Estimating Health Determinants and Outcomes in Rural Ottawa: An Integration of Geographic and Statistical Techniques

By:

Brian Mosley

A thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the degree of

Masters of Arts, Geography

Department of Geography

UNIVERSITY OF OTTAWA

Abstract

Many health geography studies, including the Ottawa Neighbourhood Study (ONS), have faced significant challenges uncovering local variation in patterns of community health in rural areas. This is due to the fact that sparsely populated rural areas make it difficult to define neighbourhoods that are representative of the social and resource utilization patterns of the individuals therein. Moreover, rural areas yield small samples from population-based regional health surveys and this leads to insufficient sample sizes for reliable estimation of health determinants and outcomes.

In response to this issue this thesis combines geographical and statistical techniques which allow for the simulation of health variables within small areas and populations within rural Ottawa. This methodological approach combines the techniques of dasymetric mapping and statistical micro-simulation in an innovative way, which will allow health geography researchers to explore health determinants and health outcomes at small spatial scales in rural areas. Dasymetric mapping is used to generate a statistical population surface over Ottawa and then estimate socio-economic (SES) variables within small neighbourhood units within rural Ottawa. The estimated SES variables are then used as correlate variables to simulate health determinant and health outcome variables form the Canadian Community Health Survey (CCHS) using statistical micro-simulation. Through this methodology, simulations of specific health determinants and outcome can be investigated at small spatial scales within rural areas.

Dasymetric mapping provided neighbourhood-level population estimates that were used to re-weight as set of SES variables that were correlates with those in the Canadian Community Health Survey (CCHS). These neighbourhood-level correlates allowed microsimulation and consequent spatial exploration of prevalence for smoking, binge drinking, obesity, self-rated mental health, and the presence of two or more chronic conditions. The methodology outlined in this paper, provides and innovative way of exploring health determinants and health outcomes in neighbourhoods for which population and health statistics are not traditionally collected at levels that would allow traditional statistical analyses of prevalence.

De nombreuses études de géographie de la santé, y compris l'étude du quartier Ottawa (ONS), ont fait face à des défis importants, découvrant des variations locales dans les profils de santé communautaire dans les zones rurales. Cela est dû au fait que les zones rurales peu peuplées rendent difficile de définir des quartiers qui sont représentatifs des tendances sociales et des ressources l'utilisation des personnes qui y sont. En outre, les zones rurales donnent des petits échantillons d'enquêtes régionales de la santé sur la population et cela conduit à des tailles d'échantillon insuffisant pour une estimation fiable des déterminants de la santé et des résultats.

En réponse à cette question, cette thèse combine des techniques géographiques et statistiques qui permettent la simulation de variables de la santé dans les petites régions et des populations dans les zones rurales Ottawa. Cette approche méthodologique combine les techniques de cartographie densimétrique et microsimulation statistique de manière innovante, ce qui permettra à santé chercheurs de géographie explorer les déterminants de la santé et de la santé à petites échelles spatiales dans les zones rurales. Cartographie densimétrique est utilisée pour générer une surface de population statistique sur Ottawa et ensuite estimer des variables socio-économiques-économiques de (SES) au sein des unités de quartier petit au sein de l'Ottawa rural. Les variables de SES estimés sont ensuite utilisés comme mettre en corrélation des variables pour simuler le déterminant de la santé et les variables de résultats de santé forment la canadienne enquête sur la santé de le communautaire (ESCC), à l'aide de statistique microsimulation. Grâce à cette méthodologie, les simulations des déterminants de la santé spécifiques et le résultat peuvent être examinées à petites échelles spatiales dans les zones rurales.

Densimétrique cartographie fourni des estimations de population de niveau de quartier qui ont servi à re-weight ensemble de variables de SES corrélats avec celles dans l'enquête sur la santé de collectivités canadiennes (ESCC) que. Ces corrélats de quartier-niveau permis microsimulation et exploration spatiale conséquente de prévalence pour fumer, la frénésie de consommation d'alcool, l'obésité, auto-évaluation de la santé mentale et la présence de deux ou plusieurs maladies chroniques. Fournit la méthodologie décrite dans cet article, et de façon novatrice d'explorer les déterminants de la santé et la santé dans les quartiers pour les statistiques de population et de la santé ne sont pas traditionnellement recueillies à des niveaux qui permettraient à des analyses statistiques traditionnels de prévalence.

Acknowledgments

I would like to take some time to thank and acknowledge some people without whom this thesis would not have been possible.

Firstly, I would like to thank my supervisor, Dr. Michael Sawada, for his guidance, advice, and support over the last two years. His expertise and encouragement were instrumental to me reaching my goal of completing this thesis. I would also like to acknowledge my committee members, Dr. Eric Crighton and Dr. Luke Copland, for their guidance and advice.

Next, I would like to express my gratitude to the Ottawa Neighbourhood Study, of which my thesis is a part. Specifically, I would like to thank Dr. Elizabeth Kristjansson for giving me the opportunity to contribute to such meaningful work. In addition, I would like to express thanks to Dr. Jean-Michael Billette, a partner of the ONS, whose work on micro-simulation was vital to the completion of this thesis. Further I would like to thank the Population Health Improvement Research Network (PHIRN) under Ronald Labonte for partially funding this research.

Furthermore, I would also be remiss if I did not thank my fellow students in the Department of Geography at the University of Ottawa. I would also like to acknowledge the academic and technical assistance given to me by the University of Ottawa, specifically the Department of Geography.

Finally, I would like to thank my parents Michael and Susan Mosley, my uncle Dr. Peter Wells, and my girlfriend Jessica Bredschnieder for their love, support, and encouragement while I completed this thesis.

Table of Contents

List of Figures

Chapter 1

Chapter 2

List of Tables

Chapter 2

List of Equations

Appendices

List of Common Acronyms

ONS – Ottawa Neighbourhood Study

SVMC – Support Vector Machine Classification

SES – Socioeconomic Status

DA – Dissemination Area

CT – Census Tract

CBPR – Community-Based Participatory Research

CCHS – Canadian Community Health Survey

List of Common Terms

Dasymetric Mapping: Dasymetric mapping is an aerial interpolation method that takes advantage of ancillary data to focus population and its characteristics on where people actually reside in geographic space. In simple terms, it distinguishes between populated and unpopulated areas.

Support Vector Machine: The SVM is a supervised classification technique used to separate data into a series of user-defined classes. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value (such as the class labels) and several attributes (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) that predicts the target values of the test data given only the test data attributes.

Participatory Mapping: Participatory mapping is part of the emerging field of community-based participatory research (CBPR). CBPR is a research approach that incorporates the "subjects", or people of the community, in the research process.

Spatial Micro-Simulation: Spatial micro-simulation is a down-scaling technique that involves simulating individual- or household-level characteristics (such as SES and Health) within small areas through linking common variables from the spatial scale of interest with detailed, large-area anonymous survey data.

Chapter 1: Thesis Introduction

1.1 Introduction

In order to understand the health of a community, it is necessary to have an accurate depiction of the neighbourhoods with which people interact (Oaks, 2004; Kristjansson et al., 2009; Mamman et al., 2009). Knowledge of these spaces allows researchers to identify and understand the contextual variables that impact the health of a community (Schwab and Syme, 1997; McMichael, 1999; Susser, 1999; Berkman, Glass, Brissette, & Seeman, 2000; O'Campo, 2003; Kristjansson et al., 2009). This kind of place-based research, however, requires defining neighbourhood units that accurately delineate the spaces with which people interact; as Schwab and Syme (1997) state, for researchers to be able to accurately assess the health of a community, they must first be able to "reflect the ecological reality of life in that population, as people experience [it]" (p. 2050). Recent studies support this statement, making it clear that not only is a person's health a result of personal compositional variables such as diet, attitude, and exercise, but that it is also fundamentally affected by spatial variables within their community. These types of contextual variables can include: neighbourhood infrastructure, neighbourhood walkability and access to healthy food (Kawachi et al., 1999; O'Campo, 2003; Leung et al., 2004; Lopez-Zetina et al., 2006; Stafford et al., 2007).

The reality, however, is that discrete spatial units like neighbourhoods that are used in health research cannot be perfectly defined. Studies have found contradictory evidence when it comes to the spatial unit at which health variables should be measured (Ross, 2004; Hayens et al., 2007; Hameed et al., 2010; Rainham et al., 2010). These studies have varied in their results, in some cases concluding that government defined enumeration units provide an acceptable unit at which to measure health; while others have found that health variables should be measured through community defined neighbourhoods. In particular, because of the sparse population density, rural areas pose the most difficulty when defining the geographic units at which to measure health (Haynes and Gale, 2000; Kristjansson et al., 2009). In rural regions, in order to capture a sample population that is large enough to estimate health variables at acceptable levels of confidence, any discrete geographic boundaries must cover large physical areas (CCHS, 2001; Haynes and Gale, 2007; Parenteau et al., 2008; Kristjansson et al., 2009). This large-area small-population issue exists in the most common types of geographic units used to study place and health: census enumeration units (Langford and Unwin, 1994) and "natural neighbourhood units" (Haynes and Gale, 2000; Haynes et al., 2007; Kristjansson et al., 2009).

By way of elaboration, Langford and Unwin (1994) note that "large enumeration units tend to have lower population densities and, conversely, smaller enumeration units tend to have higher population densities" (p. 24). As one moves from the rural periphery to the urban core of Ottawa, for example, this density effect is clear, given the purpose of census geography (Figure 1.1). The Canadian dissemination area (DA) contains between 400-700 individuals. These DAs represent the smallest area at which full census data is reported (20% sample, long-form). As such, in rural areas where population is more dispersed, the DAs must cover a larger area to capture the required number of individuals, which in turn equates to lower population densities (compared to a DA within the urban core) (Figure 1.1). Langford and Unwin (1994) further note that enumeration units may hide the data variations that occur within them because they have been defined by government bodies for statistical purposes and, therefore, may not always be associated with existing discontinuities in population density.

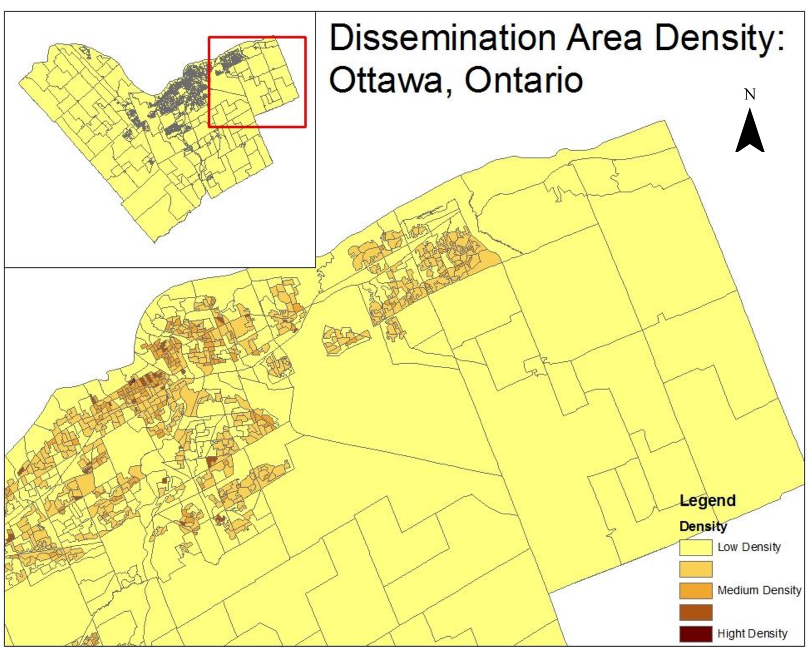


Figure 1.1: Choropleth map representing population density in dissemination areas within the City of Ottawa, Ontario, Canada.

Those studies that attempt to utilize regional survey data together with natural neighbourhood units in rural regions encounter an issue with the estimation of health variables that is akin to estimating health variables in "small-areas" (Elliotte et al., 1995; Elliotte and Wartenburg, 2004; Srebotnjak et al., 2010; Hampton et al., 2011). The so called small-area issue occurs when small census areas or neighbourhood units do not contain a large enough population size to allow accurate estimation of health variables using data collected within the purview of larger regional surveys (Pfeffermann 2002; Kristjansson et al., 2009). Briefly, the issue of small-areas arises when one wishes to estimate health outcomes or determinants[1] using data from surveys that were designed and completed for a larger scale of aggregation. Hence, small-areas do not refer to any specific scale but simply a scale that contains units that are smaller physically than those for which a given survey was designed to represent. By way of illustration, in Canada, the Canadian Community Health Survey (CCHS) is designed to collect health data at "sub-provincial levels of geography (health region or combined health regions)" (Statistics Canada, 2012). The CCHS sample design and sample sizes are sufficient to support estimation of regional variations in health determinants and outcomes. As such, the CCHS is a survey designed for estimation within large populations. Unfortunately, at smaller local scales, only a few respondents will be found within each census tract, DA or neighbourhood and so reliable statistical estimation for these units is not possible. Therefore, *prima facie*, it is not possible to explore location variations of health related variables using regional survey data. In summary, the small-area issue is a consequence of insufficient sample support for estimation and due to an insufficient population size.

Under the assumption of a uniform population density in a large region, it is fundamental to understand that a smaller population support is a consequence of the reduced size of the census area. While neighbourhood units can be large in physical area, they may still have a small population, especially if the neighbourhood is in an area of low population density resembling rural regions. This creates a paradox for health geography research: if

[1] Because this research is looking at synthetic estimation and not the relation between risk factors and health outcomes, the general term 'determinant' is adopted to imply that we do not associate our estimated risk factors with any positive or negative health outcome. In context when discussing a determinant with a particular outcome the term 'risk factor' is used.

neighbourhood unit boundaries are extended to capture a larger population, the "ecological reality" of the people living within the neighbourhood is lost (Shwab and Syme, 1997), but if the boundaries remain, accurate sample populations may not be large enough for reliable statistical estimation be performed (Kristjansson et al., 2009). As such, the small-area issue is more accurately rephrased in the context of this present research as a small population – small sample problem. Thus, the purpose of this research is to present a method to overcome the small area population issue in rural areas in order to produce health determinant and heath outcome variable prevalence estimates in rural Ottawa. The major contribution of this research focuses on combining dasymetric mapping and spatial micro-simulation methodologies as a singular approach, which can be used to investigate local-level variations in health across rural Ottawa. The simulated health variables can then serve as dependent variables for multivariate modeling of contextual health determinants in future research.

1.2 Background

The task of defining the most appropriate geographic unit at which to measure health is one that has challenged the Ottawa Neighbourhood Study (ONS). The ONS is a large-scale epidemiological study across the City of Ottawa, whose aim is to understand how the physical and social characteristics of neighbourhoods in Ottawa affect the health of residents living within them. Since 2008, the ONS has been developing neighbourhood units across the City of Ottawa (Parenteau et al., 2008). The results of the study have been well received, except in the rural regions of the city (Figure 1.2). Residents in rural areas felt that neighbourhood boundaries were too large in physical area, and as a result, that the boundaries did not accurately represent the spaces with which the residents interacted (Ottawa Neighbourhood Study, 2008).

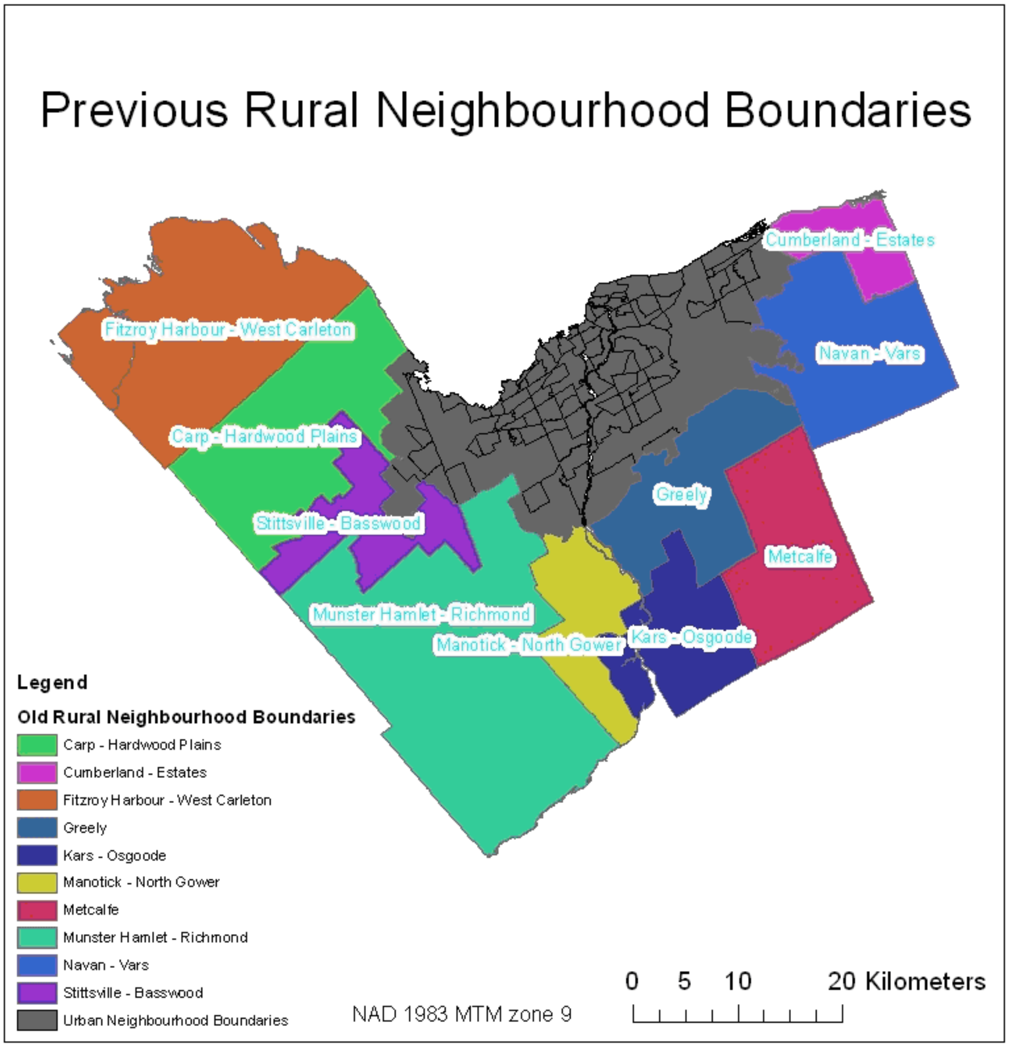


Figure 1.2: Neighbourhoods in rural Ottawa, defined in Parenteau et al. (2008).

Neighbourhoods in Rural Ottawa were large in physical area in order to capture samples that would be large enough to reliably estimate health variable prevalence in neighbourhood populations (Parenteau et al., 2008). The number of people included in a study area sample population needs to allow for an estimation that is within an acceptable confidence interval (CCHS, 2001; Sanmartin, 2006; Parenteau et al., 2008). In the rural regions of the Ottawa outside of the urban core, population density is low, except around small townships. Examples of this include the areas surrounding the towns of Osgoode, North Gower, and Cumberland (Figure 1.3). This led to the delineation of neighbourhoods that were unusually large in area because they had to capture population samples that were large enough to reliably estimate health variables. As such, neighbourhoods in rural Ottawa had to be redefined to correspond to environments with which rural residents interacted. This redefinition of neighbourhoods led to spatial units with smaller populations and insufficient sample sizes. As a consequence, a new approach to estimating health determinants and outcomes in these areas was required.

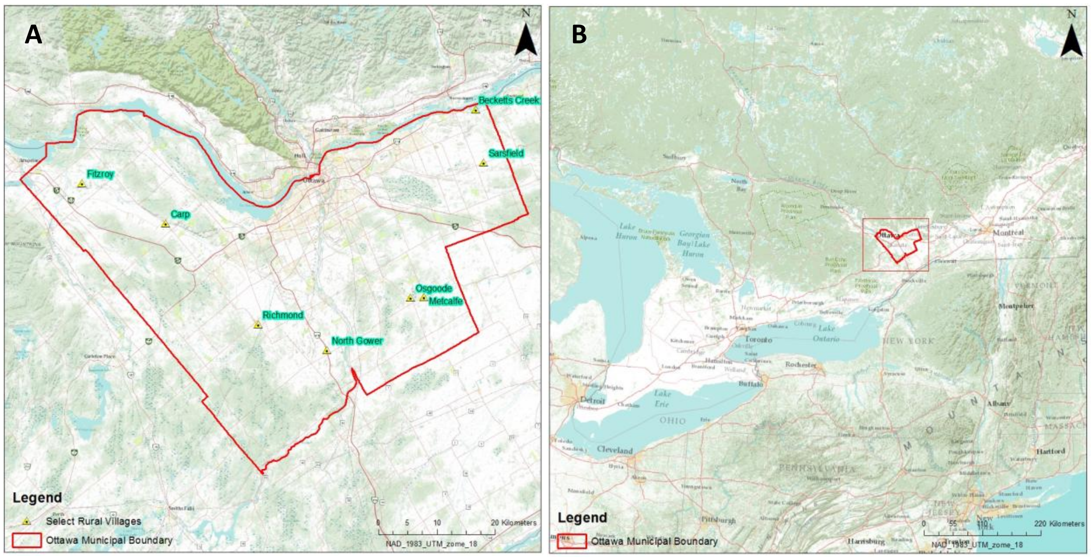


Figure 1.3: a) Select townships within the municipal boundary of the city of Ottawa, Ontario b) The location of the Ottawa municipal area, within southern Ontario, Canada.

In response to the barriers to reliably estimating health variables in rural areas, this thesis presents a combination of geographical and statistical techniques that can be used to *simulate* health variables (determinants and outcomes collected by the CCHS) in low-population rural neighbourhood units in rural Ottawa. To this end, rural Ottawa neighbourhoods, defined through a community-based participatory mapping methodology, will be used as boundaries to support the simulation of health determinants and health outcomes. Simulation support will be provided by estimation of census variables within the rural neighbourhoods. Estimation of census variables will be achieved via a population-based re-weighting of census-level variables (socioeconomic status), using a statistical population surface generated by a dasymetric mapping process. The socioeconomic variables that will be redistributed based on the dasymetrically mapped population support are from the age, sex, income, education, language and visible minority categories of the 2006 Canadian census (Appendix A). Finally, using the neighbourhood level socioeconomic variables, spatial microsimulation will be undertaken to simulate health determinants and outcomes within the neighbourhood units defined by the ONS. Three health outcomes and three health related behaviours will be micro-simulated. The behaviours are the prevalence of smoking, binge drinking, and obesity, while the health outcomes are selfrated health, selfrated mental health, and the presence of two or more chronic conditions.

The aforementioned health variables chosen for this study are from the Canadian Community Health Survey (CCHS). The CCHS is an epidemiological survey that has been conducted annually since 2007 (it was conducted bi-annually between 2000 and 2006). The six health variables were chosen for microsimulation because of two main factors. First, each one of the six health determinants and outcomes chosen had data collected for them through each cycle of the CCHS. A "cycle" refers to every time the survey was conducted (either annually or bi-annually). In this research, several cycles of the CCHS need to be available for the microsimulation to be performed. Information pertaining to some health outcomes and determinants (e.g. injuries, depression or other specific health outcomes) were not necessarily collected at each cycle of the CCHS, or it was included in an optional module that was only administered to a smaller fraction of the sample. Second, the prevalence of some other health outcomes (fibromyalgia, arthritis or other less common conditions) is very low, and estimates calculated from survey data are usually inaccurate. Thus, the six health variables chosen for microsimulation were selected because they were statistically more likely to provide accurate estimates while this methodology was being initially applied and tested (Appendix B).

1.3 Research Objectives

Thus, the main research objective addressed by this thesis can be stated as follows:

To combine dasymetric mapping and spatial microsimulation into a methodology that will provide simulated estimates of health determinants and health outcomes in irregular spatial units (community-defined neighbourhoods) with small populations in the rural regions of Ottawa. The purpose of this undertaking is to create a methodology that will provide an innovative way to study local-level variations in health determinants and outcomes in rural regions.

The main objective of this thesis can be broken down into two sub objectives:

1. *To create a dasymetric density surface over Ottawa in order to accurately facilitate the redistribution of socioeconomic status (SES) variables from the 2006 Canadian census to participatory defined neighbourhood units.*

2. *Outline an innovative approach to simulate health variable prevalences in newly defined rural Ottawa neighbourhoods using a combination of dasymetric mapping and spatial micro-simulation.*

1.4 Study Area

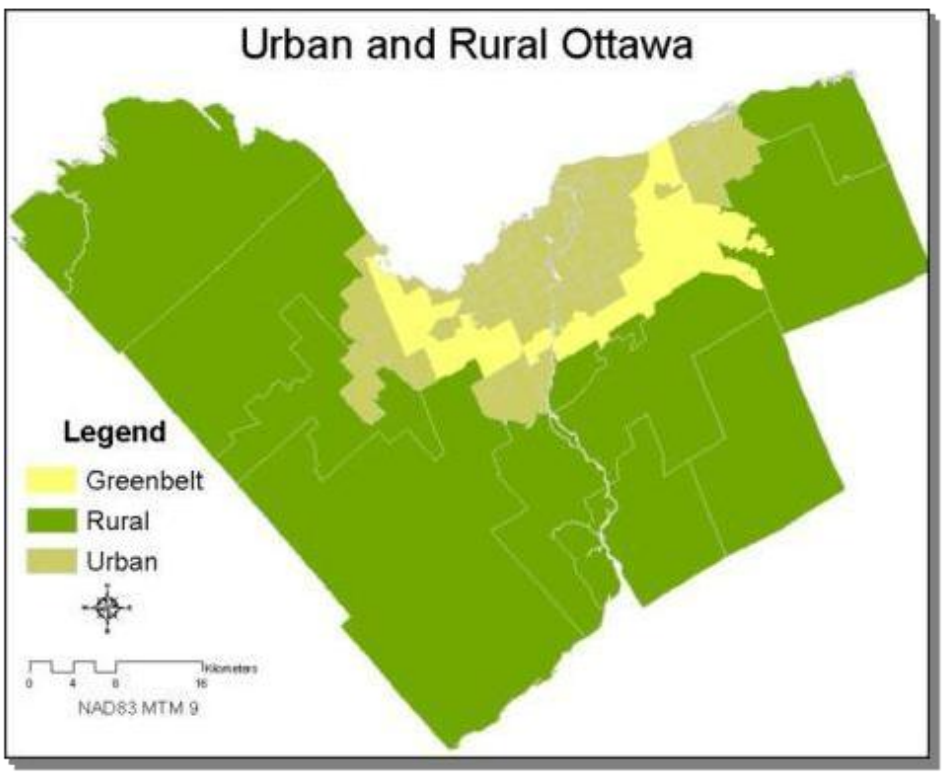


Figure 1.4: Urban and rural Ottawa and the area known as the "Greenbelt."

The rural neighbourhoods of Ottawa are the main focus of the study (Figure 1.4). The City of Ottawa is located in the far east of Ontario on the banks of the Ottawa River, across from Gatineau, Quebec (Latitude 45° 19'N and Longitude 75° 40'W)(Figure 1.3). In 2001, the City of Ottawa amalgamated with 11 municipalities (Cumberland, Gloucester, Goulbourn, Kanata, Nepean, Osgoode, Ottawa, Rideau, Rockcliffe Park, Vanier, and West Carleton), and these municipalities now comprise the rural Ottawa area. Rural Ottawa is very sparsely populated, with the exception of small villages, which contain higher population densities that are similar to those of the urban sections of the city. After amalgamation in 2001, Ottawa became the city with one of the largest municipal areas in Canada at 2,796 km^2; with just over 900,000 people, however, it is only the sixth largest population in the country (City of Ottawa, 2011). Furthermore, the City of Ottawa estimates that up to 80 percent of its municipal area is rural (City of Ottawa, 2011). Despite occupying four fifths of the city's land area, Ottawa's rural population stands at just over 85,000 (City of Ottawa, 2011), making it a very sparsely populated region.

The disperse population of rural Ottawa makes it difficult to study health variables in the area. Where populations are disperse, large study areas are needed in order to capture a large enough sample to derive statistically meaningful estimations of health outcomes/determinants from data like that of the CCHS. Of course, one could simply aggregate all of rural Ottawa in order to provide a sufficient sample size for statistical estimation methods. That approach however, becomes problematic because as the rural study area becomes larger: (a) people living there no longer identify with the region, (b) examination of spatial heterogeneities in health across the large rural area are no longer possible, and (c) the identification of spatial variations in health within manageable well defined areas are fundamental to effective health policy interventions. Thus, in large rural areas, researchers lose their ability to properly analyse contextual variables related to health (Schwab and Syme, 1997; Ottawa Neighbourhood Study, 2008). It is for these reasons that the research in this thesis is critical to studies like the Ottawa Neighbourhood Study (ONS).

