Obesity in the built environment: a spatial analysis of two Canadian Metropolitan areas

by
Charles E. Fritz III
B.Sc., Western Michigan University, 2009

Thesis Submitted In Partial Fulfillment of the Requirements for the Degree of Master of Science

in the
Department of Geography
Faculty of Environment

© Charles E. Fritz III 2013
SIMON FRASER UNIVERSITY
Spring 2013

All rights reserved. However, in accordance with the Copyright Act of Canada, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Charles E. Fritz III
Degree: Master of Science (Geography)
Title of Thesis: *Obesity in the built environment: a spatial analysis of two Canadian Metropolitan areas*

Examining Committee: Chair: Geoff Mann

Nadine Schuurman
Senior Supervisor
Professor

Valorie A. Crooks
Committee Member
Associate Professor

Meghan Winters
External Examiner
Assistant Professor, Faculty of Health Sciences,
Simon Fraser University

Date Defended/Approved: March 27th, 2013
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011
The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010
Abstract

Global prevalence of obesity and overweight has rapidly increased over the past few decades. The relative growth rate of the epidemic, particularly in more developed countries, has triggered efforts to explore environmental determinants of weight gain. Research on how the built environment affects weight gain, and health more broadly, has been widely undertaken by public health, epidemiology, and geography disciplines, yet no clear relationships have been identified. Moreover, research on the Canadian context is generally lacking. Methods native to Geographic Information Systems (GIS) and spatial epidemiology may prove effective to furthering contemporary knowledge of the built environment determinants of obesity, and overall, contribute to wider disciplines involved. The first paper of this thesis reviews literature from the spatial epidemiology discipline to glean insight from recent methodological development of spatial clustering tools and provide guidelines for practical application. The second paper explores the spatial clustering of obesity and examines the built environment for potential correlates. Both papers take a unique perspective within the respected disciplines they are informing, and thus provide novel results for future research and development.

Keywords: Obesity; built environment; public health; Geographic Information Systems (GIS); spatial epidemiology; spatial clustering
Dedication

I would like to dedicate this to my family and friends, close and distant. You have all helped a little bit along the way....
Acknowledgements

I would like to express my utmost gratitude to my supervisor, Dr. Nadine Schuurman, for providing me the opportunity to work with, and learn from her during my time as a student. I would like to thank Dr. Valorie Crooks, my committee member, for being supportive and constructive during coursework and research. I need to thank Dr. Scott Lear, from the Department of Kinesiology, for funding me through the Prospective Urban and Rural Epidemiological (PURE) study and providing valuable insight for one of the chapters in the thesis. Also, I need to express a kind thank you to Dr. Meghan Winters for reviewing this thesis and providing her thoughtful comments. Lastly, I would like to thank the staff, faculty and fellow graduate students of the Department of Geography at Simon Fraser University for creating an engaging academic atmosphere and a warm sense of community.

I must also acknowledge the funding bodies that made my academic career possible to this point. Modelling of Complex Social Systems at SFU provided me with a splendid opportunity to collaborate with a multi-disciplinary group of researchers. I also have to acknowledge the Canadian Institutes of Health Research for funding the PURE study.
# Table of Contents

Approval .................................................................................................................. ii  
Partial Copyright Licence ....................................................................................... iii  
Abstract .................................................................................................................. iv  
Dedication ............................................................................................................... v  
Acknowledgements ............................................................................................... vi  
Table of Contents ................................................................................................ vii  
List of Tables ......................................................................................................... ix  
List of Figures ....................................................................................................... x  

1. **Introduction** ................................................................................................... 1  
   1.1. Overview ...................................................................................................... 1  
   1.2. Research Objectives .................................................................................. 4  
   1.3. Background ............................................................................................... 5  
      1.3.1. Adult Obesity and Overweight ............................................................ 5  
      1.3.2. Conceptualizing Obesogenic Environments ...................................... 6  
      1.3.3. Built Environment and Obesity ......................................................... 9  
      1.3.4. Methodological Considerations for Built Environment and Obesity Research ................................................................. 12  
      1.3.5. Spatial Epidemiology ....................................................................... 15  
   1.4. Research Questions .................................................................................... 16  
   1.5. Thesis Outline ............................................................................................ 17  

2. **A Scoping Review of Spatial Cluster Analysis Techniques for Point-Event Data** ........................................................................................................ 19  
   2.1. Abstract ..................................................................................................... 19  
   2.2. Background .............................................................................................. 20  
   2.3. Materials and Methods ........................................................................... 22  
      2.3.1. Selection of Search Terms and Papers ............................................. 22  
      2.3.2. Method Testing ............................................................................... 24  
   2.4. Results ...................................................................................................... 25  
      2.4.1. K-function ....................................................................................... 27  
      2.4.2. Nearest Neighbour Statistics ........................................................... 28  
      2.4.3. Local Moran’s I ............................................................................. 30  
      2.4.4. Kernel Estimation .......................................................................... 30  
      2.4.5. Generalized Additive Model ............................................................. 31  
      2.4.6. Spatial Scan Statistic .................................................................. 33  
   2.5. Themes ..................................................................................................... 36  
   2.6. Discussion .................................................................................................. 41  
      2.6.1. Future Research Recommendations ............................................... 42  
      2.6.2. Limitations .................................................................................... 43  
   2.7. Conclusions .............................................................................................. 44  

3. **Obesity and the built environment: a spatial analysis of communities in Metropolitan Vancouver and Hamilton, Canada** ........................................... 45  
   3.1. Abstract ..................................................................................................... 45
3.2. Background .................................................................................................................. 46
3.3. Methods and Materials ............................................................................................... 49
  3.3.1. Study area .................................................................................................................. 49
  3.3.2. Participant data .......................................................................................................... 50
  3.3.3. Community Aggregation ........................................................................................... 51
  3.3.4. Built Environment Overview ................................................................................... 52
  3.3.5. Built Environment Variables ................................................................................... 52
    Local Nutrition environment and physical activity environment ........................................ 53
    Land-use and design ......................................................................................................... 54
  3.3.6. Exploratory Spatial Data Analysis ........................................................................... 56
    Spatial Relative Risk ........................................................................................................ 56
    Ordinary Least Squares Regression Analysis (OLS) ......................................................... 57
    Geographically Weighted Regression (GWR) ................................................................... 57
3.4. Results .......................................................................................................................... 58
  3.4.1. Descriptive Statistics ............................................................................................... 58
  3.4.2. Spatial Clustering ..................................................................................................... 67
  3.4.3. OLS Regression ........................................................................................................ 69
  3.4.4. GWR Results ............................................................................................................ 71
3.5. Discussion ..................................................................................................................... 76
3.6. Conclusions ................................................................................................................... 81

4. Conclusion ....................................................................................................................... 82
  4.1. Research Contributions ............................................................................................... 83
  4.2. Future Research .......................................................................................................... 85

References .......................................................................................................................... 89

Appendices .......................................................................................................................... 102
Appendix A ........................................................................................................................... 103
List of Tables

Table 1-1. ANGELO Framework: Example of intervention strategies to address obesogenic environments ................................................................. 2

Table 2-1. Table summarizes the methods used by papers in the review. Other information on methods include data type and resolution required for analysis, software available for implementing the method, and disciplines that utilized the method ........................................ 26

Table 2-2. Cuzick-Edwards global clustering method results. ........................................ 29

Table 2-3. Identified Themes .................................................................................. 37

Table 3-1. Descriptive Statistics of Individual Level Variables for Metropolitan Vancouver Communities ........................................................................... 59

Table 3-2. Descriptive Statistics of Individual Level Variables for Metropolitan Hamilton Communities ........................................................................... 60

Table 3-3. Descriptive Statistics of Built Environment Variables for Metropolitan Vancouver Communities – Mean Values Non-Obese vs. Obese ............ 62

Table 3-4. Descriptive Statistics of Built Environment Variables for Metropolitan Hamilton Communities – Mean Values Non-Obese vs. Obese ............ 63

Table 3-5. Comparison of Built Environment Variables Between Study Neighbourhoods in Metropolitan Vancouver and Hamilton .................. 64

Table 3-6. Pearson’s Correlation Coefficients for Metro Hamilton ......................... 66

Table 3-7. Pearson’s Correlation Coefficients for Metro Vancouver ..................... 67

Table 3-8. OLS regression results for Metro Vancouver ......................................... 70

Table 3-9. OLS regression results for Metro Hamilton ........................................... 71

Table 3-10. GWR results for Metro Vancouver ....................................................... 72

Table 3-11. GWR results for Metro Hamilton ........................................................ 72
List of Figures

Figure 1-1. International Obesity Task Force (IOTF) Causal Web of Obesity: Example of contrasting conceptual framework for obesogenic environments ................................................................. 3

Figure 1-2. Basic Conceptual Model for Weight Gain ................................................................. 8

Figure 2-1. Flow-chart illustrating scoping review processes .................................................. 23

Figure 2-2. K-function Global Clustering Results ................................................................. 28

Figure 2-3. Kernel-based Methods Results ........................................................................ 31

Figure 2-4. Generalized Additive Model ........................................................................ 33

Figure 2-6. Spatial Scale Example ................................................................................ 36

Figure 2-7. Exploratory Analysis Example ........................................................................ 39

Figure 3-1. Metropolitan Vancouver Study Area ................................................................ 49

Figure 3-2. Metropolitan Hamilton Study Area ...................................................................... 50

Figure 3-3. Spatial Relative Risk of Metro Hamilton Communities ...................................... 69

Figure 3-4. Spatial Relative Risk of Metro Vancouver Communities .................................... 69

Figure 3-5. Metro Vancouver #3 GWR surfaces ................................................................ 74

Figure 3-6. Metro Hamilton #3 GWR surfaces .................................................................... 75
1. Introduction

1.1. Overview

Increasing prevalence of obesity is a global epidemic affecting the livelihood of millions (Feng, Glass, Curriero, Stewart, & Schwartz, 2010). Based on recent estimates, 1.46 billion people are considered overweight or obese, with one-third of the estimate making belonging to obesity (approx., 502 million people (B. A. Swinburn et al., 2011). The meteoric rate of increase and its ubiquitous effects in countries across the development spectrum are raising global awareness about a risk factor that is largely preventable (WHO, 2012). The direct and indirect costs of overweight and obesity are also a cause for concern, with an estimated 2% - 6% share of total health-care costs in many countries (WHO, 2008). Obesity and overweight are major risk factors for several life-altering non-communicable diseases such as type-2 diabetes, cardiovascular diseases (stroke and heart diseases), osteoarthritis, and cancer (WHO, 2012). Arguably, the most perplexing facet of obesity and overweight is the rate of change since the 1970’s, as measured Body Mass Index (BMI) for adults in Canada increased from 13.8% (1978) to 25.4% (2008) of the population (Canada, 2011). While a number of genetic markers have been identified that may predispose and individual to be obese (Bell, Walley, & Froguel, 2005), the rapid rate at which obesity and overweight has increased worldwide suggests a substantial environmental influence.

Research on environmental determinants of obesity and overweight has seen an exponential increase in the last couple decades. Conceptualizations depict a multifactorial relationship between the physical and social environments that influence our lives. These environments combine to increase to “obesogenicity” of an area and are commonly referred to as obesogenic environments (Lake & Townshend, 2006; B. Swinburn, Egger, & Raza, 1999). The most frequently used conceptual framework of obesogenic environments was proposed by Swinburn and colleagues, and is called the Analysis Grid for Environments Lined to Obesity (ANGELO) framework (B. Swinburn, et
The ANGELO framework is set up with four foundational environments - physical, economic, political, and sociocultural - on the top of the grid, and two types of environment sizes – micro (settings) and macro (sectors) (Table 1-1). Though the ANGELO framework provides a researcher or user with a set of domains that contextualize the complexity of interrelationships among multiple environments, it does not represent the hierarchy of the influences. Figure 1-1 depicts the hierarchy of environments, moving left to right. This schematic is perhaps more indicative of the scale at which environments influence weight gain by providing a more visually intuitive design. Either conceptualization is instructive of the nuances involved in conceptualizing the causal pathways and mechanisms of obesity and overweight. Recent research has pinpointed the built environment has a potential factor in the complex web of interactions illustrated by these frameworks.

**Table 1-1. ANGELO Framework: Example of intervention strategies to address obesogenic environments**

<table>
<thead>
<tr>
<th>Size</th>
<th>Physical (Food and PA)</th>
<th>Economic (Food and PA)</th>
<th>Political (Food and PA)</th>
<th>Sociocultural (Food and PA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro (Settings)</td>
<td>Recreation and sports facilities</td>
<td>Canteens serving local food</td>
<td>Policies on physical education</td>
<td>Healthy food cooking workshops</td>
</tr>
<tr>
<td>Neighbourhoods</td>
<td>Safe walking paths</td>
<td>Cost of cafeteria foods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td></td>
<td></td>
<td>Policies on physical education</td>
<td></td>
</tr>
<tr>
<td>Macro (Sectors)</td>
<td>Availability of buses and bus stops</td>
<td></td>
<td>Policies and Standards on imported food quality/labeling</td>
<td></td>
</tr>
</tbody>
</table>

Note. Adapted from Swinburn, et al. (1999). PA = Physical Activity
Figure 1-1. International Obesity Task Force (IOTF) Causal Web of Obesity: Example of contrasting conceptual framework for obesogenic environments

Note. Adapted from Kumanyika, et al. (2002). Arrows illustrate suspected causal interactions between factors. Vertical and horizontal links will vary between different societies and populations.

Commonly used methods, such as multilevel modeling, or other regression-based analyses, traditionally conceptualize an individual’s neighbourhood in terms of a predefined unit of analysis, such as census tracts or dissemination areas (Papas et al., 2007). Other studies define a neighbourhood in terms of a region derived from distance calculations along a road network, called buffers (Feng, et al., 2010). Very few studies, however, target groups of individuals in the same neighbourhood to investigate micro-scale spatial patterns, or clustering of variables (Papas, et al., 2007). This research gap is due in large part to the difficulty in obtaining spatial data at a fine-grain scale, such as the address level. Spatial cluster analysis methods, native to the discipline of geography and spatial epidemiology, have the potential to contribute to the methodological tools of obesity and built environment research. The breadth of research devoted to fine-grain spatial pattern analysis and the practicality of employing these methods in an
Exploratory Spatial Data Analysis (ESDA) framework could be of interest to geographers and health researchers alike. Cataloguing and thematically evaluating the use of spatial clustering methods on address point data is necessary to identify issues related to implementation of methods and areas for future research.

Paper one will present a scoping review for spatial clustering methods for address point data. Paper two will apply a spatially explicit approach to analyze the spatial clustering of obesity and correlates of the built environment in two Canadian Metropolitan areas.

1.2. Research Objectives

The overall objective of this thesis is to contribute to two fields – spatial epidemiology and built environment and obesity research – by providing guidelines for applying spatial methods to health research through (1) a scoping review and (2) a relevant research example. The aim of the first study, presented in chapter two, was to explore and review all papers in the past ten years that have employed spatial clustering methods on address point data. This paper employs a scoping review methodology, a tool for evaluating the broad trends in a field or topic by synthesizing research findings to develop themes and highlight research gaps. This study was bolstered by practical demonstrations on a real dataset using a select suite of spatial clustering methods. The goal of the second study, presented in chapter three, was to explore the spatial trends of obesity and built environment correlates in two large Canadian Metropolitan Areas, Hamilton and Vancouver. This study sought to employ a method from paper one to evaluate spatial clustering of obesity at the address level. A spatially explicit approach was also adopted to explore potential built environment correlates of obesity. The overarching purpose of both papers was to mainly inform future research, however specific goals lie within each. The goal of paper one was to communicate, through a scoping review methodology, the practicality of using spatial clustering methods for disciplines epidemiology related disciplines. The primary goal of paper two was to contribute to the growth and knowledge base of Canadian obesity research. A second goal was to contribute to a larger study that the paper is based on by identifying any spatial relationships in the dataset.
1.3. Background

This section provides background information on both papers presented in this thesis. It begins with an overview on adult obesity and overweight followed by a discussion on obesogenic environments. Origins of the relationship between the built environment, health, and obesity are described. Methodological considerations for exploring the built environment are presented along with the Canadian context of research. The section ends by presenting the methodological foundations of spatial epidemiology followed by research questions guiding the research of thesis and the outline.

1.3.1. Adult Obesity and Overweight

Before discussing the influences of obesity it is important to touch on the basic definition of obesity, describe the biological interactions that contribute to weight gain in adults, and shed brief light on established correlates of obesity and overweight. Obesity and overweight are risk factors for several life-limiting cardiovascular diseases. The World Health Organization (WHO) defines a standard measure of weight based on Body Mass Index (BMI), as it provides a useful population-level measure for both sexes (WHO, 2012). It is defined as a person’s weight in kilograms divided by the square of his or her height in meters. A basic classification of individual BMI divides individuals in two categories; BMI greater than 30 as obese, and BMI between 30 and 25 as overweight. The complete classification of BMI is divided into six categories; underweight (<18.5 BMI), normal weight (18.5 – 24.9 BMI), overweight (25.0 – 29.9 BMI), obese class I (30.0 – 34.9 BMI), obese class II (35.0 – 39.9 BMI), and obese class III (>40 BMI) (Lau et al., 2007). At the biological level, obesity is influenced by the imbalance between energy intake and energy expenditure. Individual mediators for weight gain, such as increased physical activity and altered diets, are generally assumed to be beneficial for decreasing body size (G. Egger & B. Swinburn, 1997)(Figure1-2).

The global trends in weight gain have been rapidly increasing since the 1980s as the effects of globalized society values are diffused throughout developed and less-developed countries (LDC) (Zimmet & Alberti, 2012). This trend was particularly evident from the global presence of increasingly energy-dense and sweeter foods (Malik,
Schulze, & Hu, 2006) brought on by the westernization of the global food supply, and reduced energy expenditure caused by growth in sedentariness (Popkin, 2006). Often thought of as a developed nation crisis, the global prevalence of obesity has proved to be pervasive in LDCs as well, with countries such as Egypt having a national prevalence of female obesity of nearly 45% (B. A. Swinburn, et al., 2011). However it is more developed countries such as United States, United Kingdom, New Zealand and Canada that are home to the highest prevalence for adult males and females. For countries such as United States and Canada, adult obesity and higher BMI is associated most clearly with the socioeconomic status (SES) or social position of an individual or family, and in some cases the neighbourhoods they live in (Gordon-Larsen, Nelson, Page, & Popkin, 2006; Ross et al., 2007). Influences such as less access to health promoting amenities and the lack of inexpensive, low-fat and nutritious food options among lower SES individuals are associated with obesity (K. Morland, Wing, Diez Roux, & Poole, 2002; M. C. Wang, Kim, Gonzalez, MacLeod, & Winkleby, 2007). However this trend appears to be most common in North America, and not the other developed nations with high obesity and overweight prevalence. For instance, a study from New Zealand found that more deprived communities were more likely to have adequate access to health-promoting resources such as recreational facilities and supermarkets (Pearce, Witten, Hiscock, & Blakely, 2007), amenities that are thought to protect one from weight gain if utilized. Though inconsistencies in cited associated risk factors for low SES individuals are common when compared to US studies, low SES is still a concrete association with high BMI at a global scale (McLaren, 2007). As cross-national studies lack concordance for even basic measures, the need for studies to examine risk factors within regional contexts becomes an important step in breaking down the complex relationships therein.

1.3.2. Conceptualizing Obesogenic Environments

Conceptual frameworks for obesogenic environments were developed, like many other ecological models, in response to the Ottawa Charter for Health Promotion in 1986 (Organization, 1986). Though each conceptualization varies in terms of the types and level of influences each one acknowledges, all acknowledge the complexity of the interrelationships between factors and the pathways connecting them.
The rapid increase of obesity and overweight prevalence is concerning for a number of reasons, as not only does it signify grim projections for future prevalence but also creates a strain on health care systems. Past prevention strategies focused mainly on modifications at the individual level by prescribing educational, behavioural or pharmacological based strategies (Garner & Wooley, 1991; Kayman, Bruvold, & Stern, 1990). Though efforts are effective, the contribution of individual genetics that predispose an individual to a higher risk of weight gain cannot be the sole influence. In fact, during the peak of prevalence, the US Institute of Medicine declared that there had been no real change in the US gene pool (Thomas, Prevent, & Obesity, 1995). In response to this, research began to focus more on analysis of the contextual environments of individuals. The proverbial call to action has been immense, with a quick search on any peer-reviewed journal article database, using the terms “obesity” AND “environment” returning over 20,000 entries for the year 2012 alone.

