How and when Scale Matters: The Modifiable Areal Unit Problem and Income Inequality in Halifax

Victoria Prouse
School of Planning
Dalhousie University

Howard Ramos
Department of Sociology and Social Anthropology
Dalhousie University

Jill L. Grant
School of Planning
Dalhousie University

Martha Radice
Department of Sociology and Social Anthropology
Dalhousie University

Abstract
Methods for analyzing Canadian neighbourhoods have developed principally from studies of large cities such as Montreal, Toronto, and Vancouver. Statistical analyses on neighbourhood change in Canada, moreover, have adopted Census Tracts (CTs) as their main geographic units of analysis. However, using CTs as a proxy for neighbourhoods in smaller cities may generate misleading conclusions because such units cover too large an area and potentially mask heterogeneity of populations living within them. This phenomenon is known as the Modifiable Areal Unit Problem (MAUP) and has been investigated by geographers. Data on material, social, and structural conditions of neighbourhoods in Halifax, Nova Scotia from the 2006 Canadian Census are explored to examine the degree of the MAUP in this smaller city and to assess the usefulness of Dissemination Areas as an alternative unit of analysis for small cities. We also offer insight on how the MAUP affects analysis and make suggestions as to how planners can adjust their analyses with this in mind.

Keywords: neighbourhood, modifiable areal unit problem, MAUP, Halifax, inequality, census geography

Canadian Journal of Urban Research, Volume 23, Issue 1, Supplement pages 61-82.

ISSN: 1188-3774

Résumé
Les méthodes pour analyser les quartiers urbains au Canada se sont principalement développées à partir d'études sur les grandes villes telles que Montréal, Toronto et Vancouver. Les analyses statistiques des quartiers en mutation adoptent généralement le secteur de recensement (SR) comme l'unité d'analyse géographique de base. Toutefois, l'utilisation de l'échelle du SR dans des villes de plus petite taille peut donner lieu à des conclusions erronées, parce que cette unité couvre une trop grande superficie, masquant ainsi l'hétérogénéité démographique à l'intérieur du SR. Ce phénomène connu en tant que problème de l'aire modifiable des unités analysées (*modifiable areal unit problem* ou MAUP) a attiré l'attention des géographes. En interrogeant les données du recensement canadien de 2006 sur les conditions matérielles, sociales et structurelles des quartiers d'Halifax, en Nouvelle-Écosse, nous explorons le degré du MAUP dans cette ville de la région atlantique et nous testons l'utilité de l'aire de diffusion (AD) comme unité d'analyse alternative dans les villes à petite échelle. Nous offrons un aperçu des effets du MAUP sur l'analyse ainsi que des suggestions quant à la façon dont les urbanistes peuvent tenir compte des impacts du MAUP dans leur travail.

Mots clés: quartier, problème de l'aire modifiable des unités analysées, MAUP, Halifax, inégalité, géographie du recensement

The concept and meaning of neighbourhood presents a major dilemma for urban scholars (Chaskin 1997; Germain and Gagnon 1999), especially for those trying to understand how areas change over time, or for urban practitioners eager to identify the appropriate scale for interventions. The problem, according to Nicotera (2007, 29), lies in the need to identify objective variables and units aligned to neighbourhoods in order to assess the subjective experience of living in communities. In other words, to study neighbourhoods and changes in them, we need a clear understanding of what we are basing the notion of "neighbourhood" on, as well as a stable unit of analysis.

For the most part, Canadian researchers using statistical methods and GIS rely on Census Tracts (CTs) as the primary measure for defining a neighbourhood (Ley 1988; Ley and Dobson 2008; Cities Centre 2010; Ley and Lynch 2013; Bell et al. 2013). Most Canadian research on neighbourhoods and neighbourhood change has focused on the country's three largest cities—Toronto (Slater 2004; Walks and Maaranen 2008; Skaburskis 2012), Montréal (Langlois and Kitchen 2001; Rose and Twigge-Molecey 2013), and Vancouver (Ley and Lynch 2012). In cities with large populations and extensive districts of similar types of income or household characteristics, CTs—or groups of adjacent CTs—make sense as the unit of analysis that serves as a proxy of neighbourhood. In larger cities, CTs have remained relatively stable over the last few decades, facilitating the study of neighbourhood change. In smaller cities, however, CTs may exhibit considerable internal diversity because aggregating a population in the range that Statistics Canada prefers for the CT (2500 to 8000 inhabitants) involves encompassing a mix of housing types and household characteristics. Smaller cities may also have less stable CTs over time because population change through urban infill or suburban growth may lead Statistics Canada to change the boundaries of units: modifications to CT boundaries complicate long-term comparisons for those relying on CT-level data.

Because CTs in smaller cities are relatively large and can be unstable over time, urban planners face challenges when trying to understand long-term trends and their sociospatial dynamics. For instance, socioeconomic differences may be masked, making CTs appear more homogeneous than they actually are. As such, differences among neighbourhoods in small cities often appear less economically unequal than those in larger urban centres. In part this is because a narrowing of variability due to smaller population. We contend, however, that this is also likely related to the unit of analysis used in studies to operationalize neighbourhoods. Researchers (e.g., Gehlke and Biel 1934; Openshaw 1984; Flowerdew 2011; Nthiwa 2011; Bell et al. 2013) argue that such faulty conclusions occur because of the Modifiable Areal Unit Problem (MAUP), which results when administrative measures of an area are assumed to align with neighbourhoods but mask the internal diversity and heterogeneity within them.

Due to its spread-out geography and relatively small population Halifax Regional Municipality faces particular issues with the MAUP. The city's population is 390,328 distributed across a territory of 5,495.71 square kilometres (Statistics Canada, 2012a), leading to some very large CTs and an average population density well below that of Statistics Canada's definition of an 'urban area' (which is 400 people per km^2). Even in the urban core, densities remain relatively low and CTs relatively large. By contrast, the Toronto Census Metropolitan Area (CMA) has a population of 5,583,064 distributed across a territory of 5,905.71 square kilometres (Statistics Canada, 2012b). This means Toronto has a much higher population density per CT (945/km^2) than Halifax (71/km^2). Large CTs in cities like Halifax thus contain multiple areas of diverse composition that are hidden because of the choice of geographic unit used for analysis.

Applying Hulchanski's (2010) "Three Cities" model illustrated the obstacles of using CTs to effectively analyse socioeconomic conditions in Halifax (Prouse et al. 2014). The model examined the evolution of income inequality and polarization in Toronto's neighbourhoods, as defined by CTs, and identified three distinct trajectories: neighbourhoods becoming wealthier, those remaining about the same, and others getting poorer. In Toronto, the model revealed spatial concentrations of increasing wealth, relative stability, and increasing poverty, hence the "three cities." When applied to Halifax, results were ambiguous as increasing wealth or poverty appeared more evenly distributed throughout the city (Prouse et al. 2014). This was largely because CTs were poor proxies for neighbourhoods, since their geographic boundaries changed too dramatically over time and were also geographically too large, at times encompassing rural, suburban, and urban areas. Pampalon et al. (2009) offer a potential alternative to using CTs as geographic units of analysis. Instead of CTs, they use Dissemination Areas (DAs) in their sociospatial analysis of material and health outcomes. This is an approach that has been taken up by a number of others as well (such as Townshend 2002; Bell et al. 2013). DAs are the smallest geographic units available for analysis in public use data provided by Statistics Canada. Whether the dissemination area is a better unit of analysis for examining social and spatial patterns than the census tract remains worth investigating.

