Spatial Regression in Health: Modelling Spatial Neighbourhood of High Risk Population

Stefania Bertazzon
Department of Geography
University of Calgary
Calgary, Alberta, Canada
e-mail: bertazzs@ucalgary.ca

***Abstract*— Many health conditions affect certain individuals more than others: for example, adults over 65 years of age are more affected by cardiovascular disease than younger individuals. Therefore, the spatial pattern of the disease incidence can be modelled more effectively through the residential pattern of higher risk groups. The method is demonstrated through a spatial regression of the association of cardiac catheterization and socioeconomic determinants in Calgary (Canada). Over a 5-year interval, 45% of catheterizations are performed on seniors, that constitute 9% of the population. Seniors' residential location is therefore used as an auxiliary process to model the spatial weights of the regression model. This spatial model leads to a more realistic neighbourhood configuration, yielding more reliable regression estimates. Based on the residential location of the population at greater risk, the model presents low sensitivity to variations in the supporting geographic units. The use of a relevant auxiliary process is general and applicable to a range of conditions; it constitutes a promising alternative to the direct estimation of spatial parameters on the primary process. Overall, the spatial weights matrix based on at risk population shall increase the reliability of spatially autoregressive multivariate epidemiological models.**

Keywords- health geography; spatial regression analysis; spatial correlation; cardiovascular condition; cardiac catheterization; Seniors; risk population; residential location.

I. INTRODUCTION

Geographic information science (GIS) has been increasingly employed in population health research due to its ability to analyze interactions of health determinants in space [12], [26]. This has furthered the integration between geography and health sciences, promoting the development of more effective spatial analytical methods [13], whose reliable results can be translated into policy decisions [9], [27].

Most geographical phenomena, e.g., disease prevalence and population distribution, exhibit variations across space and self-similarity over short distances. These properties, known as spatial dependence and non-stationarity [24], are known to hamper the reliability of analytical models, by increasing the uncertainty of the estimated parameters [2]. In their presence, analytical models may lead to ineffective, or even harmful, health policy decisions. Spatial analytical methods offer a valid response to this problem; however, their ability to improve the model reliability [2], [3] depends on the representation of spatial interactions embedded in the model.

Health and its determinants interact in space [17]; hence, these interactions can be modelled by multivariate regression [2]. While local analytical methods [15] are concerned with spatial non-stationarities, spatial autoregressive methods [2] address the uncertainty stemming from spatial dependence. The specification of a spatially autoregressive model requires the definition of a neighbourhood of spatial units: the more accurate the neighbour definition, the more reliable the model estimates. Ideally, an accurate neighbourhood definition rests on a deep knowledge of the spatial process involved; more often one must estimate spatial dependencies using statistical methods, which are typically applied to the dependent variable [3], [8], [2]. In such situations, we propose the application of those statistical methods to another spatial process, which is related to the dependent, and which is better understood, if not within the health sciences, within geography or urban studies. The latter process effectively serves as an auxiliary process, in that it is used to estimate the spatial parameters that will provide a more realistic neighbourhood representation, enhancing the reliability of the regression model.

Here, a multivariate spatial regression model [2] analyses the socioeconomic determinants of cardiac catheterization cases. For a 5-year period in the study region, a large proportion of catheterizations affect seniors, i.e., individuals aged 65 years and older: over 45% of catheterizations are performed on seniors, where seniors account for 9% of the total population (12.6% of adults). While only 0.01% of adults under 65 receive catheterizations, almost 7% of seniors do. Visual observation suggests that catheterization cases are spatially associated with seniors' residential location (refer to Figure 1). *Seniors* is therefore proposed as the auxiliary process.

The association between older age and cardiovascular disease is well known and has received much attention in the health literature [29], [28], [30]. Their spatial association has received less attention, although this relationship has been examined at the neighbourhood level [16], [18], [5], [25]. The city of Calgary was chosen as an interesting study area, where seniors' residential location presents a clustered distribution, facilitating the identification of spatial associations.

In the following, Section II provides the context of this study; Section III describes the methods employed; Section

 ISBN: 978-1-61208-469-5

IV outlines the results; Section V provides a discussion, and Section VI draws the main conclusions of the study.

II. Background

One of the leading causes of death in the developed world, cardiovascular disease is known to be associated with a number of risk factors, including age and gender, limited physical activity, smoking, and diet [18], [5]. Often these factors correlate with demographic and socioeconomic characteristics, such as age, occupation, and income, which can be measured by census variables [10], [31], [7], [4].