A broad model for overweight and obesity focuses on the interaction of three influences; biology, behavior, and the environment as potential upstream indicators for mediators and moderators; total energy and physiological adjustment, respectively (Garry Egger & Boyd Swinburn, 1997) (Figure 1-2). This early conceptualization spells out the broad relationship but does not fully address interactions between upstream indicators for which little causal knowledge is known. The two most widely known frameworks for obesogenic environments are the ANGELO framework (Table 1-1) and the International Obesity Task Force framework (IOTF)(Kumanyika, Jeffery, Morabia, Ritenbaugh, & Antipatis, 2002) (Figure 1-1). A central component to each is the classification of environment types into hierarchical scales of influence. The ANGELO framework, comprised of a matrix layout similar to that of the Haddon Matrix for injury prevention (Haddon Jr, 1980), defines hierarchy of environments into two categories – Mirco-level and Macro-level (B. Swinburn, et al., 1999). Micro-level influences refer to the settings or local environments that affect the outcome, and Macro-level influences refer to the sectors of the different systems or structures that affect the outcome. IOTF framework on the other hand conceptualizes the hierarchy in terms of settings only, with systems of governance or policy originating from each of the settings – i.e., school policy on free meals existing in the school setting. In contrast, the ANGELO framework, each level – Macro or Micro – can influence any one of four environments – physical,
economic, political, and sociocultural. This structure is useful for organizational purposes but does not elucidate the directionality of multiple interdependencies. The IOTF framework represents the interactions with arrows that denote the directionality between factors, clearly categorizing them into upstream and downstream indicators – i.e., left side, upstream; right side, downstream. While both frameworks have their respective advantages and shortcomings they both achieve the goal of presenting a more nuanced strategy to inform theory-building and prevention measures.

Figure 1-2. Basic Conceptual Model for Weight Gain

Note. Adapted from Swinburn, et al. (1999). Rudimentary conceptual model for weight gain.

Incorporating the multitude of factors into research is a very difficult and complex ask. Contemporary studies focus on subsets of environments, with a partial focus on incorporating the influence of other environments. The study of the built environment has garnered immense attention in the past two decades. The built environment is defined as human-made objects or policies that define the urban design, land-use patterns, or transportation systems that the population physically interacts with (Handy, Boarnet, Ewing, & Killingsworth, 2002). A recent review on the epidemiological evidence of the effects of the built environment on obesity found 63 papers, from 2005 to 2009 alone (Feng, et al., 2010). As with other environments displayed in the figures above, the built environment shares causal pathways with several other environments. For instance, an
individual’s access to healthy foods can be conceptualized in terms of multiple environments. Using a simplistic example, the physical environment may reflect the actual presence, or absence, of food outlets within the neighbourhood setting (micro-environment), while the sociocultural environment may reflect a family’s personal attitudes toward preparing food a certain way. While controlling for covariates, or incorporating multiple interactions of suspected causal pathways is possible – i.e., complex systems approach (Galea, Riddle, & Kaplan, 2010), geosimulation (Procter, Clarke, Ransley, & Cade, 2008) - development of these methods is in the nascent stages. Therefore a nuanced approach to synthesizing these interactions in the interpretation of associations is absolutely necessary. This thesis will specifically focus on the built environment determinants of obesity.

1.3.3. **Built Environment and Obesity**

The built environment is defined as human-made objects or policies that define the urban design, land-use patterns, or transportation systems that the population physically interacts with (Handy, et al., 2002). The language of built environment has long-shared a connection with the promotion of social interaction with the spaces its theory informs. Jane Jacobs, one of the most influential urban planners, famously touted her idea that “eyes on the street” promoted a safer environment for children to recreate in nearby urban parks (Jacobs, 1992). Her phrase was in reference to the concentration of cafés and other service oriented businesses, providing for more lively social fabric. Another, famous urban thinker, William Whyte, was known for his work in theorizing the design of great public spaces conducive to generating greater foot traffic and vibrant social spaces (Whyte, 1980). Even as early as 1850’s Paris, France, a link between city planning and public health was forged by Haussamann’s plans to redevelop the city to improve air flow and alleviate unhealthy living conditions (Frank & Engelke, 2001). Owing to the growing awareness of the linkages between the built environment and health, several initiatives have been enacted that outline urban and regional guidelines for promoting healthy populations. Global programs such as WHO’s, Healthy Cities and Villages (Lafond & Heritage, 2009), and national projects, such as the US’s Health Communities Movement and the Coalition of Healthier Cities and Communities (Norris & Pittman, 2000), seek to ensure citizens can achieve optimal well-being and physical
health by improving the physical and social environments in communities. Though a concrete link between the built environment and obesity has not been strongly established, evidence further upstream, in the physical activity realm, has long suggested there is one. Obesity and physical activity have an inextricable link, and research over the past two decades has shown this.

For the past two decades research on how the built environment affects obesity and physical activity has focused on the “three D’s” of the built environment – density, diversity, and design (Cervero & Kockelman, 1997). This conceptualization was originally borne from a physical activity literature to conceptualize travel demand and mode choice among individuals. Handy (2002) laid out a more nuanced and descriptive list of built environment elements; density and intensity, land-use mix, street connectivity, street scale, aesthetic qualities, and regional structure. While still related to the three pillars of the broader built environment mentioned earlier (urban design, land-use patterns, transportation systems) these two conceptualizations combine to describe the main categories where contemporary research has been focused. An example of the earliest of evidence from the physical activity and travel behavior literature, suggested a link between increased density and land-use mix, and walking for travel (Frank & Pivo, 1994). This research presented the idea of land-use mix; in this case the measures of the degree of heterogeneity between types of different land uses in an area (Frank, 2000). In general, as neighbourhood land-use mix increases, there is an increased likelihood that there will be desirable or useful destinations in close proximity of each other, potentially resulting in an increased demand to walk, rather than drive, to access them. Since this seminal work, the application of land-use mix measures have been widely employed by physical activity and obesity researchers (For reviews see Papas, et al., 2007, Feng, et al., 2010, and Saelens & Handy, 2008). Their work has culminated in an increased body of suggested relationships based on different contexts, new measures based on evidence-based learning, and several areas for future research that constantly challenge the paradigm. Below I will briefly present some of the evidence and findings from the obesity and built environment literature, with specific interest in the outcome of adult obesity.

Evidence that explicitly links the built environment to weight gain has received substantial attention in multiple research arenas. Broadly, a number of studies have
found elements of the built environment to be significantly associated with BMI; land-use mix (Frank, Andresen, & Schmid, 2004; Li et al., 2008; Mobley et al., 2006; Rundle et al., 2007); walkability (Frank, Saelens, Powell, & Chapman, 2007; Frank et al., 2006); access to recreational facilities and parks or open spaces (Giles-Corti et al., 2003; Mobley, et al., 2006); fast-food density (Li, et al., 2008; Lopez, 2012; Mehta & Chang, 2008); population density (Lopez, 2012; Rundle, et al., 2007); presence of convenience stores (Kimberly Morland, Diez Roux, & Wing, 2006; M. C. Wang, et al., 2007). Variables with significant associations with BMI only represent a fraction of the built environment elements explored to date. A larger proportion has found weak or insignificant relationships measures such as population density, grocery store density, supermarket density, and total length of sidewalks. Solely based on the breadth of evidence, relationships between the built environment and obesity are varied and inconsistent. A closer examination of the established links reveals an even blurrier story, as inconsistencies between study designs and measures, a lack of studies in contexts other than the United States, and a lack of longitudinal associations currently plague the collective literature (Feng, et al., 2010). For instance, in a review by Turrell (2010) walkability was found to have no association with BMI in four of the six studies examined, and two having found an association with low BMI in neighbourhoods of higher walkability. As one would expect the research settings in these studies were all different (mostly metropolitan areas of the US), reducing the likelihood of cross-community comparisons. Secondly, the walkability index is a composite measure of any number of individual measures – i.e., intersection density, land-use mix, retail floor area ratio, residential or dwelling density, retail density. Additionally, some factors may be weighted differently in studies, making the effect of one element stronger than others. These issues create an obvious discord when attempting to make any generalizations about an association, let alone causation.

Studies often conceptualize place and space at different spatial units. Feng (2010) conveniently divides the different spatial units into two classifications; geographic buffers and contextual studies. Geographic buffer studies conceptualize neighbourhood space based on an a priori distance (street network or “as the crow flies”), measured from a participants home address, usually ranging from 500 m to 2 km. Contextual studies conceptualize neighbourhood space based on a predefined boundary, usually
administrative, such as census tracts, dissemination areas, postal code points, or counties. While differences in spatial scale may be caused by data availability or for confidentiality purposes, it should not necessarily be viewed as a limitation to research. The relationship between obesity and the built environment is one that has little evidence regarding generalizable etiology. For instance, take accessibility to healthy foods measured by the nearest distance to a supermarket, or similarly, density of supermarkets. If the individual owns a car, there is a likelihood they may use it to shop at other places than their neighbourhood supermarket or the nearest supermarket to their home. For this relationship, using a spatial scale that reaches much further in distance than a buffer may be appropriate to capture the accurate activity space of the individual. Recent studies have attempted to contribute to this knowledge gap by incorporating GPS tracking to access exposure to environmental characteristics (Almanza, Jerrett, Dunton, Seto, & Ann Pentz, 2012; B. W. Wheeler, Cooper, Page, & Jago, 2010). This trend represents a crucial breakthrough for providing evidence on how individuals interact with the built environment at a more precise spatial scale. However, conceptualizing space for larger sample sizes, and studies that consist of several sites where GPS tracking would be conceivably infeasible, should be approached with an understanding of the aggregation scheme imposed on the data. Geographers are well aware of this common problem, Modifiable Areal Unit Problem (MAUP), however its terminology has not been widely adopted in built environment and obesity research (A. S. Fotheringham & Wong, 1991). Paper 2 adopts a buffer approach to defining activity space.

1.3.4. Methodological Considerations for Built Environment and Obesity Research

There are several methodological caveats to exploring the relationship between the built environment and obesity. These relate to the way obesity is measured, analytical methods chosen, rationalization of built environment methods, sample sizes, individual-level variables used as controls, and lack of evidence from a variety of contexts. I will briefly outline some of the main methodological considerations below.

The dominant metric for determining adult obesity and overweight is the Body Mass Index (BMI). Several issues relate to the measurement of one’s weight. For
instance, interpreting weight for different ethnicities can be problematic as location of body fat in South Asian and Latin populations tends to build around the waist and abdomen, increasing the risk for cardiovascular diseases associated with weight gain (Pearce & Witten, 2010). To improve on this limitation waist circumference has been proposed as an alternative to accurately measure different populations (Lear, Humphries, Kohli, & Birmingham, 2012). Another issue related to weight measurement is the accuracy of BMI values derived from a person’s height and weight. The majority of BMI measures used in studies as an outcome variable derive from self-reported height and weight values, as opposed to objectively measured ones (Papas, et al., 2007). A report from Statistics Canada (Tjepkema, 2006) estimated that up to 50% of the prevalence was underestimated through misclassification. Another study estimated that up 25% of men, and 45% of women were misclassified as non-obese (Akhtar-Danesh, Dehghan, Merchant, & Rainey, 2008). Using objectively measured height and weight values to derive BMI measurements represents an important consideration for informing methodologies. Paper 2 will be using BMI measurements derived from objectively measured height and weight values.

A recent review on the epidemiological evidence from the built environment and obesity cited that only two research articles set in Canada have been published in the past ten years (Feng, et al., 2010). While this figure is most likely underestimated due to the inclusion and exclusion criteria of the review, their findings are consistent with other reviews in that Canadian studies on the built environment and obesity are rare (Papas, et al., 2007; Turrell, 2010). Evidence from an exploratory study throughout major Canadian CMAs (Census Metropolitan Areas) suggests a relationship between social position and BMI, as well metropolitan sprawl among men (Ross, et al., 2007). Another study found that a higher amount of green space was associated with a decreased likelihood of physical activity and increased odds of obesity and overweight in men (S. A. Prince et al., 2011). On the other hand, odds for obesity were lower for women as green space increased (S. A. Prince, et al., 2011). Currently no measures for land-use mix or walkability have been incorporated to studies on adult obesity in Canada. These trends slightly resemble the broader body of evidence on the built environment and obesity, that is, inconsistencies in study design and built environment variables investigated. Given the current prevalence of obesity and overweight in Canada (one in four are obese), and
its rapid increase since 1980, the relative lack of research on the built environment and obesity raises eyebrows. Recent national research initiatives from the Canadian Institute of Health Research and the Heart and Stroke Foundation of Canada, underscore Canada’s acknowledgement that this area is under-researched. The recent release of the Canadian Community Health Survey could also be a prime opportunity to perform exploratory research on the built environment context of obesity research in Canada. Paper 2 will explicitly address this research consideration by examining two metropolitan areas for correlates of obesity with neighbourhood built environment features.

Recent studies have examined the built environment and obesity using regression-based statistical procedures, such as multi-level models, logistic regression, and linear regression. The multi-level model framework represents a advancement from past techniques, as both group-level and individual-level measures can be incorporated in the analysis (Papas, et al., 2007). This technique is especially useful given the context of current evidence suggesting that some variables may contribute to area-based effects, such as deprivation or SES (Pearce, et al., 2007; Ross, et al., 2007). However, Chaix, et al. (2005) suggests that multi-level models are inept at capturing spatial variation whenever there is a geographic component to data. Their study concluded that spatially explicit models – in this case, a geo-additive model - that were centered on indicators from a participant’s home returned far stronger associations (Chaix, et al., 2005). Analyzing the built environment correlates of obesity may benefit from spatially explicit models as spatial autocorrelation in residuals may also suggest small-area effects in the data, normally not detectable by the oft-used multi-level modeling framework (Papas, et al., 2007). The use of spatial clustering methods, normally applied in the spatial epidemiology discipline, may also represent an alternative approach to analyzing the relationships between the built environment and obesity (Duncan et al., 2012; L. Huang, Tiwari, Pickle, & Zou, 2010; Mobley, Finkelstein, Khavjou, & Will, 2004; Schuurman, Peters, & Oliver, 2009). Moreover, spatially explicit modelling techniques such as geographically weighted regression (GWR) have also been used to model between area variation (Chalkias et al., 2013; D.-R. Chen & Truong, 2012; Edwards, Clarke, Ransley, & Cade, 2010; Procter, et al., 2008). In a recent review of spatial methods in epidemiology, Auchincloss (2012) cited a research gap in the lack of application of spatially explicit methods in public health based sciences. The same
statement can aptly be applied to the current research context of the built environment and obesity. Both paper 1 and paper 2 addresses this consideration.

1.3.5. **Spatial Epidemiology**

Spatial methods in health research have contributed considerably to furthering our understanding of spatial relationships in health-related data (Rushton, 2003). The development and integration of Geographic Information Systems (GIS) with spatial statistical analysis has greatly improved the capability of researchers to render, analyze, and store spatial health data used in contemporary research (Carpenter, 2011). One field that has benefitted from the blending of spatial statistics and (GIS) is spatial epidemiology (Geoffrey M Jacquez, 2000). Spatial epidemiology is broadly defined as the analysis of geographically and temporally-varying health-related events (Elliott & Wartenberg, 2004). The discipline encompasses several spatial analytical methods that are broadly grouped into three main areas; disease mapping, geographic correlation studies and spatial disease clustering and surveillance. Methods used in spatial epidemiology are not unlike the foundations of many geographic and spatial crime analysis methods; Global and Local Moran’s I, Geographical Analysis Machine, Ripley’s K-function, Kriging, Geographically Weighted Regression, Spatial Adaptive Filtering (all important methods in the history of spatial analysis). In fact, a majority of the methods are currently used today in spatial epidemiology and represent the cutting edge in methodological development.

Spatial epidemiological methods are of particular interest to public health practitioners and researchers since variations of human health have a strong spatial and temporal component (Rushton, 2003). Similar to Exploratory Spatial Data Analysis (ESDA), one of the first steps in analyzing health data is to visualize the spatial distribution of the outcome or exposure variables. As an early example of this prescribed methodology, we can look to John Snow’s foundational study of cholera in 1849 London, where cases were mapped on a canvas overlaid with broken water pumps (Johnson, 2006). Contemporary methodological frameworks still abide by the same process; explore the data with spatial cluster analysis methods or visualization tools to identify patterns, followed by tests for correlations with potential spatially-varying risk factors. Selection of the appropriate method depends on a number of factors related to the
spatial nature of the data, not least among them is the spatial unit of analysis the data was collected at – e.g., address, census tract, dissemination area, state.

Data represented at a fine spatial scale, such as home address points or postal code points, are rarely available due to confidentiality restraints (J. R. Meliker & Sloan, 2011). While the past five years has brought an increase in the amount of data released through government held disease or health insurance registries, accessing data at this resolution still remains a hurdle. In spatial cluster analysis studies this can prove to be particularly problematic since events may cluster at much finer, local scales than what an aggregated scale may reveal (Jaymie R. Meliker, Jacquez, Goovaerts, Copeland, & Yassine, 2009). As touched on in previous sections, a similar issue presents itself in the built environment and obesity literature, in the lack of etiological information for some built environment features. Given the slow adoption of spatially explicit methods in public health research, even while several studies pinpoint spatial methods as a useful alternative, or supplement to aspatial methods, spatial epidemiological methods may be able to contribute to the growth and diversity of built environment and obesity research (Auchincloss, 2012; Mobley, et al., 2004; Papas, et al., 2007). Paper 1 is a scoping review that will explore spatial clustering methods suitable for handling address point data.

1.4. Research Objectives

The overall objective of this thesis is to contribute to two fields – spatial epidemiology and built environment and obesity research – by providing guidelines for applying spatial methods to health research through a review and relevant research example. Additionally, this thesis will contribute to the growing discourse surrounding environmental determinants of adult obesity, and more broadly health, in Canada. The following research objectives provide an outline for accomplishing the overarching goals set forth:

i) Identify currently used spatial clustering methods in spatial epidemiology suitable for handling address point data through scoping review methodology.
ii) Highlight methodological considerations for research and practical applications of spatial clustering methods.

iii) Explore spatial variations of obesity in two major Canadian Metropolitan Areas using a spatial epidemiological framework.

iv) Explore potential correlates of obesity in the built environment.

In addition to these objectives, research results and data development from paper two will directly inform ongoing work in a larger study that aims to explore the perceived and objective environmental determinants of obesity, physical activity, and metabolic risk factors.

1.5. Thesis Outline

This thesis has four chapters. The first chapter introduced the core concepts to form the context for the research throughout. Methods commonly used to measure the built environment and obesity and relevant research gaps and limitations were also discussed, introducing the rationale behind the research done in chapters two and three. Chapters two and three are research papers that will be submitted for publication in peer-reviewed journals. Chapter two has already been submitted and is waiting to be reviewed, while chapter three is currently in the editing phase before being submitted.

The research presented in chapter two reviews all published material, in the past ten years, that applies spatial clustering methods to address point data. Along with a critical review of the methodological considerations for applying these methods, this paper also demonstrated how some of the methods are used, with a specific emphasis to supplement the themes highlighted in the research. Recommendations are made for future research in the area of spatial epidemiology with respect to spatial clustering method application to research and in practical settings.

Research in chapter three explores the relationships between the built environment and obesity in two Canadian Metropolitan areas. GIS is used to calculate
objective measures of the built environment. Exploratory Spatial Data Analysis is performed on a representative sample from communities in Metropolitan Hamilton and Vancouver. Spatial clustering of obesity is explored followed by a Geographically weighted regression (GWR) analysis of built environment features.

Chapter four summarizes the key findings of the research papers presented in chapters two and three, and describes research contributions of this thesis. Suggestions for future research are also presented in chapter four.
2. A Scoping Review of Spatial Cluster Analysis Techniques for Point-Event Data

2.1. Abstract

Spatial cluster analysis is a uniquely interdisciplinary endeavor, and so it is important to communicate and disseminate ideas, innovations, best practices and challenges across practitioners, applied epidemiology researchers, and spatial statisticians. In this research we conducted a scoping review to systematically search peer-reviewed journal databases for research that has employed spatial cluster analysis methods on individual-level, address location, or x and y coordinate derived data. To illustrate the thematic issues raised by our results, methods were tested using a dataset where known clusters existed. Point pattern methods, spatial clustering and cluster detection tests, and a locally weighted spatial regression model were most commonly used for individual level, address location data (n=29). The spatial scan statistic was the most popular method for address location data (n=19). Six themes were identified relating to the application of spatial cluster analysis methods and subsequent analyses, which we recommend researchers to consider; exploratory analysis, visualization, spatial resolution, aetiology, scale, and spatial weights. It is our intention that researchers seeking direction for using spatial cluster analysis methods consider the caveats and strengths of each approach, but also explore the numerous other methods available for this type of analysis. Applied spatial epidemiology researchers and practitioners should give special consideration to applying multiple tests to a dataset. Future research should focus on developing frameworks for selecting appropriate methods and the corresponding spatial weighting schemes.