In this paper we aim to evaluate how and when the scale of geographic units of analysis matters in understanding a smaller city. We first ask whether the MAUP is present in Halifax and then consider whether DAs are better geographic units of analysis for smaller cities. We systematically compare the differences in results when using CTs versus DAs in basic descriptive statistics, GIS plotting, and linear regression analysis. Our overarching goal is to assess the effect geographic units have on making policy decisions and to make recommendations about how units matter and when different scales should be applied. We begin with a literature review on the MAUP, then describe the statistical and spatial methods we use to analyze the phenomenon. Subsequently, we examine how analysis using census data aggregated at the CT and DA levels affects interpretation of the results generated from each method. We conclude with insights into how and when the scale of analysis matters in the study of neighbourhoods.

How Urban Researchers have Measured Neighbourhoods

Urban studies scholars recognize the methodological difficulties of measuring social and economic conditions and changes in trajectories of neighbourhoods (Germain and Gagnon 1999; Coulton et al. 2001). Yet scholars have not always challenged whether spatial units for which we can easily obtain data appropriately represent neighbourhoods (Ross et al. 2004; Bell et al. 2013). O'Brien (1990) suggests that researchers generally fail to assess critically the reliability of the boundaries, or the units of analysis, that are used to construct neighbourhoods for purposes of analysis.

Studies commonly use pre-established geographic units comprising aggregated administrative data to define neighbourhoods and to set spatial parameters. Flowerdew et al. (2008) caution that clusters within such geographic units typically exhibit internal heterogeneity. They argue that smaller-scale geographic units provide the strongest evidence of contextual effects at the neighbourhood level and are more homogenous than larger units.

The use of CTs as spatial surrogates for neighbourhoods was employed as early as 1910 in selected cities in the US but it was not until the 1940 US Census that CT data became part of standardized tabulations. Analytical approaches to social differentiation among CTs was advanced through the early work social area analysts such as Shref Shevky and Marilyn Williams as well as Wendell Bell (Bell 1953) and subsequently through a long tradition of factorial ecology studies, both in the US and Canada (Murdie 1969; Davies 1984; Davies and Murdie 1993; Townshend 2002). Today, Canadian researchers often rely on CTs as their geographic unit of analysis for neighbourhoods (see Walks and Maaranen 2008; Ley and Dobson 2008; Walks 2010, 2011; Bell et al. 2013). Statistics Canada defines CTs as neighbourhood-like areas comprising a population between 2,500 and 8,000 people, which are "as homogeneous as possible in terms of socioeconomic characteristics, such as similar economic status and social living conditions at the time of … creation" (Statistics Canada 2012c, remarks section). Data for smaller geographic units are also available but are used less frequently by researchers. One such alternative, DAs, are administrative units defined within CTs: they are the smallest geographic area available in Canadian "public use" census data and a number of researchers have turned to them as an alternative to CTs (c.f. Townshend 2002; Pampalon 2008; Bell et al. 2013). The boundaries of DAs follow distinctive features such as roads or waterways. Their target population range is 400 to 700 people, a figure large enough to ensure the privacy of those enumerated during statistical analysis. DAs tend to vary in surface area and therefore also vary in population density. DAs in suburban and rural areas tend to be larger and encompass a greater variety of land uses than urban DAs (Riva et al. 2008). One obstacle to using DAs is that some have zero population because they capture industrial areas (Statistics Canada, 2012d). Another issue to consider is that they are also prone to misrepresenting subsets of the population because of Statistics Canada's policy to round values of variables to multiples of five (Bell et al. 2013, 91). Even so, these smaller geographic units are potentially more socially and economically homogenous than CTs, especially in smaller cities.

Traditional neighbourhood studies that rely on political-administrative boundaries like CTs have crucial limitations. Raudenbush and Sampson (1999) caution that the appropriate scale of analysis differs for each social and demographic variable analysed (1999, 4). Chaskin (1997) also notes problems associated with boundary construction in studies dealing with neighbourhood as a spatial unit. He shows that empirical findings derived from administrative boundaries diverge from the way people actually live in neighbourhoods and do not account for the experiences of different demographic groups within them either. Chaskin (1997, 521) observes, "the delineation of neighbourhood boundaries is a negotiated and imperfect process, often driven by political considerations". The units used to study neighbourhoods often reflect bureaucratic needs and politics but not residents' on-the-ground experiences.

For these reasons, social scientists have been criticized for failing to critically interrogate the geographic units used in their analyses. Openshaw (1984, 4) contended that the "principal criteria used in the definition of these units are the operational requirements of the census… [and the] …choice of these units is often haphazard, in that considerations such as convenience rather than geographical meaning are paramount." Openshaw was the first to systematically investigate the MAUP, identified by Gehlke and Biel (1934). He understood the MAUP as an ecological fallacy yielding data discrepancies. Openshaw showed that the problem of the MAUP challenges the validity of neighbourhood studies because the geographic units selected for analysis significantly influence portraits of social and economic differences across a city. Discrepancies emerge because different scales of units of neighbourhoods change the contours of social and economic patterns. Census data are "collected for essentially non-modifiable entities (people, households) [but] they are reported for arbitrary and modifiable areal units (enumeration districts, wards, local authorities)" (Openshaw 1984, 4). The discrepancy creates opportunities for misinterpreting results.

Researchers who investigate the MAUP argue that two main forces—the scale effect and the zoning effect—contribute to a significant loss of information in the aggregation of data in large units of geography (Nthiwa 2011; Bell et al. 2013). The scale effect is found when empirical results change because of the use of different scales of data aggregation (Flowerdew 2011; Bell et al. 2013, 89). Openshaw (1984) argued that an increase in the absolute values of correlations occur as the number of observations decrease, providing evidence that data are subject to the scale effect. The zoning effect emerges when the geographical units of analysis within a study area change shape. This can occur, for instance, when CT boundaries shift over time. Flowerdew (2011) contends that the scale effect usually has greater influence on the overall impact of the MAUP than the zoning effect because there is a greater chance that different scales of units of analysis (rather than different shapes) contain significantly different data. As the geographic unit becomes smaller, a homogeneous population becomes more likely. Thus, smaller geographic units may be a better measurement of neighbourhoods because they contain less variation. Overall, when the MAUP occurs it creates artificial spatial patterns because of information loss (Hayward and Parent 2009).

Through a study of aggregated census data in the Buffalo Metropolitan Area, Fotheringham and Wong (1991) showed the dramatic consequences of the MAUP on statistical analysis of socioeconomic conditions and trajectories. Using linear regression to analyze the spatial distribution of average family income, they found significant differences in conclusions derived depending on the scale of geographic units used to capture trends in neighbourhoods. An 800-unit geographic dataset showed that a 10% increase in the proportion of the city's elderly population predicted a decrease in the average family income of $308. When data are aggregated to 25 units, however, they found that a 10% increase yielded a decrease in predicted mean family income of $2,654. Contradictory findings show clearly that the scale of the unit matters to researchers interested in identifying patterns of inequality.

Most of the MAUP literature shows that statistical estimates, such as variance and standard deviations, decline with an increase in aggregation of geography. Consequently, descriptive statistics show information loss through data smoothing when larger scales are used (Gehlke and Biehl 1934; Openshaw 1984; Fotheringham and Wong 1991; Wong et al. 1999). For example, in his study of Istanbul, Nthiwa (2011) revealed considerable variation in standard deviations for socioeconomic variables aggregated at neighbourhood and district scales. Some studies, like Flowerdew (2011), use bivariate correlations to compare relationship direction and magnitude of effect between variables at the ward and district scales. Other studies have shown dramatic differences in various forms of regression estimates because of difference in the scale of aggregation used for geographic units (Amrhein and Flowerdew 1989; Wong et al. 1999; Manley 2006; Krupka 2007; Flowerdew et al. 2008; Pawitan and Steel 2009; Shah et al. 2014).

