing personal monitoring of particulate exposure suggest that on average outdoor particulate levels contribute about 60% to an individual’s total particulate exposure (Ozkaynak et al., 1996).

To summarize the evidence to this point, it appears that particle levels are elevated outside residences near busy streets. Furthermore, individuals spend the majority of their time in their residence, and outdoor particles readily penetrate into homes to reach equilibrium of between 60% and 80% of the outdoor concentration. The relative contribution of outdoor particle levels to an individual’s total particulate exposure will vary depending upon indoor sources, but on average appears to be approximately 60%. Therefore, although the indoor residential exposure to particles from motor vehicles is variable and difficult to quantify exactly, it can be concluded that individuals residing near busy streets are likely to have elevated exposure to particles from motor vehicle emissions.

**Potential Mechanism of Action**

The final step in demonstrating the plausibility of the relationship between motor vehicle emissions and respiratory health is to explore potential physiological mechanisms for the adverse health effects of particles.
Most of the evidence around potential mechanisms of evidence is drawn from the clinical and toxicological evidence reviewed above. From a clinical perspective, the mechanism whereby respirable particles reach the lower airways is straightforward. In general, larger particles (> 5 μm in diameter) are filtered out in the nose and nasopharynx by impaction, while smaller particles behaving in a manner similar to gases penetrate down to the bronchioles and alveoli (Samet, 1996). Once particles have reached the lower portions of the respiratory system, they can exert toxicological effects. A number of mechanisms for these toxicological effects have been postulated including induction of increased airway permeability, impairment of host defenses, and induction of bronchiolar-alveolar inflammation (Bates, 1992; Schlesinger, 1995). The toxicological evidence reviewed earlier supports the mechanism of airway inflammation for both motor vehicle emissions in general (Hyde et al., 1978; Moorman et al., 1985) and particulate matter specifically (Godleski et al., 1996), and the mechanism of decreased immune function for motor vehicle emissions (Rudell et al., 1990) but not particulate matter (Anderson and Avol, 1988). Despite the apparent plausibility of these mechanisms, there is little evidence to suggest how they occur at the cellular level (Samet, 1996).

In summary, this model presents evidence to support a plausible causal relationship between exposure to motor vehicle emissions and adverse respiratory health. The evidence shows that pollutants emitted by motor vehicles are potentially harmful and that these pollutants exist at elevated concentrations near streets in urban area. In general, it is known that individuals spend the majority of their time inside their homes. Furthermore, outdoor pollutants penetrate into homes and make a considerable contribution to indoor concentrations and total pollutant exposure. Individuals in these homes are therefore likely to be exposed to pollutants from motor vehicles. Finally, there are plausible but not completely
understood mechanisms whereby these pollutants may damage the human respiratory system.

**Preliminary Investigations**

A number of aspects of this thesis have evolved from preliminary investigations completed as part of a pilot study for the thesis (Buckeridge et al., 1998). The pilot study has been reviewed along with other pertinent literature in the section on current evidence, but due to its direct bearing on the thesis, a closer examination is indicated. The following paragraphs serve to outline the similarities and differences between the pilot study and the thesis.

The pilot study and the thesis are similar with respect to the aim, study area, and overall structure. The aim of assessing the relationship between motor vehicle emissions and respiratory health arose from a review of the evolving literature in this area. Southeast Toronto was chosen as the study area due to efforts already underway to develop a respiratory health geographical information system (GIS) in this area and the resulting availability of data sets. Finally, an ecologic study design was used for reasons outlined above in the review of current evidence.

The two studies differ primarily in relation to the variables examined and the analysis. Hospital separation data is used to measure respiratory health in both studies, but the method of abstraction and the variables calculated differ. The pilot study utilized hospital separation counts abstracted by another research group and only calculated rates for all admissions in the respiratory chapter. In contrast, for the thesis, separation counts are abstracted from the raw data in a systematic manner for two measures of respiratory health and two comparison sets. In addition, rates are also calculated for first admission as well as all admissions. Systematic abstraction allows a number of observations to be made.
about the methods, and the use of comparison sets affords a greater understanding of the mechanisms underlying observed associations. Finally, differences between rates of first admission and all admissions are informative with respect to etiology and statistical model fit.

The area of greatest difference between the studies is in the assessment of exposure to motor vehicle emissions. In the pilot study, a small number of vehicle count points from a single data source were used to estimate traffic volumes for the entire study area. Given this paucity of data, the selection of streets to model was based upon street characteristics that did not include traffic counts. Emission modelling was performed in a simple manner using unvalidated conversion factors for total particulate matter. Furthermore, exposure was assessed from this data using a model that did not account for street length or shape and size of the study units. The thesis addresses these issues in a number of ways. Four sources of traffic count data are used to provide a much richer picture of traffic volume and vehicle type distribution in the study area. This data is in turn used to select streets for modelling based on traffic volume. A validated emission model is used to estimate emissions of PM 2.5 instead of all particulate matter. The exposure model is refined to consider street and study unit characteristics in the assessment of exposure. Finally, exposure is also assessed for vehicle counts to determine the influence of emission modelling.

Measurement of socio-economic status in the pilot study was performed using a modified Jarman index (Jarman, 1983). Canadian census variables were selected that most closely matched the UK variables used in the original Jarman index, and original or estimated weights were used to construct a weighted index. The thesis takes a more rigorous approach to the measurement of socio-economic status. A method described in the literature (Frohlich and Mustard, 1996) is used to construct an index by identifying important variables and calculating
appropriate weights. Furthermore, the explanatory ability of this index is compared to that of individual socio-economic variables.

Multivariate Poisson regression is used as the main method of statistical analysis in both studies, but a number of analytic differences exist between the studies. To begin with, greater attention is paid to assessing the fit of the Poisson model in the thesis. This provides important information to assist in the interpretation of modelled results, and informs future modelling endeavors. Also, while the pilot study did not take into account the spatial distribution of variables, the thesis assesses the spatial distribution of individual variables, and determines the influence of these distributions on the multivariate model.

In summary, the methods used in the thesis represent a considerable refinement of those in the pilot study. Specifically, a similar aim, study area and overall study design are employed, while improvements are made in the analysis and the measurement of respiratory health, exposure to motor vehicle emissions and socio-economic status.

**Study Area**

Southeast Toronto (SETO), the area used for this study, is a portion of the largest urban area in Canada, the City of Toronto. Given the ecologic nature of this study, the socioeconomic character of SETO has the potential to influence the results of the study. Attempts are made to control for this influence in the analysis, but such control is unlikely to be complete for a number of reasons including the inability of aggregate measures to capture the variation within a study unit. For this reason, it is important to provide a community context for the study, with a focus on the socioeconomic variability within SETO.
The former City of Toronto had been divided into 45 neighborhoods by the former Toronto Public Health for community health planning purposes. Data from the Census of Canada and other sources is used by Public Health to create detailed social, demographic and health profiles for each of the neighborhoods (McKeown, 1995). These neighborhood profiles are used below to describe and compare the 10 neighborhoods found within SETO.

In the 1991 Census of Canada, the 45 neighborhoods in the former City of Toronto contained 635,127 individuals. Ten of these neighborhoods and 121,875 individuals are within SETO. SETO borders Lake Ontario to the South, the downtown core to the West, the CPR tracks to the North, and Coxwell Street to the East. Map 1 shows major streets and neighborhood boundaries within SETO, along with enumeration area boundaries. On average, each neighborhood contains 34 enumeration areas, but this number varies with the size and population density of the neighborhood.

Selected educational, employment, income, and social indicators are displayed in Table 2 for each of the neighborhoods. Variables within each neighborhood are highlighted when their values are significantly different from the mean value for the former City of Toronto and the magnitude of the difference is 50% or greater. Examination of educational attainment variables reveals that the neighborhoods fall into three groups. A high proportion of the population in Rosedale, Yorkville, Church, Cabbagetown and St Lawrence have received a university degree (Patychuk and Yokimco, 1997a; Patychuk and Yokimco, 1997b; Patychuk and Yokimco, 1997f; Patychuk and Yokimco, 1997i; Patychuk and Yokimco, 1997j). In contrast, St James Town, Moss Park, South Riverdale and North Riverdale closely resemble the former city's average of grade 9 and university completion (Patychuk and Yokimco, 1997c; Patychuk and Yokimco, 1997d; Patychuk and Yokimco, 1997g; Patychuk and Yokimco, 1997h). Finally, a low proportion of the population in Regent Park has completed grade 9.
(Patychuk and Yokingco, 1997e). The Regent Park area contains the largest public housing project in Canada.

The lowest unemployment rates are seen in Rosedale, Yorkville, Cabbagetown and North Riverdale. Rates closer to the Toronto average are seen in Church, St James Town, St Lawrence, and South Riverdale. The highest unemployment rates are seen in Moss Park and Regent Park. The female labor force participation rate does not show a large amount of variation, but appears to be greater in Church, Cabbagetown and St Lawrence, and lower in Regent Park and Moss Park.

Income variables are distributed in a similar manner to educational attainment. Median family incomes are higher in Rosedale, Yorkville and Cabbagetown, with Rosedale and Yorkville also having a lower proportion of low income families and persons on social assistance. Interestingly, Cabbagetown has a proportion of low income families closer to the former city's average, and a higher proportion of individuals on social assistance. This reflects the diversity of Cabbagetown, which is composed of areas of 'gentrified' homes occupied by professionals, as well as areas of lower income housing. Income variables in the neighborhoods of Church, St Lawrence and North Riverdale are similar to the Toronto average. Finally, higher proportions of low income families and/or persons on social assistance are found in St James Town, Moss Park, Regent Park, and South Riverdale. Regent Park is notable for a particularly low median family income of $18214.

Like the female labor force participation rate, mobility in the last five years does not show a large amount of variation. However, it tends to be lower in Rosedale, and higher in Church, Moss Park and St Lawrence. Higher values in these three areas may be due to the presence of Ryerson Polytechnic University in that area.
Examination of social characteristics reveals that Rosedale contains a low proportion of single-parent families. The proportions in Yorkdale, Cabbagetown, South Riverdale and North Riverdale resemble the Toronto average. Significantly higher proportions are seen in Church, St James Town, Moss Park, Regent Park, and St Lawrence. When proportions of recent immigrants are compared amongst the neighborhoods, it is apparent that Rosedale and Yorkville have low proportions. With the exception of St James Town, the rest of the neighborhoods have proportions of recent immigrants comparable to the Toronto average. St James Town, with recent immigrants accounting for 13.4% of the population, is composed of mainly high rise apartment buildings.

In summary, a considerable amount of SES variation exists among SETO neighborhoods. The neighborhoods of Rosedale and Yorkville consistently rank among the most advantaged while St James Town, Moss Park, Regent Park and South Riverdale tend to be the most disadvantaged. In between these two extremes, Church, Cabbagetown, St Lawrence and North Riverdale are comparable to the former City of Toronto average. In addition to the considerable variation between neighborhoods, there also appears to be variation within some neighborhoods such as Cabbagetown. The use of the enumeration area as the unit of analysis will help to address the variation within neighborhoods. The potential influence of this community context on the study results will be discussed in the last chapter of the thesis.

To summarize the chapter, the aim was to provide a theoretical and community context for the thesis. The theoretical context was provided in part by demonstrating that the literature strongly supports what appears to be a causal relationship between respiratory health and exposure to air pollution. Of greater
relevance to the thesis, the literature also supports a relationship between respiratory health and exposure to motor vehicle emissions. The support for this relationship is inconclusive, possibly due to difficulties with exposure assessment. A theoretical model of the relationship developed in this chapter suggests that the existence of such a relationship is plausible. Finally, the community context was provided by a description of the study area that demonstrates the social and economic diversity of SETO.
Chapter Three

ORIGINAL WORK OF THESIS
The aim of this chapter is to develop a hypothesis, select an appropriate method to test this hypothesis, and report the results of implementing the method.

The chapter is divided into four sections. In the initial section, areas emerging from previous research that require further study are identified, and a hypothesis is developed. Methodological issues related to the hypothesis are reviewed in the following section, and a method to test the hypothesis is selected. The method and literature directly relevant to the method are presented in the next section. Finally, the results of the study are presented.

IDENTIFICATION OF HYPOTHESIS

Areas Emerging from Background Research

One area emerging from background research is the need for a straightforward and valid means of assessing exposure to motor vehicle emissions. In the majority of previous studies, ad hoc approaches such as measurements of traffic volume on the nearest street, and distance to the nearest street have been used to estimate exposure to motor vehicle emissions. The implicit assumptions made in these approaches are rarely examined, and the results are often difficult to interpret. In contrast, studies that employ exposure models to integrate emission and dispersion of a single pollutant based upon explicit assumptions have the potential to give results that are more meaningful. Given the evidence for respiratory effects of particulate matter (PM), this pollutant appears to be an excellent candidate for modelling in this manner. Another aspect of exposure assessment emerging from background research is that it is often only feasible to estimate exposure for areas as opposed to individuals. The use of an area measure
of exposure presents a methodological problem for individual level and ecologic studies.

Selection of the most appropriate measure of respiratory health to assess in relation to exposure is another issue emerging from background research. Exposure to emissions has been associated with respiratory hospital admissions, physical measures and symptoms of respiratory health, but not measures of respiratory health in an ambulatory setting. Furthermore, studies have found exposure to be associated with both prevalence and incidence measures of various respiratory health conditions. It is not clear from these epidemiological studies or from toxicological research if exposure is more plausibly linked with exacerbation or causation of respiratory conditions. Some studies have shown a stronger association in children, and particularly in girls. It is also interesting that one study found a simple measure of exposure to be related to hospital admissions for all causes. This finding may represent an artifact of incomplete control for confounding variables, or possibly a real association.

**Hypothesis**

The null hypothesis for the thesis is:

*Modelled area exposure to PM 2.5 from motor vehicles has no effect on hospitalization rates for selected respiratory conditions.*

The effect of modelled exposure is assessed as measurement is not feasible, and modelling can estimate the effect of a single pollutant. The pollutant chosen for modelling is PM 2.5 due to its association with respiratory health and vehicle emissions as outlined in the previous chapter.
An ancillary research question is posed to address the validity of emission modelling and the relative contribution of PM 2.5 to any observed association. This question asks if modelled exposure to traffic volume demonstrates an effect on rates of respiratory admissions. The existence of a relationship with traffic volume exposure, theoretically a crude proxy for PM 2.5 exposure, has two possible interpretations. First, it may imply that the additional steps taken to model PM 2.5 exposure (beyond those for modelling traffic volume exposure), introduce more error into exposure assessment than they remove. Second, a relationship between traffic volume and respiratory admissions could also suggest that other pollutants in emissions have greater health effects than PM 2.5.

This hypothesis is directed at an area-level analysis for a number of reasons, foremost of which is the difficulty in modelling exposure for individuals. Other reasons for this choice are also practical in nature, and they are discussed in the following section.

Hospitalization rates for selected respiratory conditions are used as a measure of respiratory health due to the availability and validity of hospitalization data relative to other sources of health data. Selected respiratory conditions are examined, as not all respiratory conditions can plausibly be associated with exposure to PM 2.5. In addition, conditions with exposure-outcome lag times greater than the temporal frame of this study (e.g. cancer) are not examined. As it is not clear if exposure is more plausibly linked with exacerbation or causation of respiratory conditions, the relationship of exposure with both prevalence and incidence of hospital admission is assessed. Given the observed association of exposure with admissions for all causes in a previous study, an ancillary research question asks if exposure to PM 2.5 has an effect on hospitalization for non-respiratory conditions. This question addresses the specificity of any observed effect on respiratory hospitalization. Unfortunately, the small number of hospitalizations precludes sub-analyses by age or gender. The use of ambulatory
data in addition to hospitalization data was examined in preparation of the thesis, but problems with completeness of coverage and validity of diagnosis make the use of this data impractical for the thesis.

To summarize, this section has identified a hypothesis for the thesis, and two ancillary research questions. The hypothesis is listed above and the two ancillary research questions are:

\[
\text{Does exposure to modelled PM2.5 have an effect on hospitalization rates for non-respiratory conditions?}
\]
\[
\text{Does modelled exposure to 'traffic volume' have an effect on hospitalization rates for selected respiratory conditions?}
\]

**Selection of Method**

This section begins by presenting an overview of the method selected to address the hypothesis. The overview introduces the method and provides a context for the review of methodological issues that follows. Issues that have influenced the selection of the method are examined in the review, along with the strengths and weaknesses of the method.

**Overview of Method**

An ecologic study design is used with the census enumeration area (EA) as the unit of analysis. The study area is defined geographically by the boundaries of South East Toronto (SETO), and the study population includes SETO residents of all ages who were admitted to an Ontario hospital between 1990 and 1992.
Respiratory health is measured from hospital admission data based upon diagnostic coding. Three-year age and sex standardized hospitalization rates are calculated for a subset of respiratory diagnoses associated with exposure to PM 2.5 air pollution. As a comparison, standardized hospitalization rates are also calculated separately for all respiratory, gastrointestinal (GI), and genitourinary (GU) admissions.

EA exposures in average daily grams of PM 2.5 are modelled from emissions of PM 2.5 and EA street frontages using a Geographical Information System (GIS) model. Emissions of PM 2.5 used in the GIS model are estimated from traffic volume and vehicle type data for major streets in SETO using motor vehicle particulate modelling software (PART5) from the United States Environmental Protection Agency (Office of Mobile Sources, 1995).

Socio-economic status is measured from 1991 Census of Canada data. Census variables, individually and in a composite index, are examined to arrive at the most appropriate means of measurement.

Following analysis of the aspatial and spatial distribution of each variable, Poisson regression is used to examine the association among variables, and test the hypothesis of the thesis. Finally, the influence of spatial distributions on the Poisson regression analysis is examined.

A graphical overview of the method for the thesis is provided in Figure 1.

A Review of Methodologic Issues

The review begins with a discussion of the study design. Issues around measurement or assessment of each of the variables are then reviewed. Finally, issues around the analysis of area-level health data are discussed.
**Study Design**

As mentioned previously, an ecologic study design is chosen for a number of practical reasons. These reasons are cost, convenience and measurement benefits over individual-level studies (Morgenstern, 1998). Cost and convenience relate to the ability to perform an ecologic study with secondary data sources, thereby avoiding the need to collect primary data. Measurement benefits relate to the difficulty of assessing individuals’ exposure to motor vehicle emissions. Exposure is routinely assessed for areas as opposed to individuals, and the attribution of area exposure to individuals is a potential source of bias (Duncan et al., 1998). An ecologic study design does not avoid this problem, but two ecologic studies have shown an association between area measures of respiratory health and area measures of exposure to motor vehicle emissions (Buckeridge et al., 1998; English, 1997).

The practical reasons favoring an ecologic design are balanced, but not completely outweighed by a number of methodologic limitations of this design. These limitations include ecologic bias, within-group misclassification, problems of confounder control, lack of available data, temporal ambiguity, and migration across groups (Morgenstern, 1998). The use of a small study unit such as the EA helps to address the potential limitations of ecological bias and within-group misclassification. The problem of confounder control is closely linked to the lack of available data. Detailed data on confounders is not routinely available, and consequently proxy measures are used, or confounders are simply not measured. This limitation is discussed below in relation to administrative databases and socio-economic status. Temporal ambiguity is a problem resulting from the cross sectional assessment of respiratory health and exposure. Consequently, it is impossible to know how long individuals have received a given exposure prior to hospitalization. Finally, migration across groups can be viewed as the spatial analogue of temporal ambiguity. Current residents of low-exposure areas may
have moved to that area after a number of years in a high-exposure area, and vice-versa. Subsequent hospitalization of these residents could then be erroneously associated with low exposure. Unfortunately, no practical approach exists for assessing the bias introduced by temporal ambiguity or mobility.

*Use of Administrative Databases to Measure Health Status*

Several benefits and limitations of the use of administrative databases for epidemiological research are consistently identified (Motheral and Fairman, 1997; Paul et al., 1993; Sorenson et al., 1996; Wray et al., 1995). These benefits and limitations are listed in Table 3.

Many of the benefits and limitations arise from similar issues. One central issue is that administrative data are collected for non-research purposes, prior to the initiation of the study. This provides a substantial benefit in minimizing study resources required for the collection of data, but leads directly to some limitations. Most of these limitations result from the lack of control the investigator is afforded around the content of the databases, and the methodology employed to collect data.

A critical content issue is the potential for bias in the selection of individuals entered into the database. This source of bias is of greater concern if the aim of a study is to describe absolute occurrence data as opposed to relative rates within the database (Sorenson et al., 1996). A study design that relies on comparisons among groups in the same database serves to minimize this problem (Motheral and Fairman, 1997). Another content issue is that databases rarely contain all the data required to address a study question. This is most frequently the case for exposures and confounding variables that may affect the outcomes routinely measured by outcomes databases. A strategy routinely employed to address this limitation is linkage of the administrative databases to other databases containing
data of interest (Paul et al., 1993). Linkage can be performed by personal identifiers, or as in this study, by geographical location.