1.5 Context

This section includes a discussion of past research conducted on the issues presented in this thesis. First, the concepts of health determinants and health outcomes are discussed, along with the difference between compositional and contextual health variables. This is followed by a review of literature discussing the increasing understanding of how contextual variables impact community health. Finally, the problems surrounding defining the best geographic unit at which to study health are examined.

1.5.1 The Importance of Contextual Variables when Analysing Health Determinants and Outcomes

1.5.1.1 The Relationship between Health Determinants and Outcomes

In this research a clear distinction is made between health determinants and health outcomes. Health determinants are the variables that affect an individual's health or the variables through which the determinant translates into health effects (Viner et al., 2001). A health outcome, however, is a state of being or effect resulting from a health determinant (Chen and Wilkins, 1991; Dunn, 2001). A situation which demonstrates the difference between health determinants and health outcomes is a smoker who gets lung cancer. In this situation, smoking is the health determinant, while the cancer is the health outcome of that activity (Dunn, 2001). Another example that demonstrates the link between health outcomes and health determinants would be an adult who performs regular aerobic activity (say jogging) and who therefore has a low risk of heart disease. In this situation, the aerobic activity is the health determinant, while the factors associated with a low risk of heart disease (such as lowered blood pressure and reduced cholesterol levels) are the outcomes (Matuk, 1996; Badets and Chui 1997).

In this present research, we will be synthetically estimating smoking, binge drinking and obesity which are determinants for a number of health outcomes like cardiovascular disease, cardiopulmonary disease, liver disease and the like. Our chosen health outcomes like selfrated health, selfrated mental health, are not outcome-specific and have known biases (Krause and Jay, 1994) but are established as predictors of morbidity/mortality (Idler and Kasl, 1991; Idler and Kasl, 1995; Fan et al., 2002). The most specific health outcome modeled in this thesis is the presence of two or more chronic conditions.

1.5.1.2 Compositional and Contextual Health Variables

When analyzing the cause of health outcomes (positive or negative), one must discern whether their determinants are a function of people's behaviour within a geographic space or the space itself. These are known as *compositional* and *contextual* characteristics, both of which present specific risks. Compositional characteristics relate to the characteristics of the individuals who live in a space, while contextual characteristics indicate the characteristics of the location itself (Macintyre et al., 2002). Macintyre et al. (2002) illustrates this concept: children in low income areas may not have a high level of physical activity (which can lead to obesity that presents risk for obesogenic health outcomes) because their families do not have gardens or the resources to take them to parks (a compositional resource-based explanation). An alternative explanation for low levels of physical activity could be that there are too few public parks in the neighbourhood and that there are no good public transport links to those that do exist (a contextual resource-based explanation) (Macintyre et al., 2002). Thus, one can see that in order to address a problem in health geography research, it is critical to understand whether the root cause is compositional or contextual. Macintyre's examples also implies the often ambiguous set of circumstances under which a given health determinant might be considered compositional vs. contextual.

1.5.1.3 The Importance of Context when Studying Health

The links between health outcomes and their determinants is a relationship that can be complicated to understand (Hook, 2001). Aggleton (1990), states that health could be defined as one being "physically and mentally fit," or that a person's health could extend beyond personal attributes and therefore be defined by the "availability of resources, both personal and societal, that help us reach our personal potential" (p. 8). Links between health outcomes are quite often tied to personal choices (Dunn, 2001), but health outcomes can also be strongly impacted by contextual health determinants in one's environment (Kawachi et al., 1999; O'Campo, 2003; Leung et al., 2004; Lopez-Zetina et al., 2006; Stafford et al., 2007). Current research suggests that contextual variables play a much bigger role in people's wellbeing than previously thought, and that they should be factored in the investigation of health outcomes (Kawachi et al., 1999; O'Campo, 2003; Leung et al., 2004; Lopez-Zetina et al., 2006; Stafford et al., 2007).

To illustrate how both compositional and contextual variables impact the health of a community, we can look at a study by Kawachi et al. (1999) that examined the causes of poor selfrated mental health in communities across 39 states in the United States. They discovered that poor selfrated health was strongly correlated with compositional health determinants such as having a low income or smoking. Moreover, poor selfrated mental health had a relationship with the type of community in which one lived. Kawachi et al. (1999) found that in communities where the state spent less money on social programs and infrastructure were "less-hospitable" environments, and by extension, contained higher levels of poor selfrated health. Poorly maintained buildings and community services in one's neighbourhood, for instance, are a factor beyond the control of the individual, but which nonetheless impacts their well being. Thus, one can see that poor selfrated mental health can be tied to both compositional determinates of health risks (low income and smoking), but also correlated to how much money was spent on making people's communities "hospitable".

Connections between health risk factors such as obesity have also been linked to contextual variables at the community level (Sherwood and Jeffery, 2000; Frank, 2000; Berrigan and Troiano, 2002; Lopez-Zetina et al., 2006; Stafford et al., 2007). In one study, it was concluded that obesity was strongly related to the built environment and an urban design that promotes the overuse of motor vehicles (Lopez-Zetina et al., 2006). Furthermore, obesity rates in communities have also been strongly tied to the availability of community resources (Stafford et al. 2007; Kristjansson, 2009). Stafford et al. (2007) found that within small towns across the United Kingdom, lower levels of obesity were correlated to the presence of swimming pools, fitness centres and major grocery stores. Conversely, the same study established that high levels of obesity in a community were strongly correlated to the absence of these same facilities. Thus, one can see that obesity is impacted by contextual variables in the community. This is not to say that personal compositional variables do not contribute to the presence of obesity, but it does show that the determinants contributing to health problems[2] are more complex than may be immediately apparent.

More recently, Hatzenbuehler et al. (2011) found that poor mental health in lesbian and gay (LG) populations is strongly correlated to social-contextual variables. In this study, Hatzenbuehler et al. (2011) found there was a clear link between general anxiety disorders in LG populations and the number of same-sex couples living in a community. The results of this study showed that LG individuals were far more likely to suffer from mood and anxiety disorders when they lived in communities that had low concentrations of same-sex couples. From these conclusions, it can be seen that the mental health of LG individuals extended beyond their own individual choices and was impacted by wider society.

Furthermore, neighbourhood contextual factors as determinants have been shown to play a role in an individual's selfrated health. In a study by Cummins et al. (2005), selfrated health was compared to neighbourhood contextual variables within 178 wards in England. The results of this study indicated that fair-to-very-bad selfrated health was strongly associated with residential environments, citing specific risk factors as poor physical quality, low political engagement, high unemployment, lower access to private transport, and lower transport wealth. These variables show how poor selfrated health can be related to ecologic contextual variables. While individuals can make personal choices to be more engaged in municipal politics in an effort to try and better their community, it is beyond their control if they live in a residential neighbourhood with poor transportation services and other public infrastructure. Thus, this study reveals that while health outcomes can result from both personal compositional choices, they can also result from contextual factors in the environment that are beyond individual-level control.

Through reviewing the above studies, we see that contextual variables at the community level are important determinants of population health. Moreover, determinants of health outcomes are not just completely explained by compositional variables. The above

[2] Obesity may be considered as a health outcome in some circumstances, such as when the condition is due to genetic, epigenetic or explicit lifestyle factors. This research considers obesity as a determinant itself since its causal pathways are not studied herein and it is considered a population level variable.

examples, however, are not meant to downplay the impact of compositional variables, but merely to highlight that health outcomes can result from both contextual and compositional determinants.

1.5.2 The Geographic Unit at which to Study Health

1.5.2.1 Enumeration Units and Natural Neighbourhood Units

In response to the increase in literature that cites contextual variables as playing a larger role in individual and community health, several studies have attempted to identify the most appropriate geographic unit at which to measure various health variables (Haynes et al., 2007; Riva, 2008). This is important because in order to properly investigate contextual variables that may be impacting the health of a community, one must be able to identify the environment with which they are interacting (Shwab and Syme, 1997; Label et. al., 2007). Traditionally, epidemiological studies have been operationalized around government defined enumeration units (Sampson et al., 2002), but recent studies have highlighted the fact that as their definition is largely influenced for statistical purposes, enumeration units may not represent actual discontinuities that would define sub-populations (Germain and Gagnon, 1999; Kawachi and Berkman, 2003; Martin, 2003; Clapp and Wang, 2006). Thus, there has been conflicting literature and a large amount of debate on which geographic unit should be used to study health variables (Ross et al., 2004; Haynes et al., 2007 Hameed et al., 2010; Rainham et al., 2010).

Several studies have found contradictory evidence when it comes to the spatial unit at which health variables should be measured. Hameed et al. (2010) recently found that the effectiveness of using SES as a predictor of traffic-related injuries varied considerably when four different neighbourhood definitions—ranging from dissemination areas and census tracts to custom boundaries—were used. Alternatively, Ross et al. (2004) found little discernable difference in the relation between health and SES in Montreal when these aspects were analyzed using natural neighbourhoods or census tracts. Rainham et al. (2010) concluded that the effect of the spatial unit in geographically based health studies still remains unclear, but from an intuitive standpoint, one could argue that well-defined neighbourhoods are better proxies than arbitrary spatial units for comparing health and place (Rainham et al. 2010).

Further, Hayens et al. (2007) conducted a study in Bristol, UK, that compared random computer generated neighbourhoods and natural neighbourhood units to the perceptions of neighbourhood and social behavior held by local residents. The study found that local residents did not identify with either set of boundaries presented by the researchers, but rather associated with much smaller areas. Thus, Haynes et al. (2007) concluded that natural neighbourhood units were not better than arbitrarily defined units for the investigation of health-related variables.

In short, the literature does not specify the single best spatial unit for studying health outcomes and/or determinants but simply implies that analyses and inferences are sensitive to the construction of the geography used.

1.5.2.2 Geographic Units in Rural Areas

If, however, one considers the geographic unit at which to study health in rural areas and urban areas separately, some insight can be gained regarding which geographic units should be used to study health. Specifically, how enumeration units are defined is critical to understanding the spatial units at which health should be studied in rural areas. Using Canadian DAs as an example, we see that the 400-700 people required to form a DA may only encompass a couple of city blocks in an urban core. Conversely, rural DAs must in some cases cover several hundred square kilometres in order to contain 400-700 people. Based on these boundaries, it has been suggested that the internal variability of rural DAs is far more likely to be greater than in urban ones (Hayens and Gane, 2000; De Marco and De Marco, 2009; Kristjansson et. al., 2009).

In urban centers, neighbourhoods tend to contain similar housing types, which in turn clusters residents of comparable SES (Hayens and Gane, 2000). Furthermore, because of the small geographic space of urban enumeration units, residents are far more likely to interact with similar spaces, both in terms of the physical landscape of the area and the services they are able to frequent (Kristjansson et al., 2009). The internal variability of rural enumeration units in terms of individual SES, community recourses, and the physical landscape is far greater because of the large areas they occupy and variation in resources utilized. This makes individual contextual variables that could be influencing health outcomes far more difficult to identify (Moore, 1995; Haynes and Gale, 2000; Kritjansson et. al., 2009). Thus, one can make the argument that high-density urban enumeration units are an acceptable geographic unit at which to study health determinants and outcomes, but rural enumeration units are not.

This conclusion also provides insight into why there have been conflicting conclusions among studies about the most appropriate spatial unit at which to measure health. Studies by authors such as Ross et al. (2004) or Riva et al. (2008) that show enumeration units to be acceptable spatial units at which to measure health were conducted in high-density urban centers (in this case, the island of Montreal). This conclusion further explains why the ONS received far more positive reviews of their conclusions about health patterns in the neighbourhoods of central Ottawa than they did for rural regions (Kristjansson et. al. 2009).

1.6 Combining Geographical and Statistical Techniques to Estimate Health Determinants and Outcomes within Small Rural Areas

In an effort to overcome the barriers of obtaining "small area" estimates of health in the large spatial units of rural Ottawa, this study proposes to combine the methods of dasymetric mapping and spatial microsimulation. These two techniques are combined in order to simulate health determinant and health outcome variables within community-defined neighbourhoods in rural Ottawa. The motivation to derive community-based neighbourhoods is simple: such neighbourhoods are self-organizing since they are the regions that people associate with. While these units may not provide sufficient support for statistical estimation because of the small population and small samples of health outcomes and determinants within data held by the CCHS, such self-identified units will be the most effective level of intervention for mitigating risk factors associated with negative health outcomes. While this research does not test whether these locally self-organized units are effective scales for intervention, the intuitive argument is convincing but their mitigation efficacy is for future research.

This section reviews the method of participatory mapping that was used by the ONS to define neighbourhood units in rural Ottawa. This is followed by research on the techniques used in this thesis. Research on dasymetric mapping and spatial microsimulation are reviewed in this section in order to highlight the benefits and shortcomings of past methodologies in each field and how they can be combined to achieve the research objectives outlined in this thesis.

1.6.1 Participatory Mapping

One way to delineate natural neighbourhood units is through the use of participatory mapping. Participatory mapping is part of the emerging field of community-based participatory research (CBPR), a research strategy that incorporates the "subjects," or people of the community, in the research process (O'Fallon and Dearry, 2002; Minkler and Wallerstein, 2003). The benefits of this approach are that the involvement of communities in the research methodology can enhance research through the contribution of local knowledge. This process can reveal variables or factors that should be taken into account in the research that would have otherwise been unknown. Involvement of communities can enhance the research by providing local context, which thereby improves the quality of the information gathered and the final results (Maantay, 2002; O'Fallon and Dearry, 2002; Beyer et al., 2010).

Participatory mapping has been shown to be useful in health geography research. In Maman et al. (2009) participatory mapping was used to help determine the most suitable locations for HIV clinics in five African countries, explaining that the participatory mapping facilitated the identification of community boundaries, [along with] the description of geographic and social separation within communities, and it identified suitable sites for the clinics (Maman et al., 2009). The participatory mapping exercise was carried out in small groups, with the participants being encouraged to draw their perceptions of where their neighbourhood boundaries lay on a map, along with other pertinent administrative boundaries. Maman et al. (2009) note that the participatory mapping was a crucial starting point for this research because community definitions of boundaries were different from what appeared on administrative maps (Maman et al., 2009).

In Sarnia, Ontario, participatory mapping was used to identify construction sites where construction workers could have been exposed to asbestos (Keith and Brophy, 2004). A study conducted in Ontario and Manitoba had casino gaming workers participate in mapping sessions where they identified hazards in their work place (Keith et al., 2001). The methodologies presented in the studies by Keith and Brophy (2004) and Keith et al. (2001) were similar in their approaches. In each method, participants were provided maps of their places of work and encouraged to highlight the areas where they felt their health was at risk or where they knew they were directly exposed to a hazard that could adversely affect their health. Maps from the participants were then compiled and analysed to identify common areas of risk. The results were then used by the participant's employers to improve health and safety in the participants' places of work.

More recently, participatory mapping was used within the Ottawa Neighbourhood Study to define new neighbourhood units in rural Ottawa. Participatory mapping was used by the ONS in rural Ottawa so that boundaries would reflect the spaces with which residents interacted. The goal of using participatory-defined neighbourhoods is that the new boundaries would clearly identify the environments that need to be investigated in order to properly understand the links between health and place in rural Ottawa. Furthermore, with the adoption of methodologies—such as statistical microsimulation in health research—simulation of health variable prevalence in small area geographic units had become increasingly possible.

The first step in participatory mapping framework of the ONS was to set up community consultations with community representatives in the rural areas that were being investigated. Ottawa's rural area was broken into four main quadrants where community consultations were performed. This was done in order to get participants to focus on the areas in which they live.

The community representatives at each consultation were from various professions, but there was a strong emphasis on having professionals from the health care sector present. A common consultation included doctors from the community, registered nurses, police officers, community group leaders, city councillors, and a participant from the City of Ottawa's rural planning office. Participants were invited to the consultations if they had expressed interest in the ONS since it began in 2006. Further, in an effort not to exclude any parties who may have valuable insight into the delineation of neighbourhoods, participants were also encouraged to bring additional guests who they felt may be able to provide insight into the mapping exercise. Examples of this type of guest included long-time residents and a former Member of Parliament (MP). The number of participants at each consultation ranged from 10—25 people.

The overall SES and demographics of the consultation region were presented, along with health trends both positive and negative in the region. Reasons for needing to delineate new neighbourhood units were also explained to participants. It was made clear to the participants that at the end of the consultation, only one set of boundaries could be drawn. ONS researchers acknowledged that there may be differing opinions among participants, but everyone was encouraged to discuss the reasons for their differing view and to come to a consensus on a boundary that they felt best represented the whole of their community.

Researchers from the ONS then initiated and guided discussions on how patterns in population density and current neighbourhood boundaries coincide with their individual conceptions of neighbourhood boundaries. The perceptions of the participants regarding where neighbourhood boundaries should be located were then recorded by allowing participants to draw on large paper maps. These maps simply provide an outline of the larger area of rural Ottawa. Roads, village locations, and water bodies were also displayed for orientation. Further, in order to keep the gathered information unbiased, ONS researchers were only allowed to explain the exercise and answer participant's questions. Researchers had to be careful not to pose any leading questions or force participation.

Each consultation lasted a couple of hours. During this time, participants were encouraged to offer their opinion as to where neighbourhood boundaries should be drawn.

Each participant was given time to draw his or her perceptions of where neighbourhood boundaries should be located. Once every participant had offered initial input, everyone was able to discuss amongst themselves both the boundaries that reflected a consensus and those where participant opinions differed. Once a consensus on the boundaries of neighbourhoods was reached, clear final boundaries were drawn on the maps.

In the next step, the boundaries identified in the neighbourhood consultations were digitized using ESRI's ArcMap 10 software. The boundaries identified on the large paper maps during the consultations are often roughly done. Thus, data layers—which delineate roads, water bodies, and other physical landmarks—were used during the digitization to ensure that the newly created neighbourhood boundaries reflected the reality of the physical environment. For example, the data layers helped ensure that the boundaries did not cross over major roads or go through the middle of a river unless the participants explicitly noted that they should.

The digitized neighbourhood boundaries were then re-distributed to the consultation participants, who had the opportunity to review the new neighbourhood boundaries and share them with colleagues, community groups, or even friends in the same neighbourhood. This process is known as "redistricting." Changes proposed by the consultation participants were subsequently reviewed. Previous consultation participants, ONS researchers, and the City of Ottawa's rural planning office were all consulted on any proposed changes. If the majority of people agreed on a change to a specific boundary, then modifications were made. Only one significant change to original participant-defined neighbourhood boundaries arose during the re-districting phase, and that was the addition of the neighbourhood of Sarsfield in East Ottawa (Figure 1.5). Other changes in the redistricting phase were relatively minor. Once all of the comments from the re-districting phase had been addressed, the new neighbourhood boundaries were finalized (Figure 1.5).

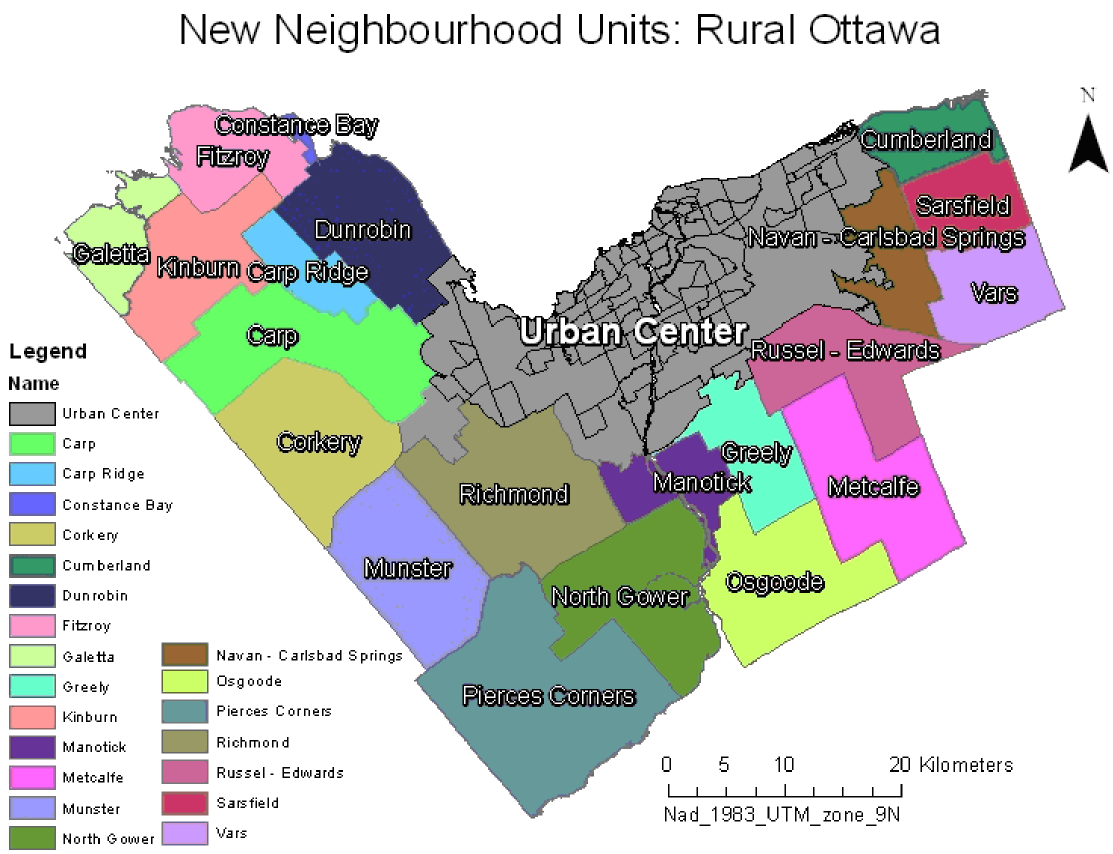


Figure 1.5: Neighbourhood units defined through the community-based participatory mapping in rural Ottawa. Defined by the ONS, 2011.

In summary, the neighbourhood boundaries defined through the ONS's participatory mapping exercise were chosen as the neighbourhoods for this study for a number of reasons. Firstly, reiterating Haynes et al. (2007), these participant-defined boundaries coincided with perceptions of residents concerning their community. It was felt that in order to present health variable estimates that would not only explore health determinants and health outcomes in a meaningful way, but that would also be useful to future research in rural Ottawa and health-based community interventions, neighbourhood delineations had to incorporate community input. As a result, the ONS did not put any population size constraints on the neighbourhood boundaries. Thus, the neighbourhoods defined by the ONS were not influenced by sample size constraints that are associated with traditional forms of estimation; rather, they were totally a product of the perceptions of community and neighbourhood by rural residents. Therefore, the neighbourhoods defined by the ONS using participatory mapping were chosen as the neighbourhoods that would be the focus of the methodology for simulation health variables outlined in this thesis.

1.6.2 Microsimulation of Health Variables

In response to the statistical barriers posed by natural neighbourhood units with low populations and insufficient numbers of health survey respondents, a technique known as "spatial microsimulation" was adopted to simulate health determinants and outcomes. The main assumption of spatial microsimulation is that the prevalence of health determinants and outcomes within our neighbourhoods will follow those estimated at the larger regional geographic scales, varying according to changing socioeconomic makeup of a set of variables common to the neighbourhoods and the large-scale survey (Lymer et al, 2009; Riva, 2012). This assumption is key to the spatial microsimulation process and to being able to leverage it in small population spaces. In this study the population of the community defined neighbourhoods in rural Ottawa range in population size from a few hundred to a couple of thousand.

Since the rural neighbourhoods of Ottawa present a small area estimation issue, the simulation of health characteristics in Rural Ottawa is a statistical down-scaling issue that can be addressed through synthetic estimation (Twigg and Moon, 2002; Lymer et al, 2009; Riva and Smith, 2010). Spatial microsimulation is a down-scaling technique that involves simulating individual- or household-level characteristics (such as health variables) within small areas (Holzer et al., 1981). The down-scaling of variables via synthetic estimation is executed through linking common variables from the spatial scale of interest with detailed, large-area, anonymous survey data like that in the CCHS. The synthetic estimation technique of spatial microsimulation can provide simulated health variables at spatial scales for which health data reporting is suppressed or not collected with sufficient sample sizes for analysis. As such, spatial microsimulation allows us to map and analyze health characteristics within small areas in the absence of spatially explicit epidemiological or social surveys. To reiterate, the key assumption here is that patterns (within the socioeconomic dimension) in health determinants and outcomes at the Health Region level in the CCHS will be reflected in rural Ottawa neighbourhood units according to variations in their socioeconomic variables.

Through this assumption, the process of microsimulation allows for health variables to be spatially explicated across rural Ottawa.

The main issue with microsimulation centres on finding socioeconomic correlates between a survey like the CCHS and the census datasets that are available within the neighbourhoods. Therein, dasymetric mapping can be a useful tool to model population counts within rural Ottawa in order to re-distribute census variables to the custom rural neighbourhood geography. A dasymetrically derived population model is critical to the subsequent estimation of the census-based SES variables, as they are used to correlate to individually detailed health survey variables recorded in the CCHS. A dasymetric population map, as opposed to one that is aerially based, can be used to re-weight SES variables to neighbourhood units defined by the ONS. The resulting neighbourhood level SES variables then support the cross-sectional spatial microsimulation to yield estimated health outcomes and determinants.

To illustrate the basic idea behind microsimulation estimation, Holzer et al. (1981) provide an example for estimating the prevalence of sickle cell anaemia in a particular county. This fictitious area has an ethnic distribution of 30% white and 70% black. A national survey has an estimated sickle cell prevalence of 10% for black residents and 0% for white residents. As such, the county-level prevalence estimate would be 7% with the affliction (Holzer et al., 1981). While simplistic, this example demonstrates the concept of coarse-level data such as national or provincial statistics being used to estimate the prevalence of a health outcome in local areas based on a single demographic variable. Analogously, this reasoning can be applied to any geographic level and estimates can be refined by multiple demographic and socioeconomic variables (and their relationships) to produce small area estimates.

A number of studies in the past decade have used statistical microsimulation to good effect in health-related research. Riva and Smith (2010) used microsimulation to estimate the prevalence of psychological distress and alcohol consumption within enumeration units (Lower Super Output Areas [LOSA]) in the UK. These estimates were obtained by linking known individuals from the Health Survey of England (HSOE) to known LOSA populations from the 2001 UK census. In the majority of LSOA areas, the simulations were within 10% of values for psychological distress and alcohol consumption (as found in the Mental Health Index in the UK).

The results in Riva and Smith (2010) indicate that microsimulation can be used to estimate health determinant and health outcome variables within geographic units for which they are not traditionally kept. Riva and Smith (2010) assume a relationship between national scale data in the HSOE and very small LSOA enumeration units. The difference in geographic scale between the national study areas in the HSOE is much larger that the geographic scale of the LSOA units. This assumption is one that could contribute to sources of error within the results.

The difference in scale between the health regions of the CCHS and the ONS rural neighbourhood units is much smaller. Thus, the assumption that the relationship between health variables and their predictors holds true for every one of the geographic units is not as presumptuous as that made in Riva et al. (2010). The smaller difference in scale in this study should increase the likelihood of having a more accurate microsimulation. However, the accuracy is not tested in this study because of the lack of independent health survey data.

Similarly, Twigg and Moon (2002) conducted a study to predict small area health-related behaviour of the prevalence of smoking at the county level in England. For this study, smoking behaviour in a large national survey was cross-tabulated against age, gender, social class, and marital status. The proportion of smokers was derived for each sub-group, and this was then applied to each sub-group in local areas (by matching variables in the national survey with those in the population census) to provide estimates of local smoking prevalence.

The methodology of microsimulation laid out in Twigg and Moon (2002), like that in Riva and Smith (2010), is very logical, but it does differ from the microsimulation process proposed in this research in one key factor. In Twigg and Moon’s study, the health determinant of smoking is being simulated in enumeration units, for which population statistics are well documented. Thus, the predictor variables for the microsimulation do not need to be estimated for the small areas for which smoking prevalence is being simulated.

This is different from the methodology laid out in this thesis, as the small area units that are the focus of the simulation have just been recently defined and do not have population statistics collected for them. Therefore the methodology of this thesis uses dasymetric mapping to perform a population based re-weighting of the variables used in the microsimulation process.

The use of dasymetric mapping for re-distributing predictor variables into small area units for microsimulation was used instead of an area based re-weighting process. This was done as it was assumed that the population surface generated by the dasymetric map would produce more accurate estimates of predictor variables than area re-weighting. This is because of the assumption of a uniform population distribution in geographic space associated with area re-weighting. It is for this reason that this thesis proposes the use of dasymetric mapping in order to provide improved estimates of predictor variables for microsimulation.

Overall, statistical microsimulation has been used for a wide range of health-related studies. Tanton and McNamara (2009) simulated poverty rates within small areas of eastern Australia. Likewise, Schneider et al. (2009) estimated county-level mammography use across the United States from state-level survey data and county-level data from the census. Konrad et al, (2009) estimated the prevalence of mental health professionals at the county level across the United States, and Ballas (2006) used microsimulation to simulate socioeconomic status and health variables across counties in Britain.