Fotheringham and Wong (1991) argue that the geographic and urban planning literatures insufficiently acknowledge the consequences of the MAUP. They lament that:

> Feeding census data into canned multiple regression programs is still a common practice and many of these applications are used to formulate urban policy. It is still rare to find references to the MAUP in textbooks which advocate regression analysis for policy formulation and even in texts on spatial analysis and spatial statistics (Fotheringham and Wong 1991, 1029).

Failing to consider the impact and definition of geographic units of analysis could lead to inaccurate results and policies.

Many researchers investigating the MAUP have concentrated on how it affects the analysis of socioeconomic segregation, racial segregation, and deprivation in cities. These are also key economic and social dimensions used by non-profit and government agencies to identify priority neighbourhoods requiring additional resources and services. Others, more recently, have looked at access to health. Thus, recognizing the effects of the MAUP is crucial for urban planning, social work, environment studies, and public health. Understanding how geographic scale affects analyses can shed light on which geographic unit of analysis is most appropriate.

Some researchers believe that the MAUP is particularly problematic for smaller cities. Krupka (2007) establishes a crucial link between city size, presence of the MAUP, and measures of economic inequality and racial segregation. His research examined the MAUP's significance in neighbourhood change studies by comparing cities of different sizes and by examining how different scales of aggregation affected analyses. Krupka showed that levels of racial segregation were similar in small and large cities but at different scales of geographic aggregation. In larger cities, CTs were relatively racially and economically homogeneous. In smaller cities, however, CTs contained considerable variation: neighbourhoods were often too small to be accurately captured by that unit of analysis (Krupka 2007, 188). Wong et al. (1999) came to similar conclusions in their analysis of spatial scales and zonal configurations in 30 American cities: they found that the scale effect was smallest in Los Angeles and largest in Newark. The smaller city, Newark, had the weakest fit between larger geographic units of analysis and on-the-ground measures of racial segregation. In a study of different approaches to defining neighbourhood boundaries in assessing socioeconomic characteristics, Lebel et al. (2007) showed that the MAUP's scale and zoning effects arise more frequently and in greater magnitude in smaller municipalities and rural areas than in large cities.

Although several geographers have examined data aggregation at varying scales of spatial units of analysis in Canada (Townshend 2002; Schuurman et al. 2007; Lebel et al. 2007; Mitra and Biuliung 2012), and some have investigated the MAUP (Bell et al. 2013; Shah et al. 2014), few have looked at the country's smaller cities. A small number of exceptions include Bourne and Barber (1971) who examined smaller centres in Ontario and Quebec as well as Shah et al. (2014) who look at Calgary, or Davies and Murdie (1993) who compare across a large number of Canadian cities. None to our knowledge, however, have looked at issues of geographic unit of analysis by focusing specifically on an Atlantic Canadian city. Much of the literature, moreover, considers aspects of measurement, and does not translate how that affects applied use of measurement for urban planners. Halifax's small population and reputation for relatively low levels of inequality compared with larger Canadian municipalities makes it an ideal case to examine. The remainder of the article examines how the MAUP affects analyses of sociospatial conditions in Halifax and offers insight on how and when the MAUP affects urban planners' applied work.

Methods: How We Measure Neighbourhood Differences Across Scales

Like other studies of the MAUP, we focus on how the analysis of individual-level socioeconomic data derived from the census is affected by using CTs versus DAs as a unit of analysis that serves, ultimately, as a proxy for neighbourhood. We also explore how it affects three types of analysis commonly used by academics, the non-profit sector, and government policy-makers: that is, we employ descriptive statistics, GIS mapping, and linear regression. If the MAUP is present, the aggregated values of measures will differ according to broader or finer geographical scales. Our analysis examines the problem through a focus on the differences between relationships and characteristics portrayed at the CT and DA.

We use data from the 2006 Canadian Census, accessed through the Canadian Census Analyzer and Statistics Canada's GeoSuite. We selected the 2006 Census because the 2011 Census asked fewer questions than in 2006 and the supplementary National Household Survey (NHS), which replaced the long form census, has been criticized for its poor data quality. Statistics Canada (2013, 2014) has noted the lack of comparability between the 2011 NHS data and 2006 Census data, especially for small geographic areas and residents of low-income neighbourhoods. Likewise, analysis of comparability and accuracy of the NHS for measures of economic inequality have been shown to yield poor estimates (Hulchanski et al. 2013).

In order to examine the MAUP, we follow Pampalon et al. (2009), who constructed a material and social deprivation index for Canada in order to facilitate sociospatial analysis and planning around health outcomes. Like others (Townshend 2002; Bell et al. 2013), their research used DAs from the census as a unit of analysis. We use the same material and social dimensions in our analysis of the MAUP but add a third dimension to incorporate structural characteristics. The dimensions and variables we selected are commonly utilized for policy analysis and can also be used to elucidate and compare differences in estimates of measures across spatial scales. For the material dimension, we look at the percentage of residents classified as low-income by Statistics Canada to elucidate the severity of relative poverty in the city (Hagaaners 1988). We measure this with the Low Income Cut Off (*LICO).* We also examine the rate of *employment* for those over the age of 15, the percentage of people over 25 without a high school diploma (*no high school)*, and average *income* of those 15 and over. For the social dimension, we examined the percentage of people who were *separated, divorced or widowed*, the percentage of individuals *living alone*, percentage of economic families classified as *single parent*, and the proportion of *visible minorities*. With respect to the structural dimension, we examine the percentage of private dwellings *owned* and dwelling *density*, which is the total number of private dwellings divided by the total land area in square kilometers. Adding the structural dimension to the analysis allows us to gain an understanding of ownership rates and the spatial patterning of housing stock and ownership rates within census tracts. Each of these measures was captured at the CT and then DA levels to assess differences between the scales.

Our analysis examines estimates and plotting of the material, social, and structural variables across the 87 CTs and 568 DAs of Halifax Census Metropolitan Area. Because of Statistics Canada's data suppression policies to ensure privacy, some DAs lacked values for *LICO, employment, no high school, income, living alone, lone parent* *families, visible minorities,* and *owned* dwellings. As a result, some of our analysis uses fewer DAs: we note each time this occurs.

We begin our analysis by examining descriptive statistics to understand how geographic scales affect means and the range of values for the different material, social, and structural variables. We follow Flowerdew (2011) in calculating the Coefficient of Variation (CoV) for each indicator. The CoV provides a standardized measure that allows comparison of the degree of differences between values produced using CTs and DAs across indicators with varying units. For example, we can compare the variation in employment rates, which are in percentages, with average individual income values, which are in dollars. We then divide the DA CoV by CT CoV for each indicator. This creates a ratio where 1.0 indicates that the dispersion is the same across DAs and CTs for the indicator. Values larger than 1.0 imply greater variability within a CT. Scores far above 1.0 indicate that CTs are masking heterogeneity occurring at the smaller scale, because the MAUP effect is present.

We follow the analysis of descriptive statistics with a spatial analysis of *LICO* and *visible minorities* using GIS to map these characteristics for CTs and DAs in Halifax. Such maps are often used by non-profits and government policy makers to visually understand sociospatial trends in a given city. We examine *LICO* and *visible minorities* in detail because the MAUP literature tends to focus on patterns of economic and/or racial segregation. We categorize each CT or DA according to the incidence of the given variable compared to the CMA average: very low, low, middle, high, and very high. If the MAUP is present, the spatial patterns revealed in the maps at the two scales will differ. The DA plotting should show heterogeneity that is missed when using the larger geographic unit of analysis.