The rationale and methodology for maintaining an administrative database will affect the quality of the data it contains. This can be seen as a benefit in that databases collected for claims purposes tend to be complete (Motheral and Fairman, 1997). More frequently however, it is seen as a limitation in that diagnostic coding systems used for administrative databases are problematic. For example, the ICD-9 coding system does not designate sidedness or level of disease, and identifies clinical entities with variable specificity and exactness (Wray et al., 1995). These limitations can be minimized by not relying upon diagnoses that require a designation of sidedness or level, and grouping similar diagnoses into diagnostic sets (Motheral and Fairman, 1997; Wray et al., 1995). Other limitations that are not as easily minimized include biases associated with comorbidity, case-mix and severity of illness. Re-abstraction studies suggest that chronic conditions tend to be underreported in hospital discharge data, especially when patients die in hospital (Romano and Mark, 1994). This is of particular importance to the thesis as many of the respiratory conditions of interest are chronic in nature. Case-mix may also affect diagnostic data as case-mix groups (CMGs) used in Canadian hospitals require the ‘most responsible diagnosis’ determined at discharge to be coded as the primary diagnosis (Pink and Bolley, 1994). Clearly, the most responsible diagnosis at the time of discharge may not be the diagnosis that led to admission. Finally, hospital discharge data does not contain information about severity of illness, and individuals admitted with the same diagnosis may have illnesses of differing severity (Taylor, 1998).

Another issue that results in both benefits and limitations is the large size of administrative databases. The direct benefit is the large sample size they usually provide, especially for relatively rare events such as hospital admissions in small areas. Limitations associated with large size are related to the interpretation of
results from poorly developed and structured studies. This limitation can be minimized by developing the study method from a clear conceptual model (Motheral and Fairman, 1997; Wray et al., 1995), and carefully documenting the steps required for the construction of analytical files (Paul et al., 1993). As with any study, the degree of confidence in the findings from a study using an administrative database is a function of the strength of the logic used to develop the study (Motheral and Fairman, 1997).

**Assessing Exposure to Motor Vehicle Emissions**

Previously in the thesis, it has been noted that studies of the relationship between respiratory health and motor vehicle emissions tend to rely on modelled as opposed to measured exposure. The following paragraphs outline the rationale for this reliance, and provide evidence to support the modelling method used in the thesis.

Measurement studies have found that little spatial variation exists in the concentration of particulate matter (PM) and other pollutants in urban areas when sampling is performed away from streets (Brook et al., 1997; Burton et al., 1996). However, when pollutant concentration is measured near streets (i.e. within 10m), considerable variation in concentration from the urban background level is seen (Brook et al., 1997; Croxford et al., 1996; McNair et al., 1996). These findings suggest that motor vehicle emissions result in an elevated concentration of pollutants in the immediate proximity of streets, but the concentration decreases to approach the background urban level as distance from the street increases.

This natural behavior of vehicle emissions makes fixed monitoring stations (which are usually located away from streets) of little use in assessing exposure to vehicle emissions in urban environments. To be meaningful, measurements must
be made near streets, and the cost of making such measurement over a large area is prohibitive. For these reasons, modelling is usually employed to assess exposure to motor vehicle emissions in urban areas.

All exposure models require motor vehicle emissions from the street network as a minimal input. Emissions are usually modelled from traffic counts and vehicle type specific emission factors (Hewison et al., 1996; Rayfield et al., 1995; Valkonen et al., 1996). It has been estimated that emissions modelled from samples of 20% or more of a street network can provide unbiased estimates of emissions modelled from the complete network (Hewison et al., 1996).

Once emission values have been calculated for streets, a dispersion model can be used to estimate the concentration of emissions at a point or area near a street. A number of complex models have been developed to perform such estimations in urban areas (Holscher et al., 1993; Kono and Ito, 1990; Valkonen et al., 1996; van den Hout et al., 1989). These models employ Fickian diffusion equations and require the input of a large number of meteorological parameters that are not routinely available. However, sensitivity analyses performed around parameters in these models reveal that estimates of concentrations depend for the most part on emission values as opposed to meteorological conditions (Glen et al., 1996; Kono and Ito, 1990). The relative importance of emission values increases when exposure is measured over longer temporal spans such as months or years. Consequently, long-term exposure trends are due primarily to emission values as opposed to meteorological conditions (Eerens and Sliggers, 1993; Fisher et al., 1992; Glen et al., 1996).

Knowledge gained from analysis of more complex models led to the development of the calculation of air pollution by road traffic (CAR) model (den Boeft et al., 1996). This model estimates exposure to a number of pollutants in motor vehicle emissions from a limited number of parameters. Exposure can be
estimated within 5 to 30 meters from an urban street, using vehicle emissions and street type as the main inputs. Streets in SETO fall into two of the five street classes used in this model. For these street classes, pollutant concentrations from vehicle emissions are estimated by the CAR model to decrease from curbside concentrations by 50% at 10m and 80% at 30m (Eerens and Sliggers, 1993). A validation study of the CAR model found modelled concentrations to be within 10% of measured concentrations (Eerens and Sliggers, 1993), and measurements in other areas support the results of the model (Nitta et al., 1993).

**Measuring Socio-Economic Status in Small Areas**

Measurement and control of the effect of socio-economic status (SES) in area based studies of health is essential to minimize potential bias from this source (Jolley et al., 1992). However, as there is no clear and universally accepted definition of SES (Carstairs and Morris, 1991), debate exists over what should be measured and controlled.

In ecological studies, SES is routinely measured for an area, but it can also be measured at the level of the individual or family. Multilevel or contextual analyses suggest that SES contributes to individual health differently at each level, but the contribution is strongly correlated between levels (DiPrete and Foristal, 1994). Empirical evidence suggests that area measures of SES can underestimate the effect of SES at the individual level (Greenwald et al., 1994; Hyndman et al., 1995). Despite differences between area and individual measures of SES, area measures have a number of benefits. These include the provision of a more stable estimate of SES, and the ability to construct population-based indices of SES from a number of individual variables (Krieger et al., 1997).

Some researchers have found that in relation to general health measures, a composite index of SES tends to rank areas more consistently and have greater
explanatory power than any single measure (Carstairs and Morris, 1989; Frohlich and Mustard, 1996). However, these indices have not been systematically evaluated or validated, and their utility in epidemiological research remains unclear (Krieger et al., 1997). On a practical note, the explanatory power of an index is influenced by the size of the area it represents (Krieger et al., 1997), and there is little evidence of how indices perform for small areas such as EAs. Other practical limitations such as the availability of data for small areas may limit the usefulness of an index for describing SES in small areas.

In some areas, census-based measures of neighborhood income have been associated with the use of health services (Cherkin et al., 1992), and similar associations have been seen with measures of poverty or low-income (Krieger, 1992). Census measures of education have also been found to be associated with community level hospital admissions (McMahon et al., 1993), and regional level health status (Frohlich and Mustard, 1996). In addition to income and education, census-derived occupational data has also been shown to be related to medical conditions (Ellen et al., 1995) and health related behaviors (Krieger, 1992).

Local expertise suggests that within SETO, the proportion of low-income individuals in an area (based upon the Statistics Canada low-income cutoff) is the most robust measure of SES at the neighborhood level for explaining health-related outcomes (Patychuk, 1998). In contrast, it appears that some area measures of education, such as the proportion with a high school diploma, can be strongly affected by age distributions within small areas. Other area measures of education, including the proportion with less than grade 9 and the proportion with a university degree appear to be more robust (Patychuk, 1998). Measures of employment class and status also tend to have variable predictive ability in SETO, perhaps because of their limitations in comparably capturing disparities in working and living conditions across divisions of ethnicity and gender (Krieger et al., 1997).
Analysis of Area Level Health Data

Several issues must be considered in the analysis of area-level health data. These issues are the choice of an appropriate statistical model, control of confounders and interactions in an ecologic design, and the influence of the spatial distribution of variables on the analysis.

Analyses of small area variations in admission rates have demonstrated that the behavior of individual-level admissions translates into the distribution of admissions at the area level. Furthermore, the most important aspect of the distribution at the individual level appears to be whether multiple admissions are possible (Cain and Diehr, 1992). If multiple admissions are possible, the variance of the count of admissions in a given area will probably be larger than it would be under a Poisson distribution. Consequently, one would expect a better fit with a Poisson regression analysis of first admissions as compared to all admissions. An alternative model that is often used in small area analyses is the logistic regression model, especially when multiple admissions are not possible. For events with a small probability of occurrence, both models should give the same parameter estimates and standard errors (Griffith and Amrhein, 1997).

The problem of confounder control in ecologic studies is mentioned earlier in this section in relation to the measurement of confounders. In addition to the problem of accurate measurement, control for confounding in the analysis of ecologic data presents a further challenge. Two methods are used to control for confounding in ecologic analyses. The first is to treat ecologic measures of confounders as covariates in the model, while the second is to standardize rates for the confounders (Morgenstern, 1995). Proper application of both of these methods requires data that is usually not available. If multiple confounders are present (e.g. age, gender and SES), the first method requires that measurements of the joint distribution of confounders within areas be included as covariates.
(Morgenstern, 1995). Unfortunately, data on such joint distributions is rarely available. Application of the second method requires all predictors in the model to be mutually standardized for the same confounders (Greenland and Robins, 1994). Again, the data necessary to perform such standardization is rarely available. On a final note, the use of product terms to model interaction effects in an ecologic study is not analogous to their use in individual-level studies. This approach should not be used unless one can be sure that the covariates of interest are uncorrelated at the individual level within areas (Greenland, 1992).

In individual-level and ecologic epidemiological studies, the spatial locations of observations are usually not measured or considered in the analysis. Ignoring the spatial dimension of data in this manner assumes that observations are distributed independently with respect to location. However, for many types of data used in epidemiological studies this assumption is not warranted. For example, it is reasonable to expect that disease rates, environmental exposures and levels of socioeconomic status (SES) might be similar in contiguous areas. This type of relationship between observations is called spatial autocorrelation, and if present, may result in biased estimates of regression coefficients and invalid results in tests of significance (Walter, 1992a). Statistical approaches exist to detect (Walter, 1992a) and control for (Richardson, 1992) this source of bias, and these approaches are discussed in the method section. In addition to the problem of spatial autocorrelation, ecologic studies are also faced with what is known as the modifiable areal unit problem (MAUP). This problem is described as the 'sensitivity of analytical results to the definition of units for which the data are collected' (Fotheringham and Wong, 1991). In other words, ecologic studies usually rely upon areas with boundaries created for other reasons, and there is evidence to suggest that regression coefficients and correlation statistics can be affected by the way that areas are defined (Amrhein, 1995). This source of bias is not addressed in the thesis, but its possible existence noteworthy.
METHOD

This section describes the steps taken to obtain the necessary data and test the hypothesis. As this thesis relies upon secondary data, the first portion of the section identifies data sources, and describes their strengths and weaknesses. The steps taken to arrive at measures of respiratory health, exposure and SES are then described. Finally, the methods of the univariate and multivariate analyses are presented.

Data Sources

The description of data sets and sources begins with the data used to calculate hospitalization rates. This is followed by a description of traffic count and vehicle type data, then an overview of data used to measure SES. Finally, the spatial data sets that give the physical location of boundaries and streets are presented.

Admission Data

Attribute data sources used to calculate standardized admission rates by EA include data from the Hospital Medical Records Institute† (HMRI) and the Postal Code Conversion File (PCCF) from the 1991 census.

A subset of the full HMRI database for the years 1990, 1991 and 1992 containing records with a patient postal code in the City of Toronto provides a starting point for this study. Hospitals in Ontario abstract data from the patient record for each hospital admission at the time of separation (i.e. discharge, transfer or death). These data are abstracted by trained individuals and diagnostic data are coded according to the International Classification of Diseases, Ninth Revision (ICD-9).

† The HMRI is now known as the Canadian Institute for Health Information (CIHI).
Once abstracted, copies of these data are sent to HMRI by each hospital. For convenience, these data are referred to as admission data.

HMRI admission data can be examined in relation to the composition of the population and the definition of the diagnoses (Lilienfeld and Stolley, 1994). As a result of universal hospital insurance in Canada, and relatively complete participation of hospitals in the HMRI database, this source most likely provides an accurate reflection of hospital admissions in the SETO population. The validity of the diagnoses in relation to other health data (e.g. ambulatory data) is expected to be high due to a physician conferring the initial diagnosis, and a trained individual abstracting the diagnosis from the chart. Nevertheless, HMRI data is not completely valid, with overall agreement between HMRI data and re-abstracted records expected to be between 80% and 95% based upon studies in the USA (Fisher et al., 1992) and Canada (Roos et al., 1993). Other problems with admission data were reviewed in the previous section.

The PCCF is a digital file that provides correspondence between the six-character postal code and Statistics Canada's standard geographical areas. The basic link is between the postal code and one or more of the 267 103 Ontario EAs in the 1991 census. The positional accuracy, logical consistency, completeness, attribute accuracy, and limitations of the PCCF are described in detail elsewhere (Statistics Canada, 1994). In brief, the PCCF is complete in that all 1991 postal codes are found in the PCCF, and all geographic codes from the 1991 census are linked to a postal code. An error analysis conducted by Statistics Canada found that matching to urban EAs resulted in a 6.7% error rate (i.e. a postal code was assigned to an incorrect EA) (Statistics Canada, 1994). This error rate includes postal codes that match to multiple EA. Such matches occur when a postal code straddles several adjacent geostatistical areas. Multiple matches occur for 3.3% of the urban postal codes in the PCCF (Wilkins, 1993).


**Traffic Count Data**

Exposure to vehicle emissions is modelled using motor vehicle counts obtained from Metro Toronto and the City of Toronto. Traffic count data are collected by both levels of municipal government using similar methods. The methods can be examined in terms of the temporal frame, measuring instruments, and location of count. Most counts are performed over 24-hours, but some counts are also performed over 8-hour periods. Automatic instruments are usually employed to perform 24-hour counts. These instruments include pneumatic tubes, inductive loops, and microwave counters. The accuracy of the counts obtained with these instruments is affected by factors such as weather and traffic speed, but is estimated to be in the area of 95% (Lasagna, 1997). Given the nature of these instruments, 24-hour counts are usually obtained at a point along a street segment for traffic in one, or both directions. In contrast, 8-hour counts are usually obtained by manual methods. This involves a team of trained observers recording the number of vehicles passing a point over an 8-hour period. This type of count is usually obtained at an intersection, with approaching traffic from all directions being counted.

**Vehicle Type Data**

Two sources are used to estimate the distribution of vehicle types in the study area. The first source of data is a 'cordon' count performed by the Metro Toronto Planning Department. At two-year intervals, the planning department performs a vehicle count at a number of points along a cordon surrounding the City of Toronto. Data for the years 1989, 1991 and 1993 are averaged for this study. The survey is taken between 0700h and 0900h on one day between May and June. Analysts record the type and direction of each vehicle that passes the cordon. Sixteen cordon count points fall within the study area. These points are shown on Map 2. The second source of data is the average distribution of vehicle types in
the province of Ontario for the year 1991. This data is obtained from the Ontario Ministry of Energy and the Environment.

*Census Data*

Parts A and B of the 1991 census are used to measure socio-economic status. The characteristics and quality of these data sources are discussed in detail elsewhere (Statistics Canada, 1993), and briefly reviewed below. Part A of the census provides basic demographic, dwelling and household data on a 100% basis. Part B provides data collected from a 20% sample of households on characteristics such as education and income.

The enumeration area (EA), the basic geographical unit of census data collection, is the area enumerated by one individual. Census profiles for an EA are not released, or are ‘suppressed’, by Statistics Canada if the global non-response rate is ≥ 40% or the population is less than 40. Another confidentiality procedure inherent in this data source is ‘random rounding’. Under this method, all figures including totals are randomly rounded either up or down to a multiple of 5 and in some cases 10. Statistics Canada asserts that this ‘does not add significant error to the census data’ (Statistics Canada, 1993), but it almost certainly introduces random error, especially when working with small areas such as EAs.

*Spatial Data*

The Street Network File (SNF) and the EA cartographic boundary file from the 1991 census are two sources of spatial data used in this study. Both sources are obtained from Statistics Canada through the Data Library at the University of Toronto. Digital layers are converted to the Universal Transverse Mercator (UTM) projection, North American Datum 1927 (NAD27) which is the provincial standard.
The SNF is used to map and model vehicle counts and emissions. This data source is discussed in detail elsewhere (Statistics Canada, 1992), and reviewed briefly here. The SNF incorporates a detailed level of geocartographic information for all major urban centers. Statistics Canada created the SNF in the early 1970s for a number of reasons including assisting in the definition of EAs. Most of the source maps for the SNF are of a scale between 1:1000 and 1:30000. Positional accuracy relative to the EA cartographic boundary file is quite good given that the EA cartographic boundary file was originally created from the SNF. The relative positional accuracy is of importance for the thesis given that these two layers are overlaid. Limited studies of absolute positional accuracy have indicated that spatial errors of up to 100 meters do occur in small portions of the Canadian SNF. These errors are related to the original source and digitization, and are not relevant to the thesis, as exact location is not referred to in any manner.

The EA cartographic boundary file is used in conjunction with the SNF to model and map exposure to vehicle counts and emissions for each EA. Other uses of the boundary file are to map admission rates and socio-economic status, and generate a binary connectivity matrix for EAs in the study area.

**Measurement of Respiratory Health**

This component of the methodology is divided into a number of sections. In the first section, diagnostic sets for respiratory and comparison conditions are defined. Admission records are then abstracted and marched to EAs by Postal Code. Following a number of quality assurance steps, age and sex standardized admission rates are calculated by EA. Figure 2 provides a graphical summary of the process to support the detailed descriptions that follow.
**Defining Diagnostic Sets**

As discussed in the second chapter of this thesis, a number of published studies have found a relationship between exposure to PM and hospital admission for respiratory diagnoses. The evidence is strongest for an association between exposure to PM and diagnoses of asthma, pneumonia and COPD. However, there is also weaker evidence that bronchitis and upper respiratory tract infections (URI) are associated with exposure to PM.

ICD-9 codes that define a subset of respiratory diagnoses associated with exposure to PM are identified from a review of studies that use diagnostic coding (Burnett et al., 1997; Burnett et al., 1995; Gordian et al., 1995; Stieb et al., 1995). Examination of the ICD-9 codes used in these studies suggests the selection of codes shown in Table 4 for asthma, bronchitis, COPD, pneumonia, and URI.

Diagnoses not selected from the respiratory chapter include disorders of the nose (470-471), allergic rhinitis (477), bronchiectasis (494), occupational lung disorders (500-505), empyema (510), pleurisy (511) and pneumothorax (512). These diagnoses should have no relationship to exposure to PM, and their inclusion in the selected subset would therefore only weaken any association with PM.

Diagnoses selected for study can be discussed in terms of which diagnoses are selected and how selected diagnoses are defined. Diagnoses listed in Table 4 are defined in accordance with previous studies. The definitions of pneumonia and URI warrant further discussion. Pneumonia is defined to not include influenza (487). This definition is used in time-series studies, possibly to avoid bias from the marked seasonal pattern of influenza. In order to be consistent, influenza is not included in the definition of pneumonia in this study. URI is defined to not include acute nasopharyngitis (460) and acute tonsillitis (463). The common cold is excluded due to the lack of specificity and sensitivity of this diagnosis, while tonsillitis is excluded to be consistent with previous studies.
Also shown in Table 4 are the ICD-9 codes for the respiratory chapter, gastrointestinal (GI) chapter and genitourinary (GU) chapter. The entire respiratory chapter is examined in addition to the subset of respiratory diagnoses for two reasons. The first is that some studies have found an association between exposure to PM and all respiratory admissions. This association appears to be weaker than that for specific diagnoses, but is worthy of examination in its own right. The second reason is to determine if the subset of diagnoses has a stronger association with PM exposure than the broader set of diagnoses encompassed by the entire respiratory chapter. The lack of a stronger association for the subset has a number of theoretical and practical implications. Theoretically, a stronger association with the subset supports the causal model postulated in the second chapter of the thesis. A stronger relationship with the broader set suggests that either one or both of the causal and exposure models are in need of revision. From a practical perspective, the lack of a stronger association with the subset may also suggest that the resources used to identify a subset are not needed in future studies of this sort.

Admissions rates for GI and GU diagnoses are examined to determine if admissions for non-respiratory conditions are associated with exposure. The selection of these ICD-9 chapters for this purpose is made after consideration of the potential for a direct or indirect relationship between the diagnoses in the chapters and exposure to PM from motor vehicle emissions. For example, cardiac admissions have been directly associated with PM exposure in a number of studies (Burnett et al., 1995; Schwartz, 1994a), and injuries could be indirectly associated through residential proximity to streets. The GI and GU chapters are two sets of diagnoses that should have little association to PM exposure or residential proximity to streets. Initially, only the GI chapter is abstracted and examined for comparison. Subsequent abstraction and comparison with the GU chapter is performed to confirm associations seen with the GI chapter.
Records with a primary diagnostic code matching those shown in Table 4 are abstracted from HMRI data for the former City of Toronto for the years 1990, 1991, and 1992. Using the SAS software program (SAS, 1996), these records are subsequently combined into one file covering the range 1990 to 1992, and unneeded fields in the database are dropped.

**Matching Hospital Records to Enumeration Areas**

As the unit of analysis for this ecological study is the EA, it is necessary to match each admission record to an EA. The PCCF is used to assign an EA to each admission record based upon the six-character postal code. The problem of multiple EA matches for a single postal code is addressed by using the ‘Single Postal Code Indicator’ in the PCCF. This indicator ensures that each record matches to only one EA by selecting the most representative EA based upon address range. Manual validation to determine the proportion of matches that are in error is not performed.