While the technique of microsimulation has been widely utilized, it faces barriers for use in small area units for which population statistics are not readily available to be used as predictor variable. This thesis seeks to improve on past methodologies by estimating predictor variables in irregular shaped geographic units through the use dasymetric mapping. The dasymetric mapping process aimed at improving the accuracy of microsimulation predictors.

1.6.3 Dasymetric Mapping

Within the above process, it is the use of dasymetric mapping that allows for the linking of common variables from large area anonymous survey data to the rural neighbourhood units. Dasymetric mapping will provide the methodology by which SES variables are re-weighted into the rural neighbourhood units.

Dasymetric mapping is an aerial interpolation method that takes advantage of ancillary data to focus population representation to where people exist in geographic space. In absolute terms, it distinguishes between populated and unpopulated areas (Poulsen and Kennedy, 2004; Langford and Higgs, 2006; Mennis and Hultgren, 2006; Hu et al., 2007) (Figure 1.6). A simple example of dasymetric mapping would be taking a land-cover map (which distinguishes between water, forest, farm land, urban centers, etc.) and overlaying it on the municipal boundaries of a town. If this town had, for example, 10,000 people, one would be able to see from the land-cover map (ancillary data) where those people live and where they cannot live within the boundaries of the town. One could eliminate areas such as water bodies, farmland, and forest as being uninhabited and have a more focused picture of where the 10,000 residents of a town were actually residing.

Figure 1.6: Contrasts the difference between how a population would be displayed in a Statistics Canada census tract (CT) in the City of Ottawa with (a) Choropleth mapping and (b) Dasymetric mapping.

For the research presented in this thesis, this method is used to create a more accurate continuous population density surface over Ottawa that facilitiates the redistribution of socioeconomic (SES) variables from the census to participatory defined neighbourhood units. The resulting simulated census data in the new neighbourhood units then forms the baseline variables used for spatial microsimulation. Using common variables between the simulated area census data and large-scale anonymized census and health survey data (CCHS), microsimulation provides a simulated population with detailed census and health characteristics for each new neighbourhood unit.

Dasymetric mapping is a key step in the methodology that is laid out in this thesis. Dasymetric mapping will be used as a means to estimate predictor (SES) variables into the neighbourhood units defined by the ONS. The estimated SES variables will then serve as a means to link large scale survey data; in this case, that means linking health variables in the CCHS to small area units. Dasymetric mapping has been utilized for health-related studies in past research (Dobson, 2002; Hay, 2004), but these studies have used dasymetric mapping in order to gain a better understanding of population distribution in populated areas for the purposes of resource allocation (Dobson, 2002; Hay, 2004). Dasymetric mapping in these studies was used by decision makers to better understand how to distribute important health services or develop emergency response plans. This thesis, however, is the first to use dasymetric mapping in conjunction with microsimulation in order to simulate health variable prevalence.

Such spatially precise information on population distribution allows for population-based re-weighting of census variables from the standard census geography to custom geographies with greater accuracy than aerial interpolation techniques can provide (Mennis, 2003). For that reason, dasymetric re-weighting should yield improved census data estimates within the custom neighbourhoods defined in this study, while also leading to improved microsimulation of health variables within those neighbourhoods. The SES census variables that are re-weighted into the new neighbourhood units will form the baseline variables for the microsimulation process. A higher level of accuracy in the re-weighting of the SES variables will in turn provide a higher likelihood of accuracy in the CCHS variables being micro-simulated.

It is important to note that there are multiple forms of dasymetric mapping and that it can be implemented in a number of ways. Eicher and Brewer (2001) summarized and compared the three main types of dasymetric mapping using a study area that included 159 counties in parts of four states and one district in the United States (Pennsylvania, West Virginia, Maryland, Virginia and the District of Columbia). The three types of dasymetric mapping include:

1. The Binary Method: In this method, ancillary data is used to redistribute population to show inhabited or uninhabited areas. The unoccupied areas are assigned a population density of 0, and the density of the inhabited areas corresponds to the density from the parent map.

2. The Three Class Method: In this method, ancillary data (typically land-cover) is used to redistribute population. In this case, the statistical surface is quantified on a percentage basis. Each one of the land classes is assigned a percentage of the population from the parent zone that will be redistributed to it.

3. The Limiting Variable Method: In this method, maximum density limits are assigned to the area class map categories. Through a process of iterative refinement, data are redistributed among the dasymetric zones to meet the maximum density thresholds set for each area class map category.

Of the three methods presented by Eicher and Brewer (2001), the limiting variable method provided the most accurate results. The binary dasymetric method only distinguishes between inhabited and uninhabited areas and the three class method only uses fixed densities, which do not take into account the area occupied by a class in a unit of analysis.

To address these issues, several researchers have developed methods of dasymetric mapping similar to the limiting variable method that correct for these shortcomings. Mennis (2003) dealt with the issue of fixed densities in the three tier classification system and suggested two new techniques that improve upon the current model. First, empirical sampling is used to determine the appropriate percentage distribution of values for each land-cover class. Secondly, area-based weighting addresses the differences in area among ancillary data classes within a given aerial unit (Mennis, 2003).

To test these methodological extensions, Mennis (2003) conducted a study to redistribute county populations in south-eastern Pennsylvania by using land-cover data that had been derived from Landsat TM imagery. His process included calculating population density fractions, area ratios, and total fractions to redistribute population among a set of grid cells. The resulting population map was a significant improvement over the original vector block-group population density map. As a result, Mennis (2003) concluded that using aerial weighting to improve the distribution of population within target zones that were based on the size of the ancillary data classes did improve accuracy, as evidenced by rural regions surrounding Philadelphia that exhibited distinct town delineations among larger open areas. The dasymetric produced by Mennis (2003) showed clear areas of high-density population in urban centers and low density in outlying rural areas bordering Philadelphia. Similar results were noted in Langford (2006).

This research adopts the methodology laid out by Mennis (2003), with the exception of the ancillary data which is different herein. Common to all the dasymetric mapping research is the use of land-use or land-cover maps to focus where populations exist. Problems arise, however, when the data is not at a sufficiently detailed spatial resolution, is out of date, or does not have enough classes for the dasymetric method (Mennis, 2003; Langford, 2006). In light of this drawback, this thesis solves the problem by performing a feature extraction from remotely sensed imagery. Specifically, high resolution imagery and a support vector machine classifier (SVMC) are used to achieve a highly spatially resolved land-use classification.

The SVMC is a supervised classification method that is used to separate data into a series of user-defined classes (Mei, 2003; Haitao, 2010). A classification task usually involves separating data into training and testing sets. As Hsu et al. (2010) explain, "each instance in the training set contains one target value (i.e. the class labels) and specific attributes (i.e. the features or observed variables). The goal of SVM is to create a model (based on the training data) which predicts the target values of the test data given only the test data attributes" (p. 1) (Figure 1.7).

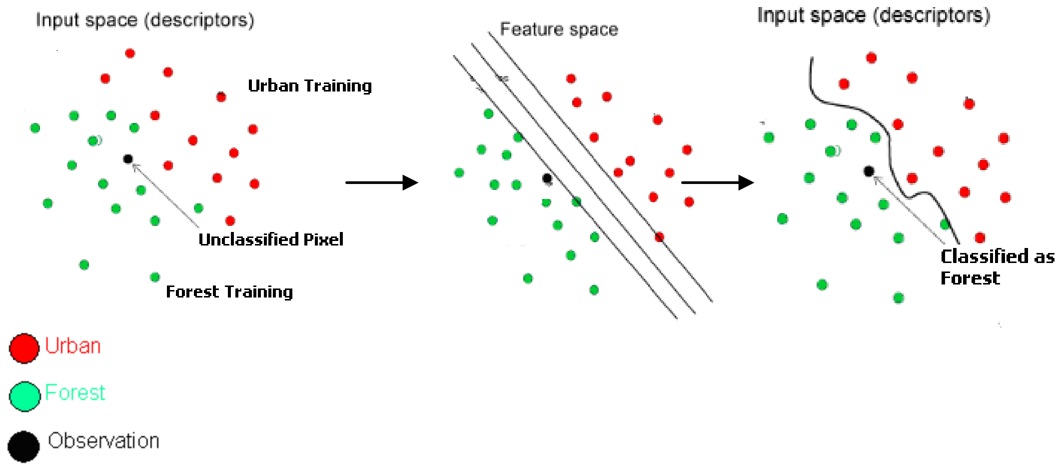


Figure 1.7: Visual interpretation of how the SVM algorithm will classify unidentified pixels into one class or another based on the plane of maximum separability derived from two training areas. Modified from StatSoft, (2011).

SVMC methodologies have become one of the primary methods of extracting land-cover classes for remotely sensed imagery. Haitao et al. (2010) conducted a study over the City of Hefei in China to test an object-oriented SVMC for land-cover classification. Quickbird multispectral imagery from 2005 with a resolution of 0.61 m was used to extract six classes of land cover from the city, including built-up, agricultural land, grassland, trail, road, and water bodies. The object-oriented classification involved identifying different features on the images and delineating them with vector polygons.

These polygons are defined through spectral and textural differences in the image. Areas of similar spectral and textural values were assumed to be a distinctive feature or land cover, and are therefore classified as a unique polygon. After the object classification of the image into polygons, a SVMC methodology was then used to classify each polygon into one of six desired classes. The SVMC methodology is a supervised classification technique, so training samples are identified by the user as a particular class. Next, the average radiance value, area, perimeter, and standard deviation of spectral values are computed for all of the selected objects (Haitao, et al., 2010). These parameters are also calculated for the rest of the objects in the image. Finally, based on their similarity, the features of a training area are assigned to a land-cover class (Haitao et. al., 2010). Using this method, a detailed land-cover map of a city can be created.

A land-use map such as the one created in Haitao et al. (2010) can be used as an ancillary dataset for dasymetric mapping. To use it, one would remove the built-up class from the land-use map and then overlay it with the population data that needs to be disaggregated. This would allow one to refine the area in which people are known to be residing within a unit of analysis. Essentially, the result would provide a binary dasymetric map.

1.7 Summary

This research proposes a methodology based on the techniques of dasymetric mapping and spatial microsimulation to simulate health variables within the set of community-defined neighbourhoods in rural Ottawa. The method is based on a two-step process of deriving neighbourhood-level variables common to regional health surveys from census data using dasymetric mapping prior to microsimulation. This work is the first to use dasymetric mapping to estimate the predictor variables that will be used in the spatial microsimulation process. Ultimately, the results are a set of maps that allow the exploration of local-level health determinant and outcome variation within geographic units that better correspond to the scale of intervention that should be most effective for the communities of Rural Ottawa.

This work is a methodological contribution and the simulated health outcomes and health determinant within the participatory defined neighbourhood units cannot be directly validated due to the lack of independent health-survey data. There are not government or third party datasets that can be used to test the accuracy of the simulation. Rather, the research focuses on innovation of the microsimulation methodology as a potentially improved way to explore health-related variables within sparsely populated rural regions. The results of the microsimulation are amenable to spatial statistical analysis such as tests of spatial dependence in outcomes and determinants (e.g., Moran's I) and, intrinsically, are well suited to the exploration of health patterns that could be further validated by primary data collection.

An additional contribution of this research lies in the use of community neighbourhood boundaries that are defined via a participatory mapping approach.

Community derived neighbourhood boundaries represent those areas with which residents actually associate and self-actualize. These neighbourhoods are, however, created with an emphasis on input from the health care sector and represent a single spatial configuration and are not "absolute" nor the only or even best definition available. Rather, these neighbourhoods present a different approach to defining the geographic unit at which health can be measured and one at which population health-based interventions should be more effective since the boundaries represent a better definition of the social connections among the constituent individuals.

Overall, this thesis integrates complex techniques from the fields of geomatics and statistics in order to present an innovative methodology and toolset for the investigation of rural health. Through this effort, health variables in rural Ottawa will be investigated at geographic scales that have not been previously possible.

1.8 Thesis Structure

The thesis is organized into the following chapters:

Chapter 2: Consists of a methodological paper outlining the process used to estimate health determinant and outcome variables in community-defined neighbourhoods within rural Ottawa. Previous studies in dasymetric mapping, support vector machine (SVM) feature extraction, and microsimulation are discussed while presenting the main methodology used.

Chapter 3: Discusses the results of the entire thesis. Conclusions about dasymetric mapping and microsimulation are also discussed. Limitations and challenges encountered during the research process are also analysed. Finally, future research opportunities are presented.

*It should be noted Chapter Two of this thesis is meant to be stand-alone document. For this reason, there are some areas of Chapter Two that repeat material from Chapters One and Three.

1.9 Contributing Authors

With the wide range of disciplines integrated into this methodology, several experts contributed to the research put forth in this thesis. Below is a list of experts and the areas of research in which they were involved:

Dr. Michael Sawada: Dr. Sawada's expertise in the geomatics field was critical in assisting with the development of the dasymetric mapping process and the acquisition of accurate ancillary data obtained through support vector machine feature classification. Dr. Sawada also oversaw the development of the entire thesis and the writing of Chapter 2 as well as editing the document.

Dr. Elizabeth Kristjansson: Dr. Kristjansson's expertise in community-based participatory research were instrumental in guiding the participatory mapping community consultations in rural Ottawa. She also contributed to the funding of this research together with Dr.Sawada with a grant from PHRN.

Dr. Jean-Michel Billette: Dr. Billette undertook the development and execution of the microsimulation process that estimated health outcomes in rural Ottawa. He also wrote the section of Chapter 2 on spatial microsimulation, including the derivation of the equations.

Dr. Mylen Riva: Dr. Riva's knowledge of the challenges of using census enumeration units in epidemiological research was very valuable in the development of the methodology put forward in this thesis. Dr. Riva's past use of microsimulation in health research was extremely helpful in understanding how to apply the technique to this thesis. She also contributed to the PHRN grant that funded this work.

References

Aggleton, P. (1990). Society Now—Health. London. Routledge.

Andresen, M. A., & Brantingham, P. L. (2008). Visualizing ambient population data within census boundaries: A dasymetric mapping procedure. *Cartographica, 43* (4), 267-275.

Badets J, Chui TWL. (1997). Canada's Changing Immigrant Population. (*Catalogue 96-311E). Ottawa: Statistics Canada.*

Ballas, D., G. Clarke, D. Dorling, H. Eyre, B. Thomas, and D. Rossiter. (2005). SimBritain: A Spatial Microsimulation Approach to Population Dynamics. *Population, Space and Place*, (11) 13-34.

Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine*, 51, 843–857.

Berrigan, D., Troiano, R.P. (2002) The association between urban form and physical activity in US adults. *American Journal of Preventive Medecine*, 23, 74–79.

Beyer, K. M. M., Comstock, S., & Seagren, R. (2010). Disease maps as context for community mapping: A methodological approach for linking confidential health information with local geographical knowledge for community health research. *Journal of Community Health, 35*(6), 635-644.

Bonham-Carter, G. F. (1994). Geographic information systems for geoscientists: Modeling with GIS. Oxford; New York: Pergamon.

Bozheva, A. M., Petrov, A. N., & Sugumaran, R. (2005). The effect of spatial resolution of remotely sensed data in dasymetric mapping of residential areas. *GIScience and Remote Sensing, 42*(2), 113-130.

Canadian Community Health Survey. (2001). http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=3226&SurvVer=0&InstaId=15282&InstaVer=1&SDDS=3226&lang=en&db=IMDB&adm=8&dis=2.

Chen, K., McAneney, J., Blong, R., Leigh, R., Hunter, L., & Magill, C. (2004). Defining area at risk and its effect in catastrophe loss estimation: A dasymetric mapping approach. *Applied Geography, 24*(2), 97-117.

Chen J, Ng E, Wilkins R. (1995). Life expectancy and health expectancy of Canadian immigrants from 1986 to 1991. In *Immigration and life expectancy in Canada.* (Statistics Canada, Catalogue 89F0084XPE) 9-22

City of Ottawa Facts. http://ottawa.ca/en/visitors/about/quick_facts/index.html

Clapp, J. M., and Y. Wang. (2006). Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban, Economics*, 59(1), 259-84.

Cornwall, A., & Jewkes, R. (1995). What is participatory research? *Social Science & Medicine*, 41(12), 1667-1676

Crampton, J. W. (2004). GIS and geographic governance: Reconstructing the choropleth map. *Cartographica, 39*(1), 41-54.

Cubbin C., LeClere F., Smith G. (2000). Socioeconomic status and injury mortality: individual and neighborhood determinants. *Journal of Epidemiology Community Health*, 54,517–524.

Cummins, S., Stafford, M., Macintyre, Sally., Marmot, M., Ellaway, A. (2005). Neighbourhood environment and its association with self rated health: evidence from Scotland and England. *Journal of Epidemiology and Community Health*, 59, 207-213.

Dent, B. (1999) Fifth Edition, Cartography: Thematic Map Design. WCB/McGraw-Hill.

Dunn JR, Dyck I. (2000). Social determinants of health in Canada's immigrant population Results from the National Population Health Survey. *Social Science and Medicine*, 51, 1573- 93.

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and aerial interpolation: Implementation and evaluation. *Cartography and Geographic Information Science, 28*(2), 125-138.

Ellen, I. G., and M. A. Turner. (1997). Does neighborhood matter? Assessing recent evidence. Housing *Policy Debate* 8(4), 833-66.

Elliott, P., Martuzzi, M., & Shaddick, G. (1995). Spatial statistical methods in environmental epidemiology: A critique. *Statistical Methods in Medical Research, 4*(2), 137-159.

Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives, 112*(9), 998-1006.

Fan, V. S., Au, D., Heagerty, P., Deyo, R. A., McDonell, M. B., & Fihn, S. D. (2002). Validation of case-mix measures derived from self-reports of diagnoses and health. Journal of Clinical Epidemiology, 55(4), 371-380.

Fisher, P. F., & Langford, M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of aerial interpolation by dasymetric mapping. *Professional Geographer, 48*(3), 299-309.

Frank, L.D. (2000). Land use and transportation interactions: implications in public health and quality of life. *Journal of Planning Education Research*, 6–22.

Evans RG, Stoddart GL. (1990). Producing health, consuming health care. *Social Science and Med*, 31(12), 1347-63.

Germain, A., and J. E. Gagnon. (1999). Is neighborhood a blackbox? A reply to Galster, Metzger, and Waite. *Canadian Journal of Urban Research* 8(2): 172.

Grengs, J. (2007). Reevaluating poverty concentration with spatial analysis: Detroit in the 1990S. *Urban Geography, 28*(4), 340-360.

Haitao, Li, Haiyan Gu, Yanshun, Han and Jinghhi, Yang. (2010). Object-oriented classification of high resolution remote sensing imagery based on an improved colour structure code and support vector machine. *International Journal of Remote Sensing*, 43(6), 1453-1470.

Hameed, S. M., Bell, N., & Schuurman, N. (2010). Analyzing the effects of place on injury: Does the choice of geographic scale and zone matter? *Open Medicine*, 4(4), 171-180.

Hampton, K. H., Serre, M. L., Gesink, D. C., Pilcher, C. D., & Miller, W. C. (2011). Adjusting for sampling variability in sparse data: Geostatistical approaches to disease mapping. *International Journal of Health Geographics, 10* (1), 54.

Hatzenbuehler, M., Keyes, K., McLaughlin, K. (2011). The protective effects of social/contextual factors on psychiatric morbidity in LGB populations. *International Journal of Epidemiology*, 40, 1071-1080.

Hay, S. I., Noor, A. M., Nelson, A., & Tatem, A. J. (2005). The accuracy of human population maps for public health application. *Tropical Medicine and International Health*, *10*(10), 1073-1086.

Haynes R., Daras, K., Reading, R. and Jones, A. (2007)Modifiable Neighbourhood Units, zone design and residents perception. *Health and Place*, 13(4): 812-825

Haynes R., and Gale S. (2000). Deprivation and poor health in rural areas: inequalities hidden by averages. *Health and Place 6*, 275-285.

Holt, J.B., C.P. Lo, and T.W. Hodler. (2004.) Dasymetric Estimation of Population Density and Aerial Interpolation of Census Data. Cartography and Geographic Information Centre, 31(2), 103-121.

Holzer, C., Jackson, D., Tweed, D. (1981). Horizontal synthetic estimation: A strategy for estimation small are health-related characteristics. *Evaluation and Program Planning*,(4) 1, 29-34

Hsu Chih-Wei, and Chih-Jen Lin, (2010). A practical guide to support vector classification. Department of Computer Science, National Taiwan University. 1-16

Hu, Z., Liebens, J., & Rao, R. (2007). Exploring relationship between asthma and air pollution: A geospatial methodology using dasymetric mapping, GIS analysis, and spatial statistics. Paper presented at the Geoinfromatics Conference *6753*(PART 2)

Idler, E. L., & Kasl, S. (1991). Health perceptions and survival: Do global evaluations of health status really predict mortality? *Journals of Gerontology,* 46(2), S55-S65.

Idler, E. L., & Kasl, S. V. (1995). Self-ratings of health: Do they also predict change in functional ability*? Journals of Gerontology* - Series B Psychological Sciences and Social Sciences, 50(6), S344-S353.

Krause, N. M., & Jay, G. M. (1994). What do global selfrated health items measure? *Medical Care,* 32(9), 930-942.

Kiwachi, I., Kennedy, B., and Glass R. (1999). Social Capital and Selfrated health: A Contextual Analysis. *American Journal of Public Health*. 89(8), 1187-1193.

Kawachi, I., and L. F. Berkman. 2004. Introduction in neighborhoods and health. In I. Kawachi and L. F. Berkman, Eds., *Neighborhoods and Health*. New York: Oxford University Press, 1, 1-19.

Konrad T., Ellis A., Thomas K., Holzer C., Morrissey J.(2009). County-Level Estimates of Need for Mental Health Professionals in the United States. *Psychiatric Services,* (60)10, 1307-1314.

Krieger N., Chen J., Waterman P., Soobader M., Subramanian S., Carson R. (2003) Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US) *Journal of Epidemiology Community Health*, 57,186–199.

Kristjansson, E., Sawada, M., Russell, K., Ali, A., He, J., McKinnon, J., Hole, D., Steeves, P., Palozzi, K., Leclair, S., and Calhoun, Ml. (2009). Neighbourhood Advantage, Social Determinants, and Health Outcomes. The Ottawa Neighbourhood Study. Invited Presentation at "The InSPQ Deprivation Index: Applications in Research, Policy and Practice." Canadian Institute of Health Information Workshop, Sept. 23, 2009, Montreal, Canada."

Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems, 30*(2), 161-180.

Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems, 31*(1), 19-32.

Langford, M., & Higgs, G. (2006). Measuring potential access to primary healthcare services: The influence of alternative spatial representations of population. *Professional Geographer, 58*(3), 294-306.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *Cartographic Journal, 31*(1), 21-26.

Lopes-Zetina, J., Kee, H., and Friis, R. (2006). The link between obesity and the built environment. Evidence from and ecological analysis of obesity and vehicle miles of travel in California, *Health and Place, 12(4), 656-664.*

Lymer S., Brown L. (2009). Predicting the need for aged care services at the small area level: The CAREMOD spatial microsimulation model. *International Journal of Microsimulation*, 2(2), 27-42.

Maantay, J. (2002). Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environmental Health Perspectives*, 110(2), 161–171.

Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Cartography and Geographic Information Science,* 34(2), 77-102.

Macintyre S., Ellaway A., and Cummins S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science and Medicine*, 55, 125-139.

Maman S, Lane T, Ntogwisangu J, Modiba P, vanRooyen H, Timbe A, Visrutaratna S, Fritz K. (2009). Using Participatory Mapping to Inform a Community-Randomized Trial of HIV Counselling and Testing. *Field Methods, 21*, 368-387.

McMichael, A. J. (1999). Prisoners of the proximate: Loosening the constraints on epidemiology in an age of change. *American Journal of Epidemiology*. 149, 887–897.

Martin, D. G. (2003). Enacting neighborhood. *Urban Geography*, *24*(5), 361-85.

Mei, T., Li, F., Qin, Q., & Li, D. (2003). Road extraction from remote sensing image using support vector machine. Paper presented at the International Society for Optical Engineering,5286(1) 299-304.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *Professional Geographer, 55*(1), 31-42.

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass, 3*(2), 727-745.

Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and its application to aerial interpolation. *Cartography and Geographic Information Science, 33*(3), 179-194.

Merz, J. (1991). Microsimulation - A survey of principles, developments and applications. *International Journal of Forecasting, 7*(1), 77-104.

M.-P. Parenteau, M. Sawada, Sam Herold, Elizabeth A. Kristjansson, R. Labonte, M. Calhoun, S. Leclair, V. Runnels, A. Musiol, R. Mawby, N. Muhajarine, D. Salisbury, Amira Ali, Marguarite Keeley, Ian Cross, Cliff Gazee (2007). A Participatory Geomatics Approach to Defining Health-Relevant Neighbourhoods in Ottawa, Canada. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 84*(6) 856.

Matuk LC. (1996) Health status of newcomers. *Canadian Journal of Public Health.* 87, 152-7.

Moore, A.J., (1995). Deprivation payments in general practice: some spatial issues in resource allocation in the UK. *Health and Place*, 1, 121-125.

Neudemark, D., Rahav, G. and Jeffe, D. (2002) Socio-economic status and binge drinking in Israel. *Drug and Alcohol Dependence*. 1(69), 15-21.

Oakes M. (2004) The (mis)estimation of neighbourhood effects: casual inference for a practicable social epidemiology. Social Science and Medicine. 58, 1929-1952.

O'Fallon, L. R., & Dearry, A. (2002). Community-based participatory research as a tool to advance environmental health sciences. *Environmental Health Perspectives*, *110*(2), 155–159

Pampalon, R., Hamel, D., & Gamache, P. (2008). Recent changes in the geography of social disparities in premature mortality in Québec. *Social Science and Medicine, 67*(8), 1269-1281.

Pampalon, R., Hamel, D., Gamache, P., & Raymond, G. (2009). A deprivation index for health planning in Canada. *Chronic Diseases in Canada, 29*(4), 178-191.

Pfeffermann, D. (2002). Small area estimation - new developments and directions. *International Statistical Review*, 70(1), 125-143

Pickett K., Pearl, M. (2001). Multilevel analyses of neighborhood socioeconomic context and health outcomes: critical review. *Journal of Epidemiology Community Health, 55*, 111–122.

Poulsen, E., & Kennedy, L. W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology, 20*(3), 243-262.

Rainham, D., McDowell, I., Krewski, D., & Sawada, M. (2010). Conceptualizing the healthscape: Contributions of time geography, location technologies and spatial ecology to place and health research. *Social Science and Medicine*, 70(5), 668-676.

Riva, M., Apparicio, P., Gauvin, L. & Brodeur, J-M. (2008). Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones. *International Journal of Health Geographics,* 7 (43)

Riva, M., Smith M. (2010). Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method. *Social Psychiatry Epidemiology*, *47*, 745-755.

Ross, N. A., Tremblay, S., & Graham, K. (2004). Neighbourhood influences on health in Montréal, Canada. *Social Science and Medicine*, *59(*7), 1485-1494.

Sampson, R. J., J. D. Morenoff, and T. Gannon-Rowley. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology 28*, 443-78.

Sanmartin C, Gilmore J. (2005). Smoking and Diabetes Care: Results from the CCHS Cycle 3.1. *(Catalogue 82-621-XIE) Ottawa: Statistics Canada.*

StatSoft, Inc. (2011). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

Schneider K., Lapane K., Clark M., Rakowski W. (2009). Using Small-Area Estimation to Describe County-Level Disparities in Mammography. *Public Health Research, Practice and Policy*, *6*(4), 1-10.

Sherwood, N.E., Jeffery, R.W. (2000) The behavioral determinants of exercise: implications for physical activity interventions. Annual Review of Nutrition 20, 21–44.

Subramaniana S., Kawachib I., Kennedy B. (2001). Does the state you live in make a difference? Multilevel analysis of selfrated health in the US. *Social Science & Medicine*, *53*, 9–19.

Srebotnjak, T., Mokdad, A. H., & Murray, C. J. L. (2010). A novel framework for validating and applying standardized small area measurement strategies. *Population Health Metrics, 8*

Stafford, M., Cummins, S., Ellaway, A., Sacker, A., Wiggins, R. abd Macintyre S. (2007) Pathways to obesity: Identifying local, modifiable determinats of physical activity and diet. 65(9), 1882-1897.

StatSoft, Inc. (2011). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

Susser, M. (1999). Should the epidemiologist be a social scientist or a molecular biologist? *International Journal of Epidemiology*. 28, 1019–1022.