We also use linear regression analysis to examine how the MAUP affects multivariate relationships. As Fotheringham and Wong (1991) point out, various forms of regression (often linear) are commonly used by researchers and urban planners to determine levels of economic and social inequality among neighbourhoods. Our models regress *LICO* on the other variables, except *income*. We removed average individual *income* from the linear regression because a Variance Inflation Factor (VIF) statistic revealed that it was too closely related to the other independent variables, leading to colinearity. All other variables were within acceptable levels of VIF. Overall, we hoped to determine whether the MAUP affected the analysis of socioeconomic measures for Halifax, a small Canadian city.

How and When Scale Matters

We begin by examining estimates for various measures of material, social, and structural dimensions of neighbourhoods. Table 1 compares descriptive statistics at the CT and DA levels. The analysis shows slight differences in mean values and greater variation in the range of values and standard deviation between the two scales of geography. The range at the DA level is considerably greater than at the CT level, providing some evidence that CT level aggregation mutes extreme values and hides social and economic polarization present at the DA level.

Standard deviations are smaller at the CT level than those at the DA level. Differences are also seen in the CoV. All CoV ratios for material, social, and structural

Table 1: Differences between CTs and DAs as Measures of Neighbourhoods

Variables	Minimum		Maximum		Mean		Standard Deviation		Coefficient of Variation		
	CT*	DA**	CT	DA	CT	DA	CT	DA	CT	DA	DA/CT
Material											
LICO	1.9	0	37.7	61.2	11.81	10.67	8.69	11.42	0.74	1.07	1.45
Employment	41.8	4.7	76.8	94	64.11	63.68	6.93	10.77	0.11	0.17	1.57
No high school	1.73	0	44.59	66.67	15.62	15.91	7.97	10.7	0.51	0.67	1.32
Income	20,099	0	72,476	122,987	35,355	34,625	8,966	12,005	0.25	0.35	1.37
Social											
Separated, divorced or widowed	7.49	1.98	34.34	66.41	16.92	16.87	5.02	6.95	0.3	0.41	1.39
Living alone	2.14	0	42.86	66.67	12.76	12.3	8.74	10.15	0.68	0.83	1.21
Single parent	3.85	0	37.5	71.43	17.17	17.06	7.85	12.08	0.46	0.71	1.55
Visible minorities	0.6	0	32.18	85.88	7.39	7.16	5.74	10.07	0.78	1.41	1.81
Structural											
Owned	3.7	0	98.35	101.96***	65.18	68.5	26.72	30.72	0.41	0.45	1.09
Dwelling density	0.99	0.51	5,229.17	32,000	835.37	1,425.26	946.58	2,265.48	1.13	1.59	1.4

* n=87

** Statistics Canada suppresses data for particularly small Dissemination Areas to protect residents' privacy. Therefore, the number of DA cases available for analysis varies by indicator. For all indicators, n=568 with the exception of SDW and DWELDEN where n=569 and LPF where n=567.

*** According to Statistics Canada all counts except for population and dwelling counts are subject to random rounding, "transform[ing] raw counts to random rounded counts" (Statistics Canada, 2013). Thus, in rare cases, the aggregate and individual category counts are rounded in a way that creates a value of over 100% of the categorical count existing within the aggregate. As such, these values were left unchanged for analysis (P. Griffith, Personal Communication, October 29, 2013).

indicators are greater than one. High CoV ratios reveal heterogeneity of socioeconomic conditions at the DA level that would be hidden by the CT scale. These consistent results reveal that DAs have greater variation around the CMA average than CTs. At the CT level, extreme anomalies do not increase the magnitude of the standard deviation. The percent of *owned dwellings* has the smallest CoV ratio. It has only 9% more variation in the proportion of *owned dwellings* when data are aggregated at the DA level compared to when they are aggregated at the CT level. The percentage of *visible minorities* has the greatest CoV ratio (1.81), which means that there is 81% more variation in the proportion of visible minorities in a neighbourhood defined by the DA level compared to the proportion found at the CT level. The data suggest that racial segregation in Halifax occurs at a smaller geographic scale than would be captured by studies of CT level data. The result is consistent with existing MAUP literature examining other cities. The DA level thus offers a more nuanced portrait of the spatial patterning of socioeconomic conditions. Analyses of descriptive statistics and CoV ratios confirm that CT level aggregation masks socioeconomic differences found at the DA level.

Examining *LICO* and *visible minorities* spatially through GIS mapping reveals interesting differences in levels of geographic units of analysis. In Figures 1a and 1b, CTs reveal strong clustering patterns of very high *LICO* emerging in the city's urban core and very low *LICO* emerging in the surrounding areas. DAs by contrast show greater variation of *LICO* levels within CTs. DAs also reveal polarized adjacencies resulting in moderated aggregate values at CT level. The phenomenon occurs when areas of very low and very high *LICO* are in close proximity. These differences reflect the finding of other studies of the MAUP looking at patterns of economic segregation.

Figures 2a and 2b show large discrepancies in the percentage of *visible minorities* in an area between CT and DA units of analysis. The overall CMA average of this variable is quite small, at 7 % of the population, so small differences can magnify effects. With that noted, in the CT map, the city looks more racially diverse than it actually is. Large portions of the city's peninsula show high concentrations of visible minorities at the CT level. However, when the data are plotted for DAs, the concentration for most "neighbourhoods" is much smaller, with some DAs showing high populations of visible minorities. The disparity is especially pronounced on the city's peninsula. These findings align with the existing MAUP literature on racial segregation in other cities.

Finally we examine the influence of the MAUP by looking at multivariate relationships. Table 2 reports results of two linear regression models of LICO on other measures at the CT and the DA levels. At the CT level, the R^2 value is 0.90, versus 0.68 at the DA level, accounting for a large amount of variation in LICO. Adjusted R^2 values are also higher at the CT level. It appears that CT offers a better model fit than the DA.