**Quality Assurance of Admission Records**

After matching each record to an EA, the admission records are limited from the City of Toronto to the SETO area. This is performed by merging the admission records with a list of EAs in the study area. The counts for some EAs are not released, or ‘suppressed’, by Statistics Canada for reasons of confidentiality. Without this count data, rates cannot be calculated for these EAs, and consequently they are not comparable to other EAs in the study area. Records from EAs with suppressed population counts are therefore deleted.

Remaining records are then checked for inconsistencies in birth date and admission date. Records with an admission date before their birthdate, or an admission date before January 1, 1990 are deleted.
The final quality assurance step involves the deletion of records without health numbers. Valid health numbers in the HMRI database are encrypted by the Ministry of Health before the records are released. The encryption is performed in a manner that ensures each encrypted health number is unique, and that each valid health number consistently encrypts to the same number. The encrypted health number can therefore be used to identify initial and repeat admissions for individuals. However, some records do not have a valid health number, and consequently the records cannot be followed in this manner. To minimize the error in the calculation of initial and repeat admission rates, records without valid health numbers are deleted.

**Calculation of Standardized Rates**

Age and sex standardized admission rates are calculated for each diagnostic set for first admissions, and all admissions. This initially involves identifying the first admission for each individual based upon the first occurrence of a health number within a diagnostic set. The number of first and all admissions is then calculated for each EA by diagnostic set with admission counts summed over three years (1990 to 1992) to increase stability of the estimates. The counts by EA are transferred to the software package Excel for the calculation of standardized rates. The method of indirect standardization is used to avoid reliance upon small, unstable admission counts in the age-sex strata of each EA. The indirect method of standardization also has the advantage of producing a ratio measure that can be used in Poisson regression analysis to capture information from EAs with no admissions (McMahon et al., 1993). In the calculation of indirectly standardized rates, observed values for each EA are taken directly from the counts of admissions. Expected values are calculated from the age-sex specific population counts for each EA from the 1991 Census and age-sex specific
admission rates for the standard population. The standard population is taken as all of SETO.

The observed number of admissions for a diagnostic set within an EA are calculated using the formula

$$Observed_i = Admissions_{1990} + Admissions_{1991} + Admissions_{1992} \quad (1)$$

Where $Admissions_{199X}$ is the number of first or all admissions in EA, in the diagnostic set of interest for the year 199X, and $Observed_i$ is the total number of first or all admissions in the diagnostic set of interest in EA, over the time 1990 to 1992.

The expected number of admissions for a diagnostic set within an EA are calculated using the formula

$$Expected_i = \sum_j p_{ij} \cdot r_j \quad (2)$$

Where $p_{ij}$ is the person-years within age-sex strata $j$ of EA, $r_j$ is the admission rate for the diagnostic set of interest in the in age-sex strata $j$ of the standard population, and $Expected_i$ is the total number of first or all expected admissions in the diagnostic set of interest in EA, over the time 1990 to 1992.

The individuals who contribute time to the denominator of an incidence rate are referred to as the 'population at risk'. When events tallied in the numerator of an incidence rate are first occurrences of disease, then the time contributed to the denominator by each individual in whom disease develops should terminate with the onset of disease (Rothman and Greenland, 1998c). Time is no longer contributed as the individual is no longer 'at risk. However, if first and subsequent occurrences of disease are included in the numerator, then time
accumulated in the denominator does not cease with the occurrence of the first event, as an additional event might occur in the same individual (Rothman and Greenland, 1998c).

In this study, the person years are calculated for rates of first admission and all admissions by multiplying the midpoint (i.e. 1991) population by 3 years. With this approach, an individual continues to contribute time to the denominator after their first admission. This results in an accurate denominator for rates of all admissions, and a slightly enlarged denominator for rates of first admission. Given the small proportion of individuals with admissions (1.5%), this should lead to a minimal error.

The indirectly standardized rate (ISR) is then calculated for a diagnostic set within an EA using the formula

$$\text{ISR}_i = \frac{\text{Observed}_i}{\text{Expected}_i} \cdot \text{CSR} = \text{SHR}_i \cdot \text{CSR}$$

Where $\text{ISR}_i$ is the indirectly standardized rate for the diagnostic set of interest in $EA_i$ and $\text{CSR}$ is the crude admission rate for the diagnostic set of interest in the standard population. As $\text{CSR}$ is a constant within a diagnostic set, the ratio of observed admissions divided by expected admissions can be compared between enumeration areas. When dealing with the outcome of mortality, this ratio is known as the standardized mortality ratio and in this study is defined as the standardized hospitalization ratio ($\text{SHR}_i$).

**Assessment of Exposure**

At the beginning of this chapter, a need is identified for a model that considers emission and dispersion in an integrated manner to arrive at an assessment of
exposure to motor vehicle emissions. This part of the thesis builds upon methods used in two previous studies to develop such a model (Buckeridge et al., 1998; English, 1997). As well as integrating aspects of exposure assessment, the model provides a simple and hopefully credible approach to exposure assessment in comparison to the complex models discussed earlier.

The assessment of EA exposure to motor vehicle emissions involves a number of steps. Conceptually, the process can be divided into two parts: modelling emissions from motor vehicles, and modelling the dispersion of the emissions to EAs. The first part of this process begins with acquiring traffic counts for SETO and converting these point counts into street segment traffic volumes. A determination is then made about which streets to model, and the traffic volumes are georeferenced to the network of modelled streets. In the final step of this part, an emission model is used to estimate the PM 2.5 emissions from the modelled streets. An overview of this part of the exposure assessment is provided in Figure 3. The second part of the process begins by modifying the street network. This involves correcting the correspondence between the street network and the EA boundary layer and constructing buffers around each street. In the final step, the modified street network is overlaid on the EA boundary layer, and exposure is assessed from the resulting layer. Figure 4 provides an overview of the second part of the exposure assessment process, and the following paragraphs describe the entire process in detail.

**Acquisition and Aggregation of Traffic Count Data**

Streets in the study area fell under the jurisdiction of two authorities. The Traffic Branch of Metropolitan (Metro) Transportation was responsible for most major roads in the study area, while Transportation Operations of the former City of Toronto was responsible for all other roads in the area. Data on traffic counts are requested from these bodies for all streets in the study area. The two bodies had
different mandates and budgets, and these factors influenced the quality of traffic count data supplied in response to a request. Neither body conducted regular, systematic traffic counts over all streets under its jurisdiction. However, Metro Transportation conducted a considerable amount of traffic analysis, and consequently gathered data in a more consistent fashion. Metro Transportation was therefore able to supply 24-hour average count data for all streets in the study area under their jurisdiction, over a relatively short time span (1990 to 1992). Of the total 73.85 km of modelled streets, 37.50 km (50.8%) is modelled from Metro data. In comparison, the City of Toronto conducted only a limited amount of traffic analysis and tends to gather data to support specific actions, or in response to complaints. The City of Toronto could therefore only supply 24-hour and 8-hour count data for some of the streets under their jurisdiction, over a relatively longer time span (1987 to 1994). Of the total 73.85 km of modelled streets, 36.35 km (49.2%) is modelled from City data. The entire street network for SETO by data availability is shown in Map 2. From this Map, it is apparent that vehicle count data are available for 104.07 km of the entire 219.03 km network (47.8%). For reasons discussed below, emissions are modelled from 73.85 km of the 219.03 km network (33.7%).

Conversion of Point Traffic Counts to Street Segment Traffic Volumes

The method of counting traffic described in the data source section converts the linear process of motor vehicle traffic to a static point count. In order to model emissions, it is necessary to convert the point counts to street traffic volumes. The approach employed for this conversion is influenced by the type of traffic count data available. The approaches employed are displayed graphically in Figure 5 and described in detail below.

The most straightforward approach is to equate the traffic volume for a street segment (i.e. a portion of a street between two intersections) to the traffic
counted at a point along that segment. This approach is used for 24-hour counts, as they are representative of the traffic volume between intersections on either side of the street segment. If multiple counts are available at one point, they are averaged to arrive at the segment value. In total, 59.52 km of the total 73.85 km modelled (80.6%) is calculated from 24-hour counts. When 24-hour count data are not available, a different approach is used for 8-hour counts.

When only 8-hour count data are available for a street, traffic volume is determined from counts at intersections along the street. This involves identifying segments of the street that have 8-hour counts available for intersections at both ends of the segment. Incoming traffic to each intersection from the street segment is summed to arrive at the total 8-hour volume for the segment. A conversion factor of 2.05 is then used to convert the 8-hour volume to a 24-hour volume (Fiodorowicz, 1997). In total, 14.33 km of the total 73.85 km modelled (19.4%) is calculated from 8-hour counts.

For most streets, data does not exist to calculate traffic volume for all segments of the street. Calculated segment volumes must therefore be applied to the remaining segments that data does not exist for. This is accomplished by assuming that segment volumes are constant between major intersections, and applying each calculated segment volume to all other segments that fall between the same major intersections.

In order to limit the computational complexity of emission and exposure modelling, segments with similar volumes are combined. Adjacent segment volumes of the same street are combined by weighted average (weighted by number of count points contributing to each segment volume) if the difference between the volumes is less than 10% of the smallest segment volume.
**Determination of Streets to Model**

The decision to model emissions for a subset of streets in the study area is made for several reasons. A practical reason is that vehicle counts are not available for every street in the area. Another reason is that a sample of streets can provide an unbiased estimate of emissions (Hewison et al., 1996), and should be sufficient for testing the hypothesis as long as a large enough range of emissions is modelled. Finally, all other published studies in this area have only considered the effect of streets with traffic counts above a given level. Given this, one would expect comparability to be greater if the same methodology is employed in this study.

The selection of streets to model for this study is based upon previous work in this area and available traffic count data. A previous study in this area (Buckridge et al., 1998) identified a network of major streets by examining aerial photos of the study area (1:5000 scale). Streets were modelled in that study if they had four or more lanes, or acted as major thoroughfares. For the current study, a more data driven approach is taken to selection. This approach involves the consideration of all available traffic count data for streets in the study area to arrive at a decision. The two approaches give approximately the same result.

All street segments with traffic volume data are examined to determine the distribution of traffic volumes by segment, and a traffic volume cut-off of 5000 vehicles per 24 hours. This cut-off is chosen for two reasons. The first reason is consistency, in that it represents the lower bound of traffic volumes modelled by other studies of this type. The second reason is to avoid selection bias, as traffic counts are only available for a small proportion of streets with lower volumes. As there is no way of discerning if streets with available volumes are different that those without, modelling lower volume streets with available data may bias the results. Emissions from all street segments with an average 24-hour volume of ≥
5000 are therefore modelled in this study. Map 2 identifies the segments from the street network that are modelled. It is apparent from the Map that traffic counts are not available for a number of streets in the study area. As count data are not available for these streets one cannot exclude the possibility that some of these streets may have traffic volumes of \( \geq 5000 \) per 24 hours. However, this is unlikely as few of these streets are thoroughfares, and most are in residential areas.

**Georeferencing Traffic Volume Data to a Spatial Street Network**

In the first step of the georeferencing process, the SNF for Metro Toronto is limited to streets in the study area by a structured query language (SQL) spatial query in MapInfo. The query identifies all streets in SETO by overlaying the boundary of SETO on the street layer, and ‘clipping’ out the streets inside the boundary polygon. Streets within the study area are then further limited to the modelled streets by performing an SQL attribute query in MapInfo based upon the names of streets to be modelled. The resulting layer of streets is then mapped and visually inspected for concordance with the streets identified by the traffic volume data. At this point extraneous streets are deleted, and missing sections of streets are added. Using MapInfo, the network is then partitioned into segments that correspond with the segments in the modelled emissions data set. This is performed by using the ‘to street’ and ‘from street’ information that accompanies both the segment data set and the street network. As each segment is partitioned on the street network, it is assigned the same unique identifier as the corresponding record in the traffic volume data set. Attribute values for street segments are then spatially related to their geographical representation (i.e. georeferenced) by way of the unique segment identifier.
Modelling of Particulate Matter Emission

Modelling of particulate matter emissions requires counts of vehicles by type, emission factors for each vehicle type, and a model to combine these two types of data. The steps taken in the modelling process are described below.

Data sets used to determine traffic volume or streets in the study area provide only aggregate data for all types of vehicles. However, as emission characteristics vary by vehicle type, the distribution of vehicle types for each segment must be determined for emission modelling. Vehicle type distribution is estimated throughout the study area from cordon count data and the average provincial distribution. Adjacent cordon count points are used to assign vehicle type distribution to 47.90 km of the 73.85 km (64.9%) of modelled street. The other 25.95 km (35.1%) of street are assigned the vehicle type distribution of the provincial average. Vehicle types from these sources are matched to the vehicle types used in the motor vehicle particulate modelling software (PART5) from the US Environmental Protection Agency (Office of Mobile Sources, 1995). Particulate matter (PM) emission factors are calculated for each standard vehicle type using the PART5 emission model. The PART5 model is a Fortran program developed by the United States Environmental Protection Agency (EPA) for modelling PM emission from motor vehicle traffic. The model is run for the year 1991, at low altitude, with a city traffic cycle and average speed of 31.5 km per hour. The calculated emission factors for particle size cut-off of 10μm (PM10) and 2.5μm (PM2.5) are shown in Table 5.

Vehicle type distributions, vehicle type emission factors and traffic volumes are used to calculate the mass of PM 10 and PM 2.5 emitted on each street segment. The equation used to calculate these values is shown below (Rayfield et al., 1995).
\[ T_j = \sum_{i=1}^{n} V_{ij} \cdot EF_i \]  

(4)

Where \( T_j \) is the total daily emission on street segment \( j \) in grams per day, \( V_i \) is the traffic volume for vehicle type \( i \) on street segment \( j \) in vehicles per kilometer per day, and \( EF_i \) is the emission factor for vehicle type \( i \) in grams per kilometer per vehicle.

Calculated values for PM 2.5 are shown on Map 3. PM 2.5 is a subset of PM 10 (accounting for nearly 90% of PM 10), and consequently the distribution of PM 10 is essentially the same as PM 2.5. For this reason, only PM 2.5 values are reported for the remainder of the thesis. Results for PM 10 are available from the author.

**Modification of Street Network**

The street network is transferred into Arc/Info format from MapInfo format for the remaining operations. When a GIS operation is referred to in the following paragraphs, the Arc/Info command used for the operation is listed in parenthesis (COMMAND). The process described in this section is summarized in Figure 4.

Initially, the spatial characteristics of the modified SNF are checked for correspondence with the EA boundaries. This is accomplished by overlaying the street network layer on the EA boundary layer. As many of the EA boundaries are coincident with streets, the overlay should result in perfect concordance between these two layers in many locations. This is generally not the case. Spatial error of this nature is inherent in the two layers given their sources and spatial accuracy as discussed above. Nevertheless, the exposure modelling algorithm is influenced by street frontage, so it is important that spatial error between these two layers be minimized to an allowable distance. The street layer is therefore
adjusted to correspond to the EA boundary layer where the EA boundaries are co-terminal with a street. This is accomplished by increasing the number of nodes in the arcs of both layers (DENSIFY), then manually ‘snapping’ the arcs of the street layer to the arcs of the EA boundary layer (SNAP). This adjustment is performed for arcs in the two layers that are within 10m of each other.

The street network is then converted to a series of polygons by creating a 10m buffer polygon around each street segment. These buffers accomplish two tasks in the exposure assessment process. The first task is to facilitate the transfer of traffic volume and emission values from the street network layer to the EA boundary layer during the overlay process that follows. A number of options were initially explored within Arc/Info in an attempt to transfer values directly from a linear street network to EAs located on both sides of the street, but none produced acceptable results. Buffers are therefore used to convert the linear street network to a polygon layer to ensure that traffic volumes and emissions are transferred to EAs on both sides of streets on the basis of street frontage. The size of the buffer is not of central importance for this step, and any buffer greater than the error between the street network layer and the EA boundary layer is sufficient. The size of the buffer is however important for the second task that it fulfills - to provide a basis for estimating EA exposure. As will be discussed below, two weights are applied to the total traffic volume and emissions transferred to each EA in an attempt to account for the size and shape of each EA. One of the weights applied to the total traffic volume and emissions is the proportion of each EA that falls within 10m of a street. This weighting approach has been applied in two previous studies (Buckeridge et al., 1998; English, 1997), and a width of 10m is selected based upon models of dispersion. As discussed earlier in this chapter, dispersion models suggest that the concentration of PM 2.5 emitted from motor vehicles decreases to half its maximal value within 10m of the street centerline (Eerens and Sliggers, 1993).
As the buffer polygons overlap where streets are closer than twice the buffer width, it is first necessary to separate the street network into a number of separate layers. The network is initially divided into two layers, EW streets and NS streets. The EW layer is in turn divided into four layers (with 15, 14, 2 and 1 segments) and the NS layer into two layers (with 18 segments in each). This results in six layers, each with a number of street segments at least 20m apart. A 10m buffer polygon is created around each segment using a resident function in Arc/Info (BUFFER). The buffer polygons are overlaid on the street segments and the segment numbers are manually transferred to the corresponding polygons (UPDATE). Segment traffic volume and emission data (in vehicles or grams per segment per 24 hours) are then joined to the buffer polygons by relating on the segment number (JOINITEM). The traffic volumes and emissions are converted to area measures for each buffer polygon by dividing through by the buffer area in square meters. In each polygon buffer layer, the traffic volume and emission fields are given unique names so that they can be identified when the layers are combined.

Buffer polygon layers are combined two at a time using an overlay function (UNION). This function maintains all original polygons from both layers, and creates a new polygon when polygons from the two input layers overlap. The new polygons are linked to attribute values for both of the overlapping polygons from the original layers. After each overlay, sliver polygons less than 10m² in area are dissolved into the adjacent polygon with the longest shared border (ELIMINATE). This overlay process is repeated five times to arrive at a single, combined buffer polygon layer. The combined buffer polygon layer contains 326 polygons, each with attribute data for all segment buffer polygons that fall within that polygon.

The total value of traffic volumes and emissions is then calculated for each polygon in this layer by summing the corresponding attribute data from all
segment buffer polygons within the polygon, and multiplying by the area of the polygon. The final layer consists of 326 buffered segments of street with associated values for traffic volume (in vehicles per square-meter per 24 hours) and emission of PM 2.5 (in grams per square-meter per 24 hours).

**Overlay of Street Network on Enumeration Areas**

The final step in this process involves an overlay of the buffered street network on the EA boundaries to determine the EA exposure by street frontage. This overlay is performed so that the resulting layer contains all features from the EA boundaries, and all features from the buffered street network that overlay the EA boundaries (IDENTITY). Using this method, the effects of the street network on areas outside of the study area are ignored.

This overlay process results in an EA exposure layer that contains 1403 polygons, each labeled by the EA it falls within. Of these 1403 polygons, 965 are also labeled by the buffered street polygon they fall within. Labeling each polygon in this manner allows data linked to the label to be accessed from a relational table. Exposure values are calculated from these data in three ways. The first approach calculates the total vehicle exposure count (in vehicles) and emission exposure values (in grams) for each EA. These values are calculated for each EA using formula 5.

\[
EA_i = \sum_{m}^{n} \text{Value}(B_m) \cdot \frac{\text{Area}(B_m \text{ in } EA_i)}{\text{Area}(B_m)}
\]  

(5)

Where \(B_m\) is the \(m^{th}\) of \(n\) buffer polygons that fall within \(EA_i\), \(\text{Value}(B_m)\) is the total count of vehicles or mass of emissions (in grams) in \(B_m\), \(\text{Area}(B_m)\) is the total area of \(B_m\) (in square meters), and \(\text{Area}(B_m \text{ in } EA_i)\) is the area of \(B_m\) that falls in \(EA_i\).
This approach directly transfers the vehicle counts and PM emissions from the street network to the surrounding EAs based on road frontage. This direct transfer of values provides an opportunity to validate the method up to this point. If the method is valid then the total vehicle count and total PM emission of the street network should be similar to the total vehicle count exposure and total emission exposure of the EA layer. The total vehicle count for the street network is 1,979,172, while the total vehicle count exposure for the EA layer is 1,848,416 (93.39%). Similarly, the total emission of PM 2.5 from the street network is 549,166.7 grams, while the total PM 2.5 exposure for the EA layer is 518,937.1 grams (94.50%). These values agree quite well, with slightly lower values for the EA layer attributable to the expected loss of counts and emissions around the outer edge of the study area.

The approach applied above does not take into account the size or shape of an EA, both of which could influence rates of motor vehicle emission related health effects within an EA. Two different weights have been applied in previous studies to address this problem (Buckeridge et al., 1998; English, 1997). The weights adjust the total EA exposure for the proportion of EA area near a street, or the total EA area. Models (Eerens and Sliggers, 1993) and measurements (Nitta et al., 1993) of PM 2.5 dispersion support the use the weight that adjusts total exposure for the proportion of EA area near street. If one assumes that population is evenly distributed within an EA, then this weight should correct for the proportion of individuals within an EA exposed to elevated concentrations of PM 2.5 from motor vehicle emissions. In short, it assumes that the emissions disperse evenly over the area near a street, but do not add to the background level in the rest of the EA. The weight that adjusts total EA exposure for the entire area of the EA is not readily supported by dispersion theory. Again, if the population is assumed to be evenly distributed throughout the EA, then this weight is equivalent to assuming that the emissions disperse evenly across the
entke EA. This weight is included for comparison as it has been applied in a previous study (English, 1997).