Tanton, M., Harding, A., McNamara, J. (2009). Urban and Rural Estimates of Poverty: Recent Advances in Spatial Microsimulation in Australia. *Geographical Research*, 41(1), 52-64.

Twigg, L., Moon, G. (2002). Prediction small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data. *Social Science and Medicine.* 54(6), 931-937

UN (2001). World Population Monitoring (2001). Population, Environment and Development. ST/ESA/SER.A/203. United Nations, New York.

Viner, R.M., Ozer, E.M., Denny, S., Marmot, M., Resnick, M., Fatusi, A., Currie, C. (2012). Adolescence and the social determinants of health. *Public Medicine* 379(9826),1641-52

Yuan, Y., Smith, R. M., & Limp, W. F. (1997). Remodeling census population with spatial information from landsat TM imagery. *Computers, Environment and Urban Systems*, *21*(3-4), 245-258.

Chapter 2: Estimating Health Outcomes in Rural Ottawa: A Geographic and Statistical Methodology.

*Target Journal: *International Journal of Health Geographics*

Authors.

Mosley, Brian; Sawada, Michael; Billette, Jean-Michel; Kristjansson, Elizabeth; Riva, Mylen

Abstract

Background: Many health geography studies, including the Ottawa Neighbourhood Study (ONS), have faced significant challenges uncovering local variation in patterns of community health in rural areas. This is due to the fact that sparsely populated rural areas make it difficult to define neighbourhoods that are representative of the social and resource utilization patterns of the individuals therein. Moreover, rural areas yield small samples from population-based regional health surveys and this leads to insufficient sample sizes for reliable estimation of health determinants and outcomes. ***Method:*** We present a methodology that combines dasymetric mapping and microsimulation to simulate health determinants and outcomes within self-identified community-defined neighbourhoods of rural Ottawa. ***Results*:** Dasymetric mapping provided neighbourhood-level population estimates that were used to re-weight a set of SES variables that were correlates with those in the Canadian Community Health Survey (CCHS). These neighbourhood-level correlates allowed microsimulation and consequent spatial exploration of prevalence for smoking, binge drinking, obesity, selfrated mental health, and the presence of two or more chronic conditions . ***Conclusions:*** The methodology outlined in this paper provides an innovative way of exploring health determinants and health outcomes in neighbourhoods for which population and health statistics are not traditionally collected at levels that would allow traditional statistical analyses of prevalence.

Contexte: De nombreuses études géographie de la santé, y compris l'étude de voisinage Ottawa (ONS), ont dû relever des défis importants découvrant les modèles en santé communautaire dans les zones rurales. Cela est dû au fait que les densités de population faibles et sporadiques en zone rurale, il est difficile de définir les quartiers, qui sont représentatifs des communautés homogènes, tandis que l'obtention d'un échantillon de la population qui sont assez grands pour une estimation fiable de variables de santé. ***Méthode***: Pour résoudre ce problème, cette étude propose une méthodologie combinant la cartographie dasymétrique et de microsimulation qui permettra aux chercheurs de simuler déterminants de la santé et les résultats dans la communauté définis par unités de voisinage dans les régions rurales d'Ottawa. ***Résultats***: la cartographie dasymétriques abouti à des estimations précises des populations et a donc été utilisée pour estimer les variables SSE au sein de la communauté définis par unités de voisinage. En utilisant les variables estimées SES en tant que corrélats entre les quartiers communautaires définies et canadienne Enquête sur la santé collectivités canadiennes (ESCC) régions sanitaires de microsimulation a été utilisé pour simuler déterminant de la santé et des prévalences résultats pour la santé. ***Conclusions***: La méthodologie décrite dans le présent document, fournit et de manière innovante d'explorer les déterminants de la santé et les résultats sanitaires dans les quartiers pour laquelle des statistiques démographiques et de santé ne sont pas traditionnellement élevés.

Background

In order to understand the health of a community, it is necessary to have an accurate depiction of the neighbourhoods with which people interact [1-4]. Knowing the spaces with which a community interacts is essential to identifying and understanding the contextual variables that impact their health [2, 5-9]. Defining neighbourhood units that accurately delineate the spaces that people identify with is therefore critical for place-based research; as Schwab and Syme (1997) state, for researchers to be able to accurately assess the health of a community, they must first be able to ‘‘reflect the ecological reality of life in that population, as people experience [it]” (p. 2051). Recent studies have supported this statement, and it has been demonstrated that a person's health is not just a result of personal compositional variables such as diet, attitude, and exercise, but is also impacted by contextual variables in their community [10-14].

Traditionally, health geography studies have set their research around government-defined enumeration units [15]. Recently, however, it has been shown that enumeration units may hide the data variations that occur within them because they have been defined by government bodies for population census purposes and, therefore, may not always be associated with existing discontinuities in population [16-19]. Thus, we see that census enumeration units may not be ideally suited to study the relationships between health and place.

In response to this issue, several studies have attempted to define "natural neighbourhood units" in an effort to create a more appropriate unit at which to study health [4, 20, 21]. Natural neighbourhood units have become increasingly utilized in health research in order to try and provide a more realistic representation of the spaces with which people interact [2, 20]. However, the use of such natural or self-identified neighbourhoods faces barriers for health research in rural areas. For example, in order to capture a sample population that is large enough to reliably estimate health variables, neighbourhood boundaries must cover large physical areas in rural regions to capture a sufficient population sample [2, 20, 22,23]. As a result, one can see that in rural neighbourhoods, the population size of a neighbourhood is critical to examining statistical relationships between health and place.

In 2008, the Ottawa Neighbourhood Study (ONS) undertook the task of defining natural neighbourhood units across the City of Ottawa [20]. The results of the study were well received with the exception of residents within the rural regions of the city [2]. This was due to the fact that rural neighbourhoods were large in physical area so residents therein did not identify with them. These large areas were necessary from an analytical perspective in order to capture a population area that would allow for the reliable estimation of health variables from regional survey data [20]. For example, the population density in the neighbourhoods of Osgoode, North Gower, and Cumberland (all of which are outside of the urban core of Ottawa) is typically low, except around small townships (Figure 2.1). The results were rural neighbourhoods that did not represent the "ecological reality of the people living in them" (p.2051) [5]. For that reason, the objective of this research focuses on combining statistical and geographic techniques that allow for the simulation of health-related statistics in rural Ottawa neighbourhoods that were defined through a participatory methodology and that have insufficient population and area to support standard statistical estimation techniques.

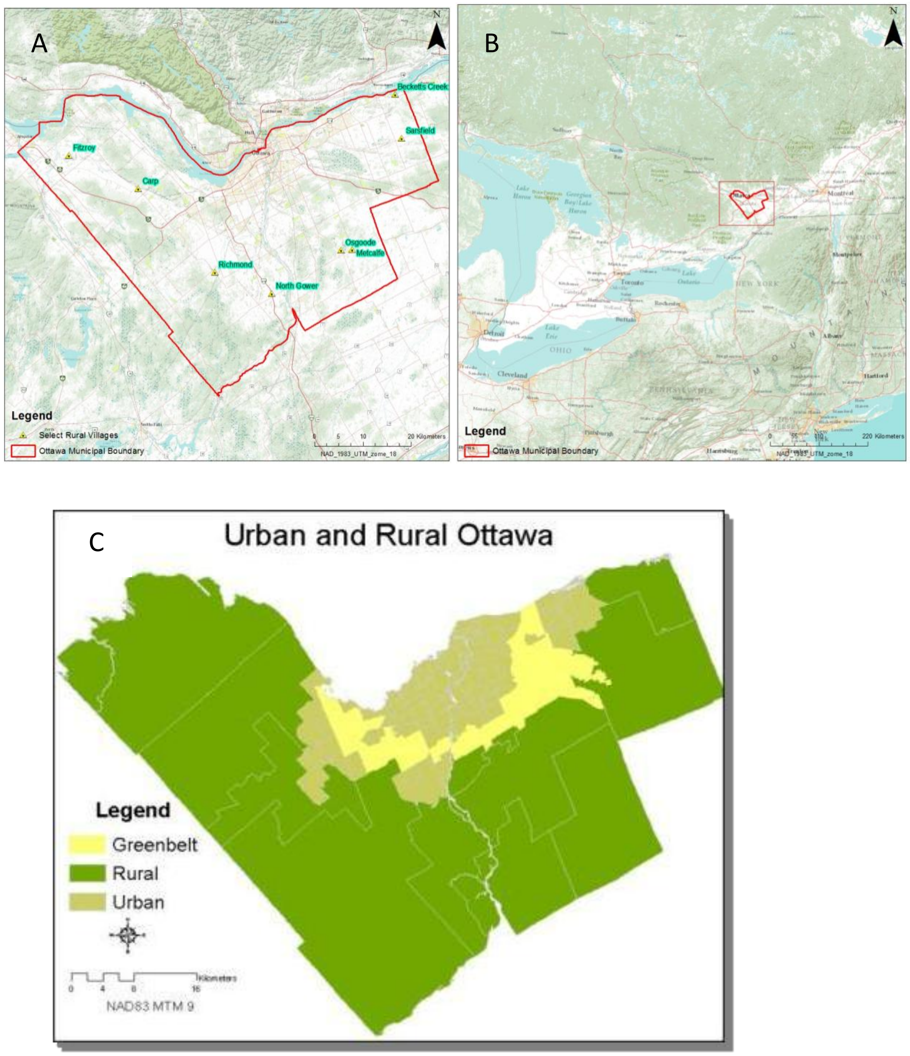


Figure 2.1: a) Select townships within the municipal boundary of the city of Ottawa, Ontario, Canada; b.) The location of the Ottawa municipal area, within southern Ontario, Canada; c) Urban and rural Ottawa and the area known as the "Greenbelt."

The City of Ottawa is located in the far east of Ontario on the banks of the Ottawa River across from Gatineau, Que. (Latitude 45° 19'N and Longitude 75° 40'W). In 2001, the City of Ottawa amalgamated with 11 municipalities (Cumberland, Gloucester, Goulbourn, Kanata, Nepean, Osgoode, Ottawa, Rideau, Rockcliffe Park, Vanier, and West Carleton), and these municipalities now comprise the majority of rural Ottawa (Figure 2.1). The rural neighbourhoods are very sparsely populated, with the exception of small villages that contain higher population densities similar to those of the urban core. After amalgamation in 2001, Ottawa became the city with the largest municipal area in Canada at 2,796 square km, but with just over 900,000 people, it is only the sixth largest population in the country [24]. Furthermore, the City of Ottawa estimates that up to 80 percent of its municipal area is rural [24]. Despite occupying four fifths of the city's land area, however, Ottawa's rural population stands at just over 85,000, and aside from village centers is very sparsely populated [24].

The geographic dispersion of population in rural Ottawa makes it very difficult to estimate local relationships between health and place. Large study areas are needed in order to capture a large enough sample size to reliably estimate health patterns in the area. This, however, is problematic, as study areas become so big that they do not represent the areas with which rural residents interact on a daily basis. Thus, researchers lose their ability to properly analyse contextual variables related to health [2, 5].

Methods

In response to the issues presented by existing census based enumeration units and natural neighbourhood units, this research presents a methodology to simulate health-related variables within the boundaries of community defined neighbourhoods that contain, for the most part, small populations too small for standard statistical estimation. To this end, population-based re-weighting of census-level variables (socioeconomic status, or SES) to the community defined neighbourhood-level is undertaken using a dasymetric mapping process. Finally, spatial microsimulation is undertaken to simulate specific health determinants and outcomes within the new neighbourhood units defined by rural communities within the purview of the Ottawa Neighbourhood Study. Dasymetric mapping and spatial microsimulation are combined in an innovative methodology to provide small area simulations of health variable prevalence within small, rural geographic units. A three step approach is taken to overcome the aforementioned issues within the rural regions of Ottawa:

- Dasymetric population mapping is undertaken for both urban and rural Ottawa;

- The dasymetric population map is used to redistribute 2006 Canadian census socioeconomic variables of age, sex, income, education, language, and visible minority categories of the into the community-defined neighbourhoods; and

- The socioeconomic data are correlates of variables in the Canadian Community Health Survey and so are used as a basis for microsimulation of health data. Three health outcomes and three health determinates are simulated. The outcomes include selfrated health, selfrated mental health, prevalence and the presence of two or more chronic conditions, while the determinants include the prevalence of obesity, smoking and binge drinking.

It is important to note that the population health determinants and health outcome variables calculated in this research are simulations and that the lack of independent data on these health variables precludes any tests of simulation accuracy. This work is methodological and illustrates how health patterns in rural neighbourhoods may be explicated so that patterns can be explored and potentially lead to further primary investigations at the community, municipalities or health professional level.

Neighbourhoods of Focus—Rural Ottawa

It has been shown that the involvement of communities can enhance studies by providing local context, thereby improving the quality of data gathered in the study and the overall results [25-27]. Participatory mapping has been used to successfully engage communities in the research process [28, 29]. Participatory mapping allows the subjects of a research project to articulate spatial components that they feel are important to the research process by drawing maps of the areas with which they identify [3, 28, 29].

In 2011, the ONS undertook a participatory mapping exercise in rural Ottawa [30]. Community consultations were set up in four districts in rural Ottawa: West, East, South and South East. Participants at the consultations were invited by ONS organizers if they had expressed interest in the ONS project since its conception in 2006. In an effort not to exclude anyone who may have valuable insight for the participatory mapping exercise, however, participants were also encouraged to bring guests. Ultimately, the consultations contained between 10 -25 individuals, with participants largely being from the health care, political and public service sectors. Some typical participants would include doctors, registered nurses, police officers, community group leaders, City of Ottawa rural planners, and elected representatives from various levels of government. The participatory mapping exercise was aimed at finding a consensus among participants on where rural neighbourhood[3] boundaries should be drawn [30].

For several reasons, the neighbourhood boundaries defined through the ONS participatory mapping exercise are used for this study (Figure 2.2). Firstly, Haynes et al. (2007) discovered that neither arbitrary neighbourhood boundaries nor expertly defined natural neighbourhood units coincided with community perceptions. As such, community defined neighbourhoods should be better to present health determinants and health outcomes in a meaningful way to rural Ottawa residents. As a consequence, meaningful neighbourhood units should provide a scale more effective for health interventions as well as future research in rural Ottawa. At the same time, however, community defined neighbourhoods cannot have any population size constraints that would ensure reliable statistical estimation. This meant that the neighbourhoods defined by the ONS were not influenced by sample size constraints associated with traditional forms of estimation, but rather that they were totally a product of the perceptions of community and neighbourhood by rural residents. Therefore, in order to study health variations among the ONS rural neighbourhoods, a new

[3] The concept of neighbourhood is used liberally in this context and the term 'community of interaction' could also be enlightening.

methodology for simulation of health variables was required. This new methodology is the focus of this research.

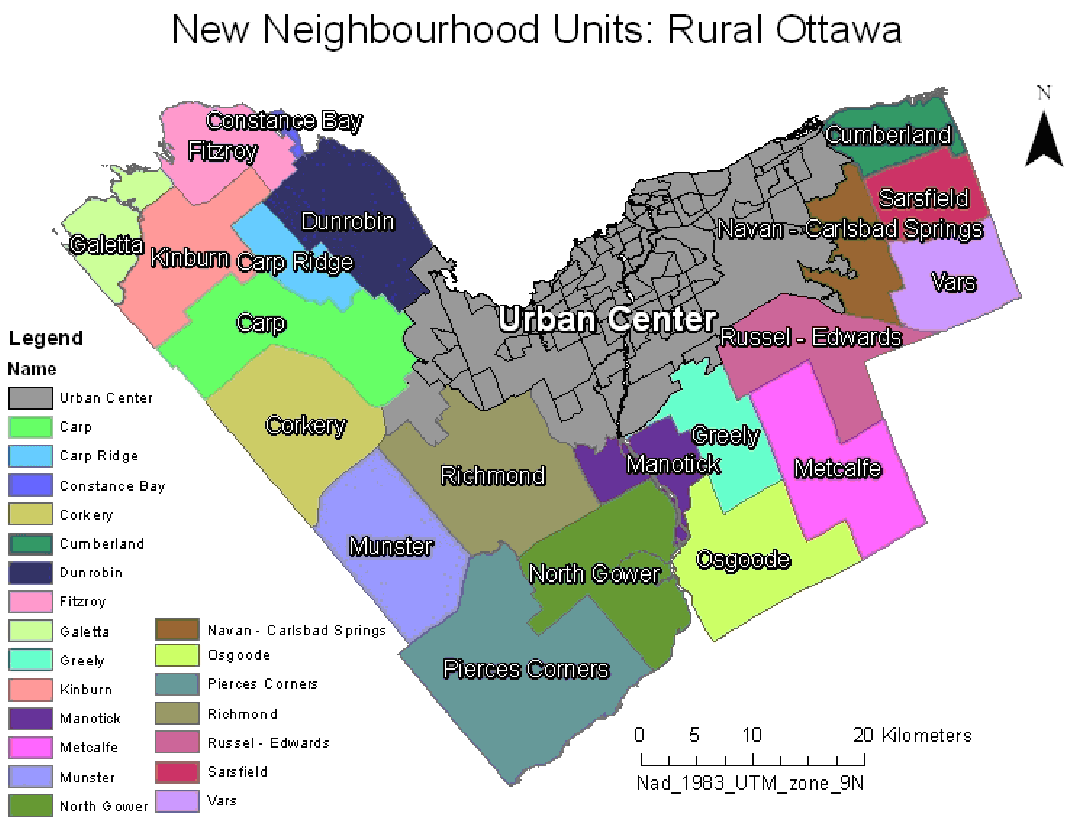


Figure 2.2: Community-defined neighbourhood units from the ONS participatory mapping of rural Ottawa. Defined by the ONS, 2011

Dasymetric Mapping

Dasymetric mapping is an aerial interpolation method that takes advantage of ancillary data to focus population representation to exactly where people live in geographic space. This is opposed to census units that assume an equal distribution of population across space. In absolute terms, dasymetric mapping distinguishes between populated and unpopulated areas at arbitrarily fine spatial scales [31-36]. This method is used herein to create a continuous population density surface over Ottawa, thereby facilitating the redistribution of socioeconomic (SES) variables from the Canadian census to community defined neighbourhood units. The resulting re-distrubuted census data in each neighbourhood forms the basis for SES variable matching with the Canadian Community Health Survey (CCHS) and therefore allows the spatial microsimulation of health variables. Using common variables between the large-scale anonymized census and health survey data (CCHS), microsimulation then provides the needed simulated health characteristics for each neighbourhood unit. We use the dasymetric mapping methodology described in Mennis (2003).

Feature Extraction

To support dasymetric mapping, feature extraction of the different land covers within the City of Ottawa is the first step to define the geographic spaces that people can inhabit. The land cover is one of two geospatial datasets that are used as ancillary data in the dasymetric mapping process; the second being City of Ottawa zoning data.

The imagery used for the feature extraction is composed of 3600 ortho-rectified, multispectral aerial photographs at a resolution of 20x20 cm with the three basic red, green, and blue spectral channels (RGB channels) (Figure 2.3). The imagery is provided by the Digital Raster Acquisition Project of Eastern Ontario (DRAPE), conducted by the Ontario Ministry of Natural Resources between 2008-2009, and they provide the highest resolution available for Ottawa. The fine spatial resolution for analysis, coupled with the sheer number of images and computational burden of object-based classification for landcover feature extaction, requires high-performance computing. As such, the analysis of the DRAPE images was conducted using feature extraction software developed by Incogna GIS Inc., whose innovative cloud-based graphics processing unit (GPU) software allows the user to perform feature extraction using a support vector machine (SVM) on extremely large datasets.

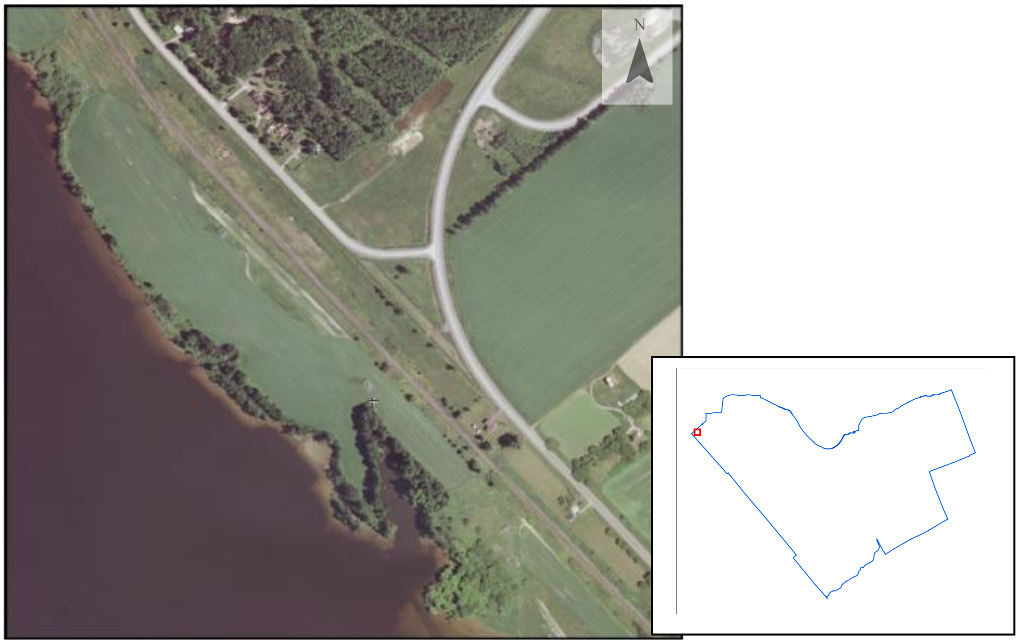


Figure 2.3: One tile of ortho-rectified, multispectral aerial imagery over western Ottawa at a resolution of 20x20cm. Source: Ontario Ministry of Natural Resources (see text for full explanation).

Using a combination of Incogna's object-oriented segregation and SVM classification, four different types of land cover were identified: water, open field or agricultural land, forest, and built-up. The object-oriented segmentation delineates polygons around pixels in the image that have similar spectral signatures (within RGB channels in this case), feature sizes, shapes, directionality, repetition, and context information. These polygons approximate unique features in the image that are different from surrounding features. The full image is tessalated into these polygons (Figure 2.4). The SVMC step is a supervised classification based on the set of training polygons identified and classified into one of the above four classes by the user (Figure 2.4). The SVMC looks for the hyper-plane[4] with the highest degree of separability between the sets of classes of data points, and then those image pixels not comprising the training set are assigned to a class based on the side of the hyper-plane on which they fall [37-40].

[4] The hyperplane is a n-dimensional plane, where the dimensions are considered here as the number of classes e.g., a 4-D hyperplane, that best separates the classes being considered.   In a two-dimension plane Cartesian plane, the hyperplane would be the line *y=mx+b* that best separates two sets of points in two-dimensional space. In dimensions larger than 3, we have hyper-dimensions, hence need hyperplanes for separability.

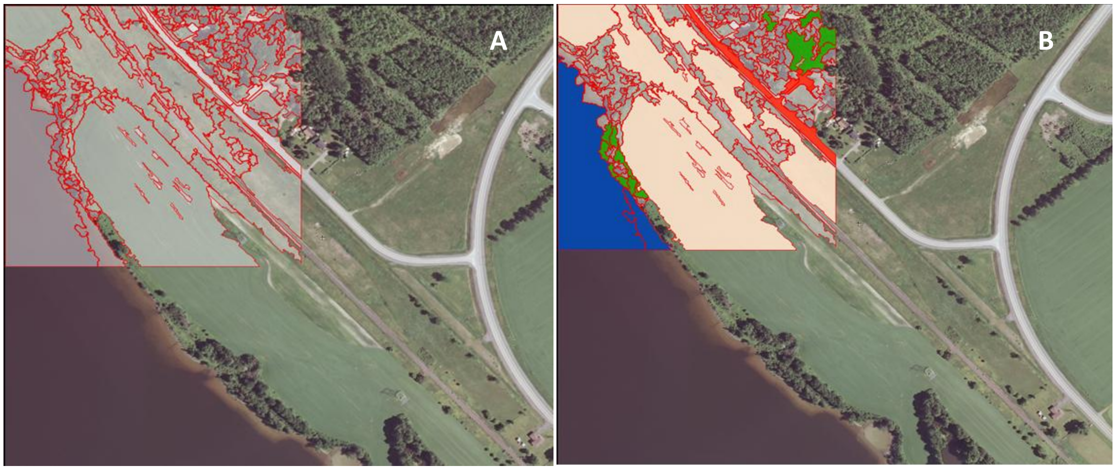


Figure 2.4: a) Polygons generated by Incogna GIS's object oriented segmentation. b) Manually classified training sites for inclusion in the SVM calcification.

Incogna's SVMC software is also unique in that every time a set of training polygons is identified, the SVM algorithm is updated and an intermediate classification is displayed over the aerial images. These intermediate classifications allow the user to visually assess the accuracy of the SVMC and decide whether additional training polygons are required to obtain an accurate classification. This feature was very useful as a preliminary gauge of the accuracy of the classification. Once a satisfactory classification was obtained, the four classes were extracted (Figure 2.5). The processing of the 3600 multispectral DRAPE images took approximately 2.5 weeks with Incogna's servers allocated to the processing of the SVM classification.

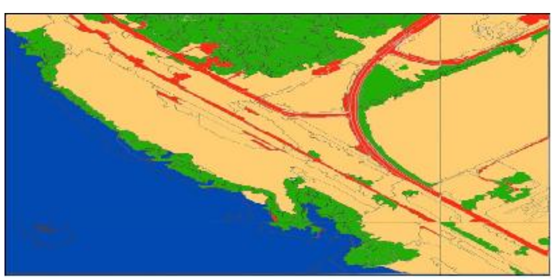


Figure 2.5: Final classification on DRAPE aerial photography defined through SVMC.

The land-cover dataset was then processed using ESRI ArcGIS 10, which involved converting SVMC results into four class integer raster format. Spatial filters were applied to remove noise. The application of spatial filters involved the use of a Euclidean allocation to fill in all of the "No Data"[5] cells with nearest neighbouring classes. The "No Data" cells appeared as a result of the conversion of the vector objects from the SVM algorithm into a raster format.

Next, a rectangular neighbourhood majority filter was run over the data to correct for small groups of misclassified cells. The rectangular filter is a 5x5 pixel filter that is laid over every pixel in the image. The filter assigns a value to that pixel based on the class that occupies the majority of the pixels in its area. Properly classified features contained several hundred pixels and were not affected by the rectangular filter, which is used to remove single or small groups of misclassified pixels.

The accuracy of the classification at a small scale was determined by comparing the classified dataset to ground truth by looking for errors of omission and commission. This validation was done by manually identifying areas on the DRAPE aerial images as one of the four classes being extracted. The areas that were manually classified were not areas that had been used as training sites during the SVMC; this was done to ensure that the accuracy of the final classification model was validated. The sites that were chosen to validate the SVMC classification were clear on the image as to their land-cover class identity. Five sample sites were chosen across the city, in both rural and urban areas. The overall accuracy for the urban areas at these sites was 95%.

For the purposes of the dasymetric methodology, however, only the built up class was required; as all other land cover classes were assumed to have no population. Thus, the built up areas were extracted from the land cover classification and overlaid back onto the DRAPE aerial imagery to see if classified built up areas coincided with visible urban areas on the imagery (Appendix C).

[5] A very small number of pixels in some images were not classified as any of the 4 classes because of the soft-margin error tolerance within the SVM classifier, e.g., a few pixels were too different from their classes to be confidently classified.

The overall accuracy of the SVMC in identifying built up areas was very accurate (Figure 2.6). The classification was not 100 percent accurate at a very fine spatial scale; however, it provided a good dataset to identify where people resided within a larger geographic area for dasymetric mapping purposes. From comparing the built up classification to the DRAPE aerial imagery, it was determined that the build up class from the SVMC provided an acceptable dataset for dasymetric mapping as it focused population representation to where people live in geographic space.

The dasymetric mapping process begins after the feature extraction is completed. The built-up class is extracted from the SVM post-processed Ottawa landcover map and serves as the definition of inhabitable areas. Populations in each census dissemination area (DA) are then redistributed to the built-up areas from the SVM landcover. The result is an example of a simple binary dasymetric map (Figure 2.6).