Keywords: Spatial Clustering, Spatial Epidemiology, Cluster Detection
2.2. Background

Epidemiologists are keenly interested in understanding disease patterns in both space and time. John Snow’s foundational investigation in 1849 London, where cholera cases were visually clustered around a water pump suspected as the source of disease (at a time when many believed cholera transmission to be airborne), is now recognized as the beginning of spatial epidemiology (Johnson, 2006). Contemporary methods in spatial epidemiology are more advanced, and the field is growing increasingly multi-disciplinary. Public health, spatial statistics and geographic information science (GIS) have contributed more recently to spatial epidemiology, creating an emphasis on interdisciplinary collaboration and knowledge translation (Beale, Abellan, Hodgson, & Jarup, 2008; Elliott & Wartenberg, 2004; Moore & Carpenter, 1999). A commonly used methodology, and one that has been greatly enhanced by these linkages, is spatial cluster analysis (Besag & Newell, 1991; Openshaw, Charlton, Wymer, & Craft, 1987). Defined by the Center for Disease Control as “an unusual aggregation, real or perceived, of health events that are grouped together in time and space,” a cluster can occur in several health classifications and data types; population-based (e.g., disease rates) (G. M. Jacquez & Greiling, 2003), event-based (e.g., point locations) (Schuurman, Peters, et al., 2009), field-based (e.g., continuously distributed observations) (Rothman, 1990) or feature-based (e.g., points aggregated to boundaries) (Mostashari, Kulldorff, Hartman, Miller, & Kulasekera, 2003). For every data type numerous spatial cluster analysis methods exist and vary broadly with respect to assumptions and interpretation (G. Jacquez et al., 2005; Kulldorff, 2006; Moore & Carpenter, 1999). It is speculated that since the development of early algorithms, hundreds of new methods, and variants of existing ones have been introduced, providing researchers with more robust statistical and analytical capabilities (Kulldorff, 2006).

Spatial epidemiology is a large tent that encompasses many disciplines (e.g., disease surveillance, public health, veterinary epidemiology, disease mapping), each of which has been separately pursuing research in cluster analysis (Luc Anselin, 1988; Brown, 1982; Clark & Evans, 1954; Gatrell, Bailey, Diggle, & Rowlingson, 1996; Getis, 2008). Building and transferring knowledge within spatial epidemiology and across other disciplines is imperative for applied researchers and practitioners to utilize the most.
recent developments in the field. The scoping review is one vehicle for such an information exchange.

Review papers (Chung, Yang, & Bell, 2004; Elliott & Wartenberg, 2004; Moore & Carpenter, 1999; Paez & Scott, 2004) act as “state of the science” reports to highlight recent innovations and trends of the discipline. In a similar vein, method comparisons (Aamodt, Samuelsen, & Skrondal, 2006; L. Duczmal, Kulldorff, & Huang, 2006; Kulldorff, Tango, & Park, 2003; Al Ozonoff et al., 2005; Yao, Tang, & Zhan, 2011), and simulations highlight parameterization caveats (M. A. Costa & Assunção, 2005; Sadahiro, 2003), statistical power (Kulldorff, et al., 2003), and practical issues and are beneficial for promoting knowledge transfer among users and developers. Though new methods are frequently tested, developed, and released as packages or standalone applications, a remaining limitation is the lack of methods available through graphical user interface-based applications and accompanying documentation for them. Moreover, implementation outside of graphical user interface applications requires experience with advanced statistical programming tools. As most emerging methods are simulated using synthetic data (Jaymie R. Meliker, et al., 2009; D. C. Wheeler, 2007), determining the efficacy of methods when tested against real-data is less certain, and performance measures and implementation issues borne from the uncertainty and variation of real-data is not frequently readily assessed (Jaymie R. Meliker, et al., 2009; Al Ozonoff, et al., 2005). Point-event data are also referred to as spatial point process data, for which many methods exist (Getis & Franklin, 2010); however few see use in epidemiological analysis since aggregated health data is more accessible to researchers – i.e., county level for US states or Health Regions in Canada. Because there are relatively few examples of point-event methods used in applied spatial epidemiological studies when compared to studies using methods based on aggregate data, it is important for researchers to know how they work and understand issues that may arise, in order to effectively evaluate and optimize method selection, parameterization, and interpretation when such datasets are available (e.g., animal health surveillance data). This paper aims to review research that has used spatial cluster analysis methods on individual-level, address location data, and highlight important issues for applied spatial epidemiology researchers and practitioners to consider when using this type of analysis.
Scoping reviews are a useful methodology for exploring a question or topic where little knowledge is currently established, highlighting research gaps and potential avenues for future studies (Arksey & O'Malley, 2005; Levac, Colquhoun, & O'Brien, 2010). As opposed to systematic reviews where quantitative analyses may be employed to glean trends in literature, scoping reviews assess the qualitative content of literature through concept and thematic mapping (Levac, et al., 2010). Our objective was to review all published literature that utilized spatial clustering techniques for point-event data. An ancillary goal was to call attention to the basis for spatial cluster analysis method selection and the issues therein by proposing several key themes recurrent throughout the papers and also illustrate the application of a selection of methods using real data (e.g., non synthetic).

2.3. Materials and Methods

2.3.1. Selection of Search Terms and Papers

Studies that fulfilled the following criteria were included in the review: the use of at least one spatial cluster analysis technique that analyzed individual-level, address location data, or data derived from real-world geographic coordinates. Operationalization of the term spatial cluster analysis is imbued with ambiguity and results would have been superfluous if specific definition of the term was not used in the selection criteria. By constraining our search to this term, we aimed to select all papers used in applied spatial epidemiology since 2000. We included all papers that employed the use of spatial statistical and geostatistical techniques. Local and global methods were included, as both classes of methods analyse the data at the individual level, and only differ in the scale at which they are evaluated. Methods were then further categorized based on the type of data: areal, point, and line. All studies that used point features to represent individual level, address location data in their analysis were included. Figure 2-1 illustrates the process of the scoping review.
Our review was constrained to papers that applied a spatial clustering method to real data as our focus was on highlighting practical issues and impediments. Methodology papers were used as a means to develop a background for comparison of each study and to generate a table illustrating common analysis themes. Search terms were extracted from a range of academic papers that employed clustering methods in order to ensure coverage of diverse disciplinary language regarding methodology. Our temporal window for acceptance was set from the year 2000 to present.

Figure 2-1. Flow-chart illustrating scoping review processes

Examples of search terms:
- "spatial cluster analysis" OR "cluster analysis" OR "local spatial autocorrelation" OR "cluster detection" OR "geographical cluster"
- AND
- "spatial epidemiology" OR "geographical information systems" OR "spatial scan" OR "georeferenced" OR "spatial stat"

1) at least one spatial or spatio-temporal clustering or point pattern analysis method was used in the study.
2) The spatial resolution of the data used was collected at the individual level.
3) Individual level data was georeferenced to real-world coordinates or address location.
4) Screen titles, abstracts, method, and figures to identify articles meeting the IFE criteria.
5) Discrepancies between definition of point address location (i.e. village centroids and postal code centroids) necessitated further reading of methods.
6) Authors met to determine article selection

1) Development of an extraction form for systematic and efficient categorization of data
2) Emerging themes from reading were iteratively developed and added to the extraction form.
3) Meetings were scheduled to discuss alignment of potential categories with prior literature, as well as overall relevance.
Once databases were selected (Medline, Web of Science, Science Direct, Academic Search Premier, Jstor, Criminal Justing Abstracts, Global Health, BIOMED Central, CINAHL, TOXNET, and Environment Complete), we started our searching process beginning with the formulation of search terms. Final results were compiled and independently reviewed by reading abstracts, and certain sections of the articles (i.e., methods, introduction, results, and figures) that revealed relevant details regarding spatial data resolution. Articles were then selected based on a set of inclusion and exclusion criteria. Systematic screening of the content was performed; we determined which articles to include in the charting and collating stage. Themes were iteratively identified and extracted based on authors' knowledge and published go-to papers (i.e., review or 'state of the science' papers). Consensus was reached for inclusion of themes over several meetings and subsequent content was collaboratively generated. Our goal in identifying themes was to summarize overall patterns of each paper's method implementation aspects. The identification of themes was also guided by the authors' expert knowledge and key review papers. A detailed description of stages can be found in Figure 2-1.

2.3.2. Method Testing

Empirical Spatial cluster analysis methods identified in the scoping review process were run on real data to supplement the thematic discussion produced by the scoping review, and to provide a visual example of some of their strengths and/or limitations. Methods were not tested for power to detect accurate clusters, or for sensitivity.

Our dataset was derived from a database of address level data of severe injuries for the province of British Columbia for years 2001-2006 called the British Columbia Trauma Registry (BCTR). Our study area was limited to Metro Vancouver, British Columbia, Canada so as to restrict the included population to predominantly urban areas. All incidents in the study area were geocoded with 95% accuracy. The BCTR codes all severe injuries using the ICD-10 classification system, categorizing injuries based on the nature of the injury (i.e., auto collision, pedestrian). Previous work by Schuurman et al. (2009) detected significant clusters of severe pedestrian injury in Metro Vancouver. Our exposure variable for this demonstration was also limited to severe
pedestrian injury. For each method, analysis using pedestrian injury data was run where applicable, allowing comparison amongst multiple methods using the same dataset. All local and global methods were compared separately.

Where controls were required, they were estimated by randomly sampling all intersections and street segment midpoints. We do not recommend the use of this method for control sampling in pedestrian injury studies, but chose it based on convenience while considering the focus of this paper.

2.4. Results

In this scoping review our initial search returned (n=945) papers. After setting the inclusion and exclusion criteria and screening each abstract and method section of the papers for key words, we extracted (n=29) papers from our initial search. We found that point pattern methods, spatial clustering and cluster detection tests, and a locally weighted spatial regression model were most commonly used for individual level, address location data (n=29). Table 2-1 provides a summary of each method outlining the application details associated with each method along with software, disciplines utilizing the method, and the relevant article citations from our search.
Table 2.1. Table summarizes the methods used by papers in the review. Other information on methods include data type and resolution required for analysis, software available for implementing the method, and disciplines that utilized the method. See Appendix 2.1 for article translation.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>DATA TYPE</th>
<th>SPATIAL RES.</th>
<th>SOFTWARE</th>
<th>DISCIPLINE USING METHOD</th>
<th>ARTICLES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-function</td>
<td>Case-event/Case-</td>
<td>Point</td>
<td>R¹, ArcGIS, MATLAB, ClusterSeer</td>
<td>Environmental Public Health; Disease Surveillance</td>
<td>26, 2, 7, 12, 4, 8, 11, 16, 19</td>
</tr>
<tr>
<td>Cuzuck-Edwards Test</td>
<td>Case-control</td>
<td>Point/Areal</td>
<td>Space Time Intelligence System</td>
<td>Cancer Epidemiology; Veterinary Epidemiology</td>
<td>15, 28, 8, 17</td>
</tr>
<tr>
<td>Nearest Neighbour Index</td>
<td>Case-event</td>
<td>Point</td>
<td>ClusterSeer, Crimestat¹</td>
<td>Veterinary Epidemiology: Injury Prevention;</td>
<td>15, 17</td>
</tr>
<tr>
<td>Nearest Neighbour Hier.</td>
<td>Case-event</td>
<td>Point</td>
<td>Crimestat</td>
<td>Injury Prevention, Population Health Surveillance</td>
<td>14, 1</td>
</tr>
<tr>
<td>Kernel Density Est.</td>
<td>Case-event/Case-</td>
<td>Point</td>
<td>Crimestat¹, ArcGIS</td>
<td>Criminology; Population Health Surveillance</td>
<td>1, 22</td>
</tr>
<tr>
<td><strong>Local Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel Intensity Method</td>
<td>Case-event/Case-</td>
<td>Point</td>
<td>R¹</td>
<td>Cancer Epidemiology</td>
<td>28</td>
</tr>
<tr>
<td>Local Moran’s I</td>
<td>Continuous</td>
<td>Point/Areal</td>
<td>GeoDa¹, ArcGIS, SpaceStat, R¹</td>
<td>Environmental Science; Epidemiology</td>
<td>10</td>
</tr>
<tr>
<td>Generalized Add. Model</td>
<td>Case-control</td>
<td>Point</td>
<td>R¹, SPSS, S-PLUS</td>
<td>Veterinary Epidemiology; Cancer Epidemiology</td>
<td>19, 21, 24, 25, 23</td>
</tr>
<tr>
<td>Spatial Scan Statistic</td>
<td>Case-event/Case-</td>
<td>Point/Areal</td>
<td>R¹, SatScan¹, ClusterSeer¹</td>
<td>Veterinary Epidemiology; Injury Prevention;</td>
<td>15, 28, 15, 11, 16, 19, 17, 1, 3, 5, 6, 9, 18, 20, 22, 28, 27, 29, 13</td>
</tr>
</tbody>
</table>

Note: 1. free software; 2. limited parameters and models; 3. refers to many types of disciplines; 4. global and local; N= strengths and limitations are derived from the articles T= Tested Method.
2.4.1. **K-function**

The K-function (Ripley, 1976) was the most commonly used global clustering method (n=8). The primary use of K-function analysis was exploring the presence and scale of spatial clustering of the selected exposure variables (Austin et al., 2005; Day & Pearce, 2011; Hillier et al., 2009). The K-function was also used to assess the spatial structure of a distribution before conducting local analyses of spatial clustering (Broman, Shum, Munoz, Duncan, & West, 2006; Epp, Argue, Waldner, & Berke, 2010; Han et al., 2004; Ngowi et al., 2010; Poljak et al., 2010; D. C. Wheeler, 2007). Knowing the scale and structure of the spatial dependency among data helps the user confirm whether local analyses are required as well as provide an approximation of spatial weight specifications. Among the reviewed papers two variations of the method were used; univariate K-function (Gatrell, et al., 1996) and K-function difference (Cuthbert & Anderson, 2002). The univariate K-function method is best suited for case-event data, and the K-function difference, or bivariate K-function, is best suited for case-control data.

Our illustration of both K-function methods shows that both datasets are clustered. Figure 2-2 illustrates the results of K-function methods applied to the BC injury data. Highlights of the outputs are the differences between results when comparing homogenous to inhomogeneous univariate k-function, and the graph as a visual utility to describe the spatial structure of the dataset. Analyses were done using the splancs (Rowlingson & Diggle, 1993) and spatstat (Baddeley & Turner, 2005) libraries of R statistical programming software (R Development Core Team, 2012).
Figure 2.2.  **K-function Global Clustering Results**

Note. Image illustrates the use of the K-function as a multiscale global spatial clustering tool where the observed curve is above the theoretical, clustering is apparent at the corresponding distances. For significance testing the inhomogeneous K-function, **(A)** assumes a non-stationary point process, whereas the ordinary K-function, **(B)** assumes a stationary point process. For a more intuitive value for K, the transformation of the values to an L function is widely applied. Simulations (n=99) were used for significance testing. **(C)** Illustrates the use of the K-function difference method when case-control data are available. When dotted blue and red curves move outside of the confidence band, significant clustering (p=0.05) is apparent in the dataset. We used two different packages in R to compare their efficacy. 99 simulations were used for Monte Carlo significance testing.

2.4.2. **Nearest Neighbour Statistics**

Overall, Nearest neighbor-based methods - Nearest Neighbour Index (NNI) (Clark & Evans, 1954), Nearest Neighbour Hierarchical (NnH) (Levine, 2006) and Cuzick
Edwards test (Cuzick & Edwards, 1990) were the 2nd most common class of global methods used in the papers reviewed. Nearly all papers were published from the health-related research disciplines (Andrade et al., 2004; Epp, et al., 2010; Lai, Low, Wong, Wong, & Chan, 2009; Jaymie R. Meliker, et al., 2009; Pasma, 2008; D. C. Wheeler, 2007). Papers that used case-event data utilized the NNI and NnH methods. Papers that used case-control data utilized the Cuzick Edwards test (Epp, et al., 2010; Jaymie R. Meliker, et al., 2009; Pasma, 2008; D. C. Wheeler, 2007). Since they are global methods, the tests do not identify locations of clustering, rather the potential scales at which the distribution may exhibit dependence or association. Overall, nearest neighbour-based methods provide a similar function to the K-function in the way of a global analysis, but differs based on the definition of spatial neighbours and scale – i.e., spatial weights.

Our test of the Cuzick Edwards method indicates significant global clustering for all levels of K nearest neighbours (k-NN) (Table 2-2). Simulations (n=999) were used for Monte Carlo significance testing. Compared to the K-function, Cuzick Edwards defines spatial relationships in terms of nearest neighbours and not distance. This method offers an alternative exploratory approach for global clustering. Analysis was done in Clusterseer 2.3 (www.biomedware.com).

**Table 2-2. Cuzick-Edwards global clustering method results.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>357</td>
<td>148.812</td>
<td>157.013</td>
<td>16.6145</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>297.625</td>
<td>326.757</td>
<td>19.7149</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>937</td>
<td>446.437</td>
<td>500.548</td>
<td>21.9266</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>1200</td>
<td>595.25</td>
<td>677.412</td>
<td>23.2354</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>1463</td>
<td>744.062</td>
<td>859.402</td>
<td>24.5241</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note. The first column (k) indicates the number of spatial neighbours used. The test statistic is denoted by T(k), and describes the amount of neighbouring cases from k. The Expected value of the test statistic E(Tk) is denoted as E(Tk). The Variance of the test statistic around the mean value is denoted as Var(T). Bonferroni and Simes corrections are applied for multiple testing. The null hypothesis of complete spatial randomness was rejected at all levels of k-nn.
2.4.3. **Local Moran’s I**

One paper utilized the Local Moran’s I (LMI) (L. Anselin, 1995) statistic (Gruebner et al., 2011). Gruebner (2011) applied a suite of Moran’s I statistics (global univariate/bivariate Moran’s I and local univariate/bivariate Moran’s I) to explore self-rated mental health and health determining factor scores at the household level in Dhaka slums. Among the variables tested, both global and local spatial autocorrelation was evident, with the intensity of the dependency decreasing as the amount of nearest neighbours increased. LMI is a powerful tool for detecting both spatial clusters and spatial clustering. Due to the type of data used for our method testing (case-control), we excluded LMI from our analysis of pedestrian injury.

2.4.4. **Kernel Estimation**

A predominate use of kernel estimation approaches was for visual exploration of a dataset (Andrade, et al., 2004; D. C. Wheeler, 2007). The visual analysis provided through the KDE (Silverman, 1986) method makes reference and communication of results intuitive (Andrade, et al., 2004). Outputs from the Kernel Density Estimation (KDE) method also provide evidence of visual ‘hotspots’ to the researcher for subsequent analyses (Lu, 2006). Similar benefits can be realized from the Kernel Intensity Function (KIF) approach (Kelsall & Diggle, 1995). Visual outputs communicate log relative risk ratios of cases and control data (D. C. Wheeler, 2007).

Our tests using these two methods (Figure 2-3) experimented with multiple bandwidth parameter settings; however for space reasons we chose to only include a few different settings. In general, results for KDE and KIF approaches are similar; both readily identified where elevated intensities of pedestrian injury were. A minor difference in outputs is seen in the KIF method however, as the underlying population distribution is included (controls), depicting a slightly different risk surface than the KDE approach. Statistical analyses were carried out using the splancs (Rowlingson & Diggle, 1993) and spatialkernel (Zheng & Diggle, 2009) libraries of the R statistical programming software (R Development Core Team, 2012). All images were imported to ArcGIS 10 (ESRI, 2009) for representation and layer overlay operations. A slight variant of the KIF
approach, the spatial relative risk function, is also accessible through the R package Sparr (Davies, Hazelton, & Marshall, 2011)

Figure 2-3. Kernel-based Methods Results

Note. Kernel density estimation (KDE) and kernel intensity function (KIF) surfaces generated using varying bandwidths (both). Representation illustrates the flexibility of using the KDE and KIF approaches serving as both visual and statistical tools. (A) KDE, (B) KDE, (C) KIF, (D) KIF. Image (D) indicates areas of elevated significant relative-risk at p-value<.015.