Calgary is one of the largest Canadian cities. Located in the foothills east of the Rocky Mountains, it covers a large and regular geographic area; its population is relatively young, affluent, and highly educated [40]. Due to its economy and history, its population presents a pseudo-concentric distribution, where age and socioeconomic status decrease as distance from the city center increases [36].

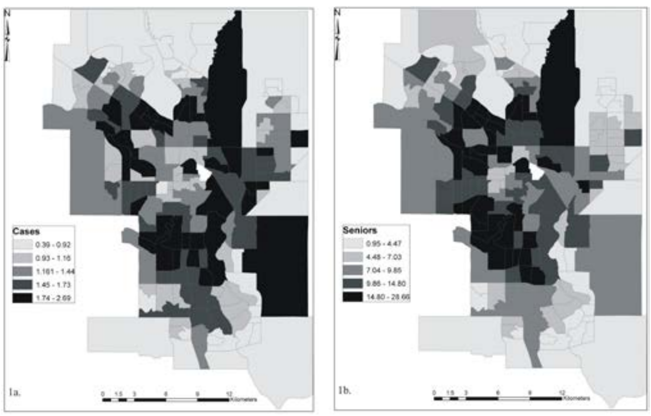

Figure 1. Distribution of Catheterization Cases and Seniors in Calgary.

The clinical data were provided by the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) initiative [22], a clinical registry begun in 1995 with the collection of cardiac catheterization data. Cardiac catheterization is a procedure performed on individuals with cardiovascular conditions [22]. The province of Alberta has a publicly funded health care system, therefore there are no financial costs associated with the procedure. We acknowledge the limitations of catheterization in representing cardiovascular disease; however, more appropriate variables, e. g., hospitalizations for acute coronary syndrome (ACS), were not systematically collected at the postal code spatial aggregation level until much more recently. Analyzing the latter, more recent data would have been problematic, due to changes in the 2011 census [39]. However, a comparison of catheterization and ACS over a two-year period of overlapping data (2006–2007) suggests that the proposed method shall be transferable to ACS data as soon as all the clinical data and census variables become available.

For this study, we considered only patients aged 20 years and older, residing in Calgary, who had one or more catheterizations between 1999 and 2003 (for multiple procedures, only the first one was retained). Patient residential addresses, at the postal code level, were spatially aggregated to match census spatial units, using postal code conversion files (PCCF +) [43]. Demographic and socioeconomic variables were drawn from the 2001 census of Canada. To match the clinical records, census variables were trimmed to represent the population aged 20 years or older. Clinical records and census variables were normalized for each spatial unit, i. e., divided by the total pertinent population and multiplied by 1,000 [34]. This variable, representing cardiac catheterization cases, is named "*Cases*" in Table 1. For the regression analysis, clinical records were age- and sex-standardized [42]: the standardized variable is named "*Standardized Cases*". The variable "*Seniors*" (Table 1) emphasizes the difference between workforce and retirees as the sum of all age groups aged 65 years and older.

All the analyses are conducted on census tracts, which are relatively small and stable spatial units, with population between 2,500 and 8,000 residents, located in a census metropolitan area [39]. For the 2001 census, the spatial database contains 181 census tracts, and the clinical database contains 11,430 catheterization cases over the 5-year interval (Figure 1).

III. Methods

Cross-correlations and spatial correlation analyses assess the association of catheterization cases and seniors' residential location. Spatial correlation, measured by bivariate Moran's I, extends this comparison to neighbouring census tracts. Spatial autocorrelation, measured by Moran's I, assesses the self-similarity of each variable over neighbouring census tracts. Moran's I ranges from –1 for negative spatial autocorrelation, to +1 for positive spatial autocorrelation, with 0 indicating spatial randomness [19]. The calculation of Moran's I requires the definition of a spatial weights matrix, W (discussed below).

Spatial autoregression aims at enhancing the reliability of estimates in the presence of spatial dependencies (1).

$$Y = X\beta + \rho WY + \varepsilon \tag{1}$$

Spatial autoregression also requires a spatial weights matrix, W, which selects the spatial units deemed spatially dependent [20], and an autoregressive parameter, ρ (rho), is estimated. This study uses a simultaneous autoregressive specification and maximum likelihood estimation [2]. Following conventional practice, the regression is computed on the age- and sex-standardized dependent variable. Indirect standardization [42] employs age and sex groups, therefore inflating the correlation between the dependent and those demographic variables. For this reason, the regression model does not include demographic variables, even though the exclusion some variables, and particularly of *Seniors*, has a large impact on the model's goodness of fit. The use of *Seniors* as an auxiliary process mitigates this impact.