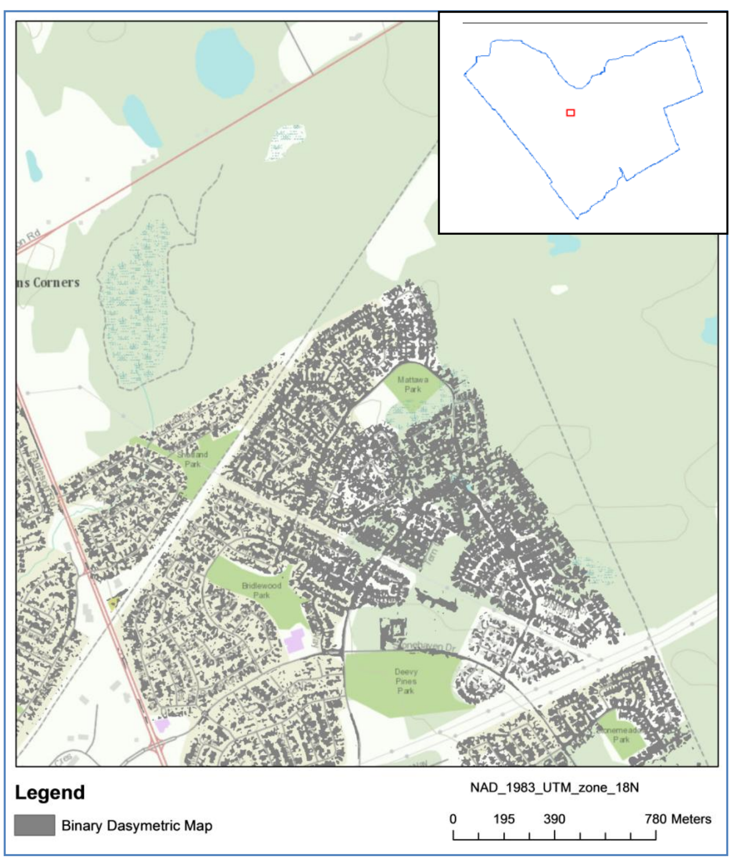


Figure 2.6: Section of binary dasymetric map in west Ottawa.

To estimate population densities, Ottawa municipal zoning (provided by the City of Ottawa) was then used to further disaggregate built-up population. The Ottawa zoning map contains 39 different classes, but only the classes that could be inhabited by people are used (Appendix D). These are identified using the definitions outlined in the *City of Ottawa Zoning By-law 2008-250*, which identifies acceptable uses for every zoning class. The acceptable zoning data is converted to a raster file and intersected with the post-processed SVMC built-up class output. Using the dasymetric method developed by Mennis (2003) (Appendix E), population density fractions, area ratios, and total fractions are calculated for each of the different zoning classes. These calculations allow for the proper assignment of populations to each land use class from an identified dissemination area (DA). Using DA-level populations from the Statistics Canada 2006 census, population count is assigned to each individual pixel on the map.

With the population density calculations completed for the City of Ottawa zoning class, a map of the City of Ottawa's population distribution can now be displayed. This map contains population for each pixel on the map, which allows for the estimation of population counts in varying units of analysis (Figure 2.7). From this map population counts were calculated for each one of the community defined rural neighbourhoods and pre-existing ONS neighbourhoods (Appendix F).

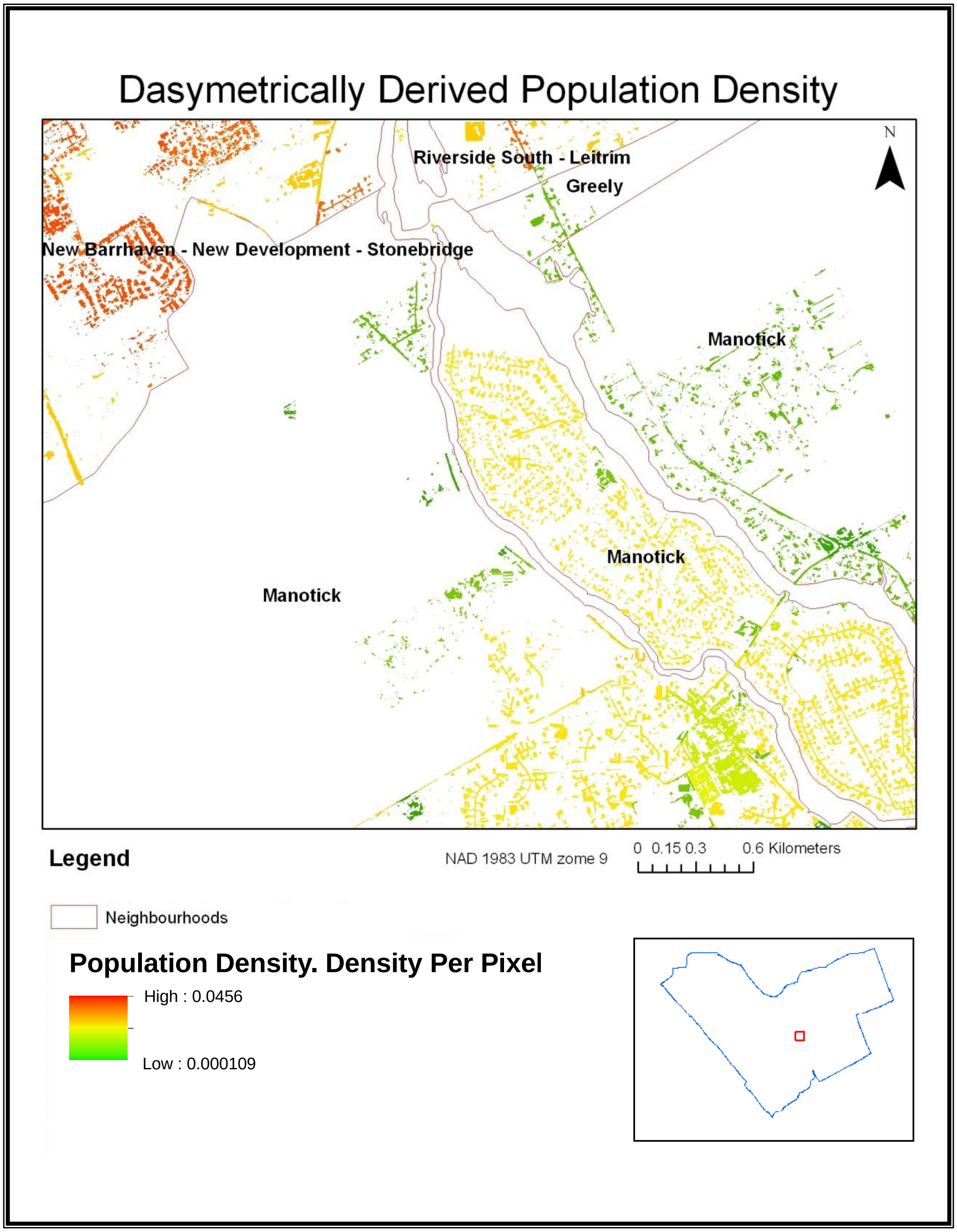


Figure 2.7: Dasymetrically derived population surface centered over the village of Manotick.

Dasymetric Validation

Prior to using the population to re-weight the CCHS SES census correlates, the accuracy of the dasymetric model is validated by calculating populations within units of analysis for which the population is already known. For our research, boundary files for the City of Ottawa and Statistics Canada census tracts (CTs) were overlaid on the dasymetric map. Using zonal statistics, dasymetrically derived populations were estimated for these areas by summing the pixels within each CT.

The estimated population values for the City of Ottawa and Statistics Canada census tracts were then compared to population counts from the 2006 census, calculating the level of agreement between the two datasets using R^2. A high level of agreement between the two datasets means that the dasymetric population surface is accurate and thus estimates of rural neighbourhood population in Ottawa will follow suit.

The dasymetrically derived population map produced accurate results. Figure 2.7 depicts the differences in densities within the rural region of Ottawa, with a medium density shown over the village of Manotick, low population densities in the sparsely populated adjacent rural areas, and a particularly high population density to the North West in the suburb of New Barrhaven-Stonbridge. For the 185 census tracts across the city, the estimated dasymetric populations had a 0.97 R^2 compared to the actual population counts (Figure 2.8). The statistical significance of the correlation was tested by calculating a two tailed *p* value for Pearson's product-moment correlation. The *p* value was equal to 2.2x10^-6, meaning the correlation is highly significant.

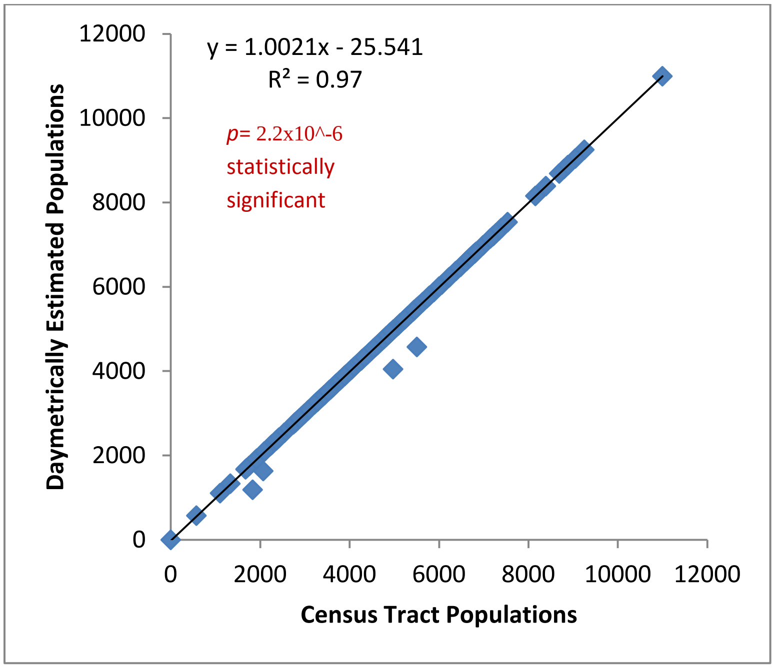


Figure 2.8: Dasymetric model population estimations within Statistics Canada census tracts compared to census tract populations recorded by statistics Canada in the 2006 census

Dasymetrically Re-weighted SES Variables

With rural neighbourhood boundaries finalized and the dasymetric population's surface validated, the process of re-weighting SES CCHS correlates into the new boundaries was undertaken. The resultant new neighbourhood units have census variables dasymetrically re-weighted into them to provide the aggregate common census variables for microsimulation. The accuracy dasymetric re-weighting of the census derived SES CCHS correlates was validated by estimating SES counts for geographic units for which the data is already known. The accuracy of using the dasymetric population density map to re-weight SES variables is tested by using the dasymetric map to estimate SES variables within all 185 census tract (CT) enumeration units. Next, all of the 185 values estimated through dasymetric re-weighting to the CT areas are compared to the actual values for the CTs obtained by Statistics Canada in the 2006 census. The level of agreement between the dasymetrically re-weighted variables and the actual CT values are then investigated by calculating their R^2 value. The relationships were again tested to see if they were statistically significant using a two tailed *p* value for Pearson's product-moment correlation. The *p* value for each of the correlations was found to be very statistically significant. If the level of the agreement is high, then the assumption will be made that the values estimated in the new neighbourhood units through the dasymetric re-weighting will provide a good starting point for the statistical microsimulation (Figures 2.9, 2.10, 2.11, 2.12, and 2.13).

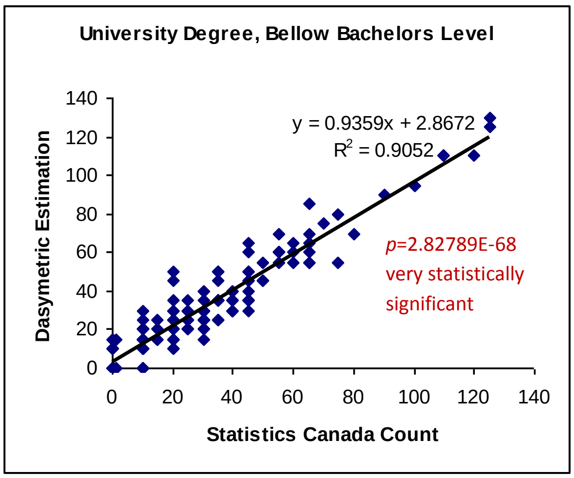


Figure 2.9: Dasymetrically re-weighted estimation of the number of people with university degrees at the Bachelor level in Statistics Canada census tracts compared to Statistics Canada counts of the number of people with university degrees below the Bachelor level at the census tract level

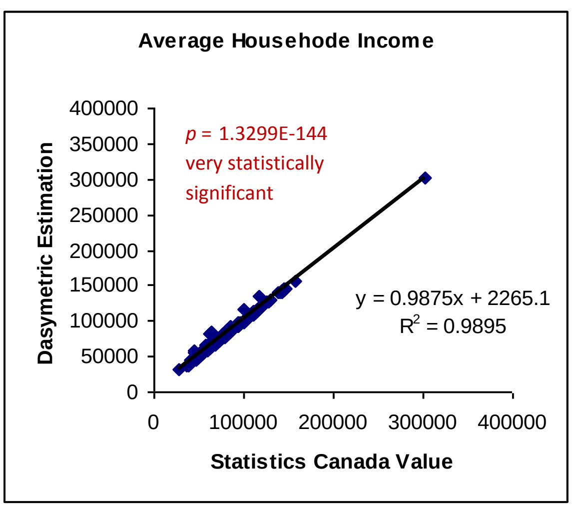


Figure 2.10: Dasymetrically re-weighted estimation of the average household income in Statistics Canada census tracts compared to Statistics Canada values of the average household income at the census tract level

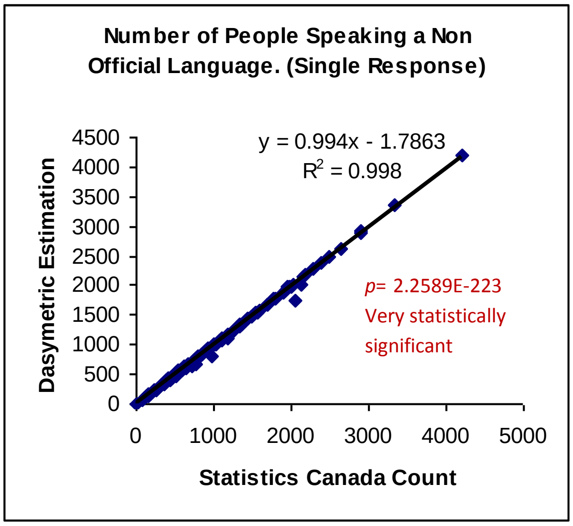


Figure 2.11: Dasymetrically re-weighted estimation of the number of people speaking a non-official language as their first language in Statistics Canada census tracts compared to Statistics Canada counts of the number of people speaking a non-official language as their first language at the census tract level

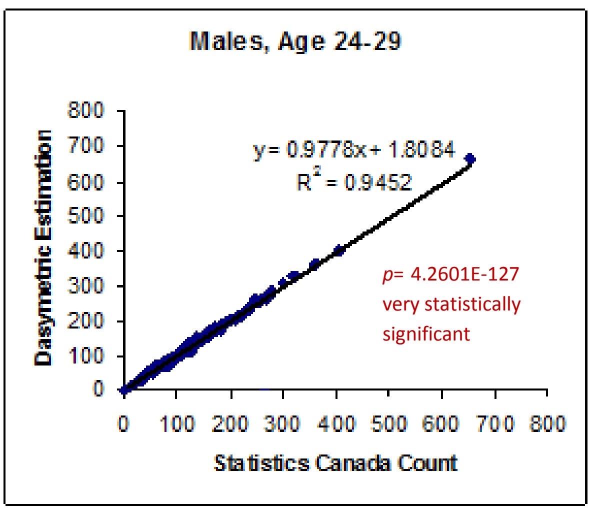


Figure 2.12: Dasymetrically re-weighted estimation of the number of males between the ages of 24-29 in Statistic Canada census tracts compared to Statistics Canada counts of the number of males between the ages of 24-29 at the census tract level

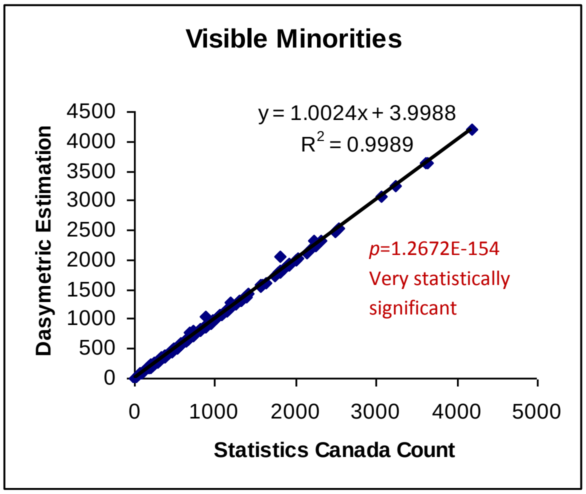


Figure 2.13: Dasymetrically re-weighted estimation of the number of people who identify as being visible minorities in Statistics Canada census tracts compared to Statistics Canada counts of the number of people who identify as being visible minorities at the census tract level

The dasymetric re-weighting of CCHS SES correlates from the Census produces estimations comparable to the data within the CT's gathered by Statistics Canada. For all of the variables tested, each had a correlation which had a statistical significance well with an acceptable confidence interval.

Spatial Microsimulation

Spatial microsimulation is a down-scaling technique that involves simulating individual- or household-level characteristics (such as SES and health) within small areas [41]. Conceptually, a small area is a geographic polygon that is too small to contain a sample size that supports statistical estimates of health variables from regional or national surveys. In the context of this research, small areas refer to ONS rural neighbourhoods in Ottawa (irrespective of actual physical area) that lack sufficient population to support estimation of health variables (most often due to a paucity of sampling) from health surveys (such as the Canadian Community Health Survey or CCHS).

As such, the simulation of health variables in Rural Ottawa is a statistical down-scaling issue that can be addressed through synthetic estimation. The down-scaling of variables via synthetic estimation is executed through linking common variables from the spatial scale of interest with detailed, large-area anonymous survey data such as the CCHS [42-44]. Spatial microsimulation allows us to map and analyze health variables within small areas in the absence of spatially explicit epidemiological or social surveys. There are a number of underlying assumptions regarding spatial microsimulation. The first assumption regards the relationship between health variables and the census-based socioeconomic variables. It is assumed therein that the available compositional variables common to the census and coarse scale surveys are adequate predictors of the required or chosen health variable. Microsimulation is largely based on matching the compositional variables within coarse resolution health surveys and census data with small-area aggregate census variables in order to create a set of weights for synthetic population estimation whose post-modeled aggregate variables match the original area-based aggregate census variables as closely as possible. The fundamental assumption is that patterns (within the socioeconomic dimension) in health determinants and outcomes at the Health Region level in the CCHS will be reflected in rural Ottawa neighbourhood units according to variations in their socioeconomic variables. The variables used in the weighting process are largely predetermined by the characteristics of the population collected in the health survey.

The Microsimulation Model

The prevalence of different health determinants and outcomes was simulated in two stages. The first stage consisted of re-weighting the original sampling weights of the CCHS in order to combine the neighbourhood-level information drawn from the census with that of the survey. The second step in the simulation process consisted of using deterministic regression equations to predict the small-area prevalence of selected health determinants and outcomes. This was done using a limited array of distributional assumptions and parameters. The first step in the microsimulation model and the notion of cross-entropy is hard to vulgarize. A summary of the steps taken in the microsimulation process are outlined bellow.

First Step: Reweighting

The calibration of survey weights is the best way to ensure that the marginal totals of different tabulations correspond to the actual population totals for different geographic units. In our research, we used socio-demographic information that had been taken from the 2006 census and dasymetrically reweighted to spatially fit within the boundaries of every neighbourhood for the calibration of weights. The following variables were taken into consideration: age, sex, education, mother tongue, and income.

Weights were calibrated using the minimum cross-entropy framework that builds and expands on the maximum entropy framework based on the well-known formula (see [45]):

$\max H(p) = -\sum_{i=1}^{n} p_i \ln(p_i)$, where the sum of all $p_i = 1$ (Equation: 2.1)

Assuming the existence of a prior probability distribution q, cross-entropy is defined as

$$I(p,q) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \qquad \text{(Equation: 2.2)}$$

and can either be minimized by a numerical solution of the first-order conditions of the Lagrangian function associated with the data matrix, or by maximization of the unconstrained dual cross-entropy function [46]:

$L(\lambda) = \sum_{j=1}^{j} \lambda_j y_j - \ln\langle \Omega(\lambda) \rangle = M(\lambda)$, where y_j is the population mean of the x_j random variable

and $\Omega(\lambda) = \sum_{i=1}^{n} q_i \exp(x_i \lambda)$. (Equation 2.3)

This function behaves like a maximum likelihood:

$$\nabla_{\lambda} M(\lambda) = y - X^{'} p \qquad \text{(Equation 2.4)}$$

The calibrated weight, therefore, can be estimated at the point where the gradient is set equal to zero. In addition, a noteworthy feature of this model is that the variances and co-variances corresponding to the negative of the Hessian of **M** are taken directly from **p**, hence guaranteeing positive definiteness and the existence of a unique solution:

$$-\frac{\partial^2 M}{\partial \lambda_j^2} = \sum_{i=1}^{n} p_i x_{ji}^2 - \left(\sum_{i=1}^{n} p_i x_{ji}^2 \right)^2 = \mathrm{var}(x_j)$$

$$-\frac{\partial^2 M}{\partial \lambda_j \lambda_k} = \sum_{i=1}^{n} p_i x_{ji} x_{ki} - \left(\sum_{i=1}^{n} p_i x_{ji} \right) * \left(\sum_{i=1}^{n} p_i x_{ki} \right) = \mathrm{cov}(x_j, x_k) \qquad \text{(Equation 2.5)}$$

For instance, let's assume that we have only three cases whose CCHS survey weights are equal to 500, 800 and 1,200 in a neighbourhood whose total population size is 2,500. As it stands, the percent values corresponding to each case are the following: 20%, 32% and 48%. Let's further assume that an examination of age and sex tables from the census has led us to conclude that the best percent values (q_i) should be: 30%, 30% and 40%. The cross-entropy value *I(p,q)* can be estimated at 0.0275 using the above equation.

The calibration process now has to ensure that this entropy is minimized with respect to the neighbourhood means for different characteristics measured from the census. Therefore, assuming that the only characteristic that we want to adjust for is income and that the average income of the neighbourhood is \$33,500, while the reported incomes of the three cases are \$50,000, \$25,000 and \$70,000, we need to find a way to make *L(λ) = M(λ)* = 33,500.

From the formula above, we can calculate that $\Omega(\lambda)$=50,500 and ln $\Omega(\lambda)$=10.83, while ln (33,500)=10.42. We therefore need a combination of weights that will recalibrate the three cases so that the natural logarithm of their weighted average becomes 10.42. The likelihood function achieves it by assigning a frequency of 0.25 to the first case, 0.70 to the second, and 0.05 to the third one. The calibrated weights have now become 500, 1400 and 100 and add up to the total population size of the neighbourhood that is equal to 2,500. The weighted average income for the neighbourhood also retains its original value of $33,500 (0.25X50,000 + 0.70X25,000 + 0.05X70,000 = 33,500).

Second Step: Predicting

Deterministic regression models are used to predict the prevalence of the following health variables: selfrated health, selfrated mental health, binge drinking, chronic conditions, smoking, and obesity. Prevalence is predicted using different link functions based on the logistic distribution and the one-parameter exponential family of distributions. For every model, a total of 500 Monte Carlo simulations are performed. Within these simulations, coefficients are not fixed and can vary within the limits of their 95% confidence interval, as it has been previously estimated with the pooled CCHS database.

Simulation error is reported for every model, and the best-fitting one is selected for prediction purposes. This process can be better illustrated with an example taken from the database. Let’s suppose that we are interested in estimating the prevalence of overweight/obesity in the rural neighborhood of Greely. In Greely, the dasymetrically estimated population for people aged 12 and older is 7,332. The following covariates were used in the regression models: sex, age and age squared, marital status, highest level of educational attainment, and physical activity index (active, moderate, and passive). The goodness-of-fit indicators associated with the different predictive models are as follows (Table 2.2):

Model		
Logistic	3119.6	3146.7
Poisson	4783.6	4889.5
Negative binomial	5093.1	5123.7
Zero-inflated Poisson	4889.0	4926.8
Zero-inflated negative binomial	5323.6	5421.9
Hurdle (Poisson-lognormal)	5147.2	5208.6
Hurdle (negative binomial-logistic)	5023.4	5166.4

Table 2.1: Goodness-of-fit indicators associated with the different predictive models for obesity in Greely.

The selected predictive model is therefore the logit and the probability threshold for being included in the overweight/obese category is 0.5. This yields a final count of 1261 overweight people, which represent 17.2% of the population of the neighborhood aged 12 and older.

Mapping and Analysis

The spatial variability of the SES variables and health outcomes from the microsimulation were mapped. Maps were produced for each health outcome for Ottawa by rural neighbourhood in order to explore the spatial variability (Figures 2.14, 2.15, 2.16, 2.17, 2.18, 2.19 and 2.20). In this step, exploratory spatial data analysis and spatial autocorrelation statistics were derived to assess the degree of spatial dependence in each mapped variable. Moran's I was then calculated for all of the estimated health outcomes, based on 40000 randomizations for the global statistics and 40000 conditional randomizations for the local statistics (Figures 2.21, 2.22). Each variable was tested for using the rate standardized values. Moran's I calculates spatial autocorrelation through assessing a variables locations and values simultaneously to determine whether values of a health variable that are at adjacent neighbourhoods are more similar than those within spatially disjunct neighbourhoods. Given a set of features and an associated attribute, it evaluates whether a mapped variable exhibits self-similarity vs. one that is randomly distributed across the ONS neighbourhoods [47-50].

Bivariate local Moran's I was used to investigate relationships between each of the different estimated health variable prevalences. Bivariate spatial statistics such as bivariate local Moran's I which is used here are concerned with investigating the presences of a 'spatial association in terms of point to point relationships across two spatial patterns' [51] (p.369). Both local Moran's I and bivariate local Moran's I are used in this study to explore the results of the simulated health variable prevalences and to demonstrate the type of insight that researchers can gain from the results of the methodology outlined in this study.