The approach that weights the total exposure value by the proportion of the EA area that falls within 10m of a street is shown in equation 6. As discussed earlier, the width of 10m is used as the concentration of emitted PM 2.5 is expected to fall to half its maximal value at that distance from the street centerline (Eerens and Sliggers, 1993).

\[
EA_i = \sum_{m}^{n} \frac{\text{Value}(B_m) \cdot \frac{\text{Area}(B_m \text{ in } EA_i)}{\text{Area}(B_m)} \cdot \frac{\text{Area}(EA_i \text{ in } B_m)}{\text{Area}(EA_i)}}
\]

Where \( \text{Area}(EA_i) \) is the total area of \( EA_i \), and \( \text{Area}(EA_i \text{ in } B_m) \) is the area of \( EA_i \) in \( B_m \).

The approach that weights the total exposure value by the area of the EA is shown in equation 7.

\[
EA_i = \sum_{m}^{n} \frac{\text{Value}(B_m) \cdot \frac{\text{Area}(B_m \text{ in } EA_i)}{\text{Area}(B_m)} \cdot \frac{1}{\text{Area}(EA_i)}}
\]

Equations 5 and 6 result in an exposure value that is expressed in vehicles or grams per 24 hours. Equation 7 gives an exposure value in vehicles or grams per square meter per 24 hours.

To summarize, the exposure assessment process produces three traffic volume exposure values and three PM 2.5 exposure values for each EA. The three values for each type of exposure represent: the total amount of traffic or PM 2.5 transferred to the EA from the street network; the total value weighted by the area of the EA within 10m of a street; and, the total value weighted by the area of
the EA. Modelling and measurement studies suggest that the most valid exposure value is the one weighted by the proportion of the EA within 10m of a street.

Measurement of SES

Data describing the socioeconomic status (SES) of EAs are obtained from the 1991 census. Following a review of the literature presented earlier in this chapter and consideration of previous work performed in SETO (Patychuk, 1998), individual SES variables thought to be associated with health status are selected for assessment. Most of the variables are taken from Part 2B of the census and consequently are based upon a 20% sample. For each variable, EAs with zero values are deleted, and the remaining EAs are entered into a bivariate rank correlation with the first admission respiratory subset standardized hospitalization ratio (SHR).

In an attempt to produce a measurement of SES with greater explanatory power than a single variable, an index is constructed using a methodology similar to one previously employed for Canadian Census data (Frohlich and Mustard, 1996). Census variables that demonstrate a significant bivariate rank correlation with the first admission respiratory subset SHR ($\alpha = 0.05$) are initially considered for inclusion in the index. Variables are excluded from consideration if more than 50% of EAs have a zero value for that variable. This step is taken to avoid the deletion of the majority of EAs in subsequent analyses. Each of the variables is then individually normalized to have a mean of zero and a standard deviation of one. The normalized census variables are entered into a multiple Poisson regression with the first admission respiratory subset SHR. Variables with non-significant regression parameters ($\alpha = 0.05$) are removed, and the regression is repeated. This process is repeated until all remaining variables have significant regression parameters. An index value is then calculated for each EA by summing
the values of the remaining standardized variables weighted by the regression coefficients.

**Analytic Method**

The analytic method employs univariate tools and multivariate tools. Tools within each of these categories can be further classified as aspatial or spatial. The aspatial tools are for the most part well defined in the epidemiology literature, and only brief descriptions of these tools are provided here. In contrast, spatial tools are not well defined in the epidemiology literature, and consequently are described in detail. The importance of considering the spatial nature of data is discussed earlier in this chapter, but the topic of spatial data analysis (SDA) in general has not been discussed. An understanding of this topic helps to place the spatial tools in perspective, and an overview of SDA is therefore provided before launching into a detailed description of the analytic method.

**Spatial Data Analysis**

Spatial data analysis is one of the main disciplines in the broader area of spatial analysis. SDA has been defined as the analysis of spatially located data where explicit consideration is given to the possible importance of the spatial arrangement in the analysis or in the interpretation of results (Bailey and Gatrell, 1995).

The behavior of spatial phenomenon can be thought of as resulting from first order and second order effects. First order effects relate to the variation of the mean value of the process in space - a global or large scale trend. Second order effects result from the spatial correlation structure, or the tendency for deviations in values of the process from its mean to follow those in neighboring sites - a local or small scale effect (Bailey and Gatrell, 1995). An example of a first order
effect is the tendency for population density to increase as one moves closer to the central business district of a city. The tendency for infectious disease cases to occur in clusters throughout a city is an example of a second order effect. In practice, a spatial phenomenon usually exhibits a mixture of first and second-order effects, but this construct is useful for modelling spatial phenomena.

If a first order effect is present then the mean value of the process is determined by the position of the observation in space. Models of first order effects are therefore relatively straightforward and involve relating values of a process against the distance of observations from a point (e.g. the city center), or the absolute location of observations (e.g. latitude and longitude). In contrast, second order effects are more difficult to model. The complexity associated with modelling second order effects is managed by making two assumptions. The first assumption is that the spatial process is stationary. Informally, this means that the statistical properties of the process are independent of absolute location, and that the covariance between values at any two sites depends only on distance and direction between the two sites, not their absolute location (Bailey and Gatrell, 1995). The second assumption, that the spatial process is isotropic, assumes that the covariance depends only on the distance between sites, and not the direction (Bailey and Gatrell, 1995). Given these assumptions, the presence and magnitude of second order effects can be assessed using tools such as the spatial autocorrelation statistics that are discussed below. If present, these effects can then be accounted for in regression analyses using spatial regression models that account for the covariance structure of observations attributable to second order effects.

**Univariate Analysis**

Univariate or exploratory data analysis plays an important role in the analytic process by clarifying the general structure of the data, providing simple
descriptive summaries, and possibly providing information to guide more sophisticated analyses (Chatfield, 1985). In this thesis, a number of tools are used to accomplish these tasks for aspatial and spatial aspects of the data. The following paragraphs outline the tools that are used and the rationale for the choice of those tools.

Aspatial

The distribution of each variable is examined in turn, and the results of the preparatory steps taken in the creation of the variable are presented. The preparatory steps are of particular importance for the hospitalization rates, as the distribution of the rates is dependent on the distribution of individual admissions. Distributions are examined using measures of central tendency and variability. In addition, observed distributions are compared to theoretical distributions used for multivariate modelling. Plots of distributions are also used to further describe the data in some cases. Information from this analysis is used to determine if variable transformation or categorization is indicated for the multivariate analysis.

Spatial

An approach to the exploratory analysis of spatial data follows from the theoretical model of spatial phenomena presented above. For each variable, this approach begins with the assessment of large scale (i.e. first order) effects, then moves on to assess localized spatial clustering of values (i.e. second order effects) (Haining, 1994). No rationale exists for the presence of a large scale effect in the data used in the thesis, and consequently plots of value against a specific point or absolute location are not made. However, the presence of large scale effects is informally assessed from spatial plots of the data (i.e. maps). Maps also allow the presence of second order effects, or spatial autocorrelation, to be visually assessed. Given that rationales exist for the presence of spatial autocorrelation in
all variables used in the thesis and the fact that visual assessment can be misleading, the presence of this local clustering of values is also examined more formally.

The literal meaning of spatial autocorrelation is self-correlation (autocorrelation) attributable to the geographical ordering of data (spatial) (Griffith, 1993). A number of statistics have been devised to measure spatial autocorrelation, and the properties of three of these statistics have been reviewed with respect to area level health data (Walter, 1992a; Walter, 1992b). Simulation studies suggest that the most robust of these with respect to data structure, population structure and population size is Moran's I statistic (Walter, 1992a). In addition, the I statistic also provides the greatest power for detecting clustering of the type expected amongst the variables used in the thesis (Walter, 1992b). The formula for I is given by

\[ I = \frac{\sum_{i} \sum_{j} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i} \sum_{j} w_{ij} \sum_{i} (y_i - \bar{y})} \]

Where there are \( n \) EAs, the data value for EA \( i \) is \( y_i \), and \( w_{ij} \) is the weight (or connectivity) for EAs \( i \) and \( j \) (Walter, 1992a). It is apparent from the formula above that I is a covariation measure similar to the traditional product moment correlation statistic, with the difference that covariation is assessed among values of the same variable weighted by their connectivity (Griffith, 1993). Connectivity weights can be defined in a number of ways, and an approach often used for area analyses is a binary measure based upon shared borders (Bailey and Gatrell, 1995). Simulations using a number of different types of weights with area health data suggest that conclusions about the strength of spatial clustering do not depend critically on the choice of weights (Walter, 1992a). For simplicity, the connectivity weighting employed in the thesis is therefore defined as
\[ w_{ij} = \begin{cases} 1 & \text{if } A_j \text{ shares a common boundary with } A_i \\ 0 & \text{otherwise} \end{cases} \]  

(9)

Where \( w_{ij} \) is the connectivity measure for area \( i (A_i) \) with area \( j (A_j) \).

If it is assumed that the \( y_i \) are observations on random variables \( Y_i \) whose distribution is normal, then if \( Y_i \) and \( Y_j \) are spatially independent \( (i \neq j) \), \( I \) has a sampling distribution which is approximately normal with expected value \( E(I) \) and variance \( \text{VAR}(I) \) as follows (Bailey and Gatrell, 1995)

\[
E(I) = -\frac{1}{(n-1)}
\]

(10)

\[
\text{VAR}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2S_0^2}
\]

(11)

where

\[
S_0 = \sum_i \sum_j w_{ij}
\]

(12)

\[
S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2
\]

(13)

\[
S_2 = \sum_i \left[ \sum_j (w_{ij} + w_{ji}) \right]^2
\]

(12)

In most cases, values for \( I \) fall in the interval (-1,1), with its actual extremes being determined by the eigenvalues of the connectivity matrix. Asymptotically \( (n \to \infty) \) the expected value of \( I \) is zero, and consequently the interpretation of \( I \) is similar to that of the traditional product moment correlation coefficient (Griffith, 1993).
After standardizing each variable, estimates of I and measures of variance are calculated with a program written in the SAS software (Griffith, 1993). The measures of variance are used to determine approximate z-values and carry out tests of significance. An adjacency matrix for EAs in SETO is calculated with a FORTRAN program that extracts the necessary spatial information from a MapInfo export file (MIF format) of the study area (Reynolds, 1998). The potential influence of deleting EAs from the study area (due to data suppression) is assessed using a FORTRAN program that performs random EA deletions from the study area using simulated data sets with varying values of I (Reynolds, 1998). These simulations suggest that measures of I obtained for this thesis are not influenced by the relatively small number of EAs deleted as a result of data suppressions by Statistics Canada.

**Multivariate Analysis**

The purpose of the multivariate analysis is to examine the relationship between variables, then select and apply the appropriate statistical tools to test the hypothesis and ancillary research questions of the thesis. The following paragraphs describe the tools used to accomplish the multivariate analysis.

**Aspatial**

If indicated by the univariate analysis, variables are transformed or categorized prior to the multivariate analysis. Relationships between variables are initially assessed using bivariate Spearman rank correlation. This statistic is used as respiratory health and exposure data are not expected to follow a normal distribution. In some cases, bivariate scatterplots are employed to further describe the relationship between variables.
Multivariate Poisson regression is used to formally test the hypothesis of the thesis and suggest answers for the ancillary research questions. Regression is used to test the hypothesis as regression tends to give less biased measures of association in ecologic studies than correlation (Nurminen, 1995), and regression effect estimates can be adjusted for confounders. Using the genmod procedure in the SAS software (SAS, 1996), the outcome variable is defined as observed admission counts, and the offset as expected admissions. Exposure is modelled as the explanatory variable, and the influence of SES is assessed by modelling it as a covariate. Parameter significance estimates and confidence intervals are obtained directly from the genmod procedure. Exponentiation of parameter estimates is used to arrive at estimates of relative risk. Model goodness of fit is assessed by comparison of the model deviance against a chi-square distribution with the appropriate degrees of freedom and examination of regression residuals (Myers, 1990). The general equation for the Poisson regression analysis is

\[
\ln(Observed_i) = \alpha + \beta_1 \text{Exposure} + \beta_2 \text{SES} + \ln(\text{Expected}_i) + \varepsilon \quad (14)
\]

Where \( Observed_i \) is the count of observed admissions for the diagnostic set of interest, \( \alpha \) is the intercept, \( \beta_1 \) is the parameter estimates for \( \text{Exposure} \), \( \beta_2 \) is the parameter estimate for \( \text{SES} \), \( \text{Expected}_i \) is the expected number of admissions (also known as the 'offset' variable), and \( \varepsilon \) is the random error component.

Two aspects of the multivariate analysis outlined above deserve further attention. The first is the possible presence of bias from incomplete control of confounding. As discussed previously, the approach used in the thesis to control for confounding cannot be expected to limit bias as all variables are not 'mutually standardized' (Greenland and Robins, 1994). Unfortunately, data on joint distributions that would improve control of confounding are not available. The analytic method presented above is used as it appears to produce reasonable
parameter estimates in other studies (Clayton et al., 1993). To assess if bias from inadequate control of confounding may be present, the degree of correlation between explanatory variables and the offset value is estimated. If lack of mutual standardization contributes to bias from confounding, one would expect to see correlations between the model covariates (exposure and SES) and the model offset (expected admissions) (Greenland, 1992). The second aspect of the multivariate analysis requiring further discussion is the interpretation of model goodness of fit. It has been observed in similar area level analyses that residual variation substantially in excess of that expected from Poisson sampling can be encountered. One approach to this problem is to model the random effects as a mixture of distributions such as a log-gamma distribution (Clayton et al., 1993). Another approach follows from the consideration that the extra-Poisson variation may be a result of spatial autocorrelation of explanatory variables, and subsequent correlation among random effects (Cook and Pocock, 1983). The latter approach is taken in this thesis and is described below.

Spatial

A first step in assessing the contribution of spatial autocorrelation among explanatory variables towards extra-Poisson variation is to assess the spatial autocorrelation of the regression residuals (Griffith, 1993; Haining, 1994). The I statistic in this situation is calculated in the same manner as described in the univariate section above. The standard error formula is more complex and numerically complicated than described above, but the aforementioned approximation appears to be a reasonable substitute (Griffith, 1993). If spatial autocorrelation is observed among the regression residuals, the next step is to fit a spatial regression model (Griffith, 1993).

The presence of spatial autocorrelation is assessed among regression residuals using Moran’s I and its standard error as described previously. In the event that
spatial autocorrelation is observed, a spatial regression model is not fit, as the complexity of this approach is beyond the scope of the thesis.

RESULTS

The results of implementing the method described in the previous section are presented in terms of univariate and multivariate results.

Univariate Results

Univariate results are presented for respiratory health data, motor vehicle emission exposure data, and SES data. Results of data preparation, aspatial and spatial analyses are presented for each of the data sets.

Respiratory Health Data

Data Preparation

The number of records abstracted from each year by diagnostic set is shown in Table 6. After the records are matched to the Ontario PCCF, 1002 records (1.4%) remain unmatched due to postal codes that are either invalid, or outside of Ontario. This proportion of unmatched records is lower than results from other similar studies in the United States (Vine et al., 1997). Table 6 shows that the unmatched proportion is similar across diagnostic sets. Geographic limitation to from the City of Toronto to SETO reduces the number of records from 69617 to 11012 (a reduction of 84.2%). Age and sex-specific population counts are not available for 32 of the 334 (9.6%) EAs in SETO, and deletion of these records reduces the total number from 11012 to 10667 (a reduction of 3.1%). The relatively small proportion of records falling in these EAs (3.1% of records in
9.6% of EAs) is most likely due to a smaller average population in these EAs as opposed to a healthier population living these EAs. This conclusion is supported by the knowledge that low population count is one of the criteria for data suppression (Statistics Canada, 1993), and the observation that suppressed EAs do not appear to be distributed throughout the study area according to any pattern. The spatial distribution of EAs is shown in Map 4. Removal of records without a valid birthdate, or with admission prior to January 1, 1990 reduces the total number of records from 10667 to 10550 (a reduction of 1.1%). Finally, deletion of records without a valid health number reduces the total number of records from 10550 to 10171 (a reduction of 3.6%).

Of the 10 171 admissions among all diagnostic sets, 8634 are first admissions. Repeat admissions therefore account for 15.1% of all admissions. However, the proportion of readmission does not appear to be uniformly distributed across all diagnostic sets. As Table 6 shows, 23.4% of respiratory subset admissions are readmissions, while the corresponding figure in the genitourinary chapter is 9.9%.

*Apatial Distribution*

Distributions and summary statistics of admissions by individuals are presented in Table 7 along with Poisson distributions for the observed means. Comparison is made to the Poisson distribution as the multivariate model used in the next section assumes that admissions are Poisson distributed. It is apparent from Table 7 that the vast majority of individuals in SETO are not admitted to hospital over the period 1990 to 1992 (98.5% of the population for the respiratory subset). Accordingly, the medians and modes for individual admissions in all diagnostic sets are zero. The frequency of admissions drops from zero in a manner that is not quite exponential, resulting in a longer tail than would be expected for a Poisson distribution. The observed distribution of admissions can be compared to the Poisson expected distribution by making use of a basic
property of the Poisson distribution, namely that the variance divided by the mean of a Poisson distribution is expected to be one. As can be seen from Table 7 the corresponding value for the observed distributions ranges from 2.61 for the respiratory subset to 1.34 for the genitourinary chapter. The extent of this extra-Poisson variation mirrors the magnitude of the re-admission rate among diagnostic sets (shown in Table 6). This observation supports the view that the lack of independence introduced into the distribution of admissions by the inclusion of multiple admissions reduces the fit of a Poisson model to admission data (Cain and Diehr, 1992).

The distributions of counts of first and all admissions by EA are presented in Figure 6 by diagnostic set along with Poisson expected distributions. It is clear from the observed distributions that neither first, nor all, admissions by EA are distributed in a truly Poisson fashion. The distribution of admissions for all diagnostic sets deviate from a Poisson shape due to the large number of EAs with no admissions (74 EAs for the respiratory subset), and the longer tail of EAs with high numbers of admissions. However, the observed distributions do resemble the expected Poisson distributions to a degree that varies among diagnostic sets. The distributions of first admissions in general, and specifically those for the respiratory subset and GU chapter appear to resemble their expected distributions most closely.

Summary statistics are presented in Table 8 for EA admission counts, crude admission rates, and standardized admission rates by diagnostic set. It is apparent that for the count data the median number of first admissions in EAs is small, ranging from 4 for the respiratory subset to 10 for the GI chapter. Median values for counts of all admissions tend to be higher by one or two admissions. Mean admission counts demonstrate a similar pattern, but tend to be somewhat larger than median values because of the long tails on the distributions. Despite lower median and mean numbers of admissions than the other diagnostic sets, the
respiratory subset and respiratory chapter have higher maximums for counts of first and all admissions. This suggests that EA admission counts are more variable in these diagnostic sets, and this suggestion is confirmed by larger relative standard deviations for the same diagnostic sets.

Summary statistics for crude admission rates closely resemble those for admission counts. The median and mean crude admission rates tend to be slightly smaller than median and mean counts, but the ordering of magnitude among diagnostic sets is the same. In contrast, the maximum values and standard deviations for crude rates tend to be higher than for admission counts, likely because of the extra variability introduced into rates by the range of population among the EAs.

To facilitate comparison between counts, crude rates and standardized rates, the latter are presented as indirectly standardized rates (ISRs) and not standardized hospitalization ratios (SHRs). The ISR is simply the SHR multiplied by the standard rate, and the mean ISRs in Table 8 are approximately the same as the standard rates for SETO. Age and sex standardization results in rates that generally have a slightly larger median value and slightly smaller mean value than crude admission rates. Changes from the values seen with crude rates are more apparent for maximum values and standard deviations, with values of both measures being smaller for the standardized rates compared to the crude rates. This effect of standardization is greatest for the respiratory subset and respiratory chapter. All of the changes seen with standardization are a result of removing variation attributable to age and sex from the crude rates. The impact is greatest on the respiratory rates as respiratory admissions in SETO demonstrate greater variability across age and sex strata than GI and GU conditions (data not shown).

To summarize, the aspatial univariate analysis of hospital admission data has revealed that distributions of admission counts for individuals and EAs resemble Poisson distributions to varying degrees. Deviations from the Poisson
distribution are a result of a high number of zero counts, and more large counts than expected. In general, first admission counts resemble Poisson distributions more closely than all admission counts, but the difference between the distributions of first and all admissions is small. Finally, it appears that distributions of indirectly standardized rates demonstrate less variability with lower maximum values than distributions of crude rates. These results suggest that a Poisson model will provide the best fit for indirectly standardized rates of first admissions.

Spatial Distribution
Counts of first respiratory subset admissions are mapped for the study area using proportional symbols in Map 5. No large scale trend is readily apparent in counts, but high levels of counts are generally seen in EAs in the St James Town, Regent Park, Moss Park and South Riverdale neighborhoods (neighborhood boundaries are shown in Map 1). Counts appear to be particularly clustered about St James Town, but as the inset map reveals, this is partly a result of the spatial concentration of EAs (in turn a reflection of high population density). First admission counts for other diagnostic groups and all admission counts are distributed in a similar manner, but are not mapped.