2.4.5. Generalized Additive Model

Generalized additive models (GAM) are a type of generalized linear model (GLM) (T. Hastie & Tibshirani, 1987; Kelsall & Diggle, 2002) that extend GLMs by adding a smoothing function to account for geographic space. GAMs have most recently been used in spatial epidemiology and disease risk mapping (Al Ozonoff, et al., 2005). Of the 29 papers reviewed, five used GAMs. All five papers utilized the GAM approach for exploratory purposes (Poljak, et al., 2010; Siqueira-Junior et al., 2008; V. Vieira, Webster, Weinberg, & Aschengrau, 2009; V. M. Vieira et al., 2010; Verónica M. Vieira,
Webster, Weinberg, & Aschengrau, 2008). A primary goal in epidemiology is the
explanation of processes generating spatial and temporal patterns of disease and
disease risks. GAMs can be applied in a space only, spatial-temporal, or time only
frameworks (Poljak, et al., 2010; V. Vieira, et al., 2009) and have been cited to be
particularly useful for analyzing longitudinal data that incorporate residential history
patterns (V. M. Vieira, et al., 2010). The spatial output of the variety of analyses is
communicated through a smoothed risk surface, aiding visual recognition of patterns, a
technique often used as an exploratory spatial data analysis (Poljak, et al., 2010). The
prime advantage of GAMs is the ability to control for the underlying population
distribution from spatial control locations – similar to the KIF - as well as covariates.

Illustration of the GAM yielded a similar visual display as the KIF and KDE,
mainly due to smoothing parameters used in the method (Figure 2-4). The smoother
chosen was the locally weighted scatter plot smoother (LOESS), and was applied to the
x and y values of each case and control location. Akaike Information Criterion (AIC) was
used to determine the optimal span size of .50. Only crude odds ratios (OR) were
calculated as covariates were not available. Statistical analysis was carried out using the
gam (Trevor Hastie, 2011) library of R statistical programming software (R Development
Core Team, 2012).
Generalized Additive Model

Note. Generalized Additive Model surface generated using optimal span size of .50. Statistically significant (p=0.05) hotspots (high odds ratios) are coloured in red tones, and coldspots are in bluer tones.

2.4.6. Spatial Scan Statistic

Over half of papers in this review applied the spatial scan statistic to examine the spatial patterns of address location data (Andrade, et al., 2004; Bautista et al., 2006; Brooker et al., 2004; Chaix et al., 2006; Epp, et al., 2010; Ernst, Adoka, Kowuor, Wilson, & John, 2006; Han, et al., 2004; L. Huang, Stinchcomb, Pickle, Dill, & Berrigan, 2009; Jaymie R. Meliker, et al., 2009; Ngowi, et al., 2010; Pasma, 2008; Polack et al., 2005; Poljak, et al., 2010; Pollack et al., 2006; Sarkar et al., 2007; Tanser, Bärnighausen, Cooke, & Newell, 2009; Warden, 2008; Westercamp et al., 2010; D. C. Wheeler, 2007; Winskill, Rowland, Mtove, Malima, & Kirby, 2011). Of the reviewed articles, 83% applied a Bernoulli model spatial scan statistic to case-control data; two of those articles also used other models in SatScan that can be applied to address location data, the Discrete Normal Continuous model (L. Huang, et al., 2009) and the Discrete Poisson Continuous
model (Ngowi, et al., 2010), Ordinal Model (Westercamp, et al., 2010) and the Multinomial Model (Westercamp, et al., 2010). A clear categorical distinction between applications was the scale of the study area in which the spatial scan statistic was used; large-scale (state, regional and metropolitan areas) versus small-scale (small areas, neighbourhoods, villages).

For different scales, the spatial scan statistic allows the user to adjust this setting based on the minimum or maximum percent of the population to include in relative risk ratio calculations; radius of the scanning window; and proximity from the center of the circle. Other papers have examined the difference between window settings (J. Chen, Roth, Naito, Lengerich, & MacEachren, 2008; Jackson, Huang, Luo, Hachey, & Feuer, 2009), but none have addressed the ability for the spatial scan statistic to detect clusters of the same phenomena at different scales.

Analysis of the injury data using the spatial scan statistic returned several significant clusters, with the most significant located in the same general region as previous tests of KDE, KIF, and GAM (Figure 2-5 and 2-6). For this test a similar approach to the KDE and KIF methods was used; exploring the results using varying definitions for maximum cluster size. We defined maximum cluster size in two ways; maximum percentage of population in scanning window and maximum scanning window radius. Cluster maps were imported to ArcGIS for overlay with streets and study area boundary.
Figure 2-5.  **Spatial Scan Statistic Results**

**Note.**  Spatial Scan Statistic surface generated using different maximum cluster size definitions based on maximum percentage of population at risk (top two images) and maximum distance of scanning window radius (bottom two images) assuming a Bernoulli probability model.  Significance of pedestrian injury cluster tested at $p=0.05$ and indicated with orange circles.
Figure 2-6. Spatial Scale Example

Note. Spatial Scan Statistic surface generated using different spatial weight conceptualizations based on maximum percentage of population at risk and maximum distance of radius assuming a Bernoulli probability model. Significance of pedestrian injury cluster is tested at $p=0.05$ and indicated with orange circles. Representation illustrates how scale of study area can affect spatial cluster analysis results, suggesting that spatial window settings should be tailored to the unique study objective in order to retain the method’s efficacy.

2.5. Themes

Our results summarize all peer-reviewed journal papers from the year 2000 to 2010 that met our selection criteria of using the term ‘spatial cluster analysis’ to examine the spatial patterning of point-event data. Application of spatial cluster analysis to point-event data is mainly carried out through means of exploratory methods, emphasizing the power of visualization. Authors seem to be aware of data resolution issues, but we maintain that consideration of other terms such as exploratory analysis, visualization, etiology, scale, spatial weights, and method selection should also be considered. These
themes are not to be taken as strict guidelines for conducting spatial cluster analysis, but rather, they are recommendations from the author’s knowledge informed from key review papers and expert knowledge. Themes are summarized in Table 2-3.

Table 2-3. Identified Themes

<table>
<thead>
<tr>
<th>METHODOLOGICAL FOCUS</th>
<th>Exploratory Analysis</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Several papers utilized multiple methods to investigate the spatial phenomena with a closer lens. Adopting an ESDA provides more in-depth analysis because multiple spatial cluster analysis methods are generally adopted. Being able to compare the dataset among various spatial methods enhances the researchers understanding of the data, better informing their inference of patterns that may arise in the dataset.</td>
<td>Spatial cluster analysis methods that incorporate visualization in their outputs are advantageous in research and practice settings. Kernel density estimation, kernel intensity function, and generalized additive models were adept for achieving this objective. Visualization is also an important step in an ESDA process.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DATA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Resolution</td>
<td>Spatial resolution is an important component in the process of selecting methods for analysis and has positive and negative implications for subsequent analysis. Acquisition of individual-level, point-event data is absolutely necessary to prevent obfuscation of local spatial clustering. MAUP and EF are two common spatial data issues that may arise due to coarseness in the dataset.</td>
</tr>
<tr>
<td>Etiology</td>
<td>Various data types and spectrums of spatial resolution can reflect the known etiology of the studied phenomena. In some situations a single address level measurement (e.g. home address) may not be an accurate surrogate for exposure. Residential histories that include data on amount of times moved, and years spent at specific locations have proved an adequate surrogate for latency in disease.</td>
</tr>
<tr>
<td>Scale</td>
<td>An overarching goal of spatial cluster analysis is to understand the spatial structure, or scale of the studied processes. Equally important is determining optimal scale to analyze the data, or study area boundary extent, as it has been suggested in the literature that variable units of space can yield different results.</td>
</tr>
<tr>
<td>Spatial Weights</td>
<td>Varying conceptualizations of space will yield different results for spatial clustering and cluster detection. Likewise, each spatial clustering method will define space with different parameters. For most methods, selecting an appropriate spatial weight conceptualization remains a decision based on researcher discretion and should be heavily considered when synthesizing results.</td>
</tr>
</tbody>
</table>

Note. Table provides an overview of themes identified across the reviewed papers. Each theme can also be interpreted as an issue to consider before choosing a spatial clustering test, as well the subsequent synthesis of results.
We categorized the themes into two divisions that may affect the way methods are employed by the researcher; (1) methodological focus and; (2) data. Methodological focus refers to the reasons for undertaking research using a particular method. Data refers to how the actual dataset may dictate the way a method is selected, or the results of subsequent analyses. Several papers utilized multiple methods to investigate spatial phenomena with a closer lens (n=24), what we are terming as an exploratory analysis. Application of a variety of methods is not too dissimilar from an exploratory spatial data analysis approach (ESDA), wherein the researcher applies multiple spatial analysis techniques to glean spatial patterns from the dataset, and identify associations to be later incorporated in to a model, or confirmatory analysis (Haining, Wise, & Ma, 1998). A caution to this approach is the idea that ‘data snooping or ‘data dredging’ of the dataset via use of multiple methods or testing may lead to spurious post-hoc conclusions about underlying processes (Selvin & Stuart, 1966; Sullivan, Timmermann, & White, 1999). There is a tradeoff between exploratory and confirmatory approaches that can be mediated by the intended objective of the analysis. Where the objective is to delineate spatial clusters to guide further study or generate hypotheses of risk factors, multiple methods may reinforce findings and provide confidence that the located areas are in fact ‘unusual’. However, if the object of analysis includes relationships among variables, for example in the context of fitting a spatial point-process model, a more restrained approach is recommended. The tendency of papers to use multiple methods suggests that the methodological focus of research using individual level, point-event data, and moreover spatial cluster analysis, is closely aligned with ESDA. Figure 2-7 illustrates the use of multiple methods on the same dataset with an emphasis on exploring the dataset. Moreover, visualization, an ancillary utility of ESDA, was also a commonly used analytical support tool or supplement to research (n=29). Among the methods recorded in the review, KDE, KIF and GAM approaches produced flexible and intuitive visual outputs, advantageous for knowledge transfer in research and practice settings. Figure 2-7 provides an example of the tested methods visualization outputs.
Figure 2-7. Exploratory Analysis Example

Note. Representation compares all visualization-capable methods applied to the same dataset. (A) Statistically significant (p=0.05) hotspots (high odds ratios) are colored in red, and coldspots are in blue. (B) Test assumes a Bernoulli probability model with a scanning window of 5% of the population at risk. Significant clusters (p=0.05) highlighted with orange circles. (C) Gaussian kernel with a bandwidth of 300, applied to cases only. (D) Bandwidth of 500 applied to cases and controls. Statistically significant regions (p=0.015) outlined by black contour line.

Spatial data resolution, etiology, scale, and spatial weights are themes that fall in the data category. Spatial resolution of data largely governs which methods can be used for analysis. Because spatial health data is normally aggregated to protect individual confidentiality, it restricts the selection of methods to those that handle data of coarser spatial resolution. It has been shown elsewhere that aggregated data can obfuscate local heterogeneity (Jaymie R. Meliker, et al., 2009; J. R. Meliker & Sloan, 2011; A. Ozonoff, Jeffery, Manjourides, White, & Pagano, 2007) caused by known spatial data issues Ecological Fallacy (EF) (Openshaw, 1984) and Modifiable Areal Unit Problem (MAUP) (A. S. Fotheringham & Wong, 1991). Etiology of a disease process or health event is related to the spatial data resolution as well, but barely acknowledged in the papers reviewed. Like spatial data resolution, etiology can affect which methods are chosen for subsequent analysis. It was suggested among some of the reviewed papers that an individuals’ home address may not be an accurate surrogate measure for some
exposures (L. Huang, et al., 2009; V. Vieira, et al., 2009), and that census tracts or dissemination areas may reflect the level of exposure more accurately. Incorporating a sense of an individual’s activity space (Orellana & Wachowicz, 2011; Zenk et al., 2011), mobility (Signorino et al., 2011) or latency period (V. M. Vieira, et al., 2010) to an exposure may paint a clearer picture of exposure as well. Similarly influential to spatial cluster analysis is the theme scale. Papers in this review were split in terms of the scale of study area for analysis; large scale (counties, metropolitan areas, provinces) and small scale (local neighbourhood areas). It has been suggested that a study area’s boundary extent can affect the likelihood for detecting a true cluster (G. M. Jacquez & Greiling, 2003; D. C. Wheeler, 2007). Papers in our review illustrate this effect, which detected clusters that span several kilometers in diameter throughout a rural area (Ngowi, et al., 2010), or multiple contiguous neighborhoods throughout a metropolitan area (Chaix, et al., 2006). Our method test with pedestrian injury illustrated this particular effect when applying the spatial scan statistic to the same dataset only changed by an increased sample size and study area size (Figure 6). Spatial weights refer to the measure of spatial neighbourhood relationships applied to spatial cluster analysis methods. In our review the most oft-cited issue with regard to spatial weights concerned the spatial scan statistic and its inability to detect irregularly shaped clusters (Bautista, et al., 2006; Brooker, et al., 2004; Chaix, et al., 2006; L. Huang, et al., 2009; Tanser, et al., 2009; D. C. Wheeler, 2007). Since early developments of the spatial scan statistic options have been added to the software that deal with this issue (Kulldorff, Huang, Pickle, & Duczmal, 2006) and other scanning window methods have been developed (Luiz Duczmal et al., 2011; Tango & Takahashi, 2005), however no reviewed papers applied them. Other methods such as KIF and GAM allow for flexible spatial conceptualizations of neighbourhood relationships through the application of smoothers (e.g., bivariate LOESS, spatial adaptive filtering), and bandwidth optimizations based on kernel functions. With regard to the spatial scan statistic, a fixed or variable scanning window setting is set a priori and its effect on cluster detection outcomes can be seen in our tests (Figure 2-5).
2.6. Discussion

Our scoping review on spatial cluster analysis methods for individual level, address location data revealed that there has been an increased use of these methods among a range of research disciplines in the last decade. Based on our initial search for academic papers that fit our broad search criteria, a return of 945 papers used methods related to spatial cluster analysis. Without analyzing the disciplines of our initial results we may be able broadly assume that the final selection of 29 papers is somewhat reflective of the distribution, differing only on spatial data resolution (i.e., boundary vs. address location). In a recent review of spatial analysis methods in spatial epidemiology (Auchincloss, 2012), the authors returned an initial total, -after applying broad search terms - of 5,641 papers, and eventually accepted 206. Their analysis drew from a variety of health disciplines and broadly surveyed all spatial analysis methods that were applied to problems in the respected area. While neither this paper nor the aforementioned provide a clear explanation for the poor ratio between papers searched and papers accepted, specific to our research objectives, we speculate this arises from data availability or inadequate methods to analyze individual level point-event data. Nonetheless the adoption of spatial cluster analysis methods within, and beyond health-related disciplines, continues to be a burgeoning trend (Marcelo Azevedo Costa & Kulldorff, 2009).

Since a 1999 review of spatial analysis methods in spatial epidemiology (Moore & Carpenter, 1999), prevailing methods NNI, Cuzick-Edwards, K-function, and spatial scan statistic remain among the most prominent for conducting spatial cluster analysis of point-event data. In particular, our review found that the spatial scan statistic was utilized for spatial cluster detection in 19 of the 29 reviewed papers. Much has been discussed about the flexibility of the spatial scan statistic (Marcelo Azevedo Costa & Kulldorff, 2009; Kulldorff, 1999; Kulldorff, Huang, & Konty, 2009; Kulldorff, et al., 2006), however, there has been equal amount of research highlighting some weaknesses of the method (Cançado et al., 2010; Neill, 2009; Tango & Takahashi, 2005). Recent uses in spatial clustering methods, GAMs (V. Vieira, et al., 2009) and KIFs (D. C. Wheeler, 2007), illustrate the advantages of generating a smooth risk surface coupled with an intuitive visual output. While there has been an increase in the type of methods being adopted for
analysis at the address level, there remain several unused methods for reasons unbeknownst to the authors (Kulldorff, 2006) (e.g., see Turnbull’s Cluster Evaluation Permutation Procedure, Besag and Newell’s R, Tango’s s flexibly shaped spatial scan statistic, Duczmal’s simulated annealing method, and Bayesian Local Likelihood method) . Furthermore, as our review clearly illustrates, a number of themes germane to spatial data analysis are seldom considered, or at least mentioned by the authors in this review, even as review papers and comparison studies routinely identify and highlight issues to consider (Beale, et al., 2008; Jackson, et al., 2009; Kulldorff, 2006; J. R. Meliker & Sloan, 2011; Moore & Carpenter, 1999; A. Ozonoff, et al., 2007; Al Ozonoff, et al., 2005) . With consideration to these issues we recommend a brief series of research areas for those involved in the discipline to ponder.

2.6.1. Future Research Recommendations

Based on the themes generated from our review we make a few recommendations for future research in spatial cluster analysis, and spatial epidemiology more broadly. Firstly, with regard to exploratory analysis, we recommend that researchers utilizing spatial cluster analysis as an exploratory tool consider using multiple tests to gain a greater understanding of the dataset. Recent approaches proposed in spatial epidemiology (Berke, 2005; G. M. Jacquez, 2009) focus on using multiple methods as a means to explore every possible avenue of the data to rule out false positives or spurious clusters. Second, researchers should not discount the utility of visualization as a supplement to analysis and enhancement to communication and dissemination of results. Most methods reviewed in this paper produce visual outputs of results, yet using visualization as an analysis tool is an alternative that has yet to influence academic research (J. Chen, et al., 2008; Grubesic, 2010).

Third, spatial data resolution and process etiology are two inherently related themes that uniquely impact the design and results of research. It has been suggested in papers from this review that some levels of data resolution are not representative of the etiology of some processes (L. Huang, et al., 2009; V. Vieira, et al., 2009). To answer this call, work on data collection techniques and spatial cluster analysis methods that incorporate a notion of an individual’s activity space and mobility (Orellana & Wachowicz, 2011; Zenk, et al., 2011), or residential history (J. R. Meliker & Sloan,
should be encouraged throughout the disciplines involved with spatial cluster analysis. Building such awareness throughout this research community will also bring more macroscopic issues to the front, such as privacy issues in health research caused by the distribution sensitive, location-specific health data (Boulos, Curtis, & AbdelMalik, 2009; Jaymie R. Meliker, et al., 2009)

Last, selection of appropriate spatial weights, and moreover, selection of a method suitable for particular datasets largely impacts the results from tests. As our test illustrated with the spatial scan statistic scanning windows, how the spatial neighbours are conceptualized can dramatically impact the location and extent of clusters. Recent papers have addressed both of these issues (G. M. Jacquez, 2009; J. R. Meliker & Sloan, 2011), and called for the use of tools to aid users in selecting appropriate methods and spatial weights. We would like to echo their recommendations in light of the large disparity between methods employed and methods available to the user. Providing users with more support around these two objectives will build a greater demand for the use of spatial cluster analysis methods on individual level, address location data.

2.6.2. Limitations

There are inherent limitations to conducting scoping reviews, as there is a subjective nature of setting search terms, and generating thematic categories (Levac, et al., 2010). This is a caveat associated with synthesizing data from various disciplines, and multiple methods of different statistical categories. With regard to search term selection, we attempted to prevent this issue by reading review papers and various methodological articles to extract key language identifiers. We felt that this was sufficient for the task, but also acknowledge there may have been articles excluded because of this (e.g., spatial point process methods, Bayesian disease mapping). It is strongly recommended that any researcher aiming to apply a scoping review methodology perform an exhaustive search for language germane to the respected literature before setting search terms.

A second limitation relates to the decision to exclude methodological papers, such as simulation studies, from the review. We sought to include all papers that have
applied methods to real-data so as to extract methodological concerns raised by researchers who do not necessarily develop some of the approaches we reviewed. Not only did this specific procedure allow us to conclude which methods are used most readily, but also provided a rich contrast between issues identified in methodology papers and issues in more practical settings. We acknowledge that by excluding methodology articles we not only left out a large portion of spatial cluster analysis methods, but also more in-depth methodological issues. A few articles that were contributed by methodology developers provided us insight to those deeper issues, complementing other articles without that focus. An interesting future study would be a review based on methodological papers, with the end goal of outlining a framework for applying spatial clustering techniques to individual-level, address location data.