Results

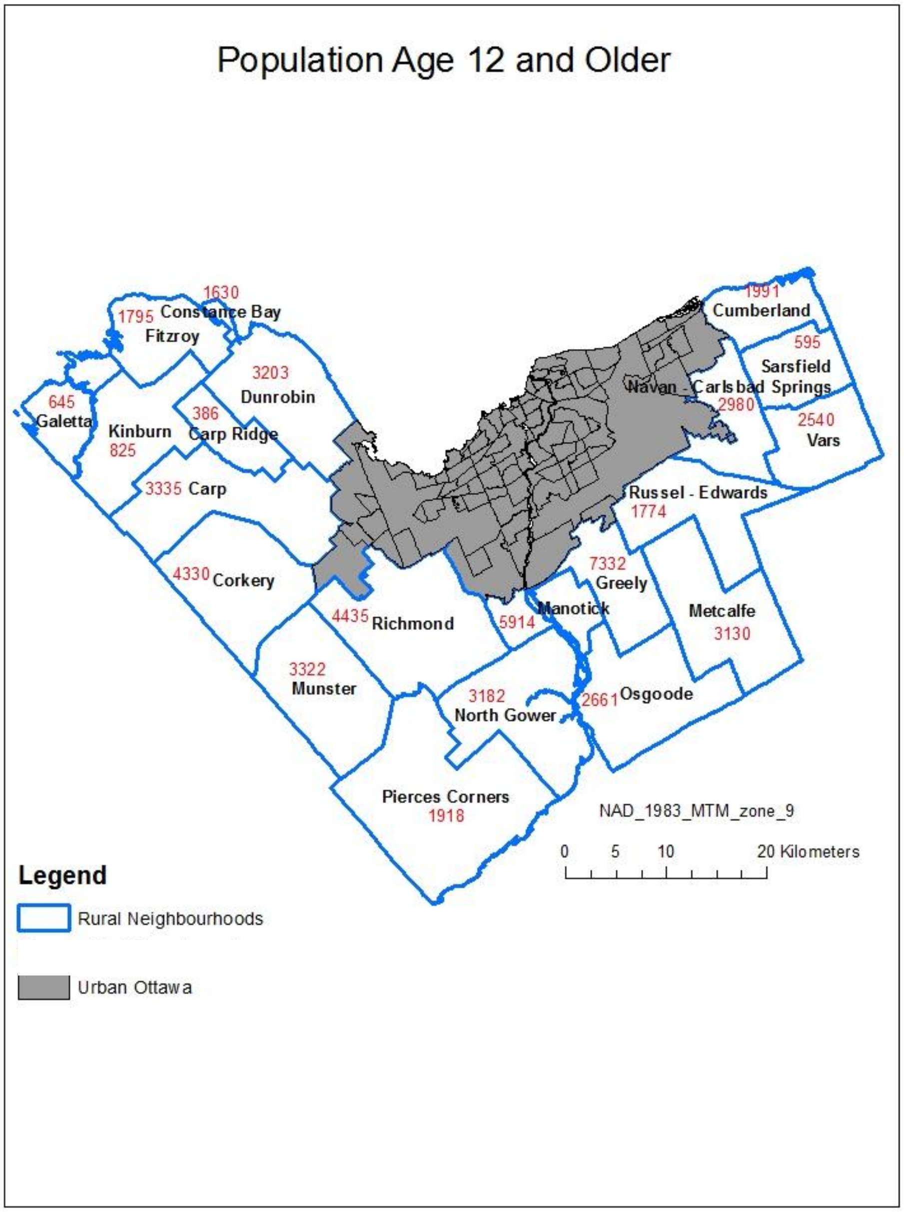


Figure 2.14: Population 12 and older within rural Ottawa Neighbrouhoods

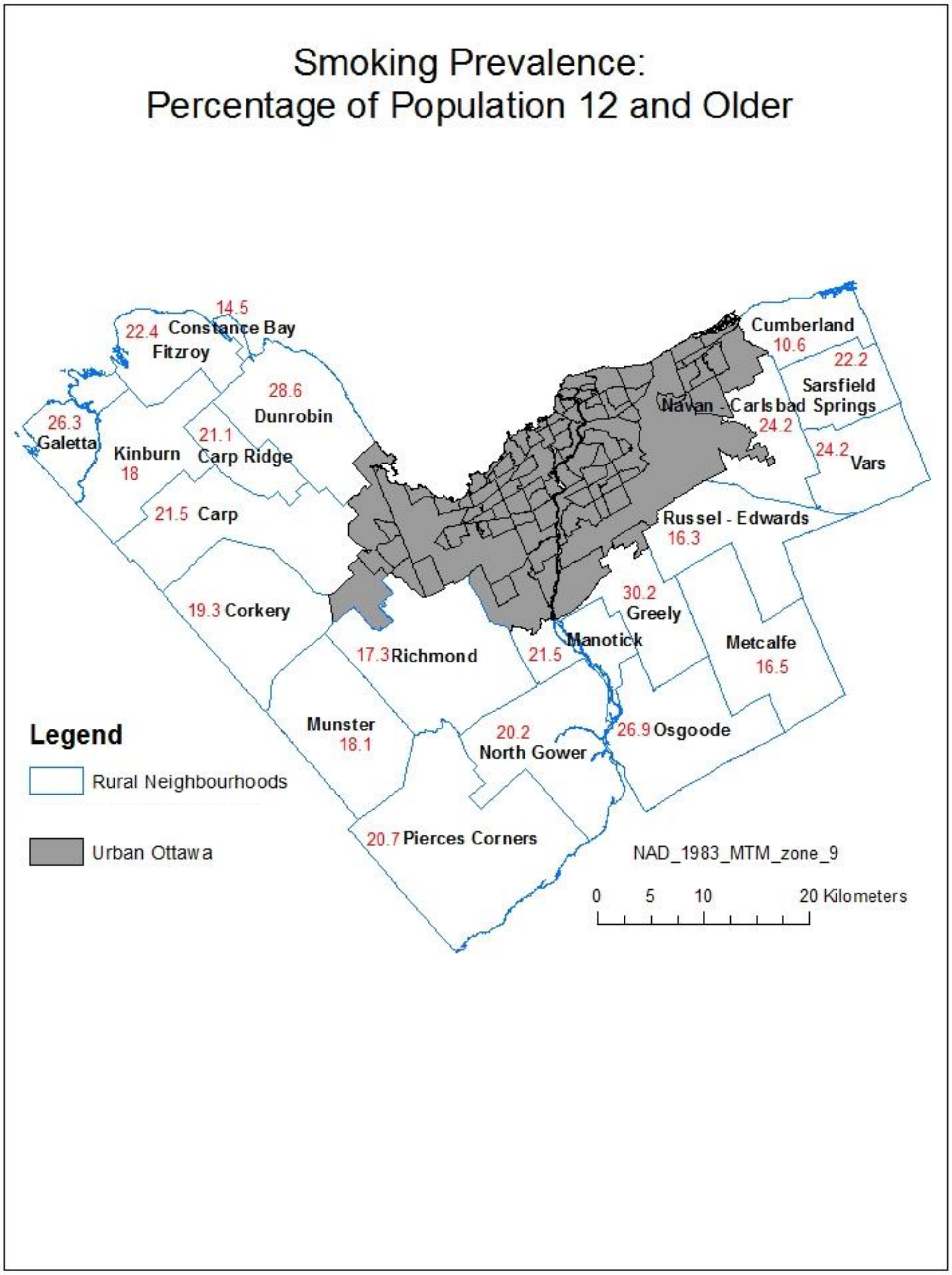


Figure 2.15: Estimated prevalence of smoking among the populations of rural Ottawa neighbourhoods

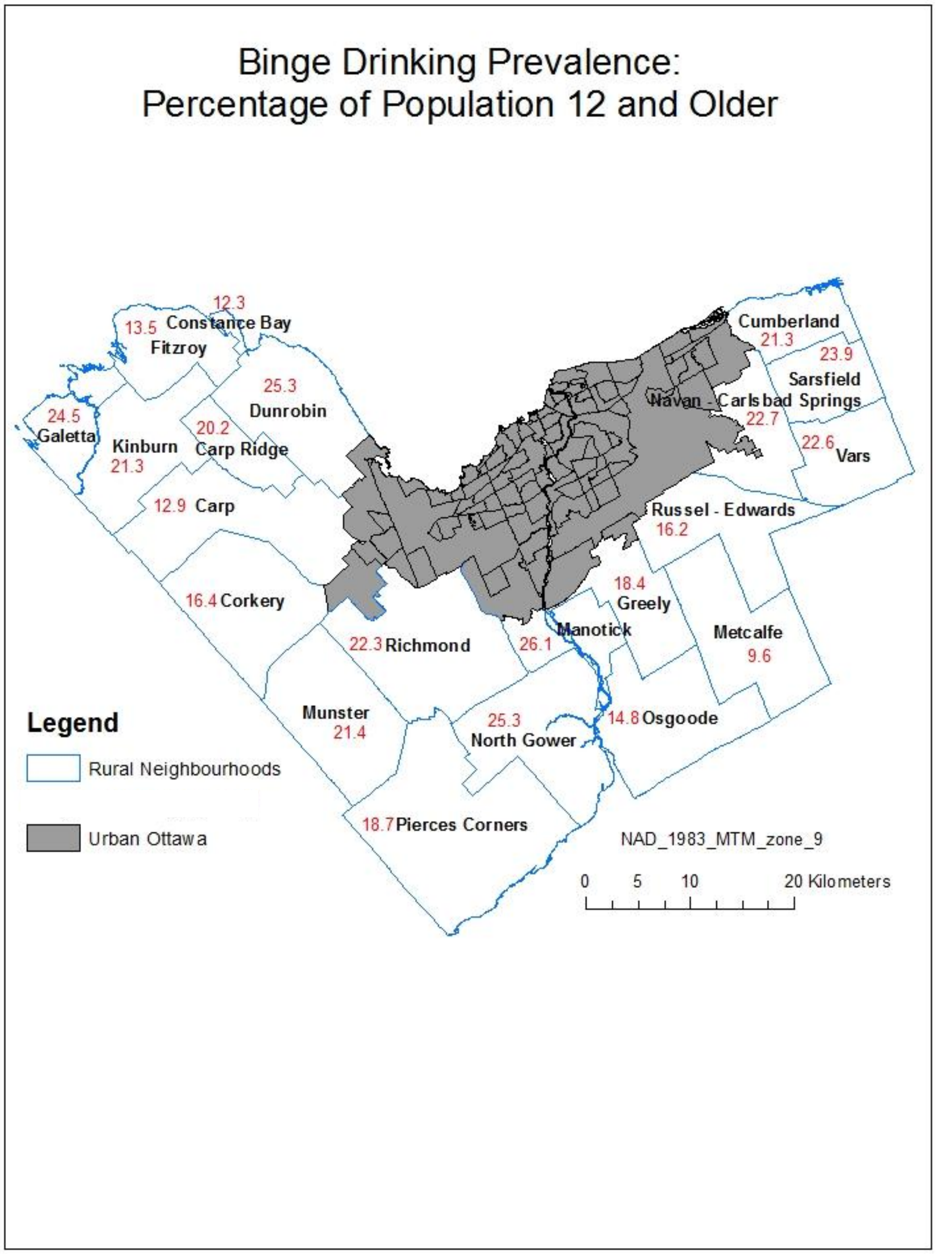


Figure 2.16: Estimated prevalence of binge drinking among the populations of rural Ottawa neighbourhoods

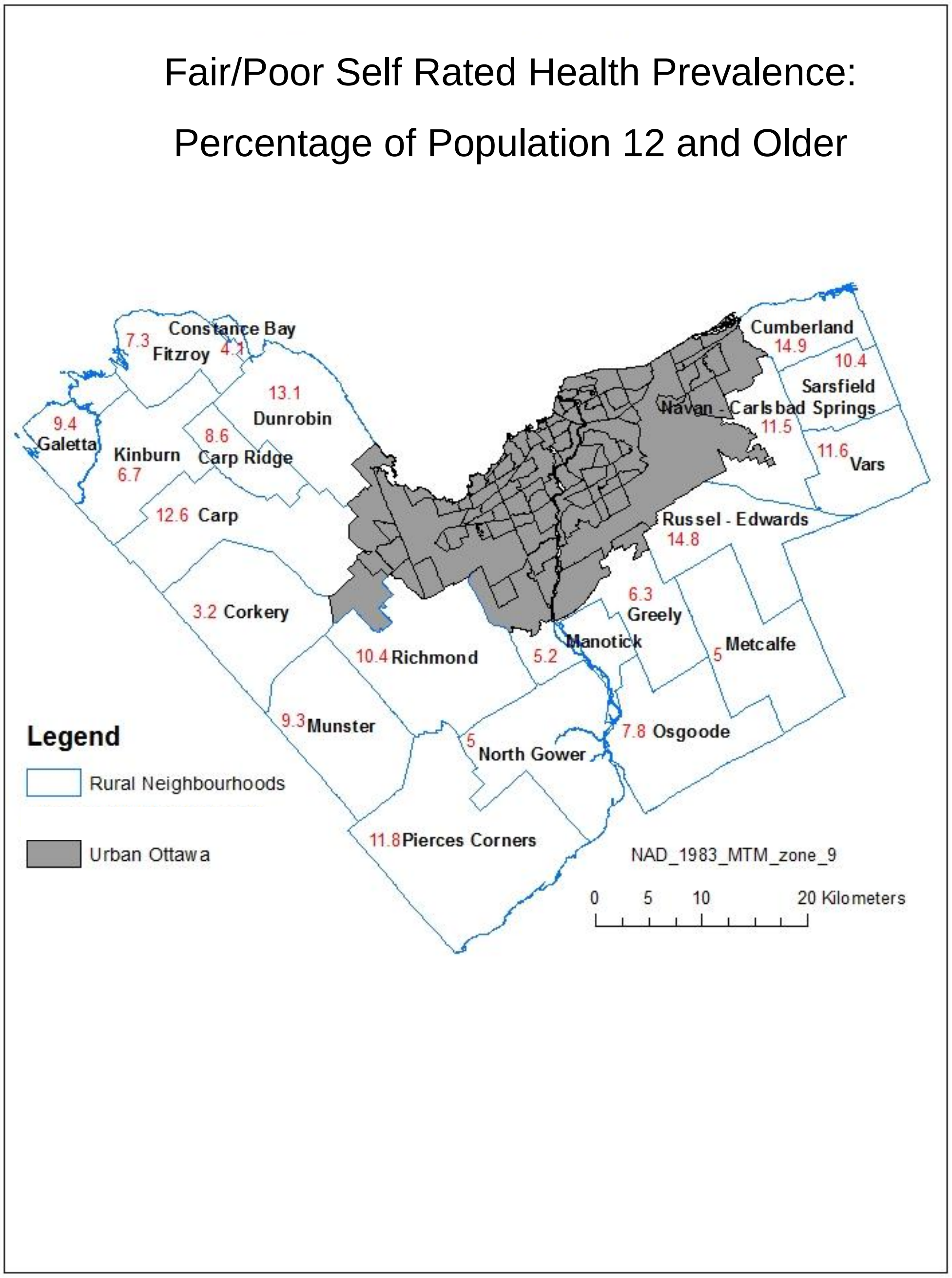


Figure 2.17: Estimated prevalence of fair and poor selfrated health among the populations of rural Ottawa neighbourhoods

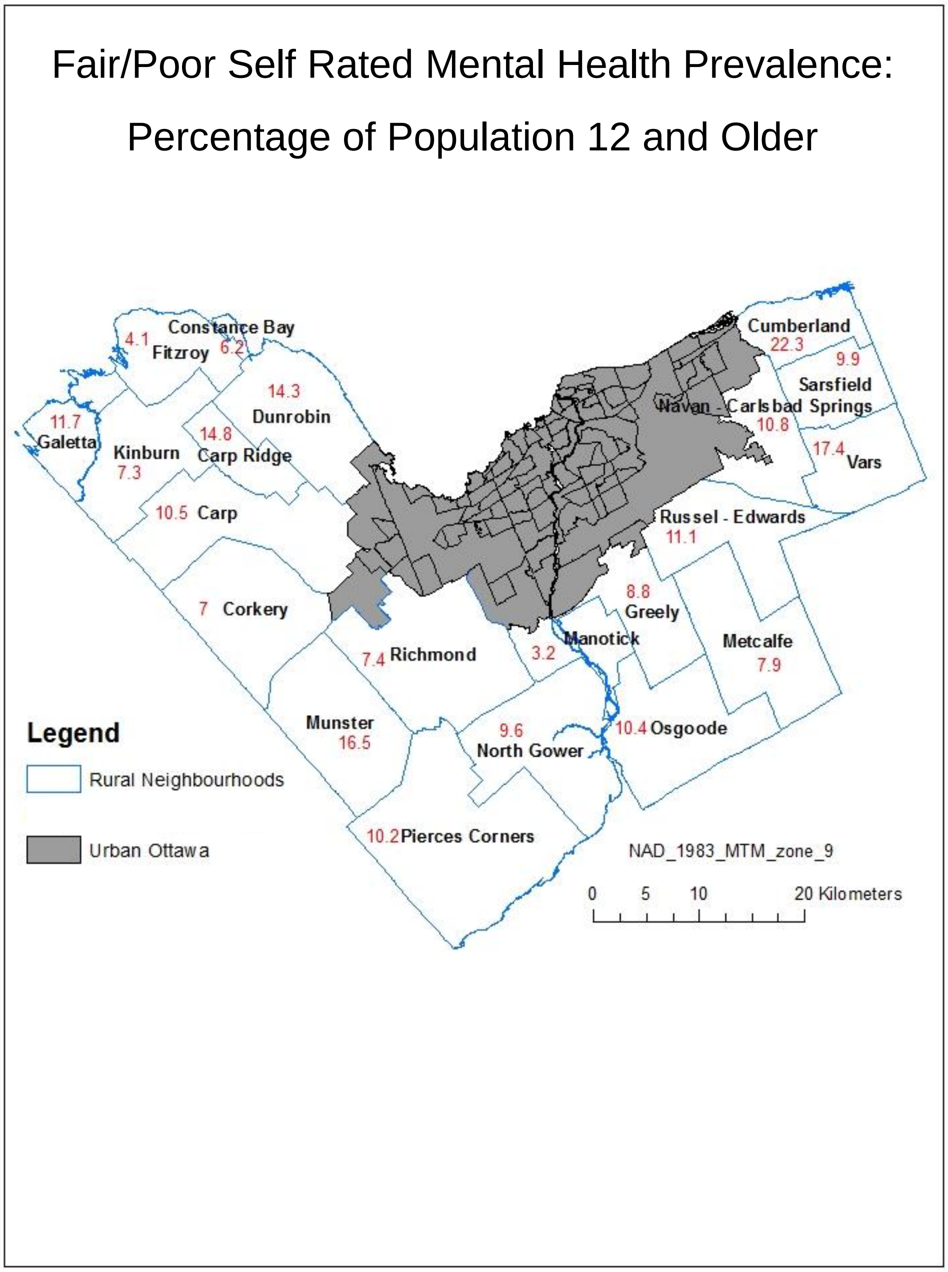


Figure 2.18: Estimated prevalence of fair and poor selfrated mental health among the populations of rural Ottawa neighbourhoods

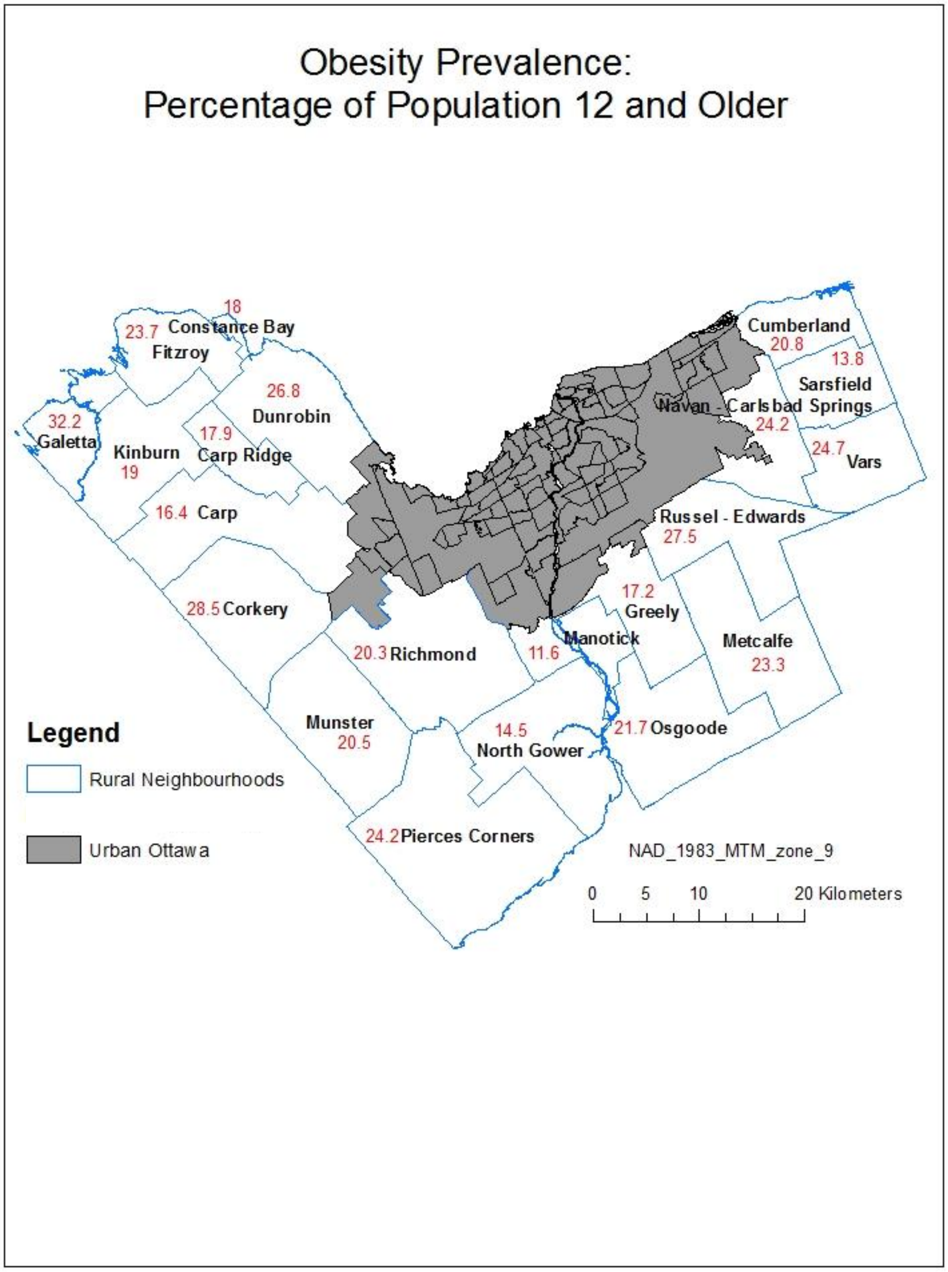


Figure 2.19: Estimated prevalence of obesity among the populations of rural Ottawa neighbourhoods

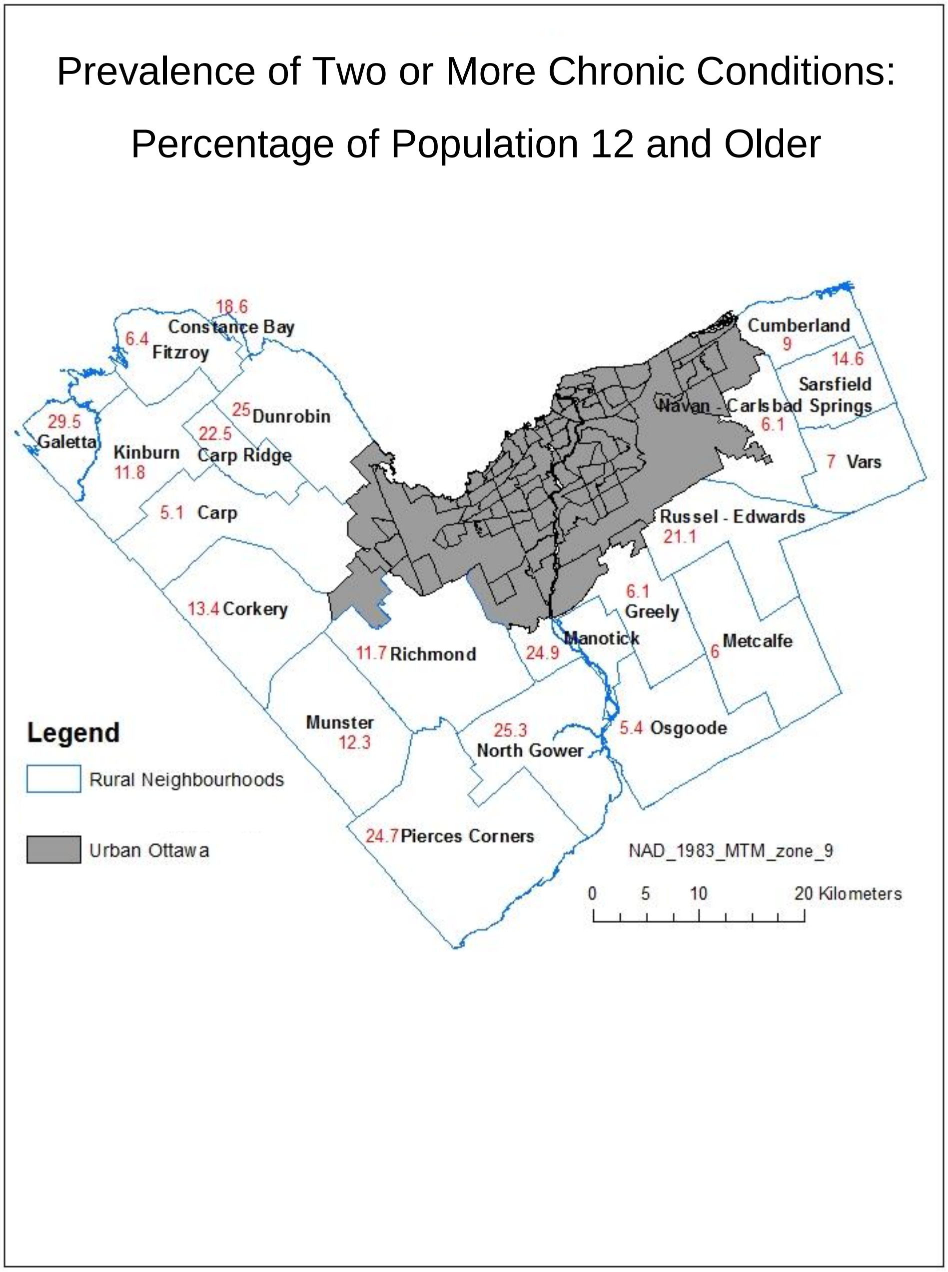


Figure 2.20: Estimated prevalence of two or more chronic conditions among the populations of rural Ottawa neighbourhoods

The more densely populated rural neighbourhoods are those on the periphery of Ottawa's urban area and contain rural villages which are well developed (Figure 2.14). The neighbourhoods of Richmond, Manotic and Greely demonstrate this pattern. The neighbourhood with the lowest prevalence of smoking is Cumberland at 10.6 percent of its population 12 and older and the highest is Greely with 30.2 of its population 12 and older (Figure 2.15). The rest of the rural neighbourhoods have prevalences ranging from 14.5- 28 (Figure 2.15). The lowest prevalence of binge drinking among the population 12 and older is Constance bay at 12.3 percent, and the highest is North Gower at 25.3 (Figure 2.16). The prevalence of fair/poor self rated health is low across the region with the highest prevalence in Cumberland at 14.9 percent of the population 12 and older (Figure 2.17). The prevalence of fair and poor self rated mental health is also low across the region, with the exception of the neighbourhood of Cumberland which has a prevalence of 22.3 percent among its population 12 and older (Figure 2.18). The neighbourhood with the lowest prevalence of obesity is Manotic at 11.6 and the highest is Galletta at 32.2 (Figure 2.19). The prevalence of two or more chronic conditions varies significantly across the rural neighbourhoods, with the lowest prevalence at 5.1 percent in Carp and the highest at 29.5 in Galetta (Figure 2.20).

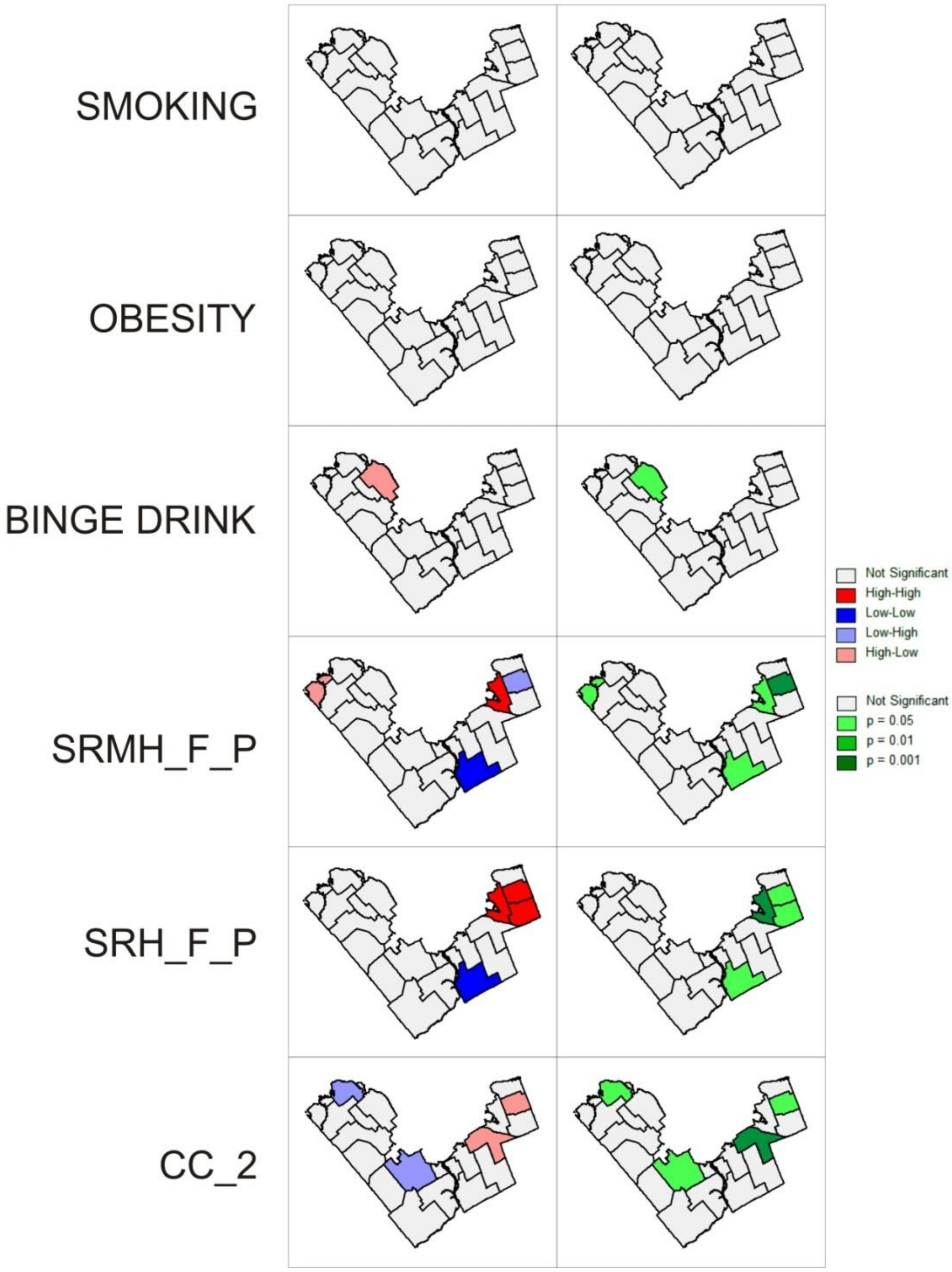


Figure 2.21: Local Moran's I results. SRMH_F_P refers to fair or poor selfrated mental health, SRH_F_P refers to fair or poor selfrated health and CC_2 refers to 2 or more chronic conditions.