Figure 1a

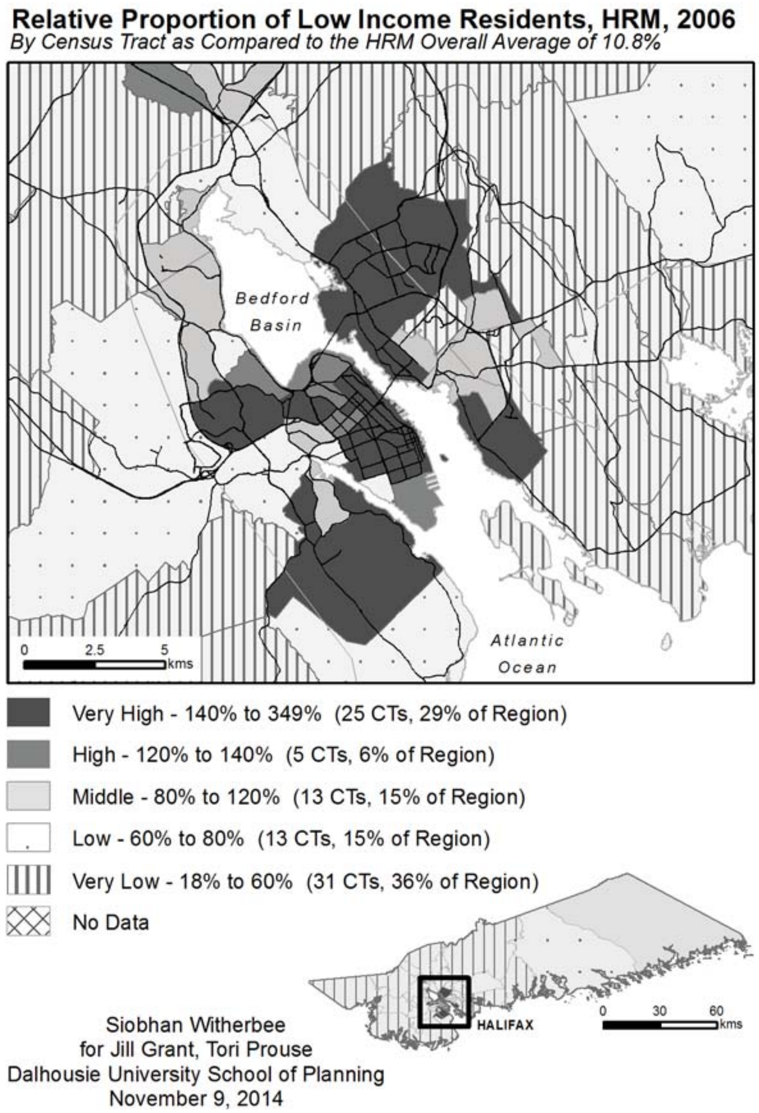
Relative Proportion of Low Income Residents, HRM, 2006
By Census Tract as Compared to the HRM Overall Average of 10.8%
Bedford
Basin
Atlantic
Ocean
0 2.5 5 kms
Very High - 140% to 349% (25 CTs, 29% of Region)
High - 120% to 140% (5 CTs, 6% of Region)
Middle - 80% to 120% (13 CTs, 15% of Region)
Low - 60% to 80% (13 CTs, 15% of Region)
Very Low - 18% to 60% (31 CTs, 36% of Region)
No Data
HALIFAX
0 30 60 kms
Siobhan Witherbee
for Jill Grant, Tori Prouse
Dalhousie University School of Planning
November 9, 2014

Figure 1b

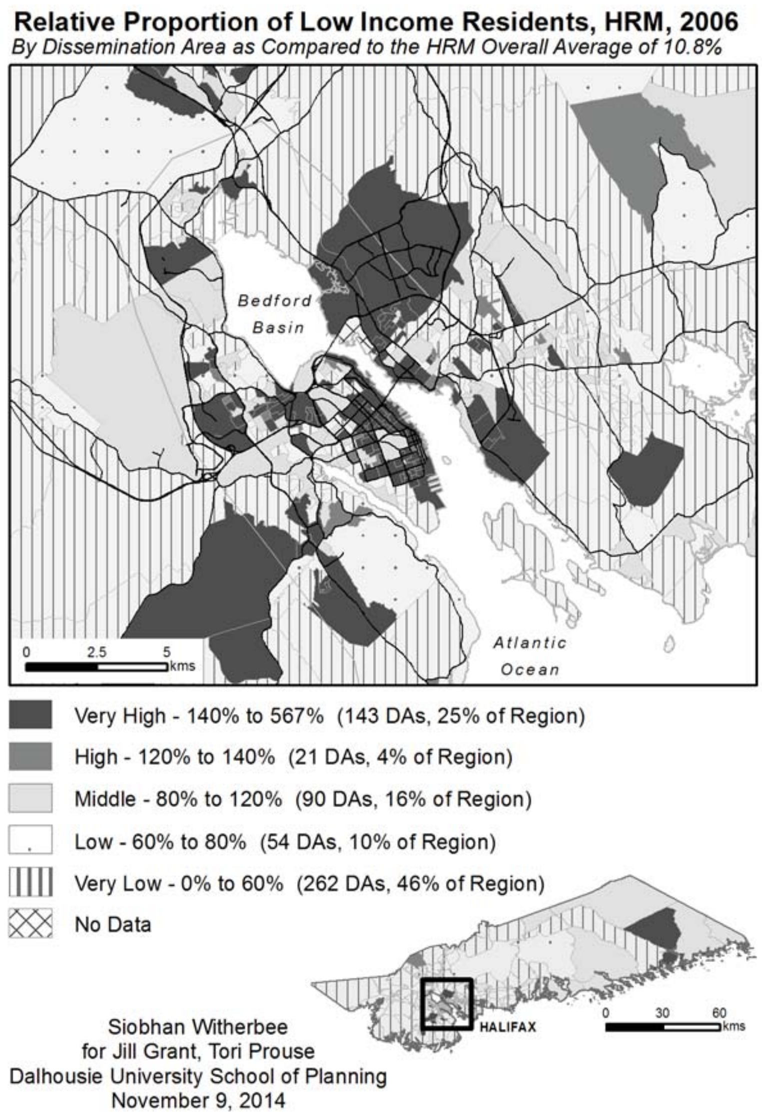
Relative Proportion of Low Income Residents, HRM, 2006
By Dissemination Area as Compared to the HRM Overall Average of 10.8%
Bedford Basin
Atlantic Ocean
0 2.5 5 kms
Very High - 140% to 567% (143 DAs, 25% of Region)
High - 120% to 140% (21 DAs, 4% of Region)
Middle - 80% to 120% (90 DAs, 16% of Region)
Low - 60% to 80% (54 DAs, 10% of Region)
Very Low - 0% to 60% (262 DAs, 46% of Region)
No Data
HALIFAX
0 30 60 kms
Siobhan Witherbee
for Jill Grant, Tori Prouse
Dalhousie University School of Planning
November 9, 2014

Figure 2a

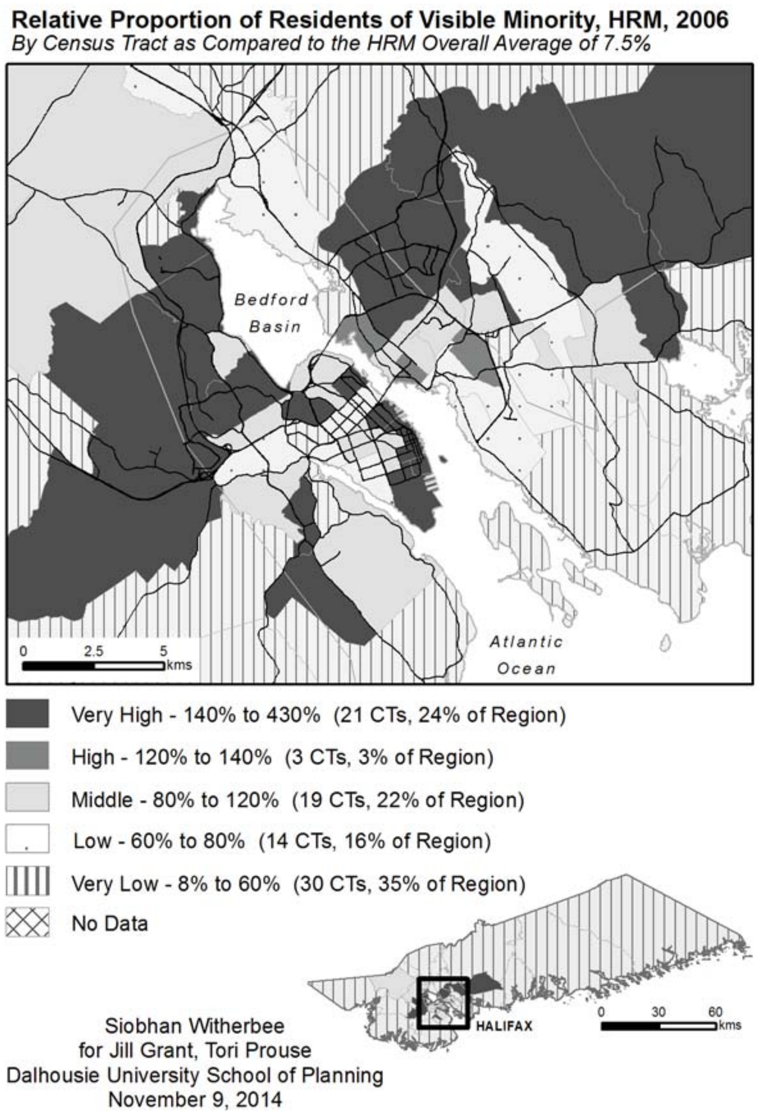
Relative Proportion of Residents of Visible Minority, HRM, 2006
By Census Tract as Compared to the HRM Overall Average of 7.5%
Bedford Basin
Atlantic Ocean
0 2.5 5 kms
Very High - 140% to 430% (21 CTs, 24% of Region)
High - 120% to 140% (3 CTs, 3% of Region)
Middle - 80% to 120% (19 CTs, 22% of Region)
Low - 60% to 80% (14 CTs, 16% of Region)
Very Low - 8% to 60% (30 CTs, 35% of Region)
No Data
HALIFAX
0 30 60 kms
Siobhan Witherbee
for Jill Grant, Tori Prouse
Dalhousie University School of Planning
November 9, 2014