2.7. Conclusions

The study of spatially dependent variables in space has a long history that spans numerous disciplines and so communication and knowledge transfer between those academic communities is assumed to be a major enabling factor. The discipline of spatial epidemiology, in itself, is an immensely interdisciplinary field, unifying researchers in statistics, public health, global health, environmental sciences, geography, and parasitology, as a small sampling of disciplines. In order to effectively search for, and select an appropriate method(s), it is therefore important to understand how data-related factors will govern the caveats and strengths of each respected approach. Our application of a scoping review technique allowed us to catalogue the approaches and summarize the issues associated with them. By compiling the literature that has applied these techniques, our results not only speak to the growth and diversity of disciplines that apply them, but also highlight the potential to communicate various approaches in spatial cluster analysis. A scoping review methodology presents itself as a useful alternative to systematic reviews, as it strives in identifying broad research gaps and qualitative themes across a narrow subset of a field. It is our intention that researchers seeking direction for using spatial cluster analysis methods, consider the caveats and strengths of each approach, but also explore the numerous other methods available for this type of analysis.
3. Obesity and the built environment: a spatial analysis of communities in Metropolitan Vancouver and Hamilton, Canada

3.1. Abstract

Obesity is a worldwide epidemic that shows no signs of slowing down. Projections in the United States predict nearly 80% of the population to be overweight or obese by 2020. Assessment of obesogenic environments that combine several determinants across the environmental spectrum is absolutely necessary in order to develop meaningful intervention and prevention measures. The built environment is suspected to influence the way humans behave by causing us to be more or less active, or more or less likely to eat nutritiously. This study explored the relationship between obesity and several built environment features across two Canadian Metropolitan areas, Vancouver, BC, and Hamilton, ON. Using individual level data from the PURE study, participants were geocoded to home address and explored for spatial clustering of obesity. The objective built environment was assessed using Geographic Information Systems (GIS). Regression analyses – ordinary least squares (OLS) and geographically weighted regression (GWR) - were run to determine any relationships between BMI and built environment variables. Spatial clustering of significantly high relative-risk for obesity was apparent in two Hamilton communities and two Vancouver communities. OLS models provided weak, but significant evidence of an association with walkability, distance to public space, distance to mixed-use centres, and distance to public recreation courts. GWR models failed to improve model fit from OLS models therefore providing little evidence of a spatially-varying relationship with built environment variables. Additionally, there were clear differences in obesity prevalence and overall built environments between Vancouver and Hamilton Communities. This study underscores the complexity of relationships between obesity and the built environment. The results suggest there is a potential link between obesity and particular built
environment features. However, relationships found in other studies were not as strong or non-existent in ours, and emphasizes the need to develop alternative measures and query different built environment settings.

3.2. Background

Historic growth in the prevalence of obesity and overweight was mostly confined to developed nations, but recent emerging trends in developing nations indicate a global influence. Under the current rate of change, it is estimated that by 2030, obese and overweight individuals will total 1.35 billion worldwide (Kelly, Yang, Chen, Reynolds, & He, 2008). In the United States, where nearly 60% of adults are obese or overweight, 15-year projections predict nearly 80% of adults to be obese or overweight at current incidence rates (Y. F. Wang, Beydoun, Liang, Caballero, & Kumanyika, 2008). As a risk factor for various cardiovascular and chronic diseases, obesity and overweight is associated with substantial direct and indirect medical costs. In Canada, adult overweight and obesity prevalence was estimated at 59% in 2004 (Tjepkema, 2006), and the financial burden of overweight and obesity was a combined (CAD) 4.3 billion, or approximately 2.2% of all healthcare costs (Katzmarzyk & Janssen, 2004). For both the present and future, overweight and obesity prevalence has the potential to create instability for population health and healthcare systems.

The biomedical model of health has been the main conceptual lens for overweight and obesity research, pinpointing individual risk factors, such as genotype, metabolism, energy intake, level of physical activity, and demographic and socioeconomic status as potential effect moderators. While these factors may partly determine individual risk for weight gain, they do not account for the rapid growth of the modern obesity epidemic, where prevalence has risen from an estimated 5% in 1985, to 18% in 2005 (Katzmarzyk, 2002). To address the disconnect of prior conceptual models of obesity and prevailing prevalence rates, research has begun to investigate how multiple systemic factors at various contextual scales of influence can combine with individual level predictors, to create what are known as obesogenic environments (Hill, Wyatt, Reed, & Peters, 2003; T. T. K. Huang & Glass, 2008; Lake & Townshend, 2006; B. Swinburn, et al., 1999). One area of the obesogenic environment is the physical or
built environment, and encompasses settings such as the home, neighbourhood, school, work, or any other settings that are human-made (Kirk, Penney, & McHugh, 2010; Papas, et al., 2007; B. Swinburn, et al., 1999). One of the more common conceptualizations of how the built environment influences individual weight gain is through how activity-promoting the surrounding built elements of a place are, usually measured by street connectivity (Nelson, Gordon-Larsen, Song, & Popkin, 2006), pedestrian infrastructure (Rundle, et al., 2007), or walkability index (Frank, et al., 2006). Our study aims to unpack the directionality and strengths of the relationships between objective measures with overweight and obesity in order to better understand the overall impact of built environment.

The Canadian context of built environment and obesity research has grown modestly over the past decade. In a recent review that catalogued all research of the directly assessed built environment and obesity, only 2 out of 63 published were from Canada (Feng, et al., 2010). Recent publications on Canadian built environments have approached the question with a similar lens as American and international counterparts by using traditional and multilevel regression to assess relationships. Research on built environments have focused on cities and regions all across Canada; Ottawa (S. A. Prince, et al., 2011; Prince et al., 2012), Quebec (Lebel, Pampalon, Hamel, & Thériault, 2009), Hamilton (Merchant, Dehghan, Behnke-Cook, & Anand, 2007), Toronto & Vancouver (Pouliou & Elliott, 2010), Edmonton (Spence, Cutumisu, Edwards, Raine, & Smoyer-Tomic, 2009) and Canada (Ross, et al., 2007). All of these studies are instructive of the current context of Canadian research and have found similarities US studies such as relationships with convenience store and fast-food outlet density (Prince, et al., 2012) and proximity (Spence, et al., 2009) and metropolitan sprawl (Ross, et al., 2007). However, mixed results are prevalent in the research as well with access to public spaces associated with increased odds in obesity for females (S. A. Prince, et al., 2011), and density and access to neighbourhood amenities not significantly related to BMI (Chaix, et al., 2006). Based on the current evidence in Canadian research, there is still a very nascent understanding of these relationships.

Most research on the built environment determinants of overweight and obesity, in both Canada and United States, are based on aspatial methods such as multi-level modelling or regression-based statistics (Papas, et al., 2007). Assuming data to be
homogenous across large administrative boundaries ignores the effect of more localized patterns such as local spatial autocorrelation (Chaix, et al., 2006), and subsequent interpretation of patterns may be distorted by spatial data bias; Modifiable Areal Unit Problem (A. S. Fotheringham & Wong, 1991) and Ecological Fallacy (Openshaw, 1984).

Our study aims to take a different approach by investigating the local level influences of the built environment among individuals residing in nested neighbourhoods. However, rather than using measures derived from the entire neighbourhood, we will use measures of both individual demographics and the objectively measured built environment.

Studies employing spatial analysis to examine obesity and environmental determinants are rare, and only a few have been attempted to date (Mobley, et al., 2004; Schuurman, Peters, et al., 2009). This approach can be advantageous over traditional aspatial methods because spatial autocorrelation in the data is leveraged as an analytical tool and not viewed as a violation of data independence (Papas, et al., 2007). Our study will employ a spatially explicit approach, first with a spatial cluster analysis method (SPARR), then a locally weighted regression technique (Brunsdon, Fotheringham, & Charlton, 1996) to examine potential correlation between BMI and the built environment.

The purpose of this study is to explore spatial patterns of BMI in adults aged 35 – 70 and determine whether those individuals are significantly spatially clustered. This is assessed by applying a visualization based spatial cluster analysis method to seven neighbourhoods of varying densities and demographics throughout Metropolitan Hamilton and Vancouver. The association between the built environment and obesity is then assessed for global and local relationships using a suite of aspatial and spatial regression techniques. Our assumption is that the built environment will have a relationship with individual BMI at the local level.
3.3. Methods and Materials

3.3.1. Study area

For the purpose of this study, communities were identified throughout Metropolitan Vancouver and Metropolitan Hamilton, Canada (Figure 3-1 and Figure 3-2). Our data were acquired from the PURE-BE study (Prospective Urban and Rural Epidemiological Built Environment), a direct assessment of built environment features throughout 55 urban and rural communities of three Metropolitan Areas of Canada: Hamilton, Vancouver, and Quebec City. Due to methodological considerations we chose to limit our analysis to a subset of 27 communities (defined as forward sortation areas (FSA)) in Metropolitan Vancouver and Hamilton that were then aggregated to 7 communities based on spatial proximity and contiguity between boundaries.

Figure 3-1. Metropolitan Vancouver Study Area

Note. Figure represents Metro Vancouver communities and surrounding study area. Boundaries outlined in black represent Census Subdivisions.
3.3.2. Participant data

Individual level data were obtained from the Prospective Urban and Rural Epidemiological (PURE) study, an international longitudinal (2006-2018) investigation of social and environmental determinants of obesity, diabetes and cardiovascular disease. Our portion of the study represents the Built Environment theme and aims to investigate objective and perceived built environmental determinants in three, large Canadian Metropolitan Areas. 8655 adults - aged 35-75 - were recruited from the PURE study and surveyed for additional information; nutrition-related, physical activity behaviour, environmental perceptions (perceived measures of BE). New individual level data were combined with a detailed dataset (International PURE dataset) of community environment, household, behavioural biological risk factors and cardiovascular disease or other chronic diseases. For this study we chose investigate a subset of individual household, behavioural, and biological variables.
Our primary dependent variable of interest was Body Mass Index (BMI), derived from measured height and weight values. BMI was calculated (BMI = weight (kg)/height (m) 2), and adults with a BMI>30 were classified as being obese, and BMI<29.9 represented all other classifications (overweight, normal weight, underweight). Individual level independent variables were also selected for descriptive statistics and regression; Canadian Household Income Classification (<$20,000(1), $20,000-$30,000(2), $30,001-$45,000(3), $45,001-$65,000(4), $65,001-$90,000(5), >$90,001(6)) , education (none, primary education, secondary or high school, trade school, college/university), a score for all types of physical activity (MET score, see IPAQ scoring guide, 2007) and time sedentary (minutes/week).

The spatial unit of analysis for this study was the home address level. We chose this unit based on the use of home address in prior studies (Boehmer, Lovegreen, Haire-Joshu, & Brownson, 2006; Burdette & Whitaker, 2004; Frank, et al., 2004; Lopez, 2012; Miles, Panton, Jang, & Haymes, 2008) and research suggesting the comparative strength of address level data in spatial analysis based research (Jaymie R. Meliker, et al., 2009; A. Ozonoff, et al., 2007). Due to constraints regarding participant confidentiality, we were unable to obtain address-level data from the Metropolitan Quebec City site, and thus omitted this site altogether from our current study. 4405 study participants were geocoded to the street network in their respective communities using ArcGIS 10 (ESRI, 2009). Our initial success rate with automatic matching was 85%. All unmatched addresses were double checked for typos or inconsistencies and rerun through the geocoding tool. Any addresses that could not be matched, due to uncommon street name, were manually coded using Google maps or a tool that provided more alternate street and road names. Manual matching was applied mostly in semi-rural areas where more than one name for a road existed, due to it also serving as a major road or highway. Once we finished geocoding (final percent correctly matched = 94%) we then aggregated communities and classified built environment features into the appropriate communities.

3.3.3. Community Aggregation

When conducting any spatial analysis, large sample sizes are needed, especially for fitting GWR models (Brunsdon, et al., 1996). To achieve larger sample sizes we
aggregated or “merged”, bordering communities with one another. Communities were examined for sample size, urbanity and contiguity with bordering candidate communities. Communities that were in rural areas, or consisted mostly of rural development (farm land or areas with lower intersection densities) were mainly excluded from the study. Communities with a heterogeneous development pattern (i.e., urban and rural) over a large land area were also excluded. This criteria was most evident in communities that had a town or village nested in it. In the end 4,124 participants out of the initial 8,655 were retained, making up 7 communities (Table 3-1 and Table 3-2).

3.3.4. Built Environment Overview

A direct assessment of the built environment in three major Canadian Metropolitan Areas (CMAs) (Vancouver, Hamilton,) was conducted for 42 urban and rural study sites, or communities where participants (N=4124) resided. Two audit tools were utilized; (1) a modified version of the Irvine-Minnesota Inventory (cite) and ; (2) a Physical Activity and Nutrition Features audit tool (for details on audit instruments see (Gasevic et al., 2011)). Each assessor was thoroughly trained on what to look for when collecting the data at a 3-day conference that involved practical application and individual trainee testing. The assessment aimed at capturing all BE features potentially related to active living and obesity within community boundaries defined by Forward Sortation Areas (FSA). To account for edge effects in communities a 500 meter buffer was applied to the edge of the entire boundary, capturing road segments that participants would potentially interact with if they lived near the edge. As mentioned above, we chose a subset of communities in Metropolitan Vancouver and Hamilton only.

3.3.5. Built Environment Variables

In order to calculate BE variables, we first had to convert raw GPS (Global Positioning System) points from .GPX(common file format for GPS data) format to .shp format (common file format for vector GIS data) for import into conventional GIS software. This was conducted for communities throughout the two CMAs. Raw data were compiled and categorized into appropriate aggregations of FSA and integrated with relevant existing datasets, such as streets and land-use (Desktop Mapping Technologies Inc.) and dissemination blocks. Variables were chosen based on key existing literature
(Brownson, Hoehner, Day, Forsyth, & Sallis, 2009; Feng, et al., 2010; Frank, et al., 2004; Papas, et al., 2007) and the availability of required primary and secondary data. Being that our spatial unit of analysis was at the individual, or address level, we also chose to calculate variables that were not derived from areal aggregations higher than dissemination block (DB) level (approximately the size of a city block). The one variable we calculated at the DB level was net population density.

Using ArcGIS 10 BE features were categorized to appropriate FSA groupings to ensure some variables were being calculated beyond proximal or local destinations. Since individual level analyses require buffers to be applied at each participant location to ensure variability in exposure measures, we selected various network buffer distances at 500 and 1000 meters. Specific variables and buffer distances were chosen based on select literature (Feng, et al., 2010) that described the efficacy of the respected parameters in statistical analyses. Calculated variables aimed to represent three anchors of the built environment: (1) local nutrition environment (2) physical activity environment and (3) land-use and design environment. Within these anchors we developed variables that would accurately reflect the density, diversity, and design the aforementioned elements.

**Local nutrition environment and physical activity environment**

All food environment types from the study were derived from the North American Industry Classification System (NAICS) for food establishments and consist of types such as specialty food stores, baked goods stores, meat or fish/seafood stores/market stores, and confectionary or nut stores. Due to the exploratory nature of this research, we chose to limit our food outlet types to places that would carry enough products for a meal. These were fruit and vegetable stores/market stores, and supermarket and grocery stores. For takeaway or eat-in types, we chose to include full-service restaurants and limited-service restaurants.

Physical activity environment features were classified based on NAICS and the study's custom classification system. In this analysis we chose a small subset of physical activity promoting features that are ubiquitous in most environments. We also included community centres or town halls, as they exist to serve local communities and tend to have multiple uses associated with them.
For both types of BE variables we calculated the nearest distance from each participant’s home address to the respected feature type. Using ArcGIS Network Analyst Extension (ESRI, 2009) we calculated network-based distance (meters) from each home address to the nearest “facility” of that type in that participant’s community. If no BE features were located in the community, distances were calculated from BE features in surrounding communities using a larger street network for that respected feature type.

**Land-use and design**

For this study land-use and design variables consisted of BE features and measures that describe an environment’s diversity and design as they relate to pedestrian connectivity, land-use mix, and street configuration. We calculated five variables; (1) land-use mix index, (2) sidewalk length per unit area, (3) intersection density, (4) population density, and (5) a walkability index.

Land-use mix (LUM) was adopted from several studies that have applied it in similar designs (Bodea, Garrow, Meyer, & Ross, 2008) (Frank, et al., 2004; Li, Harmer, Cardinal, Bosworth, & Johnson-Shelton, 2009; Mobley, et al., 2006), and is most widely known as the entropy index. Parcel-level land-use data (Desktop Mapping Technologies Inc.) for study sites were integrated within the dataset and network distance buffers of 1000 meters were calculated from each participant’s home address. We chose 1000 meters for a buffer distance as it represented an approximate average between previously prescribed buffer distances in studies of similar design. Land-use categories present within the network buffers were selected and used for parameters in calculating the LUM equation:

\[
LUM = \left( - \sum_{i} \left[ p_i \ln(p_i) \right] \right) / (\ln k)
\]

LUM examines the distribution pattern of different land uses within a neighbourhood. Where \( p_i \) is the proportion of estimated square footage attributed to land use \( i \), and \( k \) is the number of land uses. This measure represents the evenness of distribution of square footage across 6 types of land uses within a 1000 meter distance from each household. LUM values range from zero to one, with zero representing a
nearly homogeneous land use environment and one representing a perfectly even distribution of land-uses across the buffer area.

Sidewalk length per unit area, intersection density, population density and distance to major road or arterial street were all calculated for the purpose of describing the pedestrian connectivity of a participant’s locale within their community. Sidewalk length per unit area was derived from PURE-BE study/assessment, collected as presence or absence measures, and integrated with the existing road network dataset to give an estimation of sidewalk length. Network buffers (500m and 1000m) were created around each participant’s home address to represent the unit area (square kilometers). Where no sidewalks existed in any buffer distance, participants were given a score of 0. Intersection density, similarly calculated with network buffers (1000m), was considered as the total number of intersections per unit area (square kilometers). We also used intersection density as a term in the walkability index. Population density (1000 m buffer) was also used as a term to calculate the walkability index, and is also considered a proxy measure for dense built environments that are conducive to utilitarian physical activity. Using 1000 m network buffers from home address, dissemination block level population counts were aggregated for the buffer area and divided by the sum of the unit area (square kilometer). All calculations were performed using ArcGIS 10.

A walkability index was calculated as a composite measure of three already calculated variables; LUM, intersection density and population density. Walkability indices have been shown to indicate a community’s conduciveness to utilitarian physical activity and have linked to be associated with obesity (Frank, et al., 2007; Frank, et al., 2006) and physical activity (Owen et al., 2007; Sallis et al., 2009). Z-scores for individual variables were calculated and a composite score for walkability was produced (Frank, et al., 2004):

$$\text{Walkability index} = (6 \times \text{z-score of land-use mix}) + (\text{z-score of net population density}) + (\text{z-score of intersection density})$$

Walkability indices are sometimes calculated with other components such as retail floor area ratio area or residential density, but unfortunately we did not have access to data of this granularity. Walkability outputs were converted to deciles for easier interpretation.
3.3.6. Exploratory Spatial Data Analysis

Spatial Relative Risk

Our first objective was to explore the spatial distribution of obesity in all communities (n=7) while also accounting for the locations of the rest of the sample distribution, or non-obese participants. The adaptive spatial relative risk function, proposed by Davies and Hazelton (2011), is a kernel smoothing method derived from the earlier exploits of Kelsall and Diggle (Kelsall & Diggle, 1995) whereby a relative risk function is calculated based on the ratio of the distribution of cases to controls (the authors are aware obesity is not a disease and only use case/control terminology to respect the proposed methodology). Davies and Hazelton’s approach differs from more widely used methods (i.e., kernel estimation) in that the bandwidth calculation – a smoothing parameter for defining an observation’s spatial influence on proximal observations - is not fixed across all observations, but adaptive. Using an adaptive bandwidth approach has been suggested to be better suited for heterogeneous populations such as humans, particularly when related to health outcomes (Carlos, Shi, Sargent, Tanski, & Berke, 2010). The spatial relative risk function also serves as a descriptive visualization tool allowing researchers to “see” where hot or cold spots may exist in the distribution.