Indirectly standardized first respiratory subset admission rates are shown in Map 6. Rates are mapped using the choropleth method with EAs categorized into quintiles of rate. At first glance, it appears from this Map that there is a large scale trend towards higher rates in the Southeast of the study area. A closer inspection reveals that a small number of EAs with large areas are responsible for the appearance of a trend. This propensity for physically large areas to dominate the display is a recognized shortcoming of choropleth maps (Bailey and Gatrell, 1995). With the exception of the area in the immediate Southeast, no areas with local clustering of high rates are apparent. The inset map reveals that clusters of
EAs with high counts in Map 5 demonstrate nearly random ordering of rates. This is partly a result of mapping rates as opposed to counts, and partly a result of the different visual impressions from the two mapping methods. However, as with counts, EAs with the highest rates tend to be found in St James Town, Moss Park and Regent Park. The standardized respiratory subset rates for first admissions are compared to rates for the respiratory chapter, gastrointestinal chapter, and genitourinary chapter in Map 7. At the scale used in this Map, the limitations of choropleth maps are more evident, with large areas visually dominating and small areas providing little visual contribution. Nevertheless, a number of observations can be made.

To begin with, the spatial distribution of respiratory chapter rates is similar to that of respiratory subset rates. In addition, there does not appear to be any evidence of a large scale trend in the spatial arrangement of gastrointestinal chapter or genitourinary chapter rates. Local clustering is also not apparent for rates in either of these chapters. An examination of the distributions at a greater resolution (not shown here) supports this observation. Finally, at this scale there appear to be noticeable, but not extensive differences between the spatial distributions of rates for the respiratory subset and the gastrointestinal and genitourinary chapters. Spatial distributions of rates for all admissions are similar to those for first admissions and consequently are not shown.

More formal analysis of the degree of localized clustering of admission rates using Moran's I coefficient of spatial autocorrelation supports the results from inspection of the Maps. Values of I and corresponding p-values for two-tailed tests of the null hypothesis of no spatial autocorrelation are shown for each diagnostic set in Map 7. The results for the respiratory subset (I =0.005; p 0.971) and respiratory chapter (I =0.045; p 0.287) support the observation made above that the values of rates in these diagnostic sets are randomly distributed in the two-dimensional space of the study area (i.e. no spatial autocorrelation exists).
Results for the other two diagnostic sets support the observation that there is no clustering of similar rates (i.e. no positive spatial autocorrelation). On the other hand, these results also suggest that there is a slight tendency for neighboring areas to have dissimilar rates (i.e. negative spatial autocorrelation). This is most apparent for the gastrointestinal chapter (I -0.101; p 0.013), but is also seen for the genitourinary chapter (I-0.081; p 0.051). No rationale for this observation is immediately apparent. Measures of spatial autocorrelation are similar for rates of all admissions and are not presented.

In summary, the univariate spatial analysis of the respiratory health data suggests that there is no large scale trend in the spatial distribution of admission counts or rates for any of the diagnostic sets. Analysis of rates for local trends or clustering reveals that standardized admission rates do not exhibit positive spatial autocorrelation in any of the diagnostic sets. Despite the lack of positive spatial autocorrelation, EAs with the highest respiratory subset rates tend to be seen in certain neighborhoods. Interestingly, admission rates for the gastrointestinal and genitourinary chapters exhibit a mild degree of negative spatial autocorrelation.

**Exposure Data**

*Data Preparation*

The results of preparation of the exposure data are given in the methods section. This approach is taken as the results are meaningful in the context of the method, and their presentation here would necessitate repetition of a considerable portion of the method to provide that context.

Exposure data preparation results in three measures of exposure to traffic volume and three measure of exposure to PM 2.5 for each EA. The three measures represent total exposure transferred from the street network by EA street
frontage; total exposure weighted by the proportion of the EA within 10m of a street; and, total exposure weighted by the area of the EA. All three measures are similarly distributed amongst the EAs, and the choice of exposure measure does not alter the result of the multivariate analysis or the conclusions of the thesis. Accordingly, results are only presented for one measure. The measure presented is the one most strongly supported by dispersion theory, the total exposure weighted by the proportion of the EA within 10m of a street. Results for the other measures are available from the author.

Aspatial Distribution

Summary statistics for distributions of exposure to traffic volume and PM 2.5 emissions are presented in Table 9. Distributions for the 302 EAs used as the sample for the multivariate analysis are summarized separately from distributions for the 32 EAs excluded from the analysis due to the suppression of population data.

It is apparent from the range and standard deviation that considerable variation exists in the distribution of traffic volume experienced by the sample EAs. In addition, this distribution does not closely resemble a normal distribution as evidenced by the mean that is more than three times the magnitude of the median exposure value. This relationship between the median and the mean suggests that the distribution is skewed to the right despite a large number of EAs (63, 20.9%) with an exposure value of zero. The distribution of PM 2.5 exposure amongst the sample EAs demonstrates similar qualities, but tends to have less variability. In comparison to the distributions of traffic volume and PM 2.5 exposure among sample EAs, the distributions among the suppressed EAs demonstrate less variability and higher median values. The decreased relative variability appears to be a result of lower maximum values and lower proportions of EAs with exposure values of zero (3, 9.4%).
In summary, the distribution of traffic volume exposure and PM 2.5 exposure among the sample EAs demonstrates considerable variability and does not closely approximate a normal distribution. This suggests that exposure values will have to be transformed or categorized prior to a multivariate analysis. In addition, exposure values in the suppressed EAs tend to be larger and less variable than those in the sample EAs.

**Spatial Distribution**

The spatial distribution of exposure to PM 2.5 from motor vehicle emissions among the sample EAs is shown in Map 8. Exposure values are grouped by quintiles. Using this method of categorization, the distribution of exposure to traffic volume is nearly identical and is therefore not shown. This approach to mapping exposure does not demonstrate any large scale trend in the spatial distribution of PM 2.5 exposure values among EAs. However, there does appear to be a moderate amount of local spatial clustering of exposure values. Formal assessment of this apparent clustering confirms that there is indeed a significant amount of positive spatial autocorrelation amongst PM 2.5 exposure values ($I \leq 0.308; p \leq 0.001$). The highest exposure values tend to be found in EAs containing or bordering highways and narrow EAs with the long side of the EA facing a street. EAs with exposure values of zero tend to have small areas and be located away from roads. In comparison to PM 2.5 exposure, the magnitude of spatial autocorrelation is slightly larger for exposure to traffic volume ($I \leq 0.377; p \leq 0.001$). This suggests that the variation introduced into the distribution of exposure values by emission modelling decreases the tendency for elevated exposure values to cluster around streets.


**SES Data**

*Data Preparation*

The SES variables selected for consideration are listed in Table 10 under the headings of dwelling characteristics, educational attainment, employment, income, mobility, and social characteristics. It is apparent from Table 10 that several of the variables have a small number of nonzero values. This is most apparent for the variables related to dwelling characteristics and income, where only a third to a half of the values are greater than zero. Dwelling value observations of zero are clearly the result of non-response, or data suppression by Statistics Canada, as it is impossible to have an average dwelling value of zero. However, for observations of income variables, zeroes may represent missing values or true zero values. The inability to determine the validity of an observation with a value of zero requires that all observations with values of zero be deleted. Consequently, variables with low numbers of nonzero values are of limited use, as they decrease the sample size of the multivariate analysis.

Results of rank correlation between each SES variable and the standardized rate of first admission for respiratory subset conditions are shown in Table 10. Significant negative correlations are seen with university and female labor force, while significant positive correlations are seen with grade 9, unemployed, low income persons, lone parents, and female lone parents. With respect to magnitude, the strongest significant correlations are seen with unemployed ($p = 0.236$), university ($p = -0.226$), and low income persons ($p = 0.190$). Experience with measures of SES in SETO suggests that low income persons is the most robust univariate measure of SES for the study area (Patychuk, 1998). Unfortunately, this measure is available for only 111 of the 302 EAs (36.8%) and therefore another univariate measure must be selected to control for SES confounding in multivariate modelling. The most appropriate univariate measure is identified by
selecting the variable with observations for greater than half of the EAs that demonstrates the strongest rank correlation with low income persons. University (p -0.690) and unemployed (p 0.628) are the most suitable by this criteria, and university is selected for multivariate modelling as it has a larger number of nonzero observations.

Standardized SES variables initially included in the index construction process are grade 9, university, unemployed, total lone parents and female labor force. Poisson regression of these variables on standardized first admission rates for respiratory subset conditions identifies grade 9 ($\chi^2$ 4.44; p 0.035) and total lone parents ($\chi^2$ 16.04; p 0.001) as having significant explanatory power. The formula for the index constructed from these two standardized variables is

$$\text{Index}_i = (0.078 \times \text{grade 9}_i) + (0.153 \times \text{total lone parent}_i)$$

Where $\text{Index}_i$ is the value of the index for EA$_i$, and $\text{variable}_i$ is the standardized value of the given census variable for EA$_i$. In a Poisson regression analysis with standardized rates of first admission for respiratory subset conditions, this index has less explanatory power than the univariate measure of university. Furthermore, values of the index are available for 182 EAs in comparison to university that has values available for 273 EAs. Consequently, the univariate measure of university is used to control for potential confounding by SES in the multivariate analysis. The distribution of this measure is described in the following paragraphs.

_Aspatial Distribution_

The proportion of the population with a university degree ranges from 1.2% to 62.5% in the 273 EAs with nonzero values. Values between these extremes are distributed in an approximately normal manner with a median of 21.6%, mean of 23.2% and a standard deviation of 13.6%. This close approximation of the
normal distribution suggests that this variable can be included in the multivariate analysis as a continuous variable.

**Spatial Distribution**

The spatial distribution of the proportion of the population with a university degree is presented in Map 9. No large scale trend is evident in the distribution, but local clustering is evident in a number of areas. EAs with high values tend to cluster in the Rosedale, Yorkville and North Riverdale neighborhoods (neighborhood boundaries are defined in Map 1). In contrast, EAs with low values demonstrate a tendency to cluster in Regent Park and South Riverdale. The presence of a moderate degree of positive spatial autocorrelation is confirmed by a value for Moran's I of 0.352 (p < 0.001).

**Multivariate Results**

Multivariate results are presented from aspatial and spatial analyses. Results from aspatial analyses are reported for rank correlations, scatterplots, and the Poisson regression analysis. The spatial multivariate results describe the spatial distribution of residuals from the Poisson regression analysis.

**Aspatial Analysis**

**Correlation**

The first aspect of the correlation analysis examines the relationships between standardized rates of first and all admissions within the same diagnostic sets. This identifies very strong and highly significant correlations between first and all admissions with coefficients ranging from 0.989 for the GI chapter to 0.982 for the respiratory subset. Given the similarity between rates of first admissions and all admissions, results of the multivariate analysis for all admissions are not
Rates of first admissions are chosen for presentation for several reasons. First, they provide a measure of incidence that is generally easier to interpret than the prevalence measure provided by rates of all admissions (Rothman and Greenland, 1998c). In addition, it is not clear if exposure to PM 2.5 more plausibly exacerbates or causes respiratory conditions. Without such evidence, and given the relationship between incidence and prevalence, examination of the incidence of respiratory conditions (i.e. first admissions) is arguably the most appropriate approach. Finally, the distribution of rates of first admissions more closely resembles the Poisson model used in the regression analysis. Results of the multivariate analyses with first and all admissions are similar, with the exception that rates of all admissions demonstrate slightly stronger associations with exposure and SES. Results of the multivariate analysis with rates of all admissions are available from the author.

The results of the correlation analysis between the variables included in the multivariate analysis are presented in Table 11.

Examination of correlations amongst admission rates for the different diagnostic sets reveals that respiratory subset rates are very strongly correlated with respiratory chapter rates (r 0.949; p < 0.001). This relationship is expected as the respiratory subset conditions are contained within the respiratory chapter, accounting for 70% of respiratory chapter admissions. Correlations between admission rates also demonstrate that respiratory rates are strongly correlated with both GI chapter rates (r 0.745; p < 0.001), and GU chapter rates (r 0.740; p < 0.001).

Correlations between measures of exposure and admission rates demonstrate that both traffic volume and PM 2.5 exposure are significantly associated with admission rates in all diagnostic sets. The correlations are not strong, with the strongest correlation seen between exposure to PM 2.5 and respiratory subset
rates (p 0.222, p 0.001). A scatterplot of PM 2.5 with nonzero values of the Standardized Hospitalization Ratio (SHR) for respiratory subset conditions is shown in Figure 7. Note that this plot uses a log scale for SHR, and does not display one outlying EA (exposure of 1183 grams and SHR of 0.53). The plot appears to support a weak positive correlation up to exposure values of approximately 200 grams, but little relationship is apparent for outlying EAs beyond this value. In general, respiratory subset rates tend to demonstrate the strongest correlations with measures of exposure, and admission rates tend to be more strongly correlated with PM 2.5 than traffic volume. GU chapter rates are the exception to both of these trends as they are more strongly correlated with traffic volume (p 0.199, p 0.001) than respiratory subset rates (p 0.188, p 0.001), and are also more strongly correlated traffic volume than with PM 2.5 (p 0.189, p 0.001).

Given the process used to select university graduation as a measure of SES, it is not surprising that respiratory subset rates show the strongest correlation with university completion (p -0.226, p 0.001). Admission rates for other diagnostic sets demonstrate weak negative correlations with university graduation, but the correlations are not significant for GI chapter rates (p -0.085, p 0.160), or GU chapter rates (p -0.099, p 0.101). It is noteworthy that all other measures of SES examined for the thesis are also more strongly associated with respiratory subset rates than admission rates for other diagnostic sets (data not shown). Examination of correlations between the proportion of university completion and measures of exposure reveals a weak non-significant association between university and traffic volume (p 0.097, p 0.108), and no association with PM 2.5 (p 0.030, p 0.625). None of the other SES measures examined for this thesis are significantly correlated with either measure of exposure, with the exception of
average dwelling value which is significantly and negatively associated with traffic volume \((p\ -0.249, \ p\ 0.002)\) and PM 2.5 \((p\ -0.219, \ p\ 0.006)\).

In summary, the results of the correlation analysis reveal a number of important relationships between variables. To begin with, admission rates are strongly correlated between diagnostic sets. Admission rates in all diagnostic sets are also moderately correlated with both measures of exposure. Correlations with exposure tend to be strongest for respiratory subset rates, and the strongest correlation is between respiratory subset rates and PM 2.5. Finally, neither non-respiratory rates nor measures of exposure are strongly correlated with the proportion of the population graduating from university.

**Poisson Regression**

The results of the univariate analysis suggest that the counts of first admissions are distributed in an approximately Poisson manner. Additionally, the proportion of the population graduating from university in each EA appears to be distributed in an approximately normal fashion. Transformation or categorization is therefore not required for either of these variables prior to the Poisson regression analysis. In contrast, the two measures of exposure do not appear to be normally distributed, and consequently these distributions must be transformed or categorized prior to regression analysis. Regression analyses are performed using a rank transformation of exposure and categorizations of exposure by quintile and decile. An analysis using a log transformation is not used as a considerable number of EAs (63, 20.9%) have an exposure value of zero. The results of the three analyses do not differ with respect to the hypothesis or ancillary research questions. Consequently, results are only presented for the analysis with exposure categorized by quintiles. Results from the analysis using quintiles are chosen for presentation due to the ease of presentation and interpretation of these results. Results by ranks and deciles are available from the author.
The first component of the regression analysis addresses the hypothesis of the thesis. A bivariate analysis is initially performed to assess the effect of exposure to PM 2.5 on respiratory subset admission rates without adjustment for the effect of SES. This is followed by a multivariate analysis including exposure to PM 2.5 and proportion of university graduation as covariates. Results of the bivariate and multivariate analyses are presented in Table 12.

It is apparent from the results in Table 12 that exposure to PM 2.5 has a significant effect on respiratory subset first admission rates, before and after adjustment for SES. The results also indicate that the magnitude of the effect tends to increase with increasing exposure to PM 2.5, and this trend becomes more pronounced after adjustment for SES. In the SES adjusted model, the estimate of relative risk (RR) increases from 1.10 in the second quintile to 1.39 in the fifth quintile, and the estimates are significantly greater than unity for quintiles three through five. This relationship is shown in Figure 8, where the rate of the increase in RR appears to be constant through the fourth quintile, then decrease slightly between the fourth and fifth quintiles. Another observation that can be made from the results in Table 12 is that neither the bivariate nor the multivariate model provides a good fit to the data (Clayton and Hills, 1993). The expected value for model deviance over degrees of freedom is approximately one, and the observed value for both models is considerably larger than one.

Results of this portion of the analysis support the rejection of the null hypothesis of the thesis. In other words, modelled exposure to PM 2.5 from motor vehicle emissions appears to have a significant effect on hospitalization for selected respiratory conditions. However, this result must be interpreted in light of the poor fit of the Poisson model and the possible presence of bias from uncontrolled confounding.
Poor model fit is assessed in detail through examination of residuals from the regression model that adjusts for SES. A number of EAs have admission rates that are poorly fit by the model (absolute value of standardized residual greater than 3), but only seven of these exert considerable influence on the parameter estimates ($\beta$ greater than 0.2). Examination of the aspatial properties of these EAs reveals that in over half of them, high-rise dwellings are the dominant type of housing in the EA. Consequently, the contribution to the model of a variable describing proportion of high-rise dwellings is explored, but is not found to be significant. An analysis is also conducted after the deletion of these seven EAs. The result of this analysis is similar to the one summarized in Table 12, but with a somewhat stronger effect of exposure on hospitalization rates. Given these findings, results are reported for the analysis including these EAs. Spatial properties of the residuals are also assessed, and the results of this assessment are reported in the spatial section of the multivariate results.

The potential for incomplete control of confounding is assessed as the method used to control for confounders cannot be expected to reduce bias due to the lack of 'mutual standardization' of all variables (Greenland and Robins, 1994). Moderate correlations are seen between expected respiratory subset admissions (i.e. the offset variable containing information about age and population) with PM 2.5 exposure ($\rho$ 0.188, $p$ 0.002) and university graduation ($\rho$ -0.302, $p$ 0.001). These correlations suggest that control of confounding may be incomplete.

The next component of the regression analysis addresses the two ancillary research questions of the thesis. Results from analyses performed for this component are summarized in Table 13. Parameter tests of significance included in these results clearly indicate that modelled exposure to PM 2.5 has a significant

---

$^*$ The 'xbeta' measure of influence is a statistic calculated by the SAS genmod procedure that describes the change in the vector of parameter estimates when one observation is excluded from the data (SAS, 1996). It is analogous to the 'dfbetas' statistic used in ordinary least squares linear regression (Myers, 1990).
effect on admission rates in the non-respiratory diagnostic sets. The estimates of RR of admission for the GI diagnostic set are compared with estimates of RR for the respiratory subset in Figure 9. This graph reveals that the estimates of RR for GI hospitalization tend to be lower than those for respiratory subset hospitalization. However, it also reveals that the estimates of RR for GI hospitalization are within the confidence intervals around respiratory subset estimates for all but one quintile of exposure. Estimates of RR for GU hospitalization are nearly identical to those for GI and are not shown. It should also be noted that Table 13 demonstrates poor model fit is a greater problem for both GI and GU admission rates than respiratory subset rates. Results from this portion of the analysis provide an answer to the first ancillary research question, suggesting that modelled exposure to PM 2.5 has a significant effect on hospitalization rates for non-respiratory conditions. Again, this result must be interpreted in light of the poor model fit and potential bias from confounding. As with the respiratory subset, moderate correlations are seen between PM 2.5 exposure and offset values for both the GI chapter (p 0.244, p 0.001) and the GU chapter (p 0.223, p 0.001).

The second ancillary research question asks if exposure to traffic volume has a significant effect on hospitalization for selected respiratory conditions. The last model summarized in Table 13 suggests that an effect of this nature does exist. Specifically, the significance test of the traffic volume parameter in this model indicates that traffic volume has a significant effect on respiratory subset hospitalization rates. As with other models examined, this model also does not fit the data well, but it appears to have the best fit of all the models examined. The relationship between RR of respiratory subset admission and exposure to traffic volume is compared to the relationship between RR of admission and exposure to PM 2.5 in Figure 10. The graph suggests that the relationship between RR of respiratory subset admission and traffic volume is quite similar to the relationship
between RR of admission and PM 2.5. A final observation that can be made from Table 13 is that exposure to PM 2.5 also has a significant effect on respiratory chapter admission rates. However, the model fit is poorer, and the strength the effect is weaker than for respiratory subset admissions.

To summarize, a number of observations can be made from the results of the aspatial regression analysis. Most importantly, the significant effect of exposure to PM 2.5 on respiratory subset admission rates supports rejection of the null hypothesis. Estimates of RR suggest that the size of the effect increases with increasing exposure. However, the finding must be interpreted in light of poor model fit and the possibility of uncontrolled confounding. In response to the first ancillary research question, exposure to PM 2.5 also appears to have a significant effect on admission rates for non-respiratory conditions. Finally, the existence of a significant effect of traffic volume exposure on respiratory subset admission rates suggests an answer for the second ancillary research question.