Figure 2b

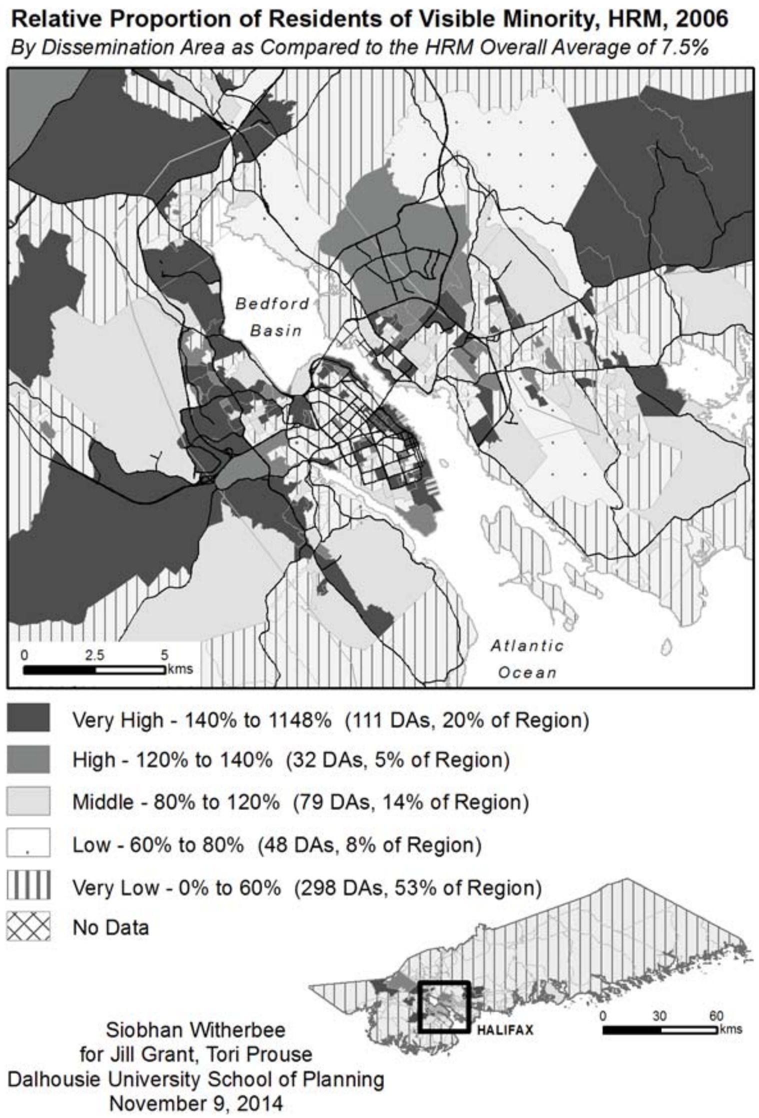
Relative Proportion of Residents of Visible Minority, HRM, 2006
By Dissemination Area as Compared to the HRM Overall Average of 7.5%
Bedford Basin
Atlantic Ocean
0 2.5 5 kms
Very High - 140% to 1148% (111 DAs, 20% of Region)
High - 120% to 140% (32 DAs, 5% of Region)
Middle - 80% to 120% (79 DAs, 14% of Region)
Low - 60% to 80% (48 DAs, 8% of Region)
Very Low - 0% to 60% (298 DAs, 53% of Region)
No Data
HALIFAX
0 30 60 kms
Siobhan Witherbee
for Jill Grant, Tori Prouse
Dalhousie University School of Planning
November 9, 2014

Table 2: Linear Regression of LICO on Material, Social and Structural Characteristics

Variables	Model 1 (CT)			Model 2 (DA)		
	Coef.	Std. Error	Beta	Coef.	Std. Error	Beta
Material						
Employment	-0.12	0.06	-0.09	-0.09	0.03	-0.08*
No high school	19.56	5.91	0.18*	20.95	3.34	0.20*
Social						
Separated, divorced or widowed	-70.28	14.64	-0.41*	-42.62	7.28	-0.25*
Living alone	31.38	10.28	0.32*	17.07	6.15	0.15*
Single parent	42.33	9.08	0.38*	21.71	3.36	0.23*
Visible minorities	1.88	7.8	0.01	6.26	3.30	0.06
Structural						
Owned	-16.62	3.16	-0.51*	-18.33	1.73	-0.49*
Dwelling density	0.00	0.00	0.13	0.00	0.00	0.14*
Constant	26.61	6.56		25.57	3.197	
n		87			567	
R^2		0.90			0.68	
Adjusted R^2		0.89			0.67	
F statistic		87.85			145.746	
p-value		0.00			0.00	

* Significant at the 0.01 level

At the CT level, the percentage of people without a *high school diploma*, the percentage of individuals *separated, divorced, or widowed*, the proportion of people *living alone*, the percentage of *single parent families*, and the *ownership* rate are all statistically significant. Each of these indicators—save the proportion of *separated, divorced, or widowed* people and the *ownership rate*—correspond with an increase in *LICO*. For DAs, all variables except percentage visible minorities reach statistical significance. The direction of the relationships is the same as with CTs. For some variables, when the size of the coefficients are examined CTs yield large estimates of effect, while for others DAs do. The effects of *no high school* diploma, *visible minorities*, and home *ownership* are all larger at the DA level. The greatest absolute difference is seen in *separated, divorced, or widowed* and the least is found in the proportion of those *employed*. When we examine beta, or standardized, coefficients we find that percentage of owned dwellings has the largest effect on LICO at both scales of geography: -0.51 at the CT level and -0.49 at the DA level. The proportion of *visible minorities* has the smallest effect, but is not statistically significant. When the absolute differences of beta coefficients are examined we find that the greatest differences are seen with *people living alone*, followed closely by *separated, divorced, or widowed*. The least difference is again found with those employed.

Overall, linear regression results show that the geographic unit of analysis, or scale of aggregation, affects how *LICO* is understood and the degree of relationships among variables. Although the direction of relationships does not change across geographic scales, we find that CTs offer a cleaner portrait of *LICO* with better model fits and fewer terms achieving statistical significance. DAs have weaker model fit, but more factors appear to influence *LICO* at this level. The scale of the unit of analysis used to capture a neighbourhood influences the size of effects on *LICO*. This is particularly the case for considering how the percentage of individuals *separated, divorced, or widowed* and percentage of individuals *living alone* affect *LICO*. Descriptive analysis and CoV ratios reveal that the MAUP is present in Halifax. Moreover, they show how the ecological fallacy—namely, that results depend on the spatial unit of analysis used rather than on the data contained within them—affects the precision of estimating outcomes. The MAUP does not, however, change overall conclusions on the main contributors to socioeconomic outcomes when linear regression analysis is used.