The precise methodology for our analysis was adopted from Davies and Hazelton’s Sparr package (Davies, et al., 2011) for the R statistical computing environment (R Development Core Team, 2012). First, we dichotomized the BMI variable for all participants; 0 for BMI < 29.999 and 1 for BMI > 30. Applying an adaptive isotropic smoothing parameter, analyses were then run for the 7 communities and log-relative risk estimates were created for a 10,000 X 10,000 pixel grid of each community extent. Asymptotic point-wise p-values were computed and tolerance contours were placed on top of the smoothed kernel density surface. The resulting images were a set of visual surfaces depicting the peaks or “hotspots” and “coldspots” of the distribution of individual’s obesity status (not obese vs. obese), providing insight as to whether there were spatial clusters of either in the environment.
Ordinary Least Squares Regression Analysis (OLS)

Using our calculated built environment variables as the independent variables and BMI and physical activity measures as the dependent variables, we ran Pearson correlation tests to ascertain direction and magnitude of any potential bivariate associations. Using a combination of visualization tools (i.e., scatter plots) and exploratory spatial data analysis techniques (e.g., Local Moran’s I) we searched the data for relationships that may present themselves in a spatial context but not in an aspatial. Once variables were selected for each respective community several models were run to evaluate the most optimal models for GWR. Residuals of each model were also mapped assessed for spatial autocorrelation, as both a diagnostic test for model misspecification, and an analytical tool to highlight potential spatial relationships not explained in the model.

Geographically Weighted Regression (GWR)

We used geographically weighted regression (GWR) to test whether relationships between BMI and other covariates – BE features among them - were spatially varying in nature. GWR is a variation of traditional regression in that it estimates local regression coefficients for each location to evaluate whether the dependent variable has a spatially varying relationship with independent variables (Brunsdon, et al., 1996). Aspatial methods, such as multilevel models or logistic regression, typically make assumptions that observations are independent and that each are nested in an aggregate spatial unit (i.e., census tract, dissemination area). On the other hand a GWR approach assumes that observations are spatially autocorrelated, and leverages this aspect to analyze relationships that may exist from local patterns. Instead of producing parameter estimates for the entire model, as with OLS or traditional regression, local parameters are calculated at each data point. This allows for the relative convenience of mapping local coefficient values, local R² values, and residuals, to reveal potential locally varying relationships not evident from the global model (OLS) (Ali, Partridge, & Olfert, 2007).

Models from the OLS selection were applied to the GWR modeling framework. An important methodological consideration when using GWR is the selection of the type and size of bandwidth parameter (A. Stewart Fotheringham, Brunsdon, & Charlton,
2002). We used a gaussian kernel to define the type of relationship with nearby points, and used the smallest corrected Akaike’s Information Criterion (AICc) to determine the size, or range of influence. Once models were run, comparisons of model strength were made by evaluating the adjusted R² values and AICc of the OLS models versus GWR. If there is a difference in AICc value by more than three, the smaller model is considered to be a better fit (A. Stewart Fotheringham, et al., 2002). Both OLS and GWR analyses were run in ArcGIS.

3.4. Results

Across all seven communities we initially geocoded a total of 4405 study participants. After cleaning the data for participants that were missing records in any of the examined fields we were left with a total of 4124. This sample represents roughly half of the total PURE study population (n=8655), and any communities identified as mostly rural have been removed. Information on age and ethnicity were also omitted due to incomplete records. Out of the 4124 participants, 780 (18%) were identified as obese (BMI≥30), and 3344 (72%) as non-obese (BMI<30).

3.4.1. Descriptive Statistics

Across all study sites, 18% of the participants were identified as obese. This figure is comfortably below the most recent national estimate for adults of 24-25% (Canada, 2011). However local level obesity prevalence painted a different picture, with 35% obesity prevalence in Metro Hamilton and 16% prevalence in Metro Vancouver. Compared to their respective provincial estimates, they are both higher; Ontario at 17%; British Columbia at 12% (Canada, 2011). At an even finer scale, the community level, differences in obesity prevalence were more pronounced. In Metro Hamilton #1, the prevalence among males and females was an alarming 45% of the study population, and in Metro Hamilton #3 the prevalence is 33%, still nearly twice the provincial average (Table 3-2). Out of all Metro Vancouver communities, communities #1 (14.6%) and #2 (13.3%) are near the provincial average, while #3 (18%) and #4 (16.5%) are slightly above. Interestingly, mean self-reported physical activity scores between the two
metropolitan study areas displayed an opposite trend, with Hamilton having a higher mean physical activity score than Vancouver.

Table 3.1: Descriptive Statistics of Individual Level Variables for Metropolitan Vancouver Communities

<table>
<thead>
<tr>
<th>Individual Level Variables</th>
<th>Metro Vancouver #1</th>
<th>Metro Vancouver #2</th>
<th>Metro Vancouver #3</th>
<th>Metro Vancouver #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N= 309</td>
<td>N=419</td>
<td>N=1159</td>
<td>N=182</td>
<td></td>
</tr>
<tr>
<td>Non-obese</td>
<td>Obese</td>
<td>Non-obese</td>
<td>Obese</td>
<td>Non-obese</td>
</tr>
<tr>
<td>Female</td>
<td>150</td>
<td>26</td>
<td>114</td>
<td>24</td>
</tr>
<tr>
<td>Male</td>
<td>19</td>
<td>32</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>264</td>
<td>62</td>
<td>138</td>
<td>36</td>
</tr>
</tbody>
</table>
| Note. Values in the same row and sub table not sharing the same subscript are significantly different at p< 0.05 in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances.
Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Bonferroni correction.

Table 3.2 Descriptive Statistics of Individual Level Variables for Metropolitan Hamilton Communities

<table>
<thead>
<tr>
<th>Individual Level Variables</th>
<th>METRO HAMILTON #1</th>
<th>METRO HAMILTON #2</th>
<th>METRO HAMILTON #3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N= 1062</td>
<td>N= 782</td>
<td>N= 211</td>
</tr>
<tr>
<td></td>
<td>Non-obese</td>
<td>Obese</td>
<td>Non-obese</td>
</tr>
<tr>
<td><strong>SEX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>306, 53.6%</td>
<td>265, 46.4%</td>
<td>317, 77.9%</td>
</tr>
<tr>
<td>Male</td>
<td>274, 56.1%</td>
<td>214, 43.9%</td>
<td>279, 74.4%</td>
</tr>
<tr>
<td>Total</td>
<td>580, 53.0%</td>
<td>479, 45.0%</td>
<td>596, 76.2%</td>
</tr>
<tr>
<td><strong>EDUCATIONAL ATTAINMENT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>1, 100.0%</td>
<td>0, 0.0%</td>
<td>1, 100.0%</td>
</tr>
<tr>
<td>Primary</td>
<td>26, 44.8%</td>
<td>32, 55.2%</td>
<td>12, 57.1%</td>
</tr>
<tr>
<td>Secondary or High School</td>
<td>203, 47.4%</td>
<td>225, 52.6%</td>
<td>126, 74.1%</td>
</tr>
<tr>
<td>Trade School</td>
<td>49, 52.7%</td>
<td>44, 47.3%</td>
<td>53, 69.7%</td>
</tr>
<tr>
<td>College/University</td>
<td>301, 62.8%</td>
<td>178, 37.2%</td>
<td>404, 78.6%</td>
</tr>
<tr>
<td><strong>CANADIAN HOUSEHOLD INCOME CLASSIFICATION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under $20k</td>
<td>35, 42.2%</td>
<td>48, 57.8%</td>
<td>3, 50.0%</td>
</tr>
<tr>
<td>$20k - $30k</td>
<td>66, 54.1%</td>
<td>56, 45.9%</td>
<td>15, 55.6%</td>
</tr>
<tr>
<td>$30k - $40k</td>
<td>106, 54.9%</td>
<td>87, 45.1%</td>
<td>52, 76.5%</td>
</tr>
<tr>
<td>$45k - $65k</td>
<td>101, 48.8%</td>
<td>106, 51.2%</td>
<td>85, 77.3%</td>
</tr>
<tr>
<td>$65k - $90k</td>
<td>125, 60.3%</td>
<td>83, 39.7%</td>
<td>140, 76.1%</td>
</tr>
<tr>
<td>Over $90k</td>
<td>146, 59.6%</td>
<td>99, 40.4%</td>
<td>301, 77.8%</td>
</tr>
</tbody>
</table>
2. Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Bonferroni correction.

Characteristics of the BE categorized into groups of non-obese/obese are listed on the bottom half of tables 3-1 and 3-2 and BE characteristics between the two study sites are in table 3-5. For the entire study area of Metro Vancouver, mean distances to nearest facility of all types were significantly different from Metro Hamilton (Table 3-5). On average, participants in Metro Hamilton had to travel nearly twice the distance to access a public recreation court in comparison to Metro Vancouver. Similarly they also had to travel just over 3 times the distance to access a fruit and vegetable market store.

This measurement may be inflated due to the low total amount of fruit and vegetable market stores in the Metro Hamilton study area, but nonetheless could be illustrative of a broader pattern in Metro Hamilton’s BE. For the BE measures that represent the design of the environment, Metro Vancouver had higher measurements for all categories but walkability. This interestingly contrasts with the fact Metro Vancouver has a higher overall land-use mix score (.37), a heavily weighted factor in calculating the walkability index. Metro Vancouver’s population density was also significantly different from Hamilton’s at 5171 people sq/km and 2914 people sq/km, respectively.
Table 3.2. Descriptive Statistics of Built Environment Variables for Metropolitan Vancouver Communities – Mean Values Non-Obese vs. Obese

<table>
<thead>
<tr>
<th>Variable</th>
<th>Metro Vancouver #1</th>
<th>Metro Vancouver #2</th>
<th>Metro Vancouver #3</th>
<th>Metro Vancouver #4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMI</strong></td>
<td>NO</td>
<td>O</td>
<td>NO</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>24.28</td>
<td>32.97</td>
<td>24.44</td>
<td>34.15</td>
</tr>
<tr>
<td>PA (MET score)</td>
<td>5192.10</td>
<td>3827.02</td>
<td>5096.13</td>
<td>4778.22</td>
</tr>
<tr>
<td>TIME SEDENTARY (MIN/WEEK)</td>
<td>2840.69</td>
<td>3312.88</td>
<td>2797.45</td>
<td>3007.41</td>
</tr>
<tr>
<td>FRUIT AND VEGETABLE STORE</td>
<td>1587.47</td>
<td>1501.55</td>
<td>768.97</td>
<td>649.76</td>
</tr>
<tr>
<td>PUBLIC RECREATION COURT</td>
<td>399.30</td>
<td>334.43</td>
<td>319.72</td>
<td>310.67</td>
</tr>
<tr>
<td>PUBLIC/GREEN SPACE</td>
<td>307.51</td>
<td>252.63</td>
<td>288.55</td>
<td>308.16</td>
</tr>
<tr>
<td>FITNESS AND RECREATION</td>
<td>562.75</td>
<td>556.24</td>
<td>755.97</td>
<td>752.72</td>
</tr>
<tr>
<td><strong>SUPERMARKET</strong></td>
<td>497.54</td>
<td>486.80</td>
<td>509.09</td>
<td>486.57</td>
</tr>
<tr>
<td><strong>COMMUNITY CENTRE</strong></td>
<td>806.39</td>
<td>761.86</td>
<td>792.91</td>
<td>693.78</td>
</tr>
<tr>
<td>LIMITED SERVICE RESTAURANT</td>
<td>426.70</td>
<td>409.20</td>
<td>385.61</td>
<td>331.61</td>
</tr>
<tr>
<td>FULL-SERVICE RESTAURANT</td>
<td>484.31</td>
<td>427.01</td>
<td>356.59</td>
<td>312.09</td>
</tr>
<tr>
<td>MIXED-USE CENTRE</td>
<td>1426.65</td>
<td>1451.78</td>
<td>1105.27</td>
<td>1140.83</td>
</tr>
<tr>
<td>WALKING TRAIL</td>
<td>259.64</td>
<td>280.60</td>
<td>326.07</td>
<td>306.64</td>
</tr>
<tr>
<td>SW LENGTH 500M</td>
<td>.054</td>
<td>.053</td>
<td>.048</td>
<td>.049</td>
</tr>
<tr>
<td>SW LENGTH 1000M</td>
<td>.042</td>
<td>.042</td>
<td>.040</td>
<td>.042</td>
</tr>
<tr>
<td>INTERSECTION DEN. (SQ. M)</td>
<td>.000131</td>
<td>.000132</td>
<td>.000172</td>
<td>.000173</td>
</tr>
<tr>
<td>LAND-USE MIX</td>
<td>.37</td>
<td>.41</td>
<td>.37</td>
<td>.38</td>
</tr>
<tr>
<td>POPULATION DENSITY (SQ. K)</td>
<td>6347.53</td>
<td>6778.91</td>
<td>6062.18</td>
<td>6239.88</td>
</tr>
<tr>
<td>WALKABILITY INDEX</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. NO: Non-Obese. O: Obese. Each column is an average. Values in the same row and subtable not sharing the same subscript are significantly different at p< 0.05 in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances.
Table 3-4. Descriptive Statistics of Built Environment Variables for Metropolitan Hamilton Communities – Mean Values Non-Obese vs. Obese

<table>
<thead>
<tr>
<th></th>
<th>Metro Vancouver #1</th>
<th>Metro Vancouver #2</th>
<th>Metro Vancouver #3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>O</td>
<td>NO</td>
</tr>
<tr>
<td>BMI</td>
<td>25.79\text{a}</td>
<td>35.69\text{b}</td>
<td>25.29\text{a}</td>
</tr>
<tr>
<td>PA (MET score)</td>
<td>7346.92\text{a}</td>
<td>7056.29\text{a}</td>
<td>6189.08\text{a}</td>
</tr>
<tr>
<td>TIME SEDENTARY (MIN/WEEK)</td>
<td>2541.10\text{a}</td>
<td>3045.03\text{b}</td>
<td>2624.21\text{a}</td>
</tr>
<tr>
<td>FRUIT AND VEGETABLE STORE</td>
<td>6441.54\text{a}</td>
<td>5741.10\text{a}</td>
<td>5469.91\text{a}</td>
</tr>
<tr>
<td>PUBLIC RECREATION COURT</td>
<td>649.85\text{a}</td>
<td>626.70\text{a}</td>
<td>784.10\text{a}</td>
</tr>
<tr>
<td>PUBLIC/GREEN SPACE</td>
<td>367.44\text{a}</td>
<td>397.14\text{a}</td>
<td>383.36\text{a}</td>
</tr>
<tr>
<td>FITNESS AND RECREATION</td>
<td>1349.60\text{a}</td>
<td>1249.78\text{a}</td>
<td>1196.98\text{a}</td>
</tr>
<tr>
<td>SUPERMARKET</td>
<td>1787.40\text{a}</td>
<td>1693.90\text{a}</td>
<td>1620.56\text{a}</td>
</tr>
<tr>
<td>COMMUNITY CENTRE</td>
<td>725.84\text{a}</td>
<td>661.42\text{a}</td>
<td>1015.28\text{a}</td>
</tr>
<tr>
<td>LIMITED SERVICE RESTAURANT</td>
<td>853.33\text{a}</td>
<td>743.92\text{b}</td>
<td>1237.61\text{a}</td>
</tr>
<tr>
<td>FULL-SERVICE RESTAURANT</td>
<td>1932.34\text{a}</td>
<td>1657.80\text{a}</td>
<td>2778.37\text{a}</td>
</tr>
<tr>
<td>MIXED-USE CENTRE</td>
<td>1664.40\text{a}</td>
<td>1804.99\text{a}</td>
<td>5291.05\text{a}</td>
</tr>
<tr>
<td>WALKING TRAIL</td>
<td>376.59\text{a}</td>
<td>421.89\text{a}</td>
<td>302.96\text{a}</td>
</tr>
<tr>
<td>SW LENGTH 500M</td>
<td>0.034\text{a}</td>
<td>0.035\text{a}</td>
<td>0.019\text{a}</td>
</tr>
<tr>
<td>SW LENGTH 1000M</td>
<td>0.024\text{a}</td>
<td>0.025\text{a}</td>
<td>0.015\text{a}</td>
</tr>
<tr>
<td>INTERSECTION DEN. (SQ. M)</td>
<td>0.000682\text{a}</td>
<td>0.000833\text{a}</td>
<td>0.000790\text{a}</td>
</tr>
<tr>
<td>LAND-USE MIX</td>
<td>3305.95\text{a}</td>
<td>3417.39\text{a}</td>
<td>2174.02\text{a}</td>
</tr>
<tr>
<td>POPULATION DENSITY (SQ. K)</td>
<td>0.36\text{a}</td>
<td>0.37\text{a}</td>
<td>0.32\text{a}</td>
</tr>
<tr>
<td>WALKABILITY INDEX</td>
<td>5\text{a}</td>
<td>5\text{a}</td>
<td>5\text{a}</td>
</tr>
</tbody>
</table>

Note: NO: Non-Obese. O: Obese. Each column is an average. Values in the same row and subtable not sharing the same subscript are significantly different at p< 0.05 in the two-sided test of equality for the same column proportions. Cells with no subscript are not included in the test. Tests assume equal variances.
2. Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Bonferroni correction.

Table 3-5. Comparison of Built Environment Variables Between Study Neighbourhoods in Metropolitan Vancouver and Hamilton

<table>
<thead>
<tr>
<th>Built Environment Variable</th>
<th>Metropolitan Area</th>
<th>Metro Hamilton</th>
<th>Metro Vancouver</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td></td>
<td>29.13&lt;sub&gt;a&lt;/sub&gt;</td>
<td>26.28&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Physical Activity Score (MET score)</td>
<td></td>
<td>6593.84&lt;sub&gt;a&lt;/sub&gt;</td>
<td>5385.86&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Time Sedentary (min/week)</td>
<td></td>
<td>2755.27&lt;sub&gt;a&lt;/sub&gt;</td>
<td>2851.24&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Fruit and Vegetable Market Store</td>
<td></td>
<td>6794.48&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1739.99&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Public Recreation Court/Centre</td>
<td></td>
<td>709.15&lt;sub&gt;a&lt;/sub&gt;</td>
<td>451.30&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Public/Green Space</td>
<td></td>
<td>373.29&lt;sub&gt;a&lt;/sub&gt;</td>
<td>290.40&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Fitness and Recreation Centre ($)</td>
<td></td>
<td>1221.34&lt;sub&gt;a&lt;/sub&gt;</td>
<td>974.17&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Supermarket</td>
<td></td>
<td>1636.46&lt;sub&gt;a&lt;/sub&gt;</td>
<td>920.66&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Community Centre</td>
<td></td>
<td>2170.56&lt;sub&gt;a&lt;/sub&gt;</td>
<td>1537.19&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Limited Service Restaurant</td>
<td></td>
<td>820.09&lt;sub&gt;a&lt;/sub&gt;</td>
<td>615.22&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Full-Service Restaurant</td>
<td></td>
<td>974.85&lt;sub&gt;a&lt;/sub&gt;</td>
<td>674.97&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Mixed-Use Centre</td>
<td></td>
<td>3323.09&lt;sub&gt;a&lt;/sub&gt;</td>
<td>2393.95&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Distance to Walking Trail</td>
<td></td>
<td>350.22&lt;sub&gt;a&lt;/sub&gt;</td>
<td>259.19&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Sidewalk Length (m)/Area of 500 m Street Buffer (m)</td>
<td></td>
<td>.028&lt;sub&gt;a&lt;/sub&gt;</td>
<td>.040&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Sidewalk Length (m)/Area of 1000 m Street Buffer (m)</td>
<td></td>
<td>.021&lt;sub&gt;a&lt;/sub&gt;</td>
<td>.030&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Intersection Density (sq. m)</td>
<td></td>
<td>.0000081&lt;sub&gt;a&lt;/sub&gt;</td>
<td>.000126&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Population Density (sq. k)</td>
<td></td>
<td>2914.47&lt;sub&gt;a&lt;/sub&gt;</td>
<td>5171.51&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Land-use Mix (entropy index)</td>
<td></td>
<td>.35&lt;sub&gt;a&lt;/sub&gt;</td>
<td>.37&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Walkability Index</td>
<td></td>
<td>6&lt;sub&gt;a&lt;/sub&gt;</td>
<td>5&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Note. Values in the same row and subtable not sharing the same subscript are significantly different at p<0.05 in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances.
2. Tests are adjusted for all pairwise comparisons within a row of each innermost subtable using the Bonferroni correction.