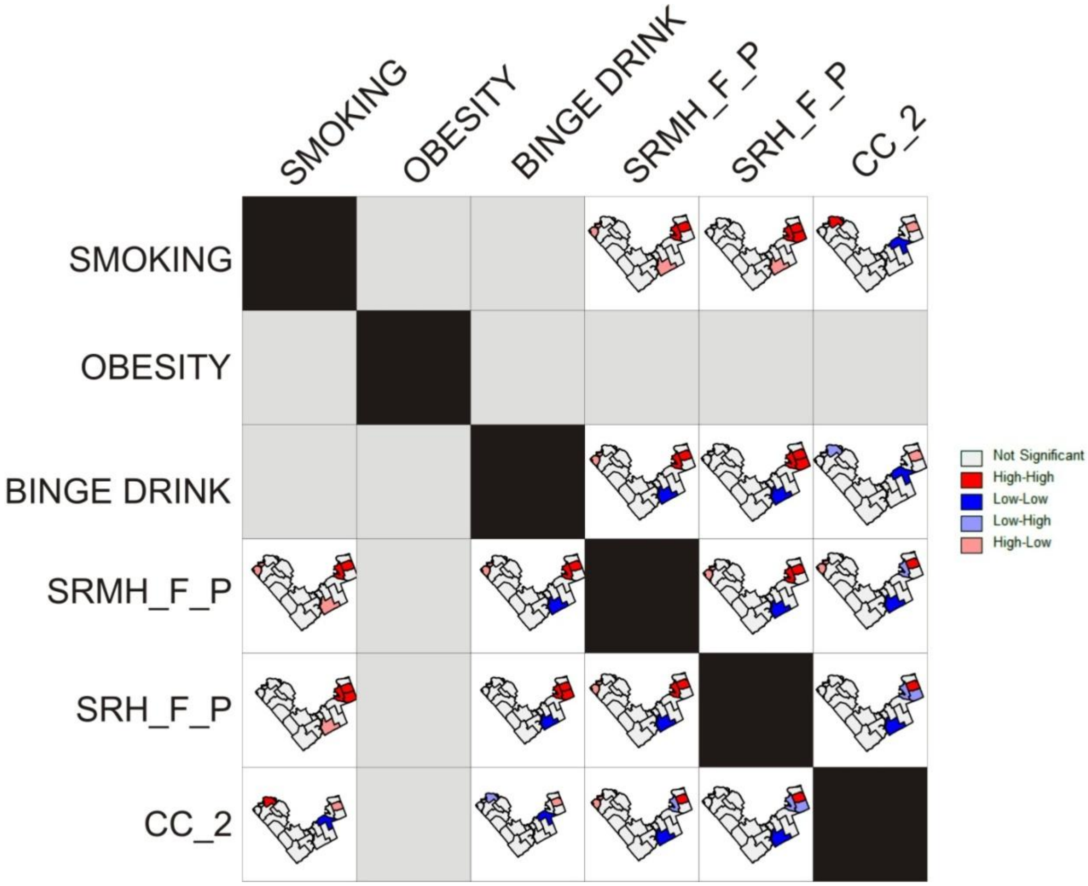


Figure 2.22: Bivariate Local Moran's I results. SRMH_F_P refers to fair or poor selfrated mental health, SRH_F_P refers to fair or poor selfrated health, and CC_2 refers to 2 or more chronic conditions.

No significant uni-variate or bivariate spatial autocorrelation was observed for any of the variables. Given that no global spatial dependence was evident, local areas of non-stationary were looked for appropriately using local Moran's I to identify whether any particular neighbourhoods exhibited significant local spatial autocorrelation (Figure 2.21).

The categories of obesity and smoking showed no significant local spatial autocorrelation (Figure 2.21). Only one neighbourhood, Dunrobin, exhibited a significant negative local spatial autocorrelation as a high-spot for binge drinking, with lower values found in adjacent neighbourhoods (Figure 2.21). Selfrated mental health exhibited more complex clusters. For example, Galetta has unusually high values of fair/poor selfrated mental health compared to surrounding neighbourhoods; Osgoode is a significantly low spot for both selfrated mental health and selfrated health; Sarsfield is low for selfrated mental health, but high for selfrated health compared with its adjacent neighbors; and Navan-Carlsbad Springs is high for both health ratings (Figure 2.21).

The results illustrate that Richmond and Fitzroy are two significant low-spots for having two or more chronic conditions, while Russell-Edwards and Sarsfield are shown to be high-spots in this category. These findings are further reflected in the bivariate local Moran's I results (Figure 2.22). Binge drinking and smoking in Sarsfield and Navan-Carlsbad Springs are significantly associated with both selfrated mental health and selfrated health, and they exhibit positive spatial autocorrelation due to high bivariate spatial autocorrelation. This is also true of chronic conditions and selfrated physical and mental health. The bivariate spatial autocorrelation for smoking and the selfrated health variables are significantly high in Osgoode; the opposite is true for binge drinking with those variables (Figure 2.22).

Discussion

Through the combination of dasymetric mapping and spatial microsimulation, this study provided insights into how health determinants and outcomes can be studied in rural areas. Each step in the process contributed to the overall objective of the study, allowing simulation of health variables in low-population under sampled rural neighbourhoods.

The dasymetric mapping methodology undertaken provided an innovative addition to the spatial microsimulation methodology by yielding a high resolution population distribution in rural Ottawa. The methodology, however, was very time consuming, due in large part to the use of SVMC feature extraction technology to provide high resolution ancillary data. Furthermore, the use of the SVMC feature extraction also required a background in remote sensing; this is an additional level of difficulty that is added to the dasymetric process used. Nevertheless, while the dasymetric process is time consuming, it provides a very detailed statistical surface of population across the City of Ottawa.

This research is methodological in nature and presents dasymetric mapping as a means of redistributing census variables to custom neighbourhood geographies using derived population weights. This technique is particularly useful when researchers do not have access to high-resolution population data. The present research could have utilized block-level population counts instead of dasymetric mapping but these are not widely available to researchers on a general basis and so dasymetric mapping provides a viable alternative. Future research should test the microsimulation outputs using both dasymetric mapping and block-level population in order to assess the merits of both on the simulated outcomes. In a related vein, the open access to the 2011 Block Population statistics in Canada should provide impetus for their use in this countries research on microsimulation in the future.

Overall, however, the use of the dasymetric mapping for the re-weighting of SES variables into the ONS community-defined neighbourhoods proved to be a promising approach. Using a dasymetric population surface for estimating predictor variables in irregular shaped geographic units brought to light a methodological approach that could be very useful in improving microsimulation methodologies. While values of the SES variables estimated within the ONS neighbourhoods could not be directly validated, the dasymetric re-weighting process was shown to be sufficiently accurate when SES variables were estimated for Statistics Canada census tracts and compared to actual census data for CTs (Figures 2.9-2.13). A dasymetric re-weighting of predictor variables into irregular shaped geographic units provides an alternative methodology to less accurate re-weighting methodologies that rely on the assumption of a uniform population distribution in geographic space.

The microsimulation and subsequent spatial autocorrelation highlighted neighbourhoods with health determinants and outcomes that had varying prevalences. While the health determinants and outcomes shown in Figures 2.14-2.20 are simulations, they serve as a good indicator of the spatial variability in the health of rural residents. The results of the microsimulation and spatial autocorrelation show that there is no single health issue of significance for all of rural Ottawa; instead, many of the neighbourhoods vary in their health determinant and outcome prevalences. The spatial variability in the maps also presents a starting point for potential community interventions and/or primary data collection to validate the simulated outcomes.

The case of Dunrobin highlights this idea, as the simulated prevalence uncovered in Dunrobin is not shared by the surrounding neighbourhoods. This is demonstrated in the obesity map (Figure 2.19), where the neighbourhoods adjacent to Dunrobin have lower obesity prevalence, while Dunrobin’s is much higher. If the obesity prevalence was analysed according to larger geographic units, the values in adjacent areas would have diluted the values in the Dunrobin area and, consequently the higher prevalence in obesity could be overlooked. This observation is also true for neighbourhoods such as Galletta. Galletta had a higher prevalence of fair/poor selfrated mental health than the surrounding neighbourhoods, but if this area had been assessed based on a larger neighbourhood unit, the Galletta "area" would have appeared to have a lower prevalence of poor selfrated health (Figure 2.18).

In select neighbourhoods, however, the simulated health variable prevalence of certain outcomes was higher than would be expected. For instance, in Cumberland the simulated prevalence of self rated mental health was 22.3 percent of the population twelve years of age and older (Figure 2.18), much higher than the rest of the simulated values. There is no independent data for comparison, however, there are a few factors which could contribute to simulated prevalence outliers: 1-The outliers are real and are the consequence of differences in lifestyle habits, age structure and demographics between the different neighbourhoods; 2-The outliers are due to small sample size from the CCHS; 3- The outliers are a consequence of an ineffective reweighting process. However, most likely, there exists a subtle mixture of all three reasons. For example, if the prevalence of a health outcome is estimated based on a small sample of 30 cases drawn from the pooled CCHS database and of those 30 cases, only one person in the sample is younger than 30, then the prevalence will be biased upward. Even if the reweighted age distribution of the sample is adjusted to better reflect census figures, there is still a dearth of information on the characteristics of people whose age is under 30.

The spatial autocorrelation in the neighbourhoods of Sarsfield and Navan-Carlsbad Springs, show that the selfrated health and selfrated mental health were strongly correlated to high binge drinking in these areas. Findings such as these are critical to the ONS because they help to improve one's understanding the determinants which contribute to the health of communities. From these results, binge drinking is serving as a health determinant, and perhaps represents a pathway to negative selfrated health among residents. This research cannot validate such an idea but does illustrate the utility of the microsimulation methodology. Upon future investigation to validate these results, one may conjecture that an intervention to improve on the overall selfrated health of Sarsfield and Navan Carlsbad Springs might target binge drinking. The results from the microsimulation and spatial autocorrelation alone, however, are not reason enough to begin to implement programs to reverse the situation. Instead, they show areas that have potentially concerning health tends, allowing researchers to focus and to investigate further.

The results presented do have some limitations. Due to the fact that the participatory neighbourhood units were only recently defined, there is no accurate data to validate the estimated health determinants and health outcomes. The simulated health outcomes, however, highlight areas of differing health variable prevalence across rural Ottawa and provide a starting point for further research in the area. The results of the microsimulation provide insight into areas of rural Ottawa that may require attention in future, health-based research.

The entire process used to simulate health variables (outcomes and determinants) provided interesting insights into different health patterns in rural Ottawa. The participatory mapping approach that was used to define new rural neighbourhood boundaries was a product of the perceptions of community and neighbourhood by rural residents. The dasymetric map proved to estimate population counts with acceptable accuracy and was able to be used to re-weight SES variables into the new neighbourhood units. Finally, the use of spatial microsimulation produced simulated health determinant and outcome variables that highlighted positive and potentially negative health patterns in rural Ottawa.

Conclusion

This study leverages important research in remote sensing, geomatics, community mapping, health geography, and microsimulation. Our method demonstrated that neighbourhood boundaries defined through participatory mapping provide an innovative geographic unit at which to study health determinants and outcomes. Using participatory mapping, new neighbourhood units were developed that represented the ecological reality of the residents of rural Ottawa. Furthermore, it was shown that a combination of dasymetric mapping and microsimulation can be used to simulate important health determinant and health outcome variables within a community-defined neighbourhood.

The final result of this study produced maps of simulated prevalence of health outcomes and determinants in areas across rural Ottawa. These prevalence maps serve as a good indication of the health variations that potentially exist across these communities. Further investigation into population patterns and perception of space is required to identify contextual variables affecting health and their true outcomes. It is important to remember that the health determinants and outcomes prevalence that were simulated are meant to uncover and highlight health patterns that may not have been apparent previously. Patterns that are cause for concern can then be investigated by municipalities or health professionals. Hopefully, this study can open the door to uncovering unique health situations that exist in rural areas and lead to the development of healthier neighbourhoods in rural areas of Ottawa and across Canada.

References

1. Riva, M., Apparicio, P., Gauvin, L. & Brodeur, J-M: **Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones**. *International Journal of Health Geographics 2008,* **7:**43.
2. Kristjansson, E., Sawada, M., Russell, K., Ali, A., He, J., McKinnon, J.,Hole, D., Steeves, P., Palozzi, K., Leclair, S., and Calhoun, Ml: **Neighbourhood Advantage, Social Determinants, and Health Outcomes. The Ottawa Neighbourhood Study**. Invited Presentation at "The InSPQ Deprivation Index: Applications in Research, Policy and Practice." *Canadian Institute of Health Information Workshop, Sept. 23, 2009*, Montreal, Canada.
3. Maman S, Lane T, Ntogwisangu J, Modiba P, vanRooyen H, Timbe A, Visrutaratna S, Fritz K: **Using Participatory Mapping to Inform a Community-Randomized Trial of HIV Counselling and Testing**. *Field Methods* 2009, **21**(4): 368-387.
4. Hameed, S. M., Bell, N., & Schuurman, N: **Analyzing the effects of place on injury: Does the choice of geographic scale and zone matter?** *Open Medicine 2010,* **4:** 171-180.
5. Schwab, M. G. and Syme, S. L: **On paradigms, community participation and the future of public health**. *American Journal of Public Health* 1997, ***12*****:** 2049-2051.
6. McMichael, A. J: **Prisoners of the proximate: Loosening the constraints on epidemiology in an age of change**. American Journal of Epidemiology 1999, **149**(10): 887-897.
7. Susser, M: **Should the epidemiologist be a social scientist or a molecular biologist?** *International Journal of Epidemiology 1999,* **28**:1019–1022.
8. Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E**: From social integration to health: Durkhe in the new millennium**. *Social Science & Medicine 2000,* **51**(6): 843–857.
9. Krieger, N: **Theories for social epidemiology in the 21st century: An ecosocial perspective.** *International Journal of Epidemiology 2001,* **30**(4)**:** 668–677.
10. Kiwachi I, Kennedy B. Flass R: Social **Capital and Self –Rated Health: A Contextual Analysis**. American Journal of Public Health 1999, **89:** 1187-1193
11. O'Campo, P: **Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health**. *American Journal of Epidemiology 2003,* **157**: 9–13.
12. Leung, M. W., Yen, I. H., & Minkler, M: **Community based participatory research: A promising approach for increasing epidemiology's relevance in the 21st century**. *International Journal of Epidemiology 2004,* **33:** 499–506.
13. Lopez-Setina J., Lee H. and Robert F: **The link between obesity and the built environment. Evidence from and ecological analysis of obesity and vehicle miles of travel in California**. *Health and Place 2006*, **12**(4): 656-664.

14. Stafford M., Cummins S., Ellaway A., Scker A., Wggins R. andMacintyre S: **Pathways to Obesity: Identifying local, modifiable determinants of physical activiety and diet**. *Social Science and Medicine 2007*, **65**(7): 1882-1897
15. Sampson, R. J., J. D. Morenoff, and T: Gannon-Rowley: **Assessing "neighborhood effects": Social processes and new directions in research**. Annual Review of Sociology 2002, **28**: 443-78.
16. Germain, A., and J. E. Gagnon: **Is neighborhood a blackbox? A reply to Galster, Metzger, and Waite**. *Canadian Journal of Urban Research 1999,* **8**(2): 172.
17. Kawachi, I., and L. F. Berkman**: Introduction in neighborhoods and health**. In I. Kawachi and L. F. Berkman, Eds., Neighborhoods and health. New York: Oxford University Press, 1, 1-19.
18. Martin, D. G: **Enacting neighborhood**. *Urban Geography2003*, **24**(5): 361-85.
19. Clapp, J. M., and Y. Wan: **Defining neighborhood boundaries: Are census tracts obsolete?** *Journal of Urban, Economics 2006,* **59**(2): 259-84.
20. M.-P. Parenteau, M. Sawada, Sam Herold, Elizabeth A. Kristjansson, R. Labonte, M. Calhoun, S. Leclair, V. Runnels, A. Musiol, R. Mawby, N. Muhajarine, D. Salisbury, Amira Ali, Marguarite Keeley, Ian Cross, Cliff Gazee : **A Participatory Geomatics Approach to Defining Health-Relevant Neighbourhoods in Ottawa, Canada.** *Journal of Urban Health*: Bulletin of the New York Academy of Medicine 2008, **84**(6): 856.
21. Haynes R., Daras, K., Reading, R. and Jones, A: **Modifiable Neighbourhood Units, zone design and residents perception**. Health and Place 2007, **13**(4): 812-825
22. Haynes R. and Gale S: **Deprivation and poor health in rural areas: inequalities hidden by averages**. *Health and Place 2000*, **6**(4): 275-285.
23. Canadian Community health Survey: **[http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=3226&SurvVer=0&InstaId=15282&InstaVer=1&SDDS=3226&lang=en&db=IMDB&adm=8&dis=2]**
24. Geography Facts: [**http://ottawa.ca/en/visitors/about/quick_facts/index.html**]
25. Maantay, J**: Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity**. Environmental Health Perspectives 2002, **110:** 161–171.
26. O'Fallon, L. R., & Dearry, A: **Community-based participatory research as a tool to advance environmental health sciences**. Environmental Health Perspectives 2002, **110:**155–159
27. Beyer, K. M. M., Comstock, S., & Seagren, R: **Disease maps as context for community mapping: A methodological approach for linking confidential health information with local geographical knowledge for community health research**. *Journal of Community Health 2010,* 35(6): 635-644.

28. Keith M and Brophy JT**: Participatory mapping of occupational hazards, disease, and injury among asbestos-exposed workers from a foundry and insulation complex in Southwestern Ontario, Canada**. *International Journal of Occupational Environment Health* 2004, **10:** 144-53.
29. Keith MM, Cann B, Brophy JT, Hellyer D, Day M, Egan S, Mayville K, Watterson A: **Identifying and prioritizing gaming workers' health and safety concerns using mapping for data collection**. *American Journal Ind Med 2001,* **39**(1):42–51.
30. Mosley, B: **Estimating Health Outcomes in Rural Ottawa: A Geographic and Statistical Methodology.** Unpublished masters dissertation 2012, University of Ottawa, Ottawa Ontario.
31. Langford, M., & Higgs, G: **Measuring potential access to primary healthcare services: The influence of alternative spatial representations of population**. *Professional Geographer 2006, **58***(3): 294-306.
32. Langford, M: **Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method.** *Computers, Environment and Urban Systems2006, **30***: 161-180.
33. Mennis, J., & Hultgren, T: **Intelligent dasymetric mapping and its application to aerial interpolation.** *Cartography and Geographic Information Science 2006,* **33**(3): 179-194.
34. Langford, M: **Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps.** *Computers, Environment and Urban Systems 2007,* **31**(1): 19-32
35. Mennis, J**: Dasymetric mapping for estimating population in small areas**. *Geography Compass 2009*, **3**: 727–745
36. Mennis, J: **Generating surface models of population using dasymetric mapping.** *Professional Geographer 2003*, **55**(1): 31-42
37. Mei, T., Li, F., Qin, Q., & Li, D: **Road extraction from remote sensing image using support vector machine**. Paper presented at the International Society for Optical Engineering, **5286**(1): 299-304.
38. Hsu Chih-Wei, and Chih-Jen Lin: **A practical guide to support vector classification**. *Department of Computer Science 2010, National Taiwan University*. 1-16
39. Haitao, Li, Haiyan Gu, Yanshun, Han and Jinghhi, Yang: **Object-oriented classification of high resolution remote sensing imagery based on an improved colour structure code and support vector machine**. *International Journal of Remote Sensing 2010,* **31**(6): 1453-1470.
40. **Surface Extraction and Classification: [**http:/http://gis.incogna.com/?p=technology#Surface]
41. Holzer, C., Jackson, D., Tweed, D**: Horizontal synthetic estimation: A strategy for estimation small are health related characteristics**. *Evaluation and Program Planning* 1981, **4**(1): 29-34
42. Twigg, L., Moon, G: **Prediction small area health related behaviour: a comparison of multilevel synthetic estimation and local survey data.** *Social Science and Medicine 2002,* **54:** 931-937

43. Tanton, M., Harding, A., McNamara, J: **Urban and Rural Estimates of Poverty: Recent Advances in Spatial Microsimulation in Australia**. *Geographical Research 2009*, **41:** 52-64.
44. Riva, M., Smith M**: Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method.** *Social Psychiatry Epidemiology 2010*, **47**(5): 745-755.
45. Shannon, C. E: **A Mathematical Theory of Communication**. *Bell System Technical Journal 1948*, **27:** 379-423.
46. Golan, A., G.G. Judge and D. Miller: ***Maximum Entropy Econometrics****: Robust Estimation with Limited Data*. Chichester, UK: Wiley.
47. Getis, Arthur, and J. K. Ord: **The Analysis of Spatial Association by Use of Distance Statistics.** *Geographical Analysis* 1992, **24:** 189-206.
48. Goodchild, Michael F: **Spatial Autocorrelation**. Catmog 47, Geo Books. 1986.
49. Griffith, Daniel: **Spatial Autocorrelation: A Primer**. *Resource Publications in Geography, Association of American Geographers 1987*. **2:** 9-26
50. Mitchell, Andy. ***The ESRI Guide to GIS Analysis***, Volume 2. ESRI Press, 2005.
51. Lee, S. **Developing Bi-vitiate Spatial Association Measure**: An Integration of Pearson's r and Moran's I. *Geographical Systems 2003,* ***3****(4): 369-385*

Chapter 3: Conclusions

3.1 Introduction

The goal of this thesis was to use dasymetric mapping and spatial microsimulation to provide insight into health determinants and health outcomes in community defined neighbourhoods of rural Ottawa. The purpose of this undertaking was to create a methodology that would enable the study of health determinants and outcomes in regions that are under sampled in regional surveys like the Canadian Community Health Survey. Moreover, this thesis presented a methodology that could be used to simulate health-related variables within custom geographic units for which population statistics are unavailable.

This thesis argued that because government enumeration units in rural areas are both large in physical area and defined for statistical purposes, they do not always properly represent the areas with which rural residents interact with on a daily basis (Germain and Gagnon 1999, Kawachi and Berkman 2003, Martin 2003, Clapp and Wang 2006). Furthermore, the assertion was made that because of statistical barriers for the estimation of health variables, past attempts to define "natural neighbourhood units" have fallen short of properly representing the environment with which rural Ottawa residents associate (Haynes et al., 2007; Parenteau et al., 2008; Kristjansson et al., 2009). These facts are problematic for geographic health research, because in order to properly identify the link between health outcomes and health determinants, one must know the spaces with which people are interacting (Shwab and Syme, 1997; Haynes et al., 2007 and Kristjansson et al., 2009). The spaces and environments with which subjects interact are critical to determining if contextual determinants are affecting a person’s health (Ross, 2004; Parenteau et al., 2008; Kristjansson et al., 2009). In response to this problem, this thesis put forward a methodology that combined geographical and statistical techniques that allow for health statistics to be simulated within neighbourhoods defined through community participatory mapping.

In order to achieve the objectives laid out in this thesis, neighbourhoods defined by the ONS through community-based participatory mapping were used as boundaries to simulate health variables in rural Ottawa. This was executed through population-based re-weighting of census-level, health determinant variables (socioeconomic status, or SES) using the statistical population surface generated by a dasymetric mapping process. The socioeconomic variables, which were redistributed, came from the, age, sex, income, education, language and visible minority categories of the 2006 Canadian census. Finally, spatial microsimulation was undertaken to estimate specific health determinants and outcomes from the Canadian Community Health Survey within the new neighbourhood units. Six health variables were micro-simulated. These included three determinants and three outcomes: smoking prevalence, binge drinking prevalence, selfrated health, selfrated mental health, obesity prevalence, and the presence of two or more chronic conditions. Through this methodology, health determinants and health outcome prevalence were simulated within community-defined neighbourhood boundaries to gain insight into the state of health variation across rural Ottawa.

This section includes a discussion on each one of the sub-objectives of this thesis. The results of dasymetric population mapping, dasymetric re-weighting and microsimulation are all analysed for the contributions they made to achieve the main objective of the thesis. Furthermore, the main objective of the thesis is discussed for its success and limitations, along with areas for future research.

3.2 Dasymetric Mapping

The dasymetric mapping component of this thesis was put forward due to the fact that the neighbourhood units defined by the ONS had been recently defined and did not follow any traditional enumeration unit boundary. Thus, in order to obtain correlate variables for microsimulation, population statistics would need to be re-weighted into the new neighbourhood units. Past methodologies that had attempted to re-weight predictor variables into the target geographic units for microsimulation had used forms of re-weighting that faced barriers because of the implicit assumption of a uniform population distribution in enumeration units (Lymer et al., 2009). This thesis, therefore, puts forward a method of dasymetric or population re-weighting to estimate census variables within ONS neighbourhoods.

The dasymetric method put forward in this study follows closely that of Mennis (2003), with two notable exceptions. Firstly, the ancillary data used for the redistribution of the population is significantly more detailed than that used in Mennis (2003). This study used two ancillary datasets: City of Ottawa zoning data and a land cover map derived from high resolution satellite imagery. These two pieces of ancillary data were combined to allow for the population to be redistributed at a much finer scale, and the land cover map derived from the SVM feature extraction was one meter resolution and had four classes. The advantage of deriving one's own land cover map as a piece of ancillary data is that one has control over the resolution of the derived map and the number of classes extracted. One can also be assured that the map is current. This is crucial, because previous studies have been limited in accuracy by using ancillary data that is of a low spatial resolution and out of date (Mennis, 2003).

Secondly, this thesis used City of Ottawa zoning data as opposed to a three class, urban, suburban and rural map used in Mennis (2003). The use of the City of Ottawa zoning map had several advantages. It served as a "filter" on the urban data extracted from the land cover map. The urban data extracted from the land cover map contained features such as factories and warehouses that could not be differentiated from houses or apartment buildings in the feature extraction process. The zoning data, however, does distinguish between these types of areas. Thus, industrial or commercial zones that were shown not to have people living in them were removed from the data set.

The zoning information was also useful in that zones that did not have people living in them were eliminated from the dataset. After the zones that did not include population were eliminated, there were still multiple zones in which population density could be calculated. This was beneficial to the dasymetric process, as it further refined the accuracy of population in certain areas. Previous studies had been limited to working with just urban, suburban, and rural classes (Mennis, 2003), but the City of Ottawa data allowed for density to be calculated in several urban and rural classes. Residential densities were differentiated along with the density of mixed commercial space.

Further, the feature extraction, while accurate at identifying urban areas, did still have a commission error. The error, however, was mitigated in the same way as undesirable urban features. Features such as farmer's fields or barren land, which may have had a similar spectral signature to urban areas, were eliminated from the ancillary dataset by the zoning dataset.

The addition of two new pieces of ancillary data to the dasymetric method used in this study provided good accuracy. The only limitation to the addition of the ancillary datasets is that the creation of a land cover map specifically for the dasymetric process was time-consuming and requires a background in remote sensing, which other dasymetric methods do not. Thus, whether or not this method of dasymetric mapping is suitable for future studies is entirely dependent on the nature of the research and the skill set of those who perform the analysis. For the purposes of this study, however, having a very detailed view of where people reside in geographic space around Ottawa was very useful for accurate health variable estimation.

The population surface generated through the dasymetric mapping process provided the basis for which the CCHS SES correlates would be re-weighted to the new neighbourhood units. The re-weighted SES variables were used as the link between CCHS areas and the new neighbourhood units for microsimulation. The dasymetric mapping step was fundamental to achieving the objectives laid out for this thesis.

3. 3 Dasymetric Re-weighting

Once the dasymetric mapping had been completed, SES census variables were re-weighted to the new neighbourhood units, and health determinants and outcomes were microsimulated. Spatial microsimulation has been used effectively in a number of studies in Europe and the UK to model health outcomes and their determinants, but these studies were simulating health variables within geographic units for which correlate variables were already available (Tantoon and McNamara, 2002; Riva and Smith, 2010). In these studies, health variables from national level health surveys were simulated into government defined enumeration units. Thus, because the target areas for microsimulation were well established enumeration units, population statistics were readily available to be used as predictor variables for microsimulation.

When microsimulation has been used to estimate health variables into irregular shaped geographic units, however, the results have been less accurate (Lymer et. al., 2009). This is due to the fact that the target geographic units in these types of studies are some type of newly defined neighbourhood unit for which there are not well-documented population statistics. Therefore, correlate variables must be re-weighted or estimated into the target geographic units for microsimulation to be performed; traditionally, some form of area re-weighting has been used. In light of this barrier, this thesis presents a population based re-weighting of population statistics using a statistical population surface generated by dasymetric mapping.