Conclusion

Our analysis of material, social, and structural measures of neighbourhoods in Halifax show that interpretations of socioeconomic conditions and trends are affected by the MAUP. Our findings are generally consistent with those of other studies assessing the implications of the MAUP (Fotheringham and Wong 1991; Krupka 2007; Schuurman *et al.* 2007; Bell et al. 2013; Shah et al. 2014). Descriptive statistics show that aggregation at the CT level masks variation and detail found in variables at the DA level. Spatial patterns revealed from mapping the variables reveal the same phenomenon. Linear regression analysis also showed differences in estimates, but did not show an improvement in model fit nor general conclusions between different geographic units of analysis. What are the implications of the results for researchers and policy makers aiming to analyse Halifax and other small cities?

On a theoretical level, the issue of whether the CT and DA level is more appropriate for conveying socioeconomic inequality and segregation in Halifax is complicated. Both scales of geography have benefits for understanding the complex nature of the phenomenon. CTs offer clearer depictions of general conditions and, in linear regression analysis, provide better model fits. Using CTs for analysis in smaller cities can help researchers and policy makers understand the general trends of what affects the material, social, and structural characteristics of the city. CTs can also help policy makers eliminate variables with smaller impacts on sociospatial trends. However, the cleaner portrait comes at a cost.

Although CTs offer a simplified account of sociospatial patterns in a city, especially for linear regression analysis, they mask the heterogeneity that occurs within the larger units. DAs offer greater nuance about what is going on in neighbourhoods and might be more useful for non-profit service providers and urban planners in smaller cities. Data at the DA level may better reflect the subjective experience of small-scale cities, in which variations in residents' circumstances are reproduced at a smaller scale than in large metropolises. Policy based on the analysis of conditions derived from larger geographic units of analysis potentially hides polarized adjacencies: that is, areas of extreme differences in the distribution of key variables. It can, for instance, make a city appear less polarized in terms of economic well-being, or more integrated in terms of racial composition. In Halifax, at least, the incidence of economic poverty and racial segregation appears as significant only at finer levels of geography. To overlook such adjacencies and nuances can lead to inefficient deployment of resources and worse, ignoring inequities that occur in smaller cities.

Consequently, our analysis suggests that smaller cities in particular should critically question the geographic units of analysis used to define neighbourhoods and should explore heterogeneities that might occur within these units. This is not to imply, of course, that neighbourhoods should only be defined as homogenous areas, nor that social diversity within neighbourhoods is undesirable. This conclusion is similar to that made by Shah et al. (2014), who saw conflicting findings between linear regression and GWR regression models in Toronto and Calgary. They agree the presence of the MAUP and its effects should be critically examined and caution that affects are likely different for each city. As noted at the outset, "neighbourhood" is a slippery concept, making it tricky to match the subjective experience or common sense definition of neighbourhoods to objective, readily available, and quantitatively measurable data. It may indeed be unwise to use the evocative term neighbourhood in this kind of study. Rather than affirming which scale best represents a "neighbourhood", we argue that the geographic unit of analysis needs to be treated with caution when using statistical analysis as a basis for place-based policy-making. If the focus of research is to best approximate what is occurring "on the ground," then mid-sized and smaller cities need smaller geographic units of analysis. If the focus is on estimating general patterns and assessing the most pressing factors affecting areas, then the larger units appear more useful. Researchers and policy makers need to be mindful of the information lost when they use larger spatial units. Our investigation of the implications of geographic scale in interpreting social and spatial conditions in Halifax offers researchers and policy makers useful insights they can engage in practice.

Acknowledgements

Funding for this research was provided by the Social Sciences and Humanities Research Council of Canada under a Partnership Grant led by J. David Hulchanski at the University of Toronto. The authors are grateful to Richard Maaranen and Siobhan Witherbee for assistance with data and mapping. We are also grateful to the comments offered by the anonymous reviewers and editor, all of which have made this article stronger.

References

Amrhein, C., and R. Flowerdew. 1989. The effect of data aggregation on a Poisson regression model of Canadian migration. In *Accuracy of Spatial Databases*, ed. M. Goodchild, S. Gopal 1989, 229-238. London: Taylor and Francis.

Bell, S., K. Wilson, L. Bissonette, and T. Shah. Access to primary health care: does neighborhood of residence matter? *Annals of the Association of American Geographers* 103 (1): 85-105.

Bell, W. 1953. The Social Areas of the San Francisco Bay Region. *American Sociological Review* 18(1):39-47.

Bourne, L.S. and G. M. Barber. 1971. Ecological Patterns of Small Urban Centres in Canada. *Economic Geography* 47 (Issue Supplement: Comparative Factorial Ecology): 258-265.

Chaskin, R. 1997. Perspectives on neighbourhood and community: A review of the literature. *Social Service Review* 71: 521-547.

Cities Centre. 2011. *Three cities of Toronto.* Neighbourhood Change Research Partnership. http://3cities.neighbourhoodchange.ca

Coulton. J. Korbin, T. Chan, and M. Su. 2001. Mapping residents' perceptions of neighbourhood boundaries: A methodological note. *American Journal of Community Psychology* 29(2): 371-383.

Davies, W.K.D., and R. Murdie. 1993. Measuring the Social Ecology of Canadian Cities. In *The Changing Social Geography of Canadian Cities.* Eds. L.S. Bourne, and D. Ley. 1993, 552–75. McGill-Queens Press.

Flowerdew, R. 2011. How serious is the modifiable areal unit problem for analysis of English census data? *Population Trends* 145: 102-114.

Flowerdew, R., D. Manley, and C. Sabel. 2008. Neighbourhood Effects on Health: Does It Matter Where You Draw the Boundaries? *Social Science Medicine* 66(6): 1241-1255.

Fotheringham, A., and D. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23(7): 1025-1044.

Gehlke, C., and H. Biehl. 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association Supplement* 29: 169-170.

Germain, A., and J.E. Gagnon. 1999. Is neighbourhood a black box? A reply to Galster, Metzger and Waite. *Canadian Journal of Urban Research* 8(2): 172-184.

Hagenaars, A., and K. de Vos. 1988. The definition and measurement of poverty. *The Journal of Human Resources* 23(2): 211-221.

Hayward, P., and J. Parent. 2009. Modeling the influence of the modifiable areal unit problem (MAUP) on poverty in Pennsylvania. *The Pennsylvania Geographer* 47(1): 120-135.

Hulchanski, J.D. 2010. *The Three Cities within Toronto: Income polarization among Toronto's neighbourhoods, 1970-2005*. Research Report, Cities Centre. http://neighbourhoodchange.ca/2011/05/12/research-paper-one/.

Hulchanski, J.D., R. Murdie, A. Walks, and L. Bourne. 2013. Canada's voluntary census is worthless. Here's why. *The Globe and Mail.* 4 October. http://www.theglobeandmail.com/globe-debate/canadas-voluntary-census-is-worthless-heres-why/article14674558/.

Krupka, D. 2007. Are big cities more segregated? Neighbourhood scale and the measure of segregation. *Urban Studies* 44(1): 187-197.

Langlois, A., and P. Kitchen. 2001. Identifying and measuring dimensions of urban deprivation in Montréal: An analysis of the 1996 census data. *Urban Studies* 38(1): 119-139.

Lebel, A., R. Pampalon, and P. Villeneuve. 2007. A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities of the Quebec City region. *International Journal of Health Geographers* 6(27): 1-15.