**Spatial Analysis**

Standardized likelihood residuals from the multivariate Poisson regression analysis used to address the hypothesis are displayed in Map 10. No large scale spatial trend is evident in distribution of the residuals. In addition, small scale or local clustering does not seem to be present in the spatial distribution. Lack of spatial autocorrelation amongst the residuals is supported by a low value of Moran’s coefficient ($I = 0.014$, $p = 0.788$). A positive residual signifies that the admission rate predicted by the model is less than the observed rate, and a negative residual signifies the opposite. Several EAs with large positive and negative residuals are located in the St James Town neighborhood. This suggests that the model does not provide a good description of the data in this area. In general, many of the EAs with large positive residuals are found in Moss Park and St James Town where respiratory admission rates tend to be higher. This
could imply that determinants of higher admission rates in these areas are not captured in the regression model.
Chapter Four

Discussion
The aim of this chapter is to discuss the results of the thesis in terms of previous work in this area. The chapter is divided into three sections. In the first section, the association between exposure to motor vehicle emissions and respiratory health is examined. The exposure assessment process used in the thesis is discussed in the second section and the final section discusses the limitations of the thesis. Conclusions are not drawn from the discussion in this chapter, but are reserved for the following chapter.

**Motor Vehicle Emissions and Respiratory Health**

The results of the analysis support an association between exposure to motor vehicle emissions and the rate of respiratory hospitalization at the level of the enumeration area (EA). In addition, the analysis estimates the effect of exposure on hospitalization by contrasting rates in exposed and unexposed population subgroups. The strength of the association and the magnitude of the effect both appear to be beyond what one can reasonably attribute to chance, and accordingly the null hypothesis is rejected. Results of statistical tests used in the analysis can establish the presence of an association, but they cannot determine if the association is causal. The broader question of causation therefore remains to be answered.

No set of necessary and sufficient criteria exists in epidemiology to discern a causal relationship from a noncausal relationship (Rothman and Greenland, 1998b). However, lists of the aspects of a relationship to consider in passing from an observation of association to a verdict of causation have been promulgated. These lists are most appropriately employed as an aid to inference (Susser, 1991), and are not intended to provide a checklist of criteria (Lanes and Poole, 1984). The most commonly used list was proposed by Hill at the inaugural meeting of
the Occupational Medicine section of the Royal Society of Medicine (Hill, 1965). The list suggests the following aspects of an association be considered in attempting to distinguish causal from noncausal associations: strength of association, consistency of results, specificity, temporality, biologic gradient, plausibility, coherence, experimental evidence, and analogy. These aspects offer a template for the ensuing discussion of the results of the thesis and previous studies.

**Strength of Association**

Hill placed strength at the top of his list as he felt that weak associations are more likely to result from bias than strong associations are. However, the potential for weak associations to be causal in nature is acknowledged by Hill, and examples of strong noncausal associations are not difficult to find (Rothman and Greenland, 1998b). In this thesis, the correlation between exposure and outcome ($p 0.222$) and the effect of exposure on outcome (RR 1.39) are both weak. Several sources of bias may have influenced these results, and these are discussed in the section on limitations. Nevertheless, these results are similar in magnitude to those reported in a case-control study of asthma admissions amongst children exposed to motor vehicle emissions (odds ratios of 1.16 to 1.52) (Edwards et al., 1994). Cross-sectional studies have also reported odds ratios in a similar range for respiratory symptoms amongst individuals exposed to emissions (Keil et al., 1996; Nitta et al., 1993; Oosterlee et al., 1996; Weiland et al., 1994; Wijkstra et al., 1993). The lack of confounder control may be a source of bias in the case-control study, but does not appear to be a problem in the cross-sectional studies.

The distinction between a relative difference and an absolute difference in rates is also discussed by Hill under the topic of strength of association. The relative difference in rates determines the degree of association, whereas the absolute difference in rates determines the practical importance of the association. In this
study, the EAs in the lowest quintile of exposure have an admission rate of 4.5 per thousand per three years. With this baseline rate, the results suggest that approximately 39 (95% confidence interval 10 to 73) of the 593 first admissions in SETO (6.6%) during each year of the study are attributable to motor vehicle emissions exposure.

Consistency of Results

Consistency refers to the repeated observation of an association in different populations under different circumstances. Lack of consistency does not rule out a causal relationship, as inconsistency may result from variations in associated risk factors between study populations, differences in methods, or random error (Rothman and Greenland, 1998b). In general, consistency across studies suggests that an association is not attributable to some factor that varies between studies. Results of the thesis are consistent with the majority of other studies that have examined the relationship between exposure to motor vehicle emissions and respiratory health. These studies use a variety of methods and outcome measures to examine the association in a variety of populations. It is noteworthy that studies with good control for confounding by SES, indoor air quality, and mobility (Nitta et al., 1993; Oosterlee et al., 1996; Wijkstra et al., 1993) report similar results to studies with poorer confounder control (Brunekreef et al., 1997; Edwards et al., 1994; Weiland et al., 1994). This implies that incomplete control for these variables may not be a large source of bias. The association seen in the thesis is inconsistent with two previous studies that report finding no association. This inconsistency may be explained in part by differences in the methods used to assess exposure. The studies that find no association use methods that likely result in a considerable amount of misclassification. If this misclassification is nondifferential it could attenuate the observed association in these individual-level studies (Livingstone et al., 1996; Waldron et al., 1995).
Specificity

The aspect of specificity is meant to assess if an exposure is associated with one effect, or many effects. Hill was quick to point out that one must not 'over-emphasize the importance of this characteristic' (Hill, 1965), and others have suggested that in this context, specificity is 'useless and misleading' (Rothman and Greenland, 1998b). The results of the thesis suggest that the association between exposure to motor vehicle emissions and respiratory hospitalization is not specific. A similar, but weaker association is seen between exposure and hospitalizations for non-respiratory diagnostic sets. Interestingly, the only other study to examine the association between this exposure and non-respiratory conditions reports an association between exposure and admission for all causes (Edwards et al., 1994).

One possible explanation for the lack of specificity is confounding. The other study to note a non-specific association does not control for confounders (Edwards et al., 1994). In the current study, the strong association between admission rates for respiratory and non-respiratory diagnostic sets may reflect a relationship between the diagnostic sets and another risk factor. If this risk factor is ecologically associated with exposure, it could confound the relationships between the diagnostic sets and exposure. Socioeconomic status (SES) could act in this manner, but no area level SES measure is strongly associated with admission rates for all diagnostic sets and exposure. Furthermore, other studies with good SES control find an association, but none of these examines the specificity of the association (Nitta et al., 1993; Oosterlee et al., 1996; Wjst et al., 1993). Given the unit of analysis in this study, the possibility of SES confounding due to gradients of SES within EAs must be considered. This could occur if disadvantaged individuals within EAs tend to live closer to busy streets, and this results in bias of within-EA effect estimates.
Another possible source of confounding is age. As with SES, attempts are made to control for this confounder by analytic methods, but correlations between expected admissions (a function of the age distribution within an EA) and exposure suggest that this control may be incomplete. Other possible confounding variables include environmental exposures associated with proximity to streets, but it is difficult to identify an environmental exposure that is plausibly related to all of the diagnostic sets examined in the thesis. Finally, a less likely, but conceivable explanation for the observed lack of specificity is that the association with non-respiratory admissions is real. Exposure to air pollution has been consistently associated with hospitalization for non-respiratory diagnostic sets such as cardiac conditions (Burnett et al., 1995; Schwartz, 1994a). In addition, one of the postulated physiological mechanisms for the action of particulate matter (PM) is through an effect on immune function, and this could have an impact on any body system.

**Temporality**

Despite Hill's statement to the contrary, temporality is a *sine qua non* for causality as a cause must precede an effect (Rothman and Greenland, 1998b). This is particularly relevant to chronic diseases where early stages of disease may lead to a modification of exposure. In the thesis, exposure assessment is based on residence at the time of admission. From hospital admission data, there is no way to determine duration of residence and by extension duration of exposure. Consequently, temporality is difficult to establish in the thesis. Duration of residence has been controlled for in three studies (Nitta et al., 1993; Oosterlee et al., 1996; Wijst et al., 1993), and these studies report an association that is similar in nature to the association found by studies that do not control for duration of residence (Brunekreef et al., 1997; Weiland et al., 1994). This could imply that the induction period is short, or that individuals in the study populations are not very mobile. Limited support exists for the first interpretation in the form of acute
exposure time-series studies that demonstrate an onset of respiratory effects within one to five days following exposure (Dockery and Pope, 1996).

**Biologic Gradient**

Biologic gradient refers to the presence of a biologically meaningful dose-response curve, traditionally monotonic in nature. This aspect of an association is not necessary for causality, and associations that exhibit a biologic gradient may not be causal (Rothman and Greenland, 1998b). The association between exposure and respiratory subset admission rates in this study demonstrates a monotonic trend in disease frequency with increasing levels of exposure. However, this trend cannot be interpreted as a true 'biologic gradient' as it exists between EA rates and EA exposures in this ecologic study. This trend only represents a true 'biologic gradient' if exists at the level of the individual (i.e. there is no ecologic bias). In addition, the monotonic nature of this trend is partly a function of the analysis. In the analysis by deciles of exposure, an increasing trend is noted, but it is not monotonic. Two other studies report observing a trend of increasing adverse respiratory health outcomes with increasing exposure (Nitta et al., 1993; Waldron et al., 1995), but neither of these trends are monotonic. As Hill notes, the difficulty in securing a satisfactory quantitative measure of exposure may influence the ability to observe a dose-response relationship (Hill, 1965).

**Plausibility**

Hill succinctly describes the contribution of plausibility to an assessment of causation. 'It will be helpful if the causation we suspect is biologically plausible. But this is a feature I am convinced we cannot demand. What is biologically plausible depends on the knowledge of the day' (Hill, 1965). A conceptual model is developed in the second chapter to demonstrate the physical and biological plausibility of an association between exposure to motor vehicle emissions and
respiratory health. It is clear from the model that biologically plausible mechanisms exist for the observed association. The exact features of these mechanisms are still unknown, but this limited knowledge of the day does not imply that the association is noncausal.

Coherence

The coherence of an association refers to whether it conflicts with what is known of the natural history and biology of the disease. The association between exposure to motor vehicle emissions and respiratory health is coherent with epidemiological studies of the association between chronic exposure to air pollution and respiratory morbidity. Large prospective cohort studies of this relationship demonstrate that chronic exposure to ambient air pollution is associated with COPD, asthma and bronchitis (Abbey et al., 1993; Dockery et al., 1993; Dockery et al., 1989). The observed association is also coherent with toxicological and clinical knowledge of the disease processes involved. Respiratory changes in animals (Hyde et al., 1978; Moorman et al., 1985) and humans (Rudell et al., 1996) experimentally exposed to motor vehicle emissions are similar to changes observed in epidemiological studies (van Vliet et al., 1997; Wijst et al., 1993). Furthermore, these respiratory changes are coherent with the increased rates of diagnoses (Edwards et al., 1994) and symptoms (Brunekreef et al., 1997; Nitta et al., 1993; Oosterlee et al., 1996) found in other epidemiological studies.

Experiment

Experiment refers to the removal of a harmful exposure through intervention or prevention. There is no published study of the effect of removing or preventing chronic exposure to motor vehicle emissions at the population level. However, an example of this sort does exist for chronic exposure to air pollution in general. In
the winter of 1986-1987 a unique natural experiment was documented in the Utah Valley when a labor dispute resulted in the closure of a steel mill (Pope, 1989). This mill was the largest source of PM emissions in the valley, and during the winter of the closure, PM concentrations fell by approximately 50 percent in comparison to previous years. In the same winter, children's hospital admissions for respiratory disease fell by over 50 percent compared with adjacent years. This situation does not involve exposure to motor vehicle emissions directly, but it does demonstrate that a reduction in chronic exposure to a constituent of emissions is associated with a reduction in adverse respiratory health outcomes.

**Analogy**

The contribution of analogy to determining causation is limited only by the imagination of scientists (Rothman and Greenland, 1998b). However, one does not need to let their imagination go too far before coming across an analogy for the association between motor vehicle emissions and respiratory health. The association between exposure to lead from motor vehicle emissions and lead toxicity is certainly analogous. Studies have demonstrated elevated blood lead levels in traffic workers and shopkeepers near busy streets (Orlando et al., 1994), as well as elevated milk lead levels in cows near busy streets (Bhatia and Choudhri, 1996).

**ASSessment of Exposure to Motor Vehicle Emissions**

The exposure assessment model used in this study represents a refinement over many of the methods used in previous studies. One area of refinement is the ability to incorporate the effects of emissions and dispersion into a single measure of exposure. This has been accomplished by only one other study in this area that used a considerably more complex model (Oosterlee et al., 1996). Other studies
in this area have either accounted for only the effect of emission (Waldron et al., 1995; Weiland et al., 1994; Wjst et al., 1993) or dispersion (Livingstone et al., 1996; Nitta et al., 1993), or attempted to account for both in an ad hoc manner (Brunekreef et al., 1997; Edwards et al., 1994). The ability to consider the contribution of more than one street to exposure is also a refinement over methods used in previous studies. With the exception of the other ecologic study in this area (English, 1997), all previous studies consider the contribution of only the one closest street to exposure. This could lead to an underestimation of exposure in areas with busy streets located in close proximity. Use of geographic information system (GIS) to implement the exposure assessment model makes it feasible to consider exposure from more than one street by automating much of the work performed manually in other studies. Automation of this process also reduces the amount of error associated with the manual estimation of residential proximity to street performed in some studies (Brunekreef et al., 1997; Edwards et al., 1994; Nitta et al., 1993). Finally, this model could easily be modified for use in individual-level studies. Assessment of exposure could be performed for any number of smaller areas or point locations to arrive at estimates of exposure for individuals in different environments.

Despite these improvements, the validity of the model has not been established. Validation studies are not feasible in this thesis, but analyses are performed to examine the sensitivity of the results to methodological choices. One sensitivity analysis examines the method of weighting emission values transferred to enumeration areas (EA) from the street network. Weighting is intended to account for the influence of the shape and size of the EA on exposure. The analysis reveals that the results of the thesis are insensitive to the choice of weight, and this implies that EA street frontage has a greater impact on exposure than EA dimensions. Another sensitivity analysis examines the impact of modelling exposure to traffic volume as opposed to emissions of PM. The
strength of association with respiratory subset admission rates is similar for both measures of exposure. One interpretation of this finding is that exposure to a mix of pollutants is responsible for the effect on respiratory rates. This interpretation is supported by the observation that exposure to individual pollutants (Brunekreef et al., 1997; Nitta et al., 1993; Oosterlee et al., 1996) and traffic volume in general (Edwards et al., 1994; Weiland et al., 1994; Wjst et al., 1993) are both associated with respiratory health. Another interpretation is that a confounding variable associated with proximity to the street is responsible for the association between traffic volume and respiratory health. This is not as likely given that a number of studies have found an association between respiratory health and traffic volume, independent of distance to the street (Brunekreef et al., 1997; Edwards et al., 1994; Weiland et al., 1994; Wjst et al., 1993).

Limitations of the exposure assessment process are due mainly to the lack of available data. No data are available on the time individuals spend at locations away from their primary residences (e.g. work, school), or the levels of exposure at those locations. Individual-level data is also not available on the time lived at the current residence. Consequently, individuals are assumed to receive the level of exposure assessed for their home address for the entire study. This assumption leads to misclassification of exposure, but it is not possible to estimate the nature or magnitude of the misclassification. However, it appears that individuals spend the majority of their time at home (Moschandreas, 1981), so the misclassification is probably less than if exposure were assessed at any other single location. Lack of data about the distribution of sources of indoor particles (e.g. smoking, gas stoves) also limits the exposure assessment process. In the absence of this data, exposure to particles from indoor sources is assumed constant within and between EAs. This is almost certainly not a valid assumption, but again, there is no way to assess the error resulting from this assumption. Another limitation results from the vehicle count data available for SETO. The available data do not
allow temporal fluctuations in exposure to be modelled, and consequently exposure is assessed from average daily levels of traffic. It is possible that exposure to peak traffic volumes poses greater risk than exposure to average traffic volumes. However, this does not appear to be the case as the only study to assess exposure to peak traffic volumes finds no association with respiratory health (Livingstone et al., 1996). Detailed data on building characteristics in the study area are also not available. Building characteristics affect that ability of outdoor air to penetrate into the building, and therefore can influence exposure to motor vehicle emission. The height of buildings is one characteristic that is variable within SETO, and several EAs with large regression residuals contain high-rise buildings. Variable building height may result in within-EA misclassification, but does not appear to be a source of ecologic confounding as a variable representing the proportion of high-rise buildings in an EA does not improve model fit. The lack of data on meteorological conditions at street level is also a potential limitation of the exposure model. Wind speed and direction can influence the dispersion of emissions, but their contribution to long term averages in dispersion are limited within urban areas (Glen et al., 1996; Kono and Ito, 1990).

LIMITATIONS OF THE STUDY

Limitations of the study are discussed under several headings. The problems of ecologic bias and confounder control are discussed first. Within-group misclassification and limitations associated with data availability are discussed next. Finally, issues of model fit and the influence of spatial distributions are discussed last.
Ecologic Bias and Confounding

The failure of ecologic effect estimates to reflect the biologic effect at the individual level is known as ecologic bias (Greenland and Morgenstern, 1989; Greenland and Robins, 1994). In this study, EA exposure is estimated to have an effect on the EA rate of respiratory hospitalization. The ecologic effect estimate is an unbiased predictor of the individual effect estimate if and only if the EA exposure level has no effect on the EA hospitalization rate given an individual’s exposure (Walter, 1991). In other words, ecologic bias is not present if individual exposure-specific hospitalization rates do not vary across EAs (Greenland and Robins, 1994). Three conditions can result in ecologic bias - within-EA bias, confounding by EA, and effect modification by EA. The first condition may bias effect estimates in any type of study and the latter two are unique to ecologic studies. For this reason, confounding by EA and effect modification by EA are sources of ‘cross-level bias’ (Morgenstern, 1998).

Within-EA bias can result from confounding, selection methods or misclassification that bias the exposure effect within EAs. Differing degrees of within-EA bias may lead to variable individual-specific hospitalization rates across EAs, and ecologic bias. It is probable that bias from confounding and exposure misclassification is variable within EAs, but without individual-level data, it is impossible to determine the magnitude and direction of within-EA bias.

Confounding by EA may occur if the background rate of hospitalization in the unexposed population is associated with exposure across EAs. This can result from the differential distribution of a risk factor between groups (e.g. SES). However, it is important to note that the differential distribution of a risk factor within an EA may not result in confounding by EA if the risk factor is not associated with exposure across EAs (Morgenstern, 1998). It may be possible to control for confounding by EA if an adequate measure of the confounder exists at the EA level. In this study, an attempt is made to control for confounding by
age and SES by standardizing rates for the effect of age and including an aggregate measure of SES in the regression analysis as a covariate. Control for multiple confounders using rate standardization can bias effect estimates (Rosenbaum and Rubin, 1984), but a repeat analysis with crude rates and age as a covariate suggests that there is no bias from this approach in the original analysis. Data on the background rate of disease in EAs are not available, so it is not possible to determine if confounding by EA contributes to ecologic bias in this study. Nevertheless, correlations between expected rates and exposure are examined in a crude attempt to assess if background rates vary across EA in a manner that is associated with exposure. Expected rates of hospitalization for all diagnostic sets are correlated with exposure and SES across EAs. This suggests, but does not confirm that confounding by EA may be a source of ecologic bias in this study.

Effect modification by EA can lead to ecologic bias if the rate difference for the exposure effect varies across EAs. Variation in access to hospitalization and use of respiratory medication may modify the effect of exposure across EAs. As with other potential sources of ecologic bias, the presence of effect modification by EA cannot be assessed without individual-level data. This inability to assess the influence of ecologic bias on the ecologic effect estimate is the major limitation of ecologic analysis for making causal inferences (Greenland and Morgenstern, 1989; Greenland and Robins, 1994).

The inability to assess the magnitude and direction of the bias demands that steps be taken to limit bias in the design of ecologic studies and the analysis of ecologic data (Nurminen, 1995). A number of such steps are taken in this study. To begin with, the smallest possible sampling unit is chosen for the study. In addition, weighted regression is used and potentially confounding covariates are included in the model. Finally, an influence analysis is conducted to demonstrate that the
deletion of influential EAs with large residuals does not alter the outcome of the analysis.

*Within-Group Misclassification*

One of the crucial assumptions made in this study is that the exposure level assigned to an EA applies to all individuals in the EA. Despite the use of the smallest possible study unit, this assumption is almost certainly not valid. It is difficult to estimate the resulting magnitude of exposure misclassification, but several factors may contribute to exposure misclassification as discussed in relation to the exposure assessment model. The principles for the interpretation of results from individual-level studies with misclassification bias do not apply to ecologic analyses (Morgenstern, 1998). Specifically, the principle that nondifferential misclassification of exposure nearly always leads to bias towards the null does not hold for ecologic studies (Brenner et al., 1992b). It is quite possible that within-EA bias from exposure misclassification has led to bias of the ecologic effect measure. As discussed previously, there is no way of determining if this is the case with the data at hand. Misclassification of socioeconomic status (SES) is likely present as well, but may not be as serious a problem. It appears that nondifferential misclassification of a confounder does not affect the ability to control for that confounder, provided there is no cross-level bias (Brenner et al., 1992a).