The Pearson correlation coefficients between BMI and the combination of individual characteristics and BE variables are shown in table 3-6 and table 3-7. These were calculated using SPSS software version 19.0 (Corp., 2010). Overall, correlations were lower than expected and tended to display an opposite directionality from their expected relationships. For instance, both land-use mix and walkability had a significant positive linear relationship with BMI in Metro Vancouver #1 (table 3-7). The same community also had a negative relationship with distance to public/green space, suggesting that the closer an individual is to public/green spaces, the higher their BMI is. Conversely, in Metro Hamilton #3 we observed a significant negative relationship between BMI and both walkability (p-value<.01) and land-use mix (p value<.05). Metro Hamilton #1 had the highest amount of significant correlations (10), and was one of two communities (Metro Hamilton #3 is other) that displayed any relationship with both education and income (both negative). Interestingly, higher physical activity scores was only associated with BMI in one community (Metro Vancouver #1), but time sedentary showed significant positive correlation with BMI in five out of seven communities.

Though illustrative of global patterns in the dataset, these results did not explain whether there was any spatial trend with respect to high BMI.
<table>
<thead>
<tr>
<th></th>
<th>Metro Hamilton #1 (N=1062)</th>
<th>Metro Hamilton #2 (N=782)</th>
<th>Metro Hamilton #4 (N=211)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>-.112&quot;</td>
<td>-.079*</td>
<td>-.006</td>
</tr>
<tr>
<td>Income</td>
<td>-.096&quot;</td>
<td>-.090*</td>
<td>-.002</td>
</tr>
<tr>
<td>Physical Activity Score (MET score)</td>
<td>-.037</td>
<td>-.007</td>
<td>-.114</td>
</tr>
<tr>
<td>Time Sedentary (min/week)</td>
<td>.168&quot;</td>
<td>.133&quot;</td>
<td>.210&quot;</td>
</tr>
<tr>
<td>Distance to Public/Green Space</td>
<td>-.011</td>
<td>-.056</td>
<td>.139*</td>
</tr>
<tr>
<td>Distance to Public Recreation Court/Field</td>
<td>-.041</td>
<td>.028</td>
<td>.084</td>
</tr>
<tr>
<td>Distance to Fitness and Recreation Centre ($)</td>
<td>-.072*</td>
<td>.053</td>
<td>.102</td>
</tr>
<tr>
<td>Distance to Supermarket and Grocery Stores</td>
<td>-.066*</td>
<td>.047</td>
<td>.021</td>
</tr>
<tr>
<td>Distance to Convenience Stores</td>
<td>-.078*</td>
<td>-.030</td>
<td>.120</td>
</tr>
<tr>
<td>Distance to Limited Service Restaurant</td>
<td>-.072*</td>
<td>-.025</td>
<td>.167*</td>
</tr>
<tr>
<td>Distance to Full Service Restaurant</td>
<td>-.106&quot;</td>
<td>.013</td>
<td>.117</td>
</tr>
<tr>
<td>Distance to Community Centre</td>
<td>-.116&quot;</td>
<td>-.061</td>
<td>.088</td>
</tr>
<tr>
<td>Distance to Mixed-Use Centre</td>
<td>.026</td>
<td>.004</td>
<td>-.160*</td>
</tr>
<tr>
<td>Distance to Fruit and Vegetable Market Stores</td>
<td>-.116&quot;</td>
<td>-.063</td>
<td>.135*</td>
</tr>
<tr>
<td>Sidewalk Length (m)/ Area of 1000 m Street Buffer (m)</td>
<td>.081&quot;</td>
<td>-.037</td>
<td>-.042</td>
</tr>
<tr>
<td>Intersection Density (sq. m)</td>
<td>.026</td>
<td>.024</td>
<td>-.140*</td>
</tr>
<tr>
<td>Land-use Mix (entropy index)</td>
<td>.047</td>
<td>-.062</td>
<td>-.165*</td>
</tr>
<tr>
<td>Walkability Index</td>
<td>.056</td>
<td>-.056</td>
<td>-.183&quot;</td>
</tr>
</tbody>
</table>

Note.  ** Correlation is significant at the 0.01 level (2-tailed).  
* Correlation is significant at the .05 level (2-tailed)
### Table 3-7. Pearson’s Correlation Coefficients for Metro Vancouver

<table>
<thead>
<tr>
<th></th>
<th>Metro Vancouver #1 (N=309)</th>
<th>Metro Vancouver #2 (N=419)</th>
<th>Metro Vancouver #3 (N=1159)</th>
<th>Metro Vancouver #4 (N=182)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDUCATION</td>
<td>-.032</td>
<td>-.068</td>
<td>.011</td>
<td>-.146*</td>
</tr>
<tr>
<td>INCOME</td>
<td>-.002</td>
<td>-.087</td>
<td>-.009</td>
<td>-.072</td>
</tr>
<tr>
<td>PHYSICAL ACTIVITY SCORE (MET score)</td>
<td>-.135*</td>
<td>-.034</td>
<td>-.038</td>
<td>.102</td>
</tr>
<tr>
<td>TIME SEDENTARY (MIN/WEEK)</td>
<td>.067</td>
<td>.146**</td>
<td>.108**</td>
<td>.091</td>
</tr>
<tr>
<td>DISTANCE TO PUBLIC/GREEN SPACE</td>
<td>-.122*</td>
<td>.090</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>DISTANCE TO PUBLIC RECREATION COURT/FIELD</td>
<td>-.082</td>
<td>.031</td>
<td>.042</td>
<td>-.196**</td>
</tr>
<tr>
<td>DISTANCE TO FITNESS AND RECREATION CENTRE ($)</td>
<td>-.010</td>
<td>.006</td>
<td>.024</td>
<td>-.150*</td>
</tr>
<tr>
<td>DISTANCE TO SUPERMARKET</td>
<td>.006</td>
<td>.016</td>
<td>-.009</td>
<td>.134</td>
</tr>
<tr>
<td>DISTANCE TO CONVENIENCE STORES</td>
<td>-.053</td>
<td>-.022</td>
<td>-.010</td>
<td>.028</td>
</tr>
<tr>
<td>DISTANCE TO LIMITED SERVICE RESTAURANT</td>
<td>-.026</td>
<td>-.020</td>
<td>-.003</td>
<td>.110</td>
</tr>
<tr>
<td>DISTANCE TO FULL SERVICE RESTAURANT</td>
<td>-.030</td>
<td>-.058</td>
<td>-.023</td>
<td>.134</td>
</tr>
<tr>
<td>DISTANCE TO COMMUNITY CENTRE</td>
<td>-.004</td>
<td>-.064</td>
<td>.046</td>
<td>-.144</td>
</tr>
<tr>
<td>DISTANCE TO MIXED-USE CENTRE</td>
<td>.003</td>
<td>-.049</td>
<td>-.080**</td>
<td>.204**</td>
</tr>
<tr>
<td>DISTANCE TO FRUIT AND VEGETABLE MARKET STORES</td>
<td>-.079</td>
<td>-.105*</td>
<td>.006</td>
<td>.054</td>
</tr>
<tr>
<td>SIDEWALK LENGTH (M)/ AREA OF 1000 M STREET BUFFER (M)</td>
<td>.020</td>
<td>-.008</td>
<td>-.037</td>
<td>-.035</td>
</tr>
<tr>
<td>INTERSECTION DENSITY (SQ. M)</td>
<td>-.014</td>
<td>.054</td>
<td>.005</td>
<td>-.036</td>
</tr>
<tr>
<td>LAND-USE MIX (ENTROPY INDEX)</td>
<td>.125*</td>
<td>.072</td>
<td>-.002</td>
<td>-.043</td>
</tr>
<tr>
<td>WALKABILITY INDEX</td>
<td>.116*</td>
<td>.082</td>
<td>.006</td>
<td>-.045</td>
</tr>
</tbody>
</table>

Note. ** Correlation is significant at the 0.01 level (2-tailed).
   * Correlation is significant at the .05 level (2-tailed)

#### 3.4.2. Spatial Clustering

The spatial relative risk function gave us a broad understanding of where, if any, spatial clustering of high BMI was evident in the data. It also produced a smoothed risk surface of log relative risk for being obese and is displayed in figure 3-3 and figure 3-4.
Raw relative risk was log-transformed to provide a more intuitive depiction of risk that is centered on 0. Hatched and solid isolines denote areas of significant log relative risk, p<.05 and p<.01, respectively. Readers will note that each map has a specific scale, with communities that have higher ranges generally pointing to higher log-relative risk throughout the respective community. As well as serving to pinpoint specific geographic regions of significantly elevated risk, this method depicts the overall spatial distribution of densities of obese people versus the densities of non-obese individuals. In general, Metro Hamilton communities exhibited spatial clustering with a higher log relative risk. Two communities in Metro Vancouver had significant clusters, whereas the other two had a stable surface of low log relative risk. Communities that had significant regions of high log relative risk were noted as distributions that may have an underlying pattern associated within clustered areas.
Figure 3-3. Spatial Relative Risk of Metro Hamilton Communities

Note: This figure represents the spatial relative risk of Metro Hamilton Communities. Each legend is unique to the relative risk ranges of that community and are in log-relative risk units. Risk over zero denotes an increase risk; below zero is a decreased risk. Hatched lines denote significantly elevated risk at the p-value < 0.05 level. Solid black line denotes significantly elevated risk at the p-value < 0.01 level. (a) Metro Hamilton #1 (b) Metro Hamilton #2 (c) Metro Hamilton #3

Figure 3-4. Spatial Relative Risk of Metro Vancouver Communities

Note. This figure represents the spatial relative risk of Metro Hamilton Communities. Each legend is unique to the relative risk ranges of that community and are in log-relative risk units. Risk over zero denotes an increase risk; below zero is a decreased risk. Hatched lines denote significantly elevated risk at the p-value < 0.05 level. Solid black line denotes significantly elevated risk at the p-value < 0.01 level. (a) Metro Vancouver #1 (b) Metro Vancouver #2 (c) Metro Vancouver #3 (d) Metro Vancouver #4.

3.4.3. OLS Regression

Results of optimal ordinary least squares regression models are displayed in table 3-8 and table 3-9. Models that did not perform well with added BE variables and were not statistically significant and are not shown for space reasons. All of our models
were statistically significant even though Adjusted $R^2$ values were very low. Models were initially explored by developing several linear regression models in SPSS (results not shown). Selection of explanatory variables was largely informed by our initial regression analyses and the corresponding correlation coefficients and a combination of exploratory spatial data analysis techniques (i.e., visualization, clustering, and scatter plots). Several models were compared against each other based on recommended diagnostics for OLS and GWR analyses. Statistically significant clustering in residuals (Global Moran’s I) indicate model misspecification and generally signal the researcher to explore for a missing variable(s) that may reduce clustering. Likewise, they may also indicate the need to model the parameters in a spatial context using GWR. With our models we experimented with several variable combinations to alleviate this problem, however, for some we could not find better variables.

### Table 3-8. OLS regression results for Metro Vancouver

<table>
<thead>
<tr>
<th>Community</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>$R^2$ Adjusted</th>
<th>AICc</th>
<th>Global Moran’s I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro Van 1</td>
<td>Physical Activity(-)*</td>
<td>Physical Activity(-)*</td>
<td>0.0147</td>
<td>1751.86</td>
<td>0.165222**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Land-use mix(+)*</td>
<td>Time sedentary(+)*</td>
<td>0.0344</td>
<td>1746.69</td>
<td>0.118038</td>
<td></td>
</tr>
<tr>
<td>Metro Van 3</td>
<td>Time Sedentary(+)*</td>
<td>Distance to Mixed Use Centre(-)*</td>
<td>0.0122</td>
<td>7026.06</td>
<td>0.12198**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance to Mixed Use Centre(-)*</td>
<td>Education(-)*</td>
<td>0.0169</td>
<td>7021.53</td>
<td>0.119431**</td>
<td></td>
</tr>
<tr>
<td>Metro Van 4</td>
<td>Education(-)*</td>
<td>Distance to Mixed Use Centre(+)*</td>
<td>0.0158</td>
<td>1063.833</td>
<td>0.019233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance to Public Recreation Courts(+)*</td>
<td>Education(-)*</td>
<td>0.0468</td>
<td>1058.984</td>
<td>0.019233</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0622</td>
<td>1056.999</td>
<td>-0.011129</td>
<td></td>
</tr>
</tbody>
</table>

Note. BMI is dependent variable. (-) signs indicate the coefficients are negative. (+) signs indicate coefficients are negative. Models used in GWR are shown in bold text

* Joint F-statistic and Joint Wald Statistic significant p-value at the .05 level
  **Statistically significant p-value at the .05 level
Table 3-9.  OLS regression results for Metro Hamilton

<table>
<thead>
<tr>
<th>Community</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>R2 Adjusted</th>
<th>AICC</th>
<th>Global Moran's I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro Hamilton 1</td>
<td>Time Sedentary (+)*</td>
<td>Time Sedentary (+)*</td>
<td>Time Sedentary (+)*</td>
<td>0.0274</td>
<td>6954.90</td>
<td>0.124542*</td>
</tr>
<tr>
<td></td>
<td>Distance to Community Centre (-)*</td>
<td>Distance to Community Centre (-)*</td>
<td>Distance to Community Centre (-)*</td>
<td>0.038193</td>
<td>6944.05</td>
<td>0.108062*</td>
</tr>
<tr>
<td>Metro Hamilton 2</td>
<td>Time Sedentary (+)*</td>
<td>Time Sedentary (+)*</td>
<td>Education (-)*</td>
<td>0.045652</td>
<td>6936.78</td>
<td>0.108062*</td>
</tr>
<tr>
<td></td>
<td>Walkability (-)*</td>
<td></td>
<td></td>
<td>0.040316</td>
<td>1363.78</td>
<td>0.0552</td>
</tr>
<tr>
<td></td>
<td>Distance to Public Space (+)*</td>
<td></td>
<td></td>
<td>0.061852</td>
<td>1360.94</td>
<td>0.036287</td>
</tr>
</tbody>
</table>

Note. BMI is dependent variable. (-) signs indicate that the coefficients are negative. (+) signs indicate coefficients are negative. Models used in GWR are shown in bold text.

* Joint F-statistic and Joint Wald Statistic significant p-value at the .05 level
**Statistically significant p-value at the .05 level

As shown in table 3-9 we included individual characteristic variables such as time sedentary and education. Though our focus was on exploring relationships with the BE these variables were included to improve model fit and strength. As can be seen in both OLS results, time sedentary, physical activity, and education all boost the goodness of fit (AICc) for the models they were included in. Separate models were developed without the non-BE variables however the resulting Joint F-statistics and Joint Wald statistics did not show any significance. In general, explanatory variables included in OLS models were among the variables with the strongest linear Pearson's correlation coefficients. Out of all BE variables, destinations that are conducive to physical activity were among the ones with greater explanatory power; distance to public recreation courts, distance to public spaces, distance to mixed-use centres, and distance to community centres. Two land-use variables, land-use mix (Metro Vancouver #1) and walkability (Metro Hamilton #3), had significant relationships with BMI, though land-use mix was in the opposite direction than expected.

3.4.4. GWR Results

Due to the low adjusted R² values from OLS models, a GWR approach provided us the potential to improve model fit on the premise that nonstationarity was present in
the data. Diagnostic statistics for the models are shown in table 3-10 and table 3-11, and also include the respective bandwidths for each community. Overall, none of our GWR models performed better than the OLS models, confirming that there is no spatial nonstationarity in any of the relationships tested and that the global models were better at explaining the variance in individual BMI.

**Table 3-10. GWR results for Metro Vancouver**

<table>
<thead>
<tr>
<th>Community</th>
<th>Model</th>
<th>R2 Adjusted</th>
<th>AICc</th>
<th>Bandwidth</th>
<th>Global Moran’s I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro Vancouver #1</td>
<td>Physical Activity(-) + Land-use mix(+)</td>
<td>0.034349</td>
<td>1748.82</td>
<td>33761.79</td>
<td>0.118015</td>
</tr>
<tr>
<td>Metro Vancouver #3</td>
<td>Time Sedentary(+) + Mixed-Use Centre(-)</td>
<td>0.016489</td>
<td>7024.29</td>
<td>30245.12</td>
<td>0.119578*</td>
</tr>
<tr>
<td>Metro Vancouver #4</td>
<td>Education(-) + Mixed-Use Centre(+) + Public Recreation(+)</td>
<td>0.062219</td>
<td>1059.34</td>
<td>33514.25</td>
<td>-0.011139</td>
</tr>
</tbody>
</table>

Note. BMI is dependent variable. (-) signs indicate the coefficients are negative. (+) signs indicate coefficients are negative.

*Statistically significant p-value at the .05 level.

**Table 3-11. GWR results for Metro Hamilton**

<table>
<thead>
<tr>
<th>Community</th>
<th>Model</th>
<th>R2 Adjusted</th>
<th>AICc</th>
<th>Bandwidth</th>
<th>Global Moran’s I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro Hamilton #1</td>
<td>Time Sedentary(+) + Community Centre(-) + Education(-)</td>
<td>0.052047</td>
<td>6936.08</td>
<td>5105.54</td>
<td>0.102661*</td>
</tr>
<tr>
<td>Metro Hamilton #3</td>
<td>Time Sedentary(+) + Walkability(-) + Public Space(+)</td>
<td>0.115624</td>
<td>1365.49</td>
<td>896.27</td>
<td>-0.036724</td>
</tr>
</tbody>
</table>

Note. BMI is dependent variable. (-) signs indicate the coefficients are negative. (+) signs indicate coefficients are negative.

*Statistically significant p-value at the .05 level.
GWR models for Metro Vancouver communities produce the least conclusive of results as nearly all communities had a bandwidth of over 30 km centered on each observation (Table 3-10). Moreover, coefficient variation and local $R^2$ values were very low. Figure 3-5 illustrates this effect more clearly. Coupled with the increase in AICc and unchanging adjusted $R^2$ values, the GWR models were less effective than the OLS models. Again, several models were tested to potentially illuminate a relationship not illustrated in the OLS models, however none proved to be stronger than the ones presented. Metro Vancouver #3 was the lone community with clustered residuals (Table 3-10) meaning that a key variable could be missing from the model. Unfortunately an exhaustive scan through our variables showed no signs of a stronger relationship. This suggests that spatial nonstationarity is not present in the Metro Vancouver data or the variables selected were not suited for accurately describing potential relationships.
Figure 3-5.  *Metro Vancouver #3 GWR surfaces*

Note. Maps illustrate our GWR model performance when model is nearly global. Point locations were smoothed using and Inverse distance weighted interpolator. Legends are customized to reflect the low variability in coefficient values (a) smoothed surface of distance to mixed land-use coefficient values. (b) smoothed surface of participant density. (c) smoothed surface of local R2 values.
Figure 3-6.  *Metro Hamilton #3 GWR surfaces*

Note.  Maps illustrate our GWR model performance when model is nearly global.  Point locations were smoothed using an Inverse distance weighted interpolator.  Legends are customized to reflect the low variability in coefficient values (a) smoothed surface of walkability index coefficient values.  (b) smoothed surface of participant density.  (c) smoothed surface of local $R^2$ values.
Although GWR models for Metro Hamilton communities also failed to improve the overall model-fit (Table 3-11), bandwidth values were lower and adjusted $R^2$ values were higher. Whereas Metro Vancouver communities exhibited bandwidth values that exceeded the study area boundary, both Metro Hamilton communities had bandwidths that potentially reflect a local trend in the data. Figure 3-6 illustrates the variability of parameter coefficients at a local level and small scale study boundary. In Metro Hamilton #3, there is a pattern of negative coefficient values for walkability near the northern corner that also align with the region of increased risk for being obese (Figure 3-3). Local Moran’s I tests also show the region to have significant clustering of low coefficient values (not shown for confidentiality reasons). In Metro Hamilton #1, coefficients range over a large scale study boundary similar to Metro Vancouver #3, and model fit based on AICc and Local $R^2$ values tend to be closer to the global OLS model. However, unlike Metro Hamilton #3, the direction and magnitude of the BE variable coefficients leave a mixed interpretation when taking into account the region of elevated risk for obesity (Figure 3-3); living closer to community centres predicts an increase in individual BMI. On the other hand, near the eastern portion of the map, individuals that lived further from community centres were predicted to be more obese. As with Metro Vancouver communities, the results of the GWR models should be interpreted with caution as they did not out-perform the OLS’ model fit.