The spatial weights matrix is viewed as a tool to enhance model reliability. In the absence of a spatial specification, the model reliability is decreased by the presence of spatial autocorrelation in the regression residuals, which inflates the variance associated with the regression parameters [2]. Therefore, the spatial weights matrix is designed to best capture the spatial autocorrelation in the dependent variable, so that most of the spatial autocorrelation can be accounted for by the model, leaving insignificant spatial autocorrelation in the residuals. The basic form of a spatial weights matrix is a binary structure, where a threshold distance, or a number of nearest neighbours, selects the neighbouring spatial units where the variable is expected to exhibit spatial autocorrelation. Often a weight is added, in order to model distance decay effects [20]. This matrix involves the specification of three parameters: distance threshold, distance metric, and distance decay function. By dynamically adjusting these three parameters, the spatial weights matrix can yield different values of the spatial autocorrelation index. Of the three parameters, distance exerts the greatest influence, by determining how many spatial units are deemed spatially autocorrelated. There are several methods to define this parameter for areal units, such as census tracts [20]. Here, we use a distance threshold based on nearest neighbours, for the following reasons: census tracts are not necessarily meaningful for the spatial pattern of cardiovascular disease; they tend to be small and pseudo-rectangular in the city center, but in the outskirts they tend to be larger and less regular (refer to Figure 1). The latter feature forms a pattern of spatial units, liable to confound the pattern of the variables recorded in those units. To reduce this confounding effect, the number of nearest neighbours is preferred. To further de-emphasize the geometry of census tracts, we consider the distance between their centroids.

The second parameter is the distance metric. Among many distance metrics used in geography, the most common is the Euclidean metric: the straight line distance measurement between two points, ‘as the crow flies’. In many North American cities, connectivity occurs over a pseudo-rectangular road pattern, better modelled by the Manhattan distance, which measures distance between points along a rectangular path with right angle turns. Connectivity over a complex or mixed network can be more accurately represented by metrics of the class known as Minkowski distance [38], which yields patterns intermediate between straight line and right angle. It is described by (2), of which Euclidean and Manhattan distances are special cases, where the key parameter, *p*, can take any value between 1 (Manhattan) and 2 (Euclidean).

$$d_{ij} = \left[\left(x_i - x_j \right)^p + \left(y_i - y_j \right)^p \right]^{1/p} \qquad (2)$$

The *p* value can be estimated to best approximate empirical distance or travel time. Within this class of metrics, an appropriate choice can refine the selection of spatial units, producing buffer shapes close to the physical, pseudo-rectangular connectivity pattern of the census tracts.

The third parameter is the distance decay function, which weights the interaction among spatial units by their distance. Interaction tends to decrease as the distance between units increases: a number of functions have been developed to model this relationship [14], [35]. Commonly the distance decay function is calibrated by a weight, often another variable, which normalizes the relationship [8].

The interaction of these parameters in the spatial weights matrix affects the estimated spatial autocorrelation and hence the reliability of the regression estimates. In the proposed method, the three parameters are calibrated on the spatial autocorrelation of the auxiliary process, as opposed to the dependent variable, or primary process.

Seniors’ residential location is understood better than the distribution of *Cases*, and can be explained by socio-economic and urban traits. Because of their high spatial association, the spatial autocorrelation of *Seniors* should be more meaningful than that of *Cases*; therefore, the method is expected to increase the reliability of the regression estimates, along with their interpretability.

Statistical analyses were conducted in TIBCO Spotfire S+ 8.2. Maps were obtained in ESRI ArcGIS 10.

IV. RESULTS

Spatial autocorrelation and cross-correlation analysis, initially run on the parameters (k = 3, p = 2), yields two important results (Table 1): *Cases* exhibits significant but moderate spatial autocorrelation (I = 0.36), whereas *Seniors* exhibits a much higher value (I = 0.54); the two variables exhibit high cross-correlation (0.60) and spatial correlation (0.35). Together, these results confirm that clustering of *Cases* tends to occur in association with the clustering of seniors’ residential location.

TABLE I. CORRELATIONS AND SPATIAL CORRELATIONS

	Clinical records		Demographhic variables		Economic variables	Education		Family status
	Cardiac cath. cases	Age & sex std. cases	Age 65 and over	Age 55 to 64	Family median income	Secondary or lower education	Non-univ. post-sec. degree	2 parents with children
	Cases	Std.Cases	Seniors	Age.