*Data Issues*

As with most ecologic studies, an important limitation is the availability of data (Walter, 1991). This limitation is exacerbated by the choice of a small study unit for the thesis. In ecologic studies, the issue of data availability is in direct conflict with the need to minimize within study unit misclassification. A decision is made to limit misclassification at the expense of data availability in this study by using
the EA as the study unit. Census derived denominator and SES data at the EA level are limited by the practice of ‘data suppression’, which sets counts to zero for an EA if confidentiality is compromised or the response rate is low (Statistics Canada, 1993). Because of data suppression, almost 10% of the EAs in SETO do not have population counts available, and SES data are not available for an even larger proportion of EAs (14% to 71%). It is difficult to assess the effect of missing EAs on the results, but the higher mean level of exposure in suppressed EAs implies that they may differ from other EAs in systematic manner. In addition, the number EAs with missing SES values limits the choice of an appropriate measure, and makes the construction of a SES index impractical. This may impact the ability to control for confounding from SES. From a spatial perspective, simulations suggest that deletion of suppressed EAs does not affect tests of spatial autocorrelation. Consideration is given to estimating suppressed population counts and SES values through areal interpolation (Flowerdew and Green, 1994), but interpolation is not used as it is not clear that this will reduce the overall error in the effect estimate.

In addition to limiting the availability of data, use of a small study unit also affects data quality. Census data at the EA-level are affected by the practice of ‘random rounding’, which rounds all counts to multiples of 5 or 10 (Statistics Canada, 1993). With a median EA population of 400, this practice can considerably affect the quality of data, especially when counts are stratified. The small size of the study unit also affects the stability of admission rates. Attempts are made to improve stability by aggregating admission counts over three years, but this still leaves 25% of the EAs with no admissions for the respiratory subset. Small counts of disease also preclude the analysis of subgroups. This is unfortunate, as previous studies have seen stronger effects in subgroups (Brunekreef et al., 1997; Oosterlee et al., 1996). Despite problems with stability of rates, the quality of admission data is thought to be good relative to other sources of health data,
especially those available at the level of the EA. Bias from diagnostic error is minimized by using a subset of similar respiratory diagnoses (Motheral and Fairman, 1997; Wray et al., 1995), but it is acknowledged that bias may exist from limitations of the ICD-9 coding system around case-mix and severity (Pink and Bolley, 1994; Taylor, 1998). Ambulatory data are also available at the EA level, but problems with completeness and diagnostic accuracy make use of these data impractical for the thesis. The prevalence of respiratory symptoms is routinely surveyed, but survey data are not available at the level of the EA. This precludes examination of the association observed in previous studies between exposure and respiratory symptoms (Brunkreef et al., 1997; Nitta et al., 1993; Oosterlee et al., 1996; Weiland et al., 1994; Wjst et al., 1993).

Data on traffic counts and vehicle type distributions are not available for some parts of SETO. However, count data are available for all major streets, and problems with selection bias among smaller streets are avoided by only modelling streets with over 5000 vehicles per day. The range of exposure values encountered in this study suggests that the decision to model a subset of streets results in sufficient exposure differences to detect an effect. Data on vehicle type distribution are only available at a few locations in the study area, and use of the average provincial distribution for the remaining areas may be a source of bias. However, the similarity between vehicle distributions in SETO and the province suggest that this is not a large problem. The quality of traffic count and vehicle type distribution data also does not appear to be a major source of bias as estimates of the validity of these data are high (Lasagna, 1997).

*Model Fit and Spatial Issues*

The fit of the data to a Poisson model is generally not good. Poor fit to the model suggests that the effect estimates should be interpreted cautiously. Model fit is not reported for any of the previous studies in this area that fit logistic regression
models to individual-level data. Lack of fit may reflect a violation of the assumption that the residual error is Poisson distributed (Greenland, 1998). The distribution of admissions by EA poorly approximates the Poisson distribution for all diagnostic sets due to the large number of EAs with no admissions. This lack of fit can be addressed in a number of ways. One approach is to model the error structure as a combination of distributions (Clayton et al., 1993). Another approach is to calculate empirical Bayes estimates (EBE) of admission rates (Langford, 1994). This approach was examined for mapping purposes, but the complexity of using EBE rates in a regression analysis is beyond the scope of the thesis. Larger study units or longer times could also be used to accrue higher counts in study units. Unfortunately, these approaches exacerbate the problems of exposure misclassification and temporal ambiguity.

Another possible reason for the lack of fit to the Poisson model is spatial autocorrelation among variables. Moderate positive spatial autocorrelation is seen for the exposure values and SES values, but not for hospitalization rates or regression residuals. This suggests that spatial autocorrelation does not contribute to poor model fit. However, spatial autocorrelation is only assessed at one spatial lag, that of the adjacent EA. Inspection of maps of hospitalization rates suggests that EAs with high rates tend to cluster in neighborhoods, but do not always have adjacent borders. A similar distribution is seen for regression residuals. The presence of spatial autocorrelation at different lags could be assessed by calculating Moran's I with connectivity matrices for a number of spatial lags then plotting spatial autocorrelation by spatial lag (Bailey and Gatrell, 1995). Clustering of high rates and residuals in neighborhoods suggests that there is an unmeasured risk factor differentially distributed by neighborhood. This analysis is not done for the thesis due to the complexity of calculating adjacency matrices at different spatial lags.
Chapter Five

CONCLUSIONS
The results of the study identify a weak ecologic association between exposure to PM 2.5 from motor vehicle emissions and the rate of hospitalization for selected respiratory diagnoses. This association is not likely to have occurred by chance, but it may have been influenced by incompletely controlled confounders. Furthermore, the lack of specificity of this association, and the poor fit of the data to the analytic model raise the possibility that the estimated measure of association and measure of effect may be biased.

If the study is examined in isolation, it is not possible to conclude that the observed ecologic association exists at the level of the individual. In the absence of individual level data, the magnitude and direction of ecologic bias cannot be estimated, and it is not possible to determine the temporal relationship of exposure and outcome. However, if this study is examined together with previous studies, the consistency of the results with individual-level studies is apparent. Given this consistency it is plausible that the ecologic association in the current study also exists among individuals.

In summary, it can be concluded that the weak association observed in this study between motor vehicle emissions and respiratory health is consistent with the majority of previous studies. This consistency adds to the body of literature supporting an association, but suggests that future studies need to carefully address the possible influence of bias from confounding. It can also be concluded from this study that motor vehicle emission exposure is not specifically associated with respiratory health among small areas. This lack of specificity has been noted in one previous study, and must be examined in future studies to elucidate the role of exposure to motor vehicle emissions in the causation of respiratory illness.
ASSESSMENT OF EXPOSURE TO MOTOR VEHICLE EMISSIONS

The lack of a valid and straightforward method of assessing exposure to motor vehicle emissions presents the greatest methodological obstacle to studies of the relationship between motor vehicle emissions and respiratory health. The exposure assessment model developed in this study represents an improvement over previous methods in a number of ways, but several improvements are still possible. Specifically, the results of the model are not validated and the use of the model in area level studies likely results in considerable misclassification. In a broader sense, development of this exposure assessment model has identified a number of assumptions and limitation of existing methods, and in doing so, identified issues to be addressed by future studies.

STUDY DESIGN

The drawbacks of an ecologic study design appear to outweigh the practical advantages of this design for studies of the association between exposure to motor vehicle emission and respiratory health. Notably, there does not appear to be a study unit that can provide a compromise between data availability and exposure misclassification. Furthermore, the current body of knowledge in this area and the limited contribution of ecologic results to causal and biological inference suggest that future studies should be at the individual level, or incorporate individual level data into a multi-level analysis. An individual-level study with an improved exposure assessment model may be feasible in South East Toronto.

A simple analysis of spatial distributions suggests that spatial autocorrelation of variables does not bias the effect estimate in this study. However, given the
observed spatial autocorrelation of variables, consideration of the spatial distribution of variables is warranted in future studies.
TABLES
Table 1  Summary of Studies Examining the Relationship between Motor Vehicle Emissions and Respiratory Health

<table>
<thead>
<tr>
<th>Author and Year of Publication</th>
<th>Study Design and Population</th>
<th>Measurement of Health</th>
<th>Assessment of Exposure Traffic Volume</th>
<th>Distance to Street</th>
<th>Measurement of Confounding Variables</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Edwards et al., 1994)</td>
<td>Case control 0 to 4 years</td>
<td>Hospital admission for asthma (case), hospital and community controls</td>
<td>Categorical</td>
<td>Home to major street in district</td>
<td>None</td>
<td>Association with Volume and Distance</td>
</tr>
<tr>
<td>(Livingstone et al., 1996)</td>
<td>Case control 2 to 64 years</td>
<td>Outpatient asthma diagnosis and medication (case), practice based controls</td>
<td>No</td>
<td>Home to nearest street with &gt; 1000 vehicles per hour</td>
<td>Demographic, ecologic SES, smoking</td>
<td>No association with Distance</td>
</tr>
<tr>
<td>(Oosterlee et al., 1996)</td>
<td>Cross sectional All ages</td>
<td>Self-reported respiratory symptoms</td>
<td>CAR model</td>
<td></td>
<td>Demographic, SES, smoking, duration of residence, IAQ</td>
<td>Association in children, stronger in females</td>
</tr>
<tr>
<td>(Brunekeef et al., 1997; van Vliet et al., 1997)</td>
<td>Cross sectional 7 to 12 years</td>
<td>Self-reported respiratory symptoms and PFTs</td>
<td>Continuous (truck and car)</td>
<td>Home to closest highway</td>
<td>Demographic, SES, smoking, IAQ</td>
<td>Association with Volume and Distance for both symptoms and PFTs, stronger in females</td>
</tr>
<tr>
<td>(Nitta et al., 1993)</td>
<td>Cross sectional Females &gt;40 years</td>
<td>Self-reported respiratory symptoms</td>
<td>No</td>
<td>Home to closest major road</td>
<td>Demographic, SES, smoking, duration of residence, IAQ</td>
<td>Association with Distance</td>
</tr>
<tr>
<td>(Wjst et al., 1993)</td>
<td>Cross sectional 9 to 11 years</td>
<td>Self-reported respiratory symptoms and PFTs</td>
<td>Continuous</td>
<td>No</td>
<td>Demographic, SES, smoking, duration of residence, IAQ</td>
<td>Association with Volume for both symptoms and PFTs</td>
</tr>
<tr>
<td>(Waldron et al., 1995)</td>
<td>Cross sectional 13 to 14 years</td>
<td>Self-reported respiratory symptoms</td>
<td>Binary (Presence of highway in ward)</td>
<td>No</td>
<td>None</td>
<td>No association with Volume</td>
</tr>
<tr>
<td>(Weiland et al., 1994)</td>
<td>Cross sectional 12 to 15 years</td>
<td>Self-reported respiratory symptoms</td>
<td>Categorical (Self-reported)</td>
<td>No</td>
<td>Demographic, SES, smoking, IAQ</td>
<td>Association with Volume</td>
</tr>
<tr>
<td>(Buckeridge et al., 1998)</td>
<td>Ecologic All ages</td>
<td>Age standardized respiratory admission rates</td>
<td>GIS model</td>
<td></td>
<td>SES</td>
<td>Association with modelled exposure</td>
</tr>
<tr>
<td>(English, 1997)</td>
<td>Ecologic Children</td>
<td>Age standardized asthma admission rates</td>
<td>GIS model</td>
<td></td>
<td>SES</td>
<td>Association with modelled exposure</td>
</tr>
</tbody>
</table>

*SES – socioeconomic status; *CAR – calculation of air pollution by road traffic; *PFT – pulmonary function tests; *IAQ – indoor air quality; *GIS – geographic information system.
Table 2  Selected Educational, Employment, Income, Mobility and Social Indicators for Neighborhoods in Southeast Toronto

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Educational Attainment</th>
<th>Employment</th>
<th>Income</th>
<th>Social Mobility</th>
<th>Social Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population</td>
<td>Less than Grade 9 (%)</td>
<td>University Degree (%)</td>
<td>Unemployment Rate</td>
<td>Female Labor Force Participation Rate</td>
</tr>
<tr>
<td>Toronto</td>
<td>635127</td>
<td>14.6</td>
<td>25.7</td>
<td>9.7</td>
<td>64.7</td>
</tr>
<tr>
<td>Rosedale</td>
<td>7495</td>
<td>1.2</td>
<td>50.7</td>
<td>5.0</td>
<td>64.0</td>
</tr>
<tr>
<td>Yorkville</td>
<td>1885</td>
<td>3.7</td>
<td>46.6</td>
<td><strong>4.3</strong></td>
<td>65.0</td>
</tr>
<tr>
<td>Church</td>
<td>15195</td>
<td>3.9</td>
<td>33.3</td>
<td>8.2</td>
<td>78.8</td>
</tr>
<tr>
<td>St James Town</td>
<td>22355</td>
<td>10.6</td>
<td>19.8</td>
<td>13.0</td>
<td>65.6</td>
</tr>
<tr>
<td>Cabbagetown</td>
<td>4410</td>
<td>7.4</td>
<td>39.3</td>
<td>6.4</td>
<td>73.2</td>
</tr>
<tr>
<td>Moss Park</td>
<td>9730</td>
<td>15.5</td>
<td>19.4</td>
<td><strong>15.5</strong></td>
<td>50.4</td>
</tr>
<tr>
<td>Regent Park</td>
<td>10620</td>
<td><strong>25.6</strong></td>
<td>6.2</td>
<td>22.7</td>
<td>40.2</td>
</tr>
<tr>
<td>St Lawrence</td>
<td>4670</td>
<td>4.0</td>
<td>29.7</td>
<td>9.4</td>
<td>74.4</td>
</tr>
<tr>
<td>South Riverdale</td>
<td>24860</td>
<td>19.4</td>
<td>15.2</td>
<td>13.4</td>
<td>61.7</td>
</tr>
<tr>
<td>North Riverdale</td>
<td>20655</td>
<td>11.8</td>
<td>29.3</td>
<td>7.6</td>
<td>69.8</td>
</tr>
</tbody>
</table>

Items in bold face are significantly different from the City of Toronto average by a magnitude of at least 50%.
Table 3  Benefits and Limitations of Using Administrative Databases for Epidemiological Research

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior collection of data minimized resources needed for study</td>
<td>Prior collection of data may result in selection bias</td>
</tr>
<tr>
<td>Data can be linked to other sources of information</td>
<td>Database usually does not contain all data required for study</td>
</tr>
<tr>
<td>Sample tends to be representative and complete, especially if database associated with payment</td>
<td>Quality of data affected by collection methods, including issues around diagnostic coding</td>
</tr>
<tr>
<td>Provides large sample size</td>
<td>Methodological rigor is required for analyses involving large sample size</td>
</tr>
</tbody>
</table>

Adapted from (Motheral and Fairman, 1997; Paul et al., 1993; Sorenson et al., 1996; Wray et al., 1995)
Table 4  Diagnoses and Associated ICD-9 Codes Used to Abstract Records for Study

<table>
<thead>
<tr>
<th>Diagnostic Set</th>
<th>Specific Diagnosis</th>
<th>ICD-9 Codes</th>
<th>Individuals</th>
<th>Admissions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asthma</td>
<td>493.0-1,493.9</td>
<td>430</td>
<td>642</td>
</tr>
<tr>
<td></td>
<td>Bronchitis</td>
<td>466.0-1,490</td>
<td>127</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>COPD</td>
<td>491.0-2, 491.8-9, 492, 496</td>
<td>238</td>
<td>411</td>
</tr>
<tr>
<td>Respiratory</td>
<td>Pneumonia</td>
<td>480.0-2, 480.8-9, 481, 482.0-4, 482.8-9, 483, 485, 486, 514</td>
<td>709</td>
<td>834</td>
</tr>
<tr>
<td>Subset</td>
<td>URI</td>
<td>461.0-3, 461.8-9, 462, 464.4, 465.0, 465.8-9, 472.0-2, 473.0-3, 473.8-9, 478.1-3, 478.7-9</td>
<td>275</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>1779</td>
<td>2322</td>
</tr>
<tr>
<td></td>
<td>Respiratory Chapter</td>
<td>460 – 519.9</td>
<td>2646</td>
<td>3316</td>
</tr>
<tr>
<td></td>
<td>GI Disorders</td>
<td>520 – 579.9</td>
<td>2669</td>
<td>4186</td>
</tr>
<tr>
<td></td>
<td>GU Chapter</td>
<td>580 – 629.9</td>
<td>2406</td>
<td>2669</td>
</tr>
</tbody>
</table>

†ICD-9 – International Classification of Diseases, 9th revision; ‡COPD – chronic obstructive pulmonary disease; URI – upper respiratory tract infection; †GI – gastrointestinal; ‡GU – genitourinary.
Table 5  Particulate Matter Emission Factors from PART5 Model and Distribution of Motor Vehicle Types

<table>
<thead>
<tr>
<th>Vehicle Type</th>
<th>Emission Factors from PART5 Model (grams per kilometer)</th>
<th>Vehicle Type Distribution (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PM 2.5&lt;sup&gt;†&lt;/sup&gt;</td>
<td>PM 10.0&lt;sup&gt;†&lt;/sup&gt;</td>
</tr>
<tr>
<td>Light Duty Gasoline Vehicle</td>
<td>0.0772</td>
<td>0.1014</td>
</tr>
<tr>
<td>Light Duty Gasoline Truck Class 1</td>
<td>0.0981</td>
<td>0.1239</td>
</tr>
<tr>
<td>Light Duty Gasoline Truck Class 2</td>
<td>0.1609</td>
<td>0.2140</td>
</tr>
<tr>
<td>Heavy Duty Gasoline Vehicle</td>
<td>0.3427</td>
<td>0.4634</td>
</tr>
<tr>
<td>Motorcycle</td>
<td>0.0483</td>
<td>0.0740</td>
</tr>
<tr>
<td>Light Duty Diesel Vehicle</td>
<td>0.8174</td>
<td>0.8850</td>
</tr>
<tr>
<td>Light Duty Diesel Truck</td>
<td>0.9043</td>
<td>0.9767</td>
</tr>
<tr>
<td>Class 2B Heavy Duty Diesel Vehicle</td>
<td>1.2357</td>
<td>1.3178</td>
</tr>
<tr>
<td>Heavy Duty Diesel Vehicle</td>
<td>3.4223</td>
<td>3.6556</td>
</tr>
<tr>
<td>Buses</td>
<td>4.0032</td>
<td>4.2622</td>
</tr>
</tbody>
</table>

<sup>†</sup>PM 2.5 / PM 10 – particulate matter smaller than 2.5 / 10 microns in diameter; <sup>§</sup>SETO – South East Toronto; <sup>†</sup>Provincial figures report these vehicle classes together with Heavy Duty Diesel Vehicle as one class.
### Table 6  
Data Reduction for HMRI\(^\dagger\) Data

<table>
<thead>
<tr>
<th>Diagnostic Set</th>
<th>Respiratory Subset</th>
<th>Respiratory Chapter</th>
<th>Gastrointestinal Chapter</th>
<th>Genitourinary Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Records by Year</td>
<td>4470</td>
<td>5002</td>
<td>4872</td>
<td>7172</td>
</tr>
<tr>
<td>Total Records</td>
<td>14344</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records Matching to Ontario EA(^\ddagger) (% Δ from above)</td>
<td>14087 (1.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records in SETO(^\S) Area</td>
<td>2596</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records with Population for EA (% Δ from above)</td>
<td>2495 (3.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records without admission or birth date errors (% Δ from above)</td>
<td>2454 (1.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records with valid health # (% Δ from above)</td>
<td>2322 (5.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Records with unique health # (% Δ from above)</td>
<td>1779 (23.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{\dagger}\)HMRI – Hospital Medical Records Institute; \(^{\ddagger}\)EA – enumeration area; \(^{\S}\)SETO – South East Toronto.
Table 7  Observed and Poisson Distributions of Hospital Admissions by Individual

<table>
<thead>
<tr>
<th>Number of Admissions</th>
<th>Respiratory Subset</th>
<th>Respiratory Chapter</th>
<th>Gastrointestinal Chapter</th>
<th>Genitourinary Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Poisson</td>
<td>Observed</td>
<td>Poisson</td>
</tr>
<tr>
<td></td>
<td>( \mu ) 0.01461</td>
<td>( \sigma^2 ) 0.03812</td>
<td>( \mu ) 0.02172</td>
<td>( \sigma^2 ) 0.04878</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2 / \mu ) 2.61</td>
<td>( \sigma^2 / \mu ) 2.24</td>
<td>( \sigma^2 / \mu ) 2.07</td>
<td>( \sigma^2 / \mu ) 1.62</td>
</tr>
<tr>
<td>0</td>
<td>120095</td>
<td>120108</td>
<td>119228</td>
<td>119257</td>
</tr>
<tr>
<td>1</td>
<td>1488</td>
<td>1754</td>
<td>2255</td>
<td>2590</td>
</tr>
<tr>
<td>2</td>
<td>182</td>
<td>13</td>
<td>264</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>57</td>
<td>0</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>0</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 8  Summary Statistics for Admission Counts, Crude Admission Rates and Indirectly Standardized Admission Rates by Enumeration Area