3.5. Discussion

Our study provides a mixed, yet thought-provoking exploration on the potential relationships between the built environment and adult obesity. Much research has been conducted to explore these relationships (Frank, et al., 2004; Li, et al., 2009; K. B. Morland & Evenson, 2009; Rundle, et al., 2007), with some papers finding strong evidence (Mobley, et al., 2004; K. B. Morland & Evenson, 2009) and others finding mixed, to no evidence of relationships to the built environment (Frank, Kerr, Sallis, Miles, & Chapman, 2008; M. C. Wang, et al., 2007). While our study echoes the efforts of those mentioned, it uniquely contributes to the conversation around obesity and the built environment with our attempt to employ a spatially explicit approach in the application of SPARR and GWR. In Metro Vancouver #1, increased land-use mix was associated with increased BMI, while in Metro Hamilton #3 a similar measure in walkability was inversely
related to BMI. Interestingly, the built environment in Metro Vancouver #1 is quite urbanized and the average income is relatively high (Table 3-1). Whereas the geographical region in Metro Hamilton #3 is characterized by a built environment that is considered suburban; it is dominated by curvilinear street networks, lacking a central commercial district, and isolated by large arterial streets and freeways. These two communities speak to the incongruent findings among some of research to date.

This study found interesting relationships between BMI and the distance to community spaces, such as community centres, mixed-use centres, public spaces, and public recreation courts. In the case of distance to public space and public recreation courts, two communities (Metro Hamilton #3 and Metro Vancouver #4) had a positive relationship with BMI. These two variables are similar in what they support for individuals, with one providing a cost-free opportunity for organized activity (public recreation courts) and the other providing a cost-free opportunity for unorganized activity, or play. The assumed effect on physical activity is two pronged as well, the trip to the facility and the subsequent activity after arriving. While the relationship between public recreation courts and public space were the same between Vancouver and Hamilton, relationships between mixed-use centres and community centres with BMI were mixed in both metropolitan areas. In this study, mixed-use centres were defined as community spaces that offered multiple uses besides a gathering space, such as a workout area, swimming pool, library, learning centre. It is possible that the difference in the definition of these facilities varies between the two areas, or that the qualitative features of the two may vary drastically. This could also hold true for any built environment feature in this study. Future studies should zoom in on the built environment to assess the differences in quality of these features.

Much of the research to date concerning BMI and the built environment adopts an aspatial approach that often stratifies samples into classes of weight status, income, urbanization level or sex and bases confirmatory analysis on non-contiguous geographic areas. While these approaches are instructive of broad patterns that may exist between a set of variables and BMI they fail to assess whether there is an elevated risk of higher BMI based on agglomerations of obese individuals, or clusters (Schuurman, Peters, et al., 2009). Our approach started with a spatial cluster analysis method, SPARR. SPARR identified and delimited elevated levels of log relative risk for obesity in four out of seven
communities, and provided a visual representation of overall spatial distribution of obesity throughout the communities. This information helped to contextualize the underlying patterns of obesity before moving forward with modeling. Though it was only an initial method in our study the importance of these results may carry useful information in application of other exploratory approaches, such as scanning the microenvironment for qualitative impressions of ground-level influences (Frank & Engelke, 2001).

Pearson’s correlation tests may have illustrated the most about the underlying influences of obesity. In two Hamilton communities, multiple BE variables had a linear relationship with BMI (up to 11 in Metro Hamilton #1). Conversely, the highest amount of BE variables Vancouver communities showed any relationship with was three. This trend may be chalked up to random factors, but as well, could be illustrative of the difference in built environments between the two regions. For instance, out of all communities in the Metro Vancouver study area there was a total of 23 fruit and vegetable market stores, compared to just 6 throughout Metro Hamilton. Sidewalk length per unit area, a measure of pedestrian connectivity was also much smaller (both buffer sizes) throughout Metro Hamilton, adding to a moderate base of evidence that the built environment between the two study sites is different. In light of a recent study (Pouliou & Elliott, 2010) in Vancouver and Toronto which found walkability to be an accurate indicator of the built environment in both cities, Hamilton and Vancouver may not be fair comparisons.

OLS and GWR models presented modest evidence of multivariate linear relationships between BMI, and individual level and BE variables. A non-spatial variable, time sedentary, had the most relationships across all communities and is important to highlight even though it is not a BE variable. Time sedentary is considered to be an activity distinct from physical activity, in that it entails different health risks brought on by activities therein (Tremblay, Colley, Saunders, Healy, & Owen, 2010). Combined with being obese and living in a BE not conducive to an active lifestyle, such as poor access to physical activity centres, may put an individual at a higher risk for associated health outcomes. For some individuals, time sedentary may even reflect the lack of opportunity or desirability to walk in their neighbourhood. This question was outside the scope of this study but is certainly an important aspect to consider in future studies.
Various walkability indices have been used to predict BMI (Frank, et al., 2008; Frank, et al., 2006; Kligerman, Sallis, Ryan, Frank, & Nader, 2007; Raja et al., 2010) however results are often either contradictory of each other or found to be in favor of predicting increases or decreases in a physical activity variable. For one community (Metro Hamilton #3) in our study area we found walkability to have a statistically significant inverse linear relationship with BMI – as walkability decreases BMI increases. The GWR results also showed a spatial relationship with the walkability and BMI in this area. Conversely, an area of elevated risk for obesity in Vancouver (Metro Vancouver #3) was also ranked as one of the most walkable areas, with a variety of food type outlets, and proximity to rapid mass transit. Again, this inconsistency begs the question of apparent differences in the built environment between the two Metropolitan areas, as well as walkability having only a random influence on individual BMI. Inconsistent results similar to ours have been found elsewhere (Frank, et al., 2008), and raise important questions on variable selection and the relative importance of the contribution of the BE to individual BMI.

Our study had important limitations that must be considered for future studies. First, running spatial cluster analysis on samples from a population can produce bias in the results as locations are not representative of the entire spatial boundary being investigated. Many spatial clustering algorithms are suited for case-control data derived from population based disease registries, and as such, are developed for locating disease clusters (Auchincloss, 2012). Our community aggregations were based on representative samples of the population and aimed to replicate variation in population densities across the BE continuum. Additionally, our cases and controls were retrospectively identified as, meaning, cases were not matched with controls. Also to note, our definition of obesity as being a ‘case’ or having a clear cut-off between obese on non-obese using BMI may be an inappropriate measure for obesity, as body fat percentage has been shown to be higher in South Asians when compared to other ethnicities of the same BMI (Lear, et al., 2012).

A second limitation is our selection and calculation of some BE variables. With regard to selection, we failed to capture densities or counts of establishments that were either supportive or unsupportive of active and healthy lifestyles. Several studies have found fast-food density to be associated with increased BMI in adults (Li, et al., 2009;
Rundle, et al., 2007). Our BE dataset limited us from calculating this variable because we limited service restaurants entail a collection of low-nutrition food outlets – e.g., cafes, sandwich shops. We did test limited service density within a 1000m network buffer but the results showed little difference. It should be noted however, that to date no Canadian study has found an absence of significance between BMI and density of fast-food, convenience stores, grocery stores and recreation facilities (Pouliou & Elliott, 2010; S. A. Prince, et al., 2011). The calculation of objectively measured BE variables may also be a source of weaker than expected relationships. Walkability index and land-use mix were derived from a nationally available land-use dataset that is not as up to date as some of our measures and respective levels may be under-calculated due to growth in the study areas, particularly Vancouver. Additionally, applying the same walkability index to two vastly different urban environments may have caused issues since the quality of stores within classified zones is not being assessed. This issue could be addressed following a micro-scale scan of the environment, as touched on earlier.

A common limitation with cross-sectional studies, and built environment research as a whole, is self-selection bias (Handy, et al., 2002). Self-selection bias posits that individuals may choose to live in more walkable, or health-promoting neighbourhoods, because that suites their lifestyle. Any interpretation of a measured relationship with the built environment then, must factor in the possibility that the individual may play a greater role than the effect of a walkable or high entropy neighbourhood may. It is not immediately apparent that our results suggest an occurrence of this, but future studies using this dataset may encounter it.

As the study was focused on determinants of high BMI in the BE, our study did not account for perceived factors related to the BE (i.e., walkability, safety from crime) or the individual (i.e., quality of life indicators). A study of rural US adults found perceived food and activity environments supportive of healthy lifestyles to be associated with a decreased likelihood of being obese (Casey et al., 2008). Investigating the effect modification of the perceived built environment versus the objective may provide clearer insight into potential intervention strategies for public health policies. Given the lack of explanatory power from our objectively measured BE variables, an inquiry into the perceived environment may bring about greater influence.
3.6. Conclusions

Our study explored two Canadian metropolitan areas for spatial clustering of obesity and corresponding relationships with built environmental determinants. Exploratory results provided moderate evidence of elevated spatial relative-risk for obesity. Global and local regression offered a unique perspective on conceptualizing built environment relationships with BMI, finding significant relationships in global OLS models. The results of the global models suggest that communities across both, metropolitan Vancouver and Hamilton, live in differentially structured built environments that could potentially promote or prevent active and healthy lifestyles. Walkability, land-use mix, distance to recreational courts, mixed-use centres, community centres and public spaces displayed a statistically significant linear relationship with BMI across various communities. These results convey possible interactions that individuals in certain study areas may have with the objective built environment and provide important results for future studies associated with this dataset. This research also adds to the growing body of Canadian studies tackling the BE and obesity (Pouliou & Elliott, 2009, 2010) (S. A. Prince, et al., 2011; Prince, et al., 2012; S.A. Prince et al., 2011), as studies from the United States continue to dominate the evidence. Future research should aim to refine variable selection and hone the unit of analysis to a lower aggregation – e.g., forward sortation areas – to gain a clearer picture of small-scale effects.
4. Conclusion

The spatial and aspatial effects of places on health are complex and require alternative modalities of thought and methodology to adequately address the multitude of factors at play. The research in this thesis presents a potential approach for analyzing health effects across space and place, while also contributing knowledge to a rapidly growing arena of research across the globe. The first paper, presented in chapter two, provides health researchers with an overview of methods for spatial cluster analysis of address point data. The second paper, presented in chapter three, informs the Canadian context of built environment and obesity research, and also raises awareness about spatial methods in the field more broadly.

Paper one sought to provide an overview and analysis of spatial clustering methods suitable for analyzing data at a fine spatial resolution. The paper also addressed important methodological caveats and strengths to consider when deciding on which method(s) to choose for spatial cluster analysis. Several different methods were found to be suitable for analyzing address location data based on published papers in the last ten years. While several other known methods exist, only a fraction were represented by our review, and one, the spatial scan statistic, accounted for over half of the published material. The review also identified six themes associated with methodological implementation, derived from the published material and our research team’s expert knowledge. Themes suggest that prospective researchers consider the areas of exploratory analysis, visualization, spatial resolution, aetiology, scale, and spatial weights. This paper is intended to underscore the complexities of applying spatial clustering methods to data of any resolution, but specifically address location data.

The second paper applied a spatial model to explore the clustering of obesity and potential correlates of the built environment in two Canadian metropolitan areas. The first objective applied a spatial cluster analysis method that combined relative risk calculations and kernel density mapping. To accomplish the second objective, several
methods were used to explore potential underlying relationships associated with the clustering of high BMI. Pearson’s correlation coefficient was used to examine bivariate associations before selecting parameters for the ordinary least squares regression model. Geographically weighted regression was applied to the ordinary least squares models to assess whether BMI increase or decrease was predicted by spatially varying built and non-built environmental variables. Results indicated that high BMI did form spatial clusters, in both Hamilton and Vancouver communities. BMI associations with the built environment were mixed throughout both metropolitan areas. Evidence from one Hamilton community suggested that low walkability was associated with an increase in BMI, while an area in Vancouver that was walkable, or at least heterogeneous in land-use, had a positive association with increased BMI. Geographically weighted regression did not improve model fit, and thus, built environment variables did not vary spatially with fluctuations in BMI.

4.1. Research Contributions

The research in this thesis has important implications for future studies related to spatial analysis of health data in both a Canadian, and global context. The results from paper one illustrated important concepts that spatial health researchers can apply to aid in the selection of an appropriate spatial clustering method. These results not only raise awareness of spatial thinking in health research by highlighting key concepts, they also compile a synthesis of peer-reviewed papers on a very specialized type approach of spatial analysis with address point data. The second paper contributes to the growing knowledge of how the built environment affects obesity in Canada - a study area that is nascent at the moment. For instance, areas in the communities that displayed spatial clustering of obesity could be looked at in more depth for finer scale built or non-built environmental features. The results of paper two may also provide a greater impetus for Canadian-based research as the relationship between obesity and the environment is even less clear compared to other research settings, such as the United States.

Paper two represents an important step for Canadian studies looking at obesity and potential correlates with the built environment. Although the relationships displayed in our results did not have the greatest explanatory power for built environment
relationships, obesity was shown to form clusters, or contiguous areas of significantly elevated risk. Studies looking at the spatial clustering of obesity are rare and results are usually mixed (Mobley, et al., 2004; Schuurman, Peters, et al., 2009). Additionally, this study evaluated spatial clustering at the finest of spatial resolutions, address point level, indicating that spatial patterns of obesity may exhibit a local characteristic in the way they are formed. This analytical framework could also be extended to locate areas or clusters of low risk for obesity, or high and low physical activity.

Paper two’s contribution to the Canadian context of obesity research is influential, but not necessarily based on the highlights of the results. Establishing causal factors of obesity is challenging, particularly when the explanatory variables are environmental and the study population potentially interacts with several different contexts outside of their immediate environment on a daily basis. This study found that between the two metropolitan areas, the built environment had fewer relationships with BMI in Vancouver than it did in Hamilton. This broad finding may suggest that participants from their respective study area may interact with the built environment in a different way, or that the built environment may be qualitatively different between the two study areas. For comparison, the population density throughout all Metro Vancouver communities is twice that of Metro Hamilton’s. In addition to Metro Vancouver’s higher growth rate in the past ten years and more advanced transportation infrastructure, on the surface, Metro Vancouver and Hamilton make up two very different built environments. The results from this study support this pattern as well. For instance increased walkability was associated with high BMI in a Hamilton community while increased land-use was associated with high BMI in Vancouver. Our results also showed no consistent built environment variable that can be associated with high or low BMI. Both of these findings speak to the idea that environments may be utilized differently depending on context.

Though obesity relationships with built environmental variables were not strong, there are contributions for both policy and for other studies. Policies for obesity prevention and control are generally categorized into three streams; (1) health services and clinical interventions that target individuals, (2) community-level interventions that directly influence behaviours, and (3), public policies that target broad social or environmental determinants (Sacks, Swinburn, & Lawrence, 2008). Results from this
study would directly impact policy decisions at the third level as the built environment is heavily impacted by such policies as land-use zoning, transportation planning, and local design guidelines. While direct policy change will not be a linear outcome from this thesis, the results bolster the need for continued discussion and subsequent investigation of the built environmental determinants of obesity and physical activity at the private and public level. A continued funding stream from agencies such as the Heart and Stroke Foundation of Canada and Canadian Institutes of Health Research and more community based research from university programs will continue to push knowledge growth, and transfer, throughout local governments and research programs.

Paper one’s results mark a more research-oriented contribution, and perhaps highlight the methodological contributions of paper two. The application of spatial clustering techniques is not a new field to spatial epidemiology, however the use of address point level data is rare and methods are more sensitive to spatial data issues. Results from paper one inform the many disciplines that seek to employ spatial clustering analysis methods on address point level data about those issues. Some central findings of the paper was that studies generally lacked discussions on marquee issues related to spatial data handling and over half of the studies relied on one method (Spatial Scan Statistic), while several others are known in the literature. Paper one’s findings also tied into the methodological contributions of paper two: strengthening argument for the use of spatially explicit methods in public health research. Both papers showed how combining spatial statistical methods (e.g., spatial clustering analysis, spatial relative risk) and GIS based methods (e.g., kernel density estimation, geographically weighted regression), public health data can be explored, visualized, and analyzed in a different way. Results from the scoping review provided a guideline for public health researchers or epidemiologists looking to employ spatially explicit methods.

4.2. Future Research

This thesis has revealed several areas where the potential for future research will be possible – and desirable. Both papers highlighted the advantages of employing a spatially explicit approach to analyzing health data. Paper two employed a spatially explicit approach to investigate the spatial clustering of obesity and potential built
environmental correlates, and highlighted key areas for continued enquiry of the Canadian context of the discipline. Together, these two papers illustrate a potential approach that researchers can take to analyze public health data, specifically when applied to obesity and the built environment.

Paper one highlighted several areas for future research, some of which are important to touch on again. For instance paper one emphasized the importance of considering multiple spatial clustering tools when conducting exploratory analysis, or when attempting to delineate true spatial clusters, a recommendation made elsewhere in the field of spatial epidemiology. The use of visualization to aid analysis and interpretation was underutilized in the papers reviewed. This is important to note because the size and shape of some clusters may be unrealistic for some spatial contexts and scales. Geovisual analytics is an approach that integrates visual analytics with geospatial and cartographic analysis to incorporate perceptual and cognitive reasoning with scientific enquiry (J. Chen, et al., 2008). Studies that employ this technique have the potential to illuminate visualized patterns that non-visual methods do not have the capability of. Paper two employed methods capable of producing a visual output in the spatial relative risk function and geographically weighted regression.

Another recommendation from paper one is for the continued research into etiology of disease or spatial phenomena processes. If more information is known about how the process under investigation spreads or interacts with other variables delineating and subsequently selecting the appropriate spatial resolution to represent the data will greatly increase the accuracy of some results. To achieve this end, future research on data collection techniques such as the use of GPS units to track participants interaction with the environment, will aid researchers in asking better defined questions or parameterizing models with greater accuracy. This future research consideration could also impact future studies on the built environment and obesity, as a current limitation in some studies is selecting the appropriate spatial data resolution to capture an individual’s interactions with built environment features in different spaces such as distal communities where one may spend a significant portion of time or the work environment.

In the second paper the results underscored the need to further the Canadian understanding of built environment and obesity relationships, both in working with this
study and broader Canadian endeavors. Foremost, work in other Canadian metropolitan areas should be conducted in order to capture a full range of built environments. For instance, future work based in rural or peri-urban areas would be beneficial as rural populations in the Canada have been shown to have higher prevalence of obesity than urban populations (Canada, 2011; Shields & Tjepkema, 2006). An initial focus of this study was to investigate differences in the built environment determinants with obesity across the development spectrum. However, inconsistencies in sample sizes, accuracy and variety of data, and varying spatial scales, prohibited these direct comparisons. Data from the PURE study was collected in rural areas throughout all three study sites (Metropolitan Vancouver, Hamilton, and Quebec City) and future research will explore these patterns, making it the first of its kind in Canada.

Continued work on Canadian urban areas would benefit the broader knowledge of obesity and built environment relationships. This thesis explored the micro-scale environment of communities that were aggregated together to increase the sample size for spatial regression. Analyzing the relationships at an even finer scale – single FSA boundary - and characterizing each FSA based on a multitude of built environment features may reveal more about micro-scale relationships at the neighbourhood level. This type of analysis would most likely be aspatial and consist of multilevel models more suitable for small numbers; however the growth of Canadian knowledge should not be discouraged by this fact. Moreover, research in other major Canadian urban centres, where development patterns may have produced differently structured built environments, should be conducted.

Work on the perceived built environment has recently taken off in the United States. Identifying whether variations in BMI are explained better by perceived or objective measures will help pin point meaningful prevention measures that target health promotion and awareness programs at the population level. For instance, a study found that adults in rural communities were less likely to be obese if perceived food and physical activity environments supported healthy behaviours (Casey, et al., 2008). The PURE study also has the ability to conduct these investigations as smaller subsamples of community populations were asked questions about the perceived walkability environment.
Lastly, further development of a spatially explicit approach to examining the spatial variation of obesity in the built environment should be pursued. Using spatial cluster analysis to explore variations in obesity may reveal patterns both related to the perceived (e.g., neighbourhood social capital) and objective built environment (e.g., density of fast-food restaurants). Likewise, spatial cluster analysis could also be conducted on other variables that are non-spatial such as perceived quality of life, or various health statuses. Analysis that takes place after identifying a cluster could also be of a mixed variety, either opting for a confirmatory approach with regression, or a qualitative approach that inquires about the qualitative environmental features. The use of GIS methods in combination with ethnographic methods may reveal more insightful data on frequency of visit, and utilization of built environment features (Kwan & Ding, 2008; Matthews, Detwiler, & Burton, 2005).
References


Doi 10.1186/1748-5908-5-69


10.1016/j.amepre.2005.12.001


Signorino, G., Pasetto, R., Gatto, E., Mucciardi, M., La Rocca, M., & Mudu, P. (2011). Gravity models to classify commuting vs. resident workers. An application to the analysis of
residential risk in a contaminated area. *International Journal of Health Geographics, 10*(1), 11.


100


Appendices
Appendix A.

Reference Translation List for Table 2-1