The dasymetric process used in this thesis was more time-consuming than other processes that have been used in past research, but the additional accuracy proved valuable because re-weighting census variables based on the dasymetric map were shown to be sufficiently accurate. The neighbourhoods defined by the ONS through participatory mapping had just been established; thus, the accuracy of the dasymetric re-weighting was validated by estimating populations within neighbourhoods for which (SES) predictor variables were already known. The process of dasymetric re-weighting SES variables was therefore executed for a wide selection of SES variables into Statistics Canada CT units. When this validation was performed, the dasymetric re-weighting proved to accurately estimate SES variables. An R^2 of .80 or greater was recorded between the 2 datasets (once errors had been eliminated), with very good statistical significance values (p value>0.05). From the results of the validation, the assumption was made that the SES variables re-weighted to the new neighbourhood would serve as the best possible correlate variables for the microsimulation process. This step was beneficial, as the re-weighted SES variables provided the baseline variables used for spatial microsimulation.

3.4 Spatial Microsimulation

Finally, the microsimulation and subsequent spatial analyses highlighted the variation in health outcomes and determinants across rural Ottawa. While the health determinants and outcomes shown are simulations, they serve as a good indicator of potential health prevalence within different areas across rural Ottawa. The results of the micro- simulation and spatial autocorrelation show that there is no single health issue of significance for all of rural Ottawa; instead, health determinant and health outcome prevalences vary across the region, demonstrating areas of potential further study and intervention.

3.5 Limitations

The results of this thesis provide some very interesting insights into the prevalence of health variables across rural Ottawa and a methodology to simulate health variables within geographic units for which they are not traditionally kept. There are, however, a couple of limitations to the study that should be acknowledged when looking at the results and considered in future research.

Firstly, the health determinant and health outcome variables that were simulated within the new community defined neighbourhoods could not be directly verified. This is because the boundaries of these neighbourhoods had just been recently defined by the ONS and no third party statistics are available for verification of the simulations. Thus, the results of the simulation of health variables cannot be directly verified. This limitation, however, was expected, as the goal of this thesis was to provide a methodology that would provide insight into health variable prevalence in irregular shaped geographic units for which health statistics are not traditionally kept. Rather the results of the simulation should be taken as a preliminary insight into rural neighbourhood health variable prevalences, and as a guide for future research in the area.

Secondly, it must be acknowledged that the community defined neighbourhood units around which this study was focused should not be taken as the ideal geographic unit at which to study health, as no such unit exists. The neighbourhood units defined by the ONS and used in this study were created in an attempt to define boundaries that would represent the average daily activities of given communities. These neighbourhoods presented a different approach to defining the geographic unit at which health can be measured. While the consultations that were undertaken by the ONS yielded good consensus among participants, it is acknowledged that the boundaries presented inevitably will not be perfect for every rural resident.

3.6 Contributions

The methodology for simulating health variables in this thesis drew together methods from several different fields. The techniques of participatory mapping, dasymetric mapping, support vector machine feature classification, and spatial microsimulation were all combined into one singular approach that allowed for the simulation of health variables within communities across rural Ottawa. While none of the techniques were individually new, this thesis does represent the first time they have all been combined to create one methodological approach. It is through the integration of each one of these techniques that the results of this thesis were obtained.

To begin, while the technique of spatial microsimulation had been used to good effect in several health geography studies in the United Kingdom (Twigg and Moon, 2002; Ballas, 2006; Riva and Smith, 2010), the version presented in this thesis was the first to use dasymetrically re-weighted population variables as the correlates between geographic areas. This approach was taken because the community defined neighbourhoods that were the focus of the simulation had just been recently defined and did not have government or third-party population statistic available for the area. Thus, population statistics needed to be estimated within the boundaries of the new neighbourhood units in order to obtain the correlate variables required for the spatial microsimulation process. The dasymetric re-weighting of population variables was chosen in response to past studies that had not achieved optimal results when using other forms of re-weighting to estimate correlate variable in irregularly shaped study areas. Thus, while it has been noted that the results of the spatial micro-simulation presented in this thesis cannot be directly validated, this thesis does provide a theoretical framework for the estimation of health variable prevelances within irregularly shaped geographic units.

Furthermore, the dasymetric mapping methodology used in this thesis differs from past dasymetric methodologies (Mennis, 2003; Langford, 2006) in that it uses SVMC to create a high resolution land-cover map to be used as a piece of ancillary data specifically for the dasymetric approach. This was done to overcome limitations noted in past dasymetric methodologies that cited out-of-date or low-resolution ancillary data as a limitation to the overall accuracy of the dasymetric population surface generated. This thesis thus integrated the use of detailed feature extraction approaches as a means to improve upon existing dasymetric methodologies.

Finally, the results of the methodology carried out in this thesis provide some very interesting insights into the prevalence of health variables in neighbourhoods across rural Ottawa. Firstly, the neighbourhoods defined through the community participatory mapping exercise allow for health variable prevalences to be viewed within geographic units that had not been previously available. The community defined neighbourhood units in this thesis were defined largely through community input, and thus the simulated health prevalences are representative of those unique communities. Overall the results of this thesis provide preliminary insight into the prevalence of smoking, binge drinking, obesity, chronic conditions, selfrated health and selfrated mental health in neighbourhoods across rural Ottawa.

3.7 Future Research

The research presented in this thesis provides potential method for simulating health outcomes in rural areas, but there are a few areas of research that could be extended. Firstly, the edge effect in the participatory mapping methodology impacts the ability of researchers to define true, natural neighbourhood units along the city boundary. Neighbourhood boundaries along the edge of the city likely should not follow municipal boundaries perfectly. As such, it is vital that the participatory mapping methodology be continued in the counties surrounding Ottawa. Participatory mapping consultations carried out in the counties of Lanark, Grenville, Dundas, and Russell could be combined with the participatory mapping done within Ottawa to create neighbourhoods that are not bound by any administrative or political boundaries.

With the development of new neighbourhoods outside of Ottawa, dasymetric mapping and microsimulation would need to be performed in these areas. If completed, the methodology outlined in this thesis could be used to estimate health outcomes for a large portion of Eastern Ontario.

Furthermore, while dasymetric mapping was shown to create a highly accurate picture of population density across the City of Ottawa, this methodology could be extended to include a temporal scale. This means that while the current dasymetric methodology represents a static population, people move to schools and work places during the day, thus changing the population density of the city. As such, a "daytime dasymetric map" could be created to represent the population density of Ottawa during the day. This would be known as "ambient population density." The most successful form of this type of dasymetric mapping was completed in the LandScan USA project (Bhaduri et al., 2007).

3.8 Conclusion

Overall, combining dasymetric mapping and spatial microsimulation provided a helpful and innovative methodology for simulating health outcomes in neighbourhoods across rural Ottawa. The process of estimating population and SES variables through dasymetric re-weighted provided the needed data to support spatial microsimulation.

The process presented in this thesis overcame several barriers related to health research in rural Ottawa. With the new neighbourhood units, dasymetric mapping and spatial microsimulation were able to be combined in order to depict health patterns in an alternative way. Several of the new neighbourhood units had quite different health variable prevalence when compared to the neighbourhoods around them. This reveals that should health be analysed at large spatial scales, the prevalence of specific health phenomena could be diluted. The results also demonstrate that geographical and statistical techniques such as dasymetric mapping and spatial microsimulation can provide simulations that highlight patterns in health that otherwise may have gone unnoticed.

The resulting estimated health variables will serve as an extremely useful tool for the Ottawa Neighbourhood Study. Concerning health patterns that were uncovered in this thesis will be further investigated by the ONS to determine their exact cause and what should be done to reverse them. There is a strong need to validate the results of this new methodological approach in rural Ottawa. Ultimately, it is hoped that the findings presented in this thesis will transform the way spatial health research is performed and lead to better health care services, not only in Ottawa, but around the country.

References

Ballas, D., G. Clarke, D. Dorling, H. Eyre, B. Thomas, and D. Rossiter. (2005). SimBritain: A Spatial Microsimulation Approach to Population Dynamics. *Population, Space and Place,* (11) 13-34.

Bhaduri, B., Bright, E., Coleman, P., Urban, M. (2007) LandScan USA: A High Resolution Geospatial and Temporal Modeling Approach for Population Distribution and Dynamics *GeoJournal* (69) 103-117.

Clapp, J. M., and Y. Wang. (2006). Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban, Economics,* 59(1), 259-84.

Germain, A., and J. E. Gagnon. (1999). Is neighborhood a blackbox? A reply to Galster, Metzger, and Waite. *Canadian Journal of Urban Research* 8(2): 172.

Haynes R., Daras, K., Reading, R. and Jones, A. (2007)Modifiable Neighbourhood Units, zone design and residents perception. *Health and Place*, 13(4): 812-825

Kawachi, I., and L. F. Berkman. (2004). Introduction in neighborhoods and health. In I. Kawachi and L. F. Berkman, Eds., Neighborhoods and health. New York: Oxford University Press, 1, 1-19.

Kristjansson, E., Sawada, M., Russell, K., Ali, A., He, J., McKinnon, J.,Hole, D., Steeves, P., Palozzi, K., Leclair, S., and Calhoun, Ml. (2009). Neighbourhood Advantage, Social Determinants, and Health Outcomes. The Ottawa Neighbourhood Study. Invited Presentation at "The InSPQ Deprivation Index: Applications in Research, Policy and Practice." Canadian Institute of Health Information Workshop, Sept.

Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems, 30*(2), 161-180.

Lymer S., Brown L. (2009). Predicting the need for aged care services at the small area level: The CAREMOD spatial microsimulation model. *International Journal of Microsimulation*. 2(2), 27-42.

Martin, D. G. (2003). Enacting neighborhood. *Urban Geography*, 24(5), 361-85.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *Professional Geographer*, *55*(1), 31-42.

M.-P. Parenteau, M. Sawada, Sam Herold, Elizabeth A. Kristjansson, R. Labonte, M. Calhoun, S. Leclair, V. Runnels, A. Musiol, R. Mawby, N. Muhajarine, D. Salisbury, Amira Ali, Marguarite Keeley, Ian Cross, Cliff Gazee (2007). A Participatory Geomatics Approach to Defining Health-Relevant Neighbourhoods in Ottawa, Canada. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 84*(6) 856.

Riva, M., Smith M. (2010). Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method. *Social Psychiatry Epidemiology*, *47*, 745-755.

Ross, N. A., Tremblay, S., & Graham, K. (2004). Neighbourhood influences on health in Montréal, Canada. *Social Science and Medicine*, *59(*7), 1485-1494.

Schwab, M. G. and Syme, S. L (1997). On paradigms, community participation and the future of public health. American Journal of Public Health, (12) 2049-2051.

Tanton, M., Harding, A., McNamara, J. (2009). Urban and Rural Estimates of Poverty: Recent Advances in Spatial Microsimulation in Australia. *Geographical Research*, 41(1), 52-64.

Twigg, L., Moon, G. (2002). Prediction small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data. *Social Science and Medicine.* 54(6), 931-937

Appendix A

Listed below are all of the Socio-Economic Status (SES) variables from the 2006 Canadian census, which were dasymetrically re-weighted and used as baseline values for the statistical microsimulation.

Income Variables
Total income in 2005 of population 15 years and over—20% sample data / Median income $
Total income in 2005 of population 15 years and over—20% sample data / Average income $
Family income in 2005 of economic families—20% sample data
Household income in 2005 of private households—20% sample data
Total income in 2005 of population 15 years and over—20% sample data
Median family income $
Average family income $
Median household income $
Average household income $

Language/Minority Variables
Multiple responses / English and French
Multiple responses / English and non-official language
Multiple responses / French and non-official language
Total population by mother tongue—20% sample data
Single responses / English
Single responses / French
Single responses / Non-official languages
Total population by visible minority groups—20% sample data
Total population by visible minority groups—20% sample data / Total visible minority population

Education Variables
Total population 25 to 64 years by highest certificate, diploma or degree—20% sample data
Total population 25 to 64 years by highest certificate, diploma or degree—20% sample data / No certificate, diploma or degree

Total population 25 to 64 years by highest certificate, diploma or degree—20% sample data / Certificate, diploma or degree

Certificate, diploma or degree / High school certificate or equivalent

Certificate, diploma or degree / Apprenticeship or trades certificate or diploma

Certificate, diploma or degree / College, CEGEP or other non-university certificate or diploma

Certificate, diploma or degree / University certificate, diploma or degree

University certificate, diploma or degree / University certificate or diploma below Bachelor level

University certificate, diploma or degree / University certificate or degree

University certificate or degree / Bachelor degree

University certificate or degree / University certificate or diploma above Bachelor level

University certificate or degree / Degree in medicine, dentistry, veterinary medicine or optometry

University certificate or degree / Master degree

University certificate or degree / Earned Doctorate

Total population 65 years and over by highest certificate, diploma or degree—20% sample data

Total population 65 years and over by highest certificate, diploma or degree—20% sample data / No certificate, diploma or degree

Total population 65 years and over by highest certificate, diploma or degree—20% sample data / Certificate, diploma or degree

Certificate, diploma or degree / High school certificate or equivalent

Certificate, diploma or degree / Apprenticeship or trades certificate or diploma

Certificate, diploma or degree / College, CEGEP or other non-university certificate or diploma

Certificate, diploma or degree / University certificate, diploma or degree

University certificate, diploma or degree / University certificate or diploma below Bachelor level

University certificate, diploma or degree / University certificate or degree

University certificate or degree / Bachelor degree

University certificate or degree / University certificate or diploma above Bachelor level

University certificate or degree / Degree in medicine, dentistry, veterinary medicine or optometry

University certificate or degree / Master degree

University certificate or degree / Earned Doctorate

Age/Sex variables

Population, 2006—100% data

Total population by sex and age groups—100% data

Total population by sex and age groups—100% data / Male, total

Male, total / 0 to 4 years

Male, total / 5 to 9 years

Male, total / 10 to 14 years

Male, total / 15 to 19 years

Male, total / 20 to 24 years

Male, total / 25 to 29 years

Male, total / 30 to 34 years

Male, total / 35 to 39 years

Male, total / 40 to 44 years

Male, total / 45 to 49 years

Male, total / 50 to 54 years

Male, total / 55 to 59 years

Male, total / 60 to 64 years

Male, total / 65 to 69 years

Male, total / 70 to 74 years

Male, total / 75 to 79 years

Male, total / 80 to 84 years

Male, total / 85 years and over

Total population by sex and age groups—100% data / Female, total

Female, total / 0 to 4 years

Female, total / 5 to 9 years

Female, total / 10 to 14 years

Female, total / 15 to 19 years

Female, total / 20 to 24 years

Female, total / 25 to 29 years

Female, total / 30 to 34 years

Female, total / 35 to 39 years

Female, total / 40 to 44 years

Female, total / 45 to 49 years

Female, total / 50 to 54 years

Female, total / 55 to 59 years

Female, total / 60 to 64 years

Female, total / 65 to 69 years

Female, total / 70 to 74 years

Female, total / 75 to 79 years

Female, total / 80 to 84 years

Female, total / 85 years and over

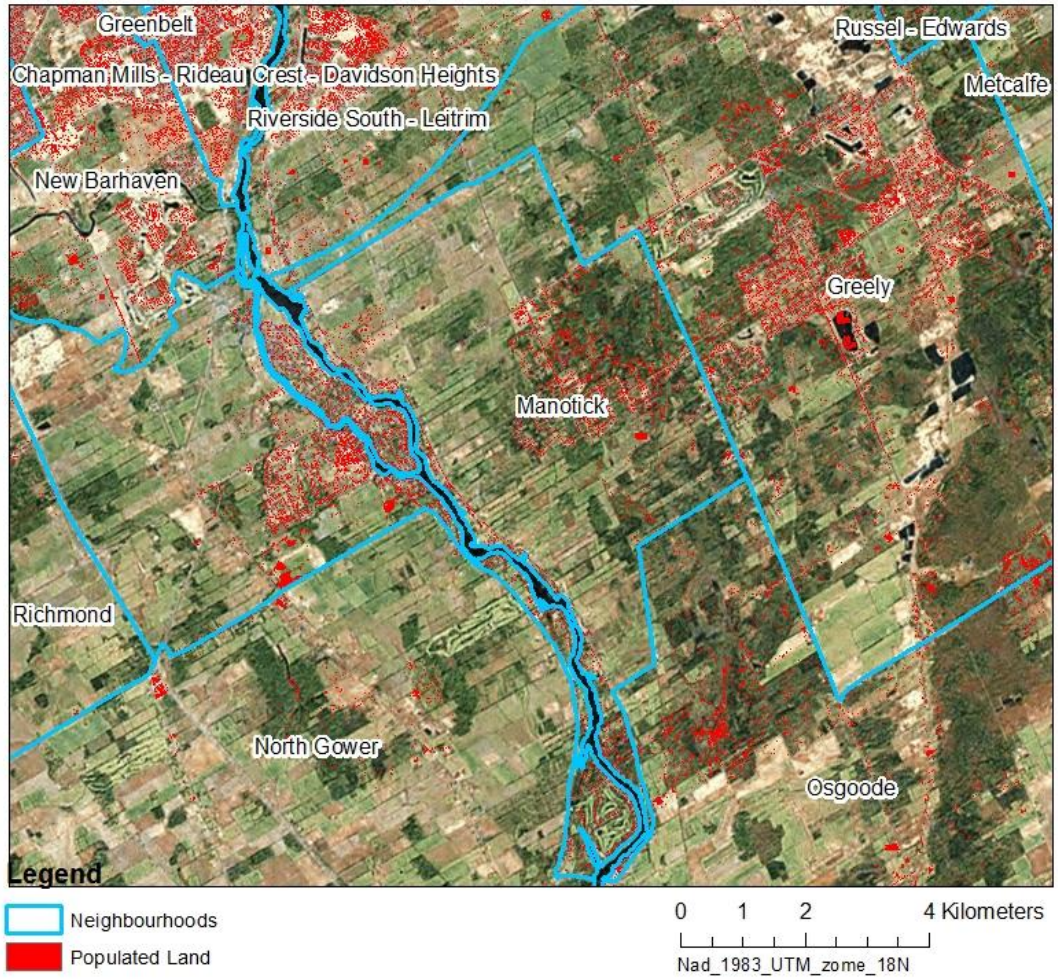
Appendix B

The selected health determinants and outcomes from the CCHS used in this thesis, along with their corresponding sample sizes at each cycle of the CCHS. Further, a select number of other health outcomes from the CCHS are listed. This table is displayed to demonstrate the statistical reasoning for choosing: binge drinking, smoking, obesity, Self rated health, self rate mental health and chronic conditions for inclusion in this study.

	2001 Cases	Sample	2003 Cases	Sample	2005 Cases	Sample	2007 Cases	Sample	2008 Cases	Sample	2009 Cases	Sample	Total Cases	Sample
Binge drinking	294	1488	294	1580	305	1561	168	788	134	760	138	749	1333	6926
Smoking	414	1935	335	2103	389	1962	162	994	159	967	139	939	1598	8900
Chronic conditions	630	1836	737	2043	743	1968	319	997	302	965	371	940	3102	8749
Self-perceived health	225	1936	211	2045	224	1974	88	998	102	968	122	940	972	8861
Self-perceived mental health	N/A	N/A	113	2005	101	1946	47	980	64	955	54	926	379	6812
Obesity	189	1342	233	1799	235	1921	123	954	141	908	137	899	1058	7823
Diabetes	62	1935	99	2045	96	1972	45	997	50	966	53	939	405	8854
Injuries	298	1935	299	2038	330	1943	N/A	N/A	N/A	N/A	153	940	1080	6856
Heart disease	84	1935	102	2047	102	1972	53	997	29	965	51	940	421	8856
Fibromyalgia	20	1935	41	2046	34	1973	N/A	N/A	N/A	N/A	N/A	N/A	95	5954

Appendix C

Built up land cover extracted through SVMC and overlaid on top of aerial imagery to show how populations were focused to where they existed in geographic space for dasymetric mapping.



Appendix D

Aggregation of the zoning classes according to the City of Ottawa Zoning By-law 2008-250 Consolidation. (modified from, Parenteau, 2008)

Original Zoning Type	New Zoning Type
Agricultural Zone	Nopop
Arterial Mainstreet Zone	AM
Development Reserve Zone	DR
Environment Protection Zone	EP
General Mixed Use Zone	GM
Minor Institutional Zone	NoPop
Major Institutional Zone	NoPop
General Industrial Zone	IG
Heavy Industrial Zone	NoPop
Light Industrial Zone	IL
Business Park Industrial Zone	IP
Community Leisure Facility Zone	NoPop
Major Leisure Facility Zone	NoPop
Central Experimental Farm Zone	NoPop
Local Commerical Zone	R3
Mixed Use Centre Zone	MC
Mixed Use Downtown Zone	MD
Mineral Extraction Zone	ME
Mineral Aggregate Reserve Zone	ME
Parks and Open Space Zone	O1
Residential First Density Zone	R1
Residential Second Density Zone	R2
Residential Third Density Zone	R3
Residential Fourth Density Zone	R4
Residential Fifth Density Zone	R5

Rural Commercial Zone	NoPop
Rural General Industrial Zone	NoPop
Rural Heavy Industrial Zone	NoPop
Rural Institutional Zone	NoPop
Mobile Home Park Zone	RM
Rural Residential Zone	RR
Rural Countryside Zone	RU
Air Transportation Facility Zone	NoPop
Ground Transportation Facility Zone	NoPop
Traditional Mainstreet Zone	TM
Village Mixed Use Zone	VillRes
Village Residential First Density Zone	VillRes
Village Residential Second Density Zone	VillRes
Village Residential Third Density Zone	VillRes

Appendix E

The calculations needed to perform dasymetric population redistribution (Mennis, 2003).

Step in the Dasymetric Mapping Process	Formula	Description
Sample population Density	***PD = POP / LX*** PD = Population Density POP = Population Lx = Area of a given Land cover class	The sample population density is calculated by obtaining a population count for a small sample area within each class and dividing it by the area that the population falls in. Sample populations were obtained from 2006 dissemination areas from Statistics Canada.
Population Density Fraction	**PDF = PD / (PDa + PDb + PDc…PDx)** PDF = population density fraction PD = population density of land cover classes within given unit of analysis.	The population density fraction is calculated by dividing a given zone's population density by the sum of the population density values for all of the zones found within the unit of analysis. This calculation provides a percentage value. This percentage is the percent of the total population of your unit of analysis that should be allocated to the specific class.
Area Ratio	***AR = Lx / Lt*** AR = Area Ratio Lx = Area of a given land cover class Lt = Total area of unit of analysis	The population density fraction, however, assumes that there is an equal area of each one of the zoning classes within the unit of analysis. One can assume that this is most likely not the case. This error is corrected by calculating an area ratio. The area ratio is very simply calculated by dividing a zoning class’s area by the total area in the unit of analysis.
Total Fraction	***TF = (PDFa *ARa) / ((PDFa*ARa)+(PDFb*ARb)+*** ***(PDFc*ARc)+….(PDFx*ARx)*** TF = Total Fraction PDFx = Population Fraction Ar = Area Ratio	The population density fraction and the area ratio formulas can be integrated into one single statement that calculates the proper percentage of population for a class within a unit of analysis. The total fraction may be calculated by multiplying the population density fraction and area ratio of a given zoning class, and dividing that result by the result of that same expression for all zoning classes in that block group.
Population Distribution	***PP = (TF*TP)/LX***	The final calculation is the population distribution. This is calculated by multiplying the total

	PP = population per pixel TF = Total Fraction TP = Total population LX = Area of given land cover class	fraction of a class by the total population of your unit of analysis and then dividing it by the area of the given zoning class at which you are looking. This calculation gives the number of people that should occupy every pixel across your unit of analysis. Using these values, population estimates can be obtained for a single house, a street, or a neighbourhood.

Appendix F

Listed below are all the neighbourhoods defined by the Ottawa Neighbourhood Study and their corresponding population estimation.

Neighbourhood Name	Population
Fitzroy	2428.682617
Constance Bay	2456.566894
Galetta	717.313781
Carp Ridge	968.397766
Cumberland	4149.861816
Munster	3514.995849
Pierces Corners	1784.368041
Corkery	3159.856445
Kinburn	1637.672973
North Gower	4575.27246
Manotick	8532.229492
Osgoode	5604.345214
Metcalfe	4851.610839
Russel - Edwards	2790.368652
Greely	7507.250488
Vars	2682.259033
Sarsfield	913.758422
Dunrobin	4915.07666
Carp	4565.03955
Navan - Carlsbad Springs	3529.228271
Stittsville	19846.66406
New Barhaven	21782.3418
Richmond	6152.122558
Barrhaven West - Old Barrhaven	18030.78516
Bayshore	7983.603515
Beacon Hill South - Cardinal Heights	6946.221679
Beaverbrook	5247.665527
Bells Corners Commercial - Bells Corners East	4435.4458
Bells Corners West	4448.298339
Billings Bridge - Alta Vista	11808.75098
Blossom Park - Blossom Park West - Sawmill Creek*	7212.324218
Borden Farm - Stewart Farm - Parkwood Hills - Fi*	10184.92773
Braemar Park - Bel Air Heights - Copeland Park	7324.920898
Briar Green - Leslie Park	4932.006835
Byward Market - Parliament Hill	8182.483398
Carleton Heights - Rideauview	6694.214843
Carleton University	310.808868

Carlington	10133.84766
Carlingwood - McKellar Park - Laurentien View	9515.45996
Carlingwood West - Glabar Park - McKellar Heights	5345.399414
Carson Grove - Carson Meadows	7994.314941
Beechwood Cemetery	150.698913
Centrepointe	7277.993164
Centretown - Downtown	20514.76172
Orleans Chapel Hill	8529.426757
Cityview - Skyline - Fisher Heights	6548.353027
Civic Hospital - Experimental Farm - Central Park	9442.611328
Crestview - Meadowlands	8478.933593
Emerald Woods - Sawmill Creek	5160.012207
Glebe - Dows Lake	10883.41602
Golden Triangle - Old Ottawa East - Ottawa South	13065.55762
Greenboro East	10210.78711
Hawthorne Meadows - Sheffield Glen	6474.240234
Hintonburg - Mechanicsville	9561.943359
Hunt Club - Ottawa Airport	4395.274902
Hunt Club East - Western Community	8562.541992
Iris	6831.413085
Island Park	5193.034179
Katimavik - Hazeldean	14781.79688
Lebreton Development	1.805603
Ledbury - Heron Gate - Ridgemont - Elmwood	13753.17969
Lindenlea - New Edinburgh	5318.980957
Lower Town	8172.361328
Orleans Village - Chateauneuf	12912.80859
Overbrook West - McArthur	11524.125
Pineview	5987.884277
Playfair Park - Lynda Park - Guildwood Estates	6049.057128
Qualicum - Redwood Park	4382.852539
Orleans Queenswood Heights	13499.16406
Riverside Park	4624.115722
Sandy Hill - Ottawa East - University of Ottawa *	10834.40918
Sheahan Estates - Trend Village - Arlington Woods	3908.858154
South Keys - Heron Gate - Greenboro West	3892.871582
Tanglewood	4933.922851
Vanier South	7008.798339
West Centre Town - Little Italy - Civic Hospital*	11918.46875
Westboro	9432.545898
Whitehaven - Queensway Terrace North	11297.91602
Woodvale - Craig Henry - Manordale - Estates of *	8669.117187

Woodroofe - Lincoln Heights	4369.362792
Crystal Bay – Lakeview Park – Britannia Village *	11929.57324
Rockcliffe - Manor Park	5292.277832
Canadian Army Base - NRC	5246.595703
Notre-Dame Cemetery	3.917568
Vanier North	9037.70996
Orleans North West	12276.73242
Industrial Orleans	156.981933
Industrial Hunt Club South	571.807434
Orleans Central	3618.042724
Hunt Club Park	8680.974609
Industrial East	2956.141113
Elmvale - Eastway - Riverview - Riverview Park W*	18463.92188
Cummings	8686.326171
Merivale Gardens - Grenfell Glen - Pineglen - Co*	2815.248046
Rothwell Heights - Beacon Hill North	10492.35156
Orleans Chatelaine Village	3923.380371
Glen Cairn - Kanata South Business Park	9167.977539
Hunt Club Woods - Quintarra - Revelstoke	5749.319824
Bridlewood - Emerald Meadows	19255.4043
Orleans Chapel Hill South	5560.021972
Blackburn Hamlet	8526.108398
Chapman Mills - Rideau Crest - Davidson Heights	16142.30273
Greenbelt	2102.387695
Kanata Lakes - Marchwood Lakeside - Morgan's Gra*	19403.78516
Orleans Avalon - Notting Gate - Fallingbrook - G*	34853.42578
Riverside South - Leitrim	8789.961914