Ley, D. 1988. Social upgrading in six Canadian cities. *The Canadian Geographer* 32(1): 32-45.

Ley, D., and C. Dobson. 2008. Are there limits to gentrification? The contexts of impeded gentrification in Vancouver. *Urban Studies* 45(12): 2471-2498.

Ley, D., and N. Lynch. 2012. Divisions and disparities in Lotus-land: Sociospatial income polarization in Greater Vancouver, 1970-2005. *University of Toronto Cities Centre Research Paper 223.* http://neighbourhoodchange.ca/cities/vancouver/

Manley, D. 2006. *The modifiable areal unit phenomenon: An investigation into the scale effect using UK census data.* PhD Thesis. University of St. Andrews: St. Andrews, UK.

Mitra, R., and R. Buliung. 2012. Built environment correlates of active school transportation: Neighbourhood and the modifiable areal unit problem. *Journal of Transport Geography* 20: 51-61.

Nakaya, T. 2000. An information statistical approach to the modifiable areal unit problem in incidence rate maps. *Environment and Planning* 32(1): 91-109.

Nicotera, N. 2007. Measuring neighbourhood: A conundrum for human service researchers and practitioners. *American Journal of Community Psychology* 40: 26-51.

Nthiwa, A. 2011. Modeling scale and effects of the modifiable areal unit problem on multiple deprivations in Istanbul, Turkey. Master's Thesis: Faculty of Geo-Information Science and Earth Observation of the University of Twente, Enschede, The Netherlands. http://www.itc.nl/library/papers_2011/msc/upm/nthiwa.pdf

O'Brien, R. 1990. Estimating the reliability of aggregate-level variables based on individual-level characteristics. *Sociological Methods & Research* 18: 473-504.

Openshaw, S. 1984. *The modifiable areal unit problem.* Norwich: Geo Books.

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

Pawitan, G., and D. Steel. 2009. Exploring the MAUP from a spatial perspective. *Working Paper 20-09*. Centre for Statistical and Survey Methodology, University of Wollongong: Wollongong, Australia.

Prouse, V., J. L. Grant, M. Radice, H. Ramos, and P. Shakotko. 2014. *Neighbourhood change in Halifax Regional Municipality, 1970 to 2010: Applying the "Three Cities" Model.* http://theoryandpractice.planning.dal.ca/neighbourhood/working-papers.html

Raudenbush, S., and R. Sampson. 1999. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighbourhoods. *Sociological Methodology* 29: 1-41.

Riva, M., P. Apparaicio, L. Gauvin, and J. Brodeur. 2008. Establishing the soundness of administrative spatial units for operationalizing the active living potential of residential environments: an exemplar for designing optimal zones. *International Journal of Health Geography* 7: 43-55.

Rose, D., and A. Twigge-Molecey. 2013. A city-region growing apart? Taking stock of income disparity in Greater Montréal, 1970-2005. Cities Centre, University of Toronto, Research Paper 22.

Ross, N. A., S. Tremblay, and K. Graham. 2004. Neighborhood influences on health in Montreal, Canada. *Social Science & Medicine* 59 (7): 1485–94.

Schuurman, N., N. Bell, J. Dunn, and L. Oliver. 2007. Deprivation indices, population health and geography: An evaluation of the spatial effectiveness of indices at multiple scales. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 84(4): 591-603.

Shah, T.I., L. Aspen, and S. Bell. 2014. Assessment of Choice of Units of Analysis for Studying Associations between Geographic Accessibility to PHC Services and Sociodemographic Factors. *Spatial Knowledge and Information - Canada* 2.

Skaburskis, A. 2010. Gentrification and the rent gap in the context of 'risk society'. *Environment and Planning A* 42(4): 895-913.

Skaburskis, A. 2012. Gentrification and Toronto's changing household characteristics and income distribution. *Journal of Planning Education and Research* 32(2): 191-203.

Slater, T. 2004. Municipally-managed gentrification in South Parkdale, Toronto. *The Canadian Geographer* 48(3): 303-325.

Statistics Canada. 2014. *Persons living in low-income neighbourhoods. http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-014-x/99-014-x2011003_3-eng.cfm*

Statistics Canada. 2013. Confidentiality (non-disclosure) rules. *Data quality and confidentiality standards and guidelines*. http://www12.statcan.ca/census-recensement/2011/ref/DQ-QD/conf-eng.cfm.

Statistics Canada. 2012a. *Focus on geography series, 2011 Census (Halifax)* Statistics Canada Catalogue no. 98-310-XWE2011004. Ottawa, Ontario. Analytical products, 2011 Census. Last updated October 24, 2012. https://www12.statcan.gc.ca/census-recensement/2011/as-sa/fogs-spg/Facts-cma-eng.cfm?LANG=Eng&GK=CMA&GC=205.

Statistics Canada. 2012b. *Focus on geography series, 2011 Census (Toronto)*. Statistics Canada Catalogue no. 98-310-XWE2011004. Ottawa, Ontario. Analytical products, 2011 Census. Last updated October 24, 2012. https://www12.statcan.gc.ca/census-recensement/2011/as-sa/fogs-spg/Facts-cma-eng.cfm?LANG=Eng&GK=CMA&GC=535.

Statistics Canada. 2012c. Census tract (CT). *Census dictionary.* http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo013-eng.cfm.

Statistics Canada. 2012d. Dissemination area (DA). *Census dictionary.* http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo021-eng.cfm.

Statistics Canada. 2012e. *Overview of the Census population.* http://www12.statcan.gc.ca/census-recensement/2011/ref/overview-apercu/pop1-eng.cfm#int1.

Statistics Canada. 2007. *GeoSuite. 2006 Census (Geography products: Geographic data products).* Statistics Canada Catalogue no. 92-150-XBB. Ottawa, Ontario.

Townshend, I.J. 2002. Monitoring Community Dimensions: City-Wide Characteristics and Differentiation by Social Region. In *Monitoring Cities: International Perspectives*, eds W.K.D. Davies and I.J. Townshend 2002, 435-459. Brussels: International Geographical Union, Urban Commission.

Walks, R. A. 2011. Economic restructuring and trajectories of socio-spatial polarization in the Twentieth Century Canadian city. In *Canadian regions: Trajectories* *of growth and change* ed. L. Bourne, T. Hutton, R. Shearmur, and J. Simmons. Toronto: Oxford University Press, 125-159.

Walks, R.A. 2010. New divisions: Social polarization and neighbourhood inequality in the Canadian city. In *Canadian cities in transition: New directions in the twenty-first century,* ed. T. Bunting, P. Filion, and R. Walker. Toronto: Oxford University Press, 170-190.

Walks, R. A., and L. S. Bourne. 2006. Ghettos in Canada's cities? Racial segregation, ethnic enclaves, and poverty concentration in Canadian urban areas. *The Canadian Geographer* 50(3): 273-297.

Walks, R. A., and R. Maaranen. 2008. The timing, patterning, and forms of gentrification and neighbourhood upgrading in Montréal, Toronto, and Vancouver, 1961 to 2001. *Centre for Urban and Community Studies Research Paper 211.* University of Toronto. http://www.urbancentre.utoronto.ca/pdfs/publications/RP211Walk-Maaranen-Gentrification1960-2001.pdf.

Wong, D. 1997. Spatial dependency of segregation indices. *The Canadian Geographer* 41(2): 128-136.

Wong, D., H. Lasus, and R. Falk. 1999. Exploring the variability of segregation indexes with scale and zonal systems: An analysis of thirty US cities. *Environment and Planning A* 31: 507-522.