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>All</td>
<td>First</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EA</strong>&lt;sup&gt;†&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission Counts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory Subset</td>
<td>4</td>
<td>5</td>
<td>5.9</td>
<td>7.7</td>
</tr>
<tr>
<td>Respiratory Chapter</td>
<td>6</td>
<td>8</td>
<td>8.8</td>
<td>11.0</td>
</tr>
<tr>
<td>Gastrointestinal Chapter</td>
<td>10</td>
<td>12</td>
<td>11.9</td>
<td>13.9</td>
</tr>
<tr>
<td>Genitourinary Chapter</td>
<td>7</td>
<td>8</td>
<td>8.0</td>
<td>8.8</td>
</tr>
<tr>
<td><strong>Crude EA</strong>&lt;sup&gt;†&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission Rates (per 1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory Subset</td>
<td>3.4</td>
<td>4.0</td>
<td>5.8</td>
<td>7.4</td>
</tr>
<tr>
<td>Respiratory Chapter</td>
<td>5.3</td>
<td>6.3</td>
<td>8.4</td>
<td>10.4</td>
</tr>
<tr>
<td>Gastrointestinal Chapter</td>
<td>8.6</td>
<td>9.6</td>
<td>11.9</td>
<td>13.9</td>
</tr>
<tr>
<td>Genitourinary Chapter</td>
<td>5.8</td>
<td>6.6</td>
<td>7.5</td>
<td>8.3</td>
</tr>
<tr>
<td><strong>EA ISR</strong>&lt;sup&gt;‡&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(per 1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory Subset</td>
<td>3.8</td>
<td>4.5</td>
<td>5.4</td>
<td>7.0</td>
</tr>
<tr>
<td>Respiratory Chapter</td>
<td>5.7</td>
<td>6.7</td>
<td>8.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Gastrointestinal Chapter</td>
<td>8.7</td>
<td>10.0</td>
<td>11.4</td>
<td>13.7</td>
</tr>
<tr>
<td>Genitourinary Chapter</td>
<td>6.0</td>
<td>6.4</td>
<td>7.8</td>
<td>8.2</td>
</tr>
</tbody>
</table>

<sup>†</sup>EA – Enumeration Area; <sup>‡</sup>ISR – Indirectly Standardized Rate
Table 9  Summary Statistics for Exposure to Traffic Volume and PM 2.5 Emissions

<table>
<thead>
<tr>
<th></th>
<th>Traffic Volume Exposure (vehicles per 24 hours)</th>
<th>PM 2.5(^\dagger) Exposure (grams per 24 hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Number of EAs(^\dagger)</td>
<td>302</td>
<td>32</td>
</tr>
<tr>
<td>Maximum</td>
<td>4179.6</td>
<td>681.7</td>
</tr>
<tr>
<td>75(^{th}) Percentile</td>
<td>166.8</td>
<td>195.7</td>
</tr>
<tr>
<td>Median</td>
<td>57.4</td>
<td>118.9</td>
</tr>
<tr>
<td>25(^{th}) Percentile</td>
<td>8.9</td>
<td>9.3</td>
</tr>
<tr>
<td>EAs with Zero Values # (%)</td>
<td>63 (20.9)</td>
<td>3 (9.4)</td>
</tr>
<tr>
<td>Mean</td>
<td>180.0</td>
<td>160.9</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>450.4</td>
<td>191.4</td>
</tr>
</tbody>
</table>

\(^\dagger\) PM 2.5 – particulate matter smaller than 2.5 microns in diameter; \(^*\)EA – Enumeration Area
Table 10  Census Variables Examined for SES Measurement

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Description</th>
<th>EAs(^{+}) with Non-Zero Value Number (%)</th>
<th>Correlation with Outcome(^{+}) Coefficient (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dwelling Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dwelling Value</td>
<td>Average value of dwellings in $</td>
<td>159 (52.6)</td>
<td>-0.014 (0.862)</td>
</tr>
<tr>
<td><strong>Educational Attainment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 9</td>
<td>% of population who have less than Grade 9 education</td>
<td>225 (74.5)</td>
<td>0.158 (0.018)</td>
</tr>
<tr>
<td>University</td>
<td>% of population who have a university degree</td>
<td>273 (90.4)</td>
<td>-0.226 (0.001)</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>% of population &gt; 25 unemployed</td>
<td>248 (82.1)</td>
<td>0.236 (0.001)</td>
</tr>
<tr>
<td>Female Labor Force</td>
<td>% of female population &gt; 25 participating in the labor force</td>
<td>286 (94.7)</td>
<td>-0.175 (0.003)</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Income</td>
<td>Average income of households</td>
<td>209 (69.2)</td>
<td>-0.086 (0.216)</td>
</tr>
<tr>
<td>Low Income Families</td>
<td>% of families with low income</td>
<td>97 (32.1)</td>
<td>0.194 (0.057)</td>
</tr>
<tr>
<td>Low Income Persons</td>
<td>% of persons with low income</td>
<td>111 (36.8)</td>
<td>0.190 (0.046)</td>
</tr>
<tr>
<td><strong>Mobility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobility</td>
<td>% of population that moved in the previous year</td>
<td>285 (94.4)</td>
<td>-0.009 (0.885)</td>
</tr>
<tr>
<td><strong>Social Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lone Parents</td>
<td>% of families with children headed by a lone-parent</td>
<td>252 (83.4)</td>
<td>0.162 (0.010)</td>
</tr>
<tr>
<td>Female Lone Parents</td>
<td>% of families with children headed by a female lone-parent</td>
<td>243 (80.5)</td>
<td>0.164 (0.011)</td>
</tr>
<tr>
<td>Immigration</td>
<td>% of population that immigrated to Canada (from 1988-1991)</td>
<td>206 (68.2)</td>
<td>-0.001 (0.987)</td>
</tr>
</tbody>
</table>

\(^{+}\)EA – enumeration area; \(^{+}\)Values of rank correlation with standardized first respiratory subset hospitalization rate.
<table>
<thead>
<tr>
<th></th>
<th>Standardized Admission Rates</th>
<th>Exposure</th>
<th>SES†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Resp‡ Subset</td>
<td>Resp‡ Chapter</td>
<td>GI‡ Chapter</td>
</tr>
<tr>
<td>Respiratory Subset</td>
<td>1</td>
<td>0.949 (0.001)</td>
<td>0.745 (0.001)</td>
</tr>
<tr>
<td>Respiratory Chapter</td>
<td>0.949 (0.001)</td>
<td>1</td>
<td>0.778 (0.001)</td>
</tr>
<tr>
<td>GI Chapter</td>
<td>0.745 (0.001)</td>
<td>0.778 (0.001)</td>
<td>1</td>
</tr>
<tr>
<td>GU Chapter</td>
<td>0.740 (0.001)</td>
<td>0.784 (0.001)</td>
<td>0.791 (0.001)</td>
</tr>
<tr>
<td>Traffic Volume</td>
<td>0.188 (0.001)</td>
<td>0.181 (0.002)</td>
<td>0.160 (0.005)</td>
</tr>
<tr>
<td>PM 2.5</td>
<td>0.222 (0.001)</td>
<td>0.206 (0.001)</td>
<td>0.177 (0.002)</td>
</tr>
<tr>
<td>University Graduation</td>
<td>-0.226 (0.001)</td>
<td>-0.184 (0.160)</td>
<td>-0.085 (0.101)</td>
</tr>
</tbody>
</table>

†SES – socioeconomic status; ‡Resp – respiratory; §GI – gastrointestinal; §GU – genitourinary; §PM 2.5 – particulate matter smaller than 2.5 microns in diameter.
### Table 12  Results of Bivariate and Multivariate Poisson Regression Analyses to Address Hypothesis

<table>
<thead>
<tr>
<th>Model Deviance / Degrees of Freedom</th>
<th>Parameter Name</th>
<th>Parameter $\chi^2$</th>
<th>Parameter p-value</th>
<th>Parameter Level</th>
<th>Relative Risk(^*)</th>
<th>95% CI(^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.154  PM 2.5(^1)</td>
<td>17.26</td>
<td>0.002</td>
<td>2</td>
<td>1.05</td>
<td>0.88 to 1.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1.15</td>
<td>0.96 to 1.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1.30</td>
<td>1.10 to 1.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1.28</td>
<td>1.09 to 1.50</td>
</tr>
<tr>
<td>4.837  PM 2.5</td>
<td>22.66</td>
<td>0.001</td>
<td>2</td>
<td>1.10</td>
<td>0.92 to 1.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1.24</td>
<td>1.04 to 1.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1.34</td>
<td>1.14 to 1.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1.39</td>
<td>1.19 to 1.63</td>
</tr>
<tr>
<td>University</td>
<td>89.80</td>
<td>0.001</td>
<td>1 %</td>
<td>0.82</td>
<td>0.78 to 0.85</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Risk is relative to lowest quintile for PM 2.5 and relative to a 1% increase for university graduation; \(^*\)CI – confidence interval; \(^\dagger\)PM 2.5 – particulate matter smaller than 2.5 microns in diameter.
<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Model Deviance / Degrees of Freedom</th>
<th>Parameter Name</th>
<th>Parameter $\chi^2$</th>
<th>Parameter p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastrointestinal Chapter</td>
<td>7.806</td>
<td>PM 2.5</td>
<td>25.11</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>22.29</td>
<td>0.001</td>
</tr>
<tr>
<td>Genitourinary Chapter</td>
<td>5.316</td>
<td>PM 2.5</td>
<td>33.11</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>26.69</td>
<td>0.001</td>
</tr>
<tr>
<td>Respiratory Chapter</td>
<td>6.254</td>
<td>PM 2.5</td>
<td>19.95</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>85.06</td>
<td>0.001</td>
</tr>
<tr>
<td>Respiratory Subset</td>
<td>4.738</td>
<td>Traffic Volume</td>
<td>49.08</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>104.47</td>
<td>0.001</td>
</tr>
</tbody>
</table>

$^{1}$PM 2.5 – particulate matter smaller than 2.5 microns in diameter.
FIGURES
Figure 1: Overview of Method and Data Sources

Method

HMRI Data by Postal Code

Vehicle Type Distribution

PART5 Emission Model

HMRI Data by Enumeration Area

Census Data

Spatial Street Network

PM Exposure Model

Spatial EA Boundaries

1991 Census Data

Street Vehicle Count Data

Variables Calculated for Each EA

Respiratory Subset Hospitalization Rate

Average Daily Exposure to PM 2.5

Respiratory Chapter Rate

Vehicle Count Exposure

Gastrointestinal Chapter Rate

Individual Variables

Genitourinary Chapter Rate

Socio-Economic Status Index

LEGEND

Data Source

Derived Data

Data Model

ABBREVIATIONS

HMRI - Hospital Medical Records Institute
PCCF - Postal Code Conversion File
PM - Particulate Matter
EA - Enumeration Area
Figure 2     Hospital Admissions Data Flow

LEGEND

Data Source  Derived Data  Data Process

ABBREVIATIONS
HMRI – Hospital Medical Records Institute
PCCF – Postal Code Conversion File
SETO – South East Toronto
EA – Enumeration Area

1990-92 Admissions by Postal Code

Ontario PCCF

1990-92 Admissions by Enumeration Area

Limit to SETO Area

Limit to Full Records

Identify First Admission

Count Admissions by EA

Age-Sex Standardized Admission Rates by Enumeration Area

Census Data
Figure 3  Motor Vehicle Emissions Data Flow

LEGEND

Data Source

Data Process

Derived Data

Data Model

Vehicle Count Data
City 24 and 8 hour
Metro 24 hour

Aggregate / Calculate

Point to Line

Limit to 5000 per day

Combine Adjacent

Assign Segment ID

24 Hour Traffic Volume by Street Segment

Spatial Street Network

Limit to study area

Limit to modelled

Limit to 5000 per day

Verify Spatial Data

Assign Segment ID

Spatial Network of Street Segments

Vehicle Type Distribution in Study Area

PART5 Emission Model

24 Hour Emission Values by Street Segment

Spatial Network of Street Segments

Vehicle Count Data
City 24 and 8 hour
Metro 24 hour

Aggregate / Calculate

Point to Line

Limit to 5000 per day

Combine Adjacent

Assign Segment ID

24 Hour Traffic Volume by Street Segment

Spatial Street Network

Limit to study area

Limit to modelled

Limit to 5000 per day

Verify Spatial Data

Assign Segment ID

Spatial Network of Street Segments

Vehicle Type Distribution in Study Area

PART5 Emission Model

24 Hour Emission Values by Street Segment

Spatial Network of Street Segments
Figure 4  Exposure Estimation Data Flow Diagram

Spatial Network of Street Segments

Corrected Spatial Street Network

DENSIFY  SNAP

EW Streets

NS Streets

BUFFER UPDATE  UNION ELIMINATE  UNION ELIMINATE  UNION ELIMINATE

15 Arcs  15 Polys

14 Arcs  14 Polys

2 Arcs  2 Polys

1 Arc  1 Poly

30 Polys

52 Polys

6 Polys

40 Polys

Buffered Street Network (326 Polys)

Spatial EA Boundary Layer (335 Polys)

IDENTITY Overlay

Emission and Count Values by Segment

Average Daily Exposure to PM and Vehicle Counts (335 Polys)

Combined Street Network and EA Layer (1403 Polys)

Legend:
- Data Source
- Derived Data
- Data Model

Abbreviations:
- Arc – Linear Street Segment
- Poly – Polygon
- PM – Particulate Matter
- EA – Enumeration Area
- COMMAND – Arc/Info Command

Calculate and Weight

Statistics
24 Hour Point Counts to 24 Hour Segment Volumes

- Traffic passing point in both directions is counted
- Automatic counting methods usually employed

Traffic Volume (Segment One) = \( \frac{11000 + 10500}{2} \)
= 10750 Vehicles per 24 Hr

8 Hour Point Counts to 24 Hour Segment Volumes

- Traffic approaching intersection from all directions is counted
- Manual counting methods usually employed

Traffic Volume (Segment Two) = \( (3000 + 2200) \times 2.05^\dagger \)
= 10660 Vehicles per 24 Hr

\(^\dagger\) This value is the factor used to convert 8 hour counts to 24 hour counts by the City of Toronto (Lasagna, 1997)
Figure 6  Observed Distribution of Hospital Admissions and Poisson Distribution of Hospital Admissions in South East Toronto by Enumeration Area (Admission counts greater than 50 shown as 50)

Number of Enumeration Areas

Number of Admissions [Max 82]

- Respiratory Subset (All Admissions) [Mean 7.69 Var 92.59]
- Respiratory Subset (First Admission) [Mean 5.89 Var 47.80]
- Poisson (All Admissions)
- Poisson (First Admission)

Number of Admissions [Max 94]

- Respiratory Chapter (All Admissions) [Mean 10.99 Var 149.60]
- Respiratory Chapter (First Admission) [Mean 8.76 Var 83.06]
- Poisson (All Admissions)
- Poisson (First Admission)

Number of Admissions [Max 85]

- GI Chapter (All Admissions) [Mean 13.86 Var 147.70]
- GI Chapter (First Admission) [Mean 11.86 Var 102.10]
- Poisson (All Admissions)
- Poisson (First Admission)

Number of Admissions [Max 55]
Figure 7  Scatterplot of PM 2.5 Exposure from Motor Vehicle Emissions and Standardized Hospitalization Ratio (SHR) for First Respiratory Subset Admissions in South East Toronto by Enumeration Area (N = 229)

Outlying EA with exposure of 1183 grams and SHR of 0.53 not shown. PM 2.5 - particulate matter smaller than 2.5 microns in diameter.
Figure 8  Relative Risk of Admission for Respiratory Subset (Rate of First Admissions) in South East Toronto by PM 2.5 Exposure Quintile

Adjusted for University Graduation. 95% Confidence Intervals around relative risk estimates. PM 2.5 - particulate matter smaller than 2.5 microns in diameter.
Figure 9  Relative Risk of Admission for Respiratory Subset and GI Chapter (Rate of First Admissions) in South East Toronto by PM 2.5 Exposure Quintile

Adjusted for University Graduation. 95% Confidence Intervals around relative risk estimates for respiratory subset. PM 2.5 - particulate matter smaller than 2.5 microns in diameter.
Figure 10  Relative Risk of Admission for Respiratory Subset (Rate of First Admissions) in South East Toronto by PM 2.5 and Traffic Volume Exposure Quintile

Adjusted for University Graduation. 95% Confidence Intervals around relative risk estimates for PM 2.5 exposure. PM 2.5 - particulate matter smaller than 2.5 microns in diameter.
Map 1  Neighbourhood and Enumeration Area Boundaries in South East Toronto

Neighbourhood Boundary

Enumeration Area Boundary
Map 2  Street Network in South East Toronto by Traffic Data Status with Location of Cordon Count Points

Streets by Vehicle Count Status
Total Length of All Streets in Class (km)

- Vehicle Count Unknown (115 km)
- Vehicle Count Modeled (74 km)
- Vehicle Count Not Modeled (30 km)

🚗 Cordon Count Point

Scale 1 : 50 000

0 km 0.5 km 1 km 1.5 km 2 km
Map 3  Average Daily Emissions of PM 2.5 from Motor Vehicles in South East Toronto for Streets with Over 5000 Vehicles per day

PM 2.5 Emissions
grams per km

- 11,600 to 30,200
- 5,500 to 11,600
- 3,500 to 5,500
- 1,900 to 3,500

PM 2.5 - particulate matter smaller than 2.5 microns in diameter
Map 4  Enumeration Area Boundaries and Suppressed Enumeration Areas (EA) in South East Toronto

<table>
<thead>
<tr>
<th>EA Property</th>
<th>Mean</th>
<th>Variance</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (square km)</td>
<td>0.0618</td>
<td>0.1109</td>
<td>5.8659</td>
<td>0.0002</td>
<td>20.6577</td>
</tr>
<tr>
<td>Population</td>
<td>404</td>
<td>55832</td>
<td>1255</td>
<td>45</td>
<td>121875</td>
</tr>
</tbody>
</table>

/ Enumeration Area Boundary

Suppressed Enumeration Area N=32

Enlarged Map Scale 1 : 25 000

Main Map Scale 1 : 50 000
Map 5  Counts of First Respiratory Subset Admissions in South East Toronto by Enumeration Area (EA)

Count of First Admissions in EA
Size of circle proportional to count

Enumeration Area Boundary

Suppressed Enumeration Area (N=32)
Map 6  Indirectly Standardized First Respiratory Subset Admission Rates in South East Toronto by Enumeration Area

Respiratory Subset Admission Rates per 1000 grouped by Quintile

- 8 to 55
- 5 to 8
- 3 to 5
- 1 to 3
- 0

/ Enumeration Area Boundary

Suppressed Enumeration Area (N=32)
Map 7  Indirectly Standardized First Admission Rates for Respiratory Subset, Respiratory Chapter, Gastrointestinal Chapter, and Genitourinary Chapter in South East Toronto by Enumeration Area

Respiratory Subset
I - 0.005
p 0.971

Respiratory Chapter
I - 0.045
p 0.287

Gastrointestinal Chapter
I - 0.101
p 0.013

Genitourinary Chapter
I - 0.081
p 0.051

First Admission Rates by Quintile

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 20th</td>
<td></td>
</tr>
<tr>
<td>20th to 40th</td>
<td></td>
</tr>
<tr>
<td>40th to 60th</td>
<td></td>
</tr>
<tr>
<td>60th to 80th</td>
<td></td>
</tr>
<tr>
<td>80th to 100th</td>
<td></td>
</tr>
</tbody>
</table>

Scale 1 : 100 000

0 km 1.0 2.0 km 3.0 4 km

Enumeration Area Boundary

Suppressed Enumeration Area (N=32)
**Map 8**  
**Average 24 Hour Exposure to PM 2.5 from Motor Vehicles in South East Toronto by Enumeration Area**

Exposure to PM 2.5 from Motor Vehicles  
grams per 24 hours grouped by quintile

- 75 to 1,200
- 40 to 75
- 18 to 40
- 1 to 18
- 0

PM 2.5 - particulate matter smaller than 2.5 microns in diameter

---

**Enumeration Area Boundary**

**Suppress Enumeration Area (N=32)**

---

**Enlarged Map**  
Scale 1:25 000

**Main Map**  
Scale 1:50 000
Map 9  Proportion of the Population with a University Degree in South East Toronto by Enumeration Area

Population with University Degree
percent grouped by quintile

- 34.5 to 62.5
- 26.5 to 34.5
- 17.0 to 26.5
- 10.5 to 17.0
- 1.0 to 10.5

Enumeration Area Boundary

Suppressed Enumeration Area (N=61)

Enlarged Map
Scale 1: 25 000

Main Map
Scale 1: 50 000

0 km 0.25 km 0.5 km 0.75 km 1 km
0 km 0.5 km 1 km 1.5 km 2 km
Map 10: Standardized Likelihood Residuals from Poisson Regression Analysis Addressing Hypothesis
REFERENCES


Ozkaynak, H. et al., 1996. Personal exposure to airborne particles and metals: results from the particle team study in Riverside, California. Journal of Exposure Analysis and Environmental Epidemiology, 6: 57-78.


Patychuk, D. and Yokingco, C., 1997b. Neighbourhood Profiles, Church/Yonge, City of Toronto Community Services Public Health, Toronto.


Patychuk, D. and Yokingco, C., 1997g. Neighbourhood Profiles, South Riverdale, City of Toronto Community Services Public Health, Toronto.

Patychuk, D. and Yokingco, C., 1997h. Neighbourhood Profiles, St James Town, City of Toronto Community Services Public Health, Toronto.

Patychuk, D. and Yokingco, C., 1997i. Neighbourhood Profiles, St Lawrence, City of Toronto Community Services Public Health, Toronto.


Roos, L.L. et al., 1993. Registries and administrative data: organization and accuracy. Medical Care, 31(3): 201-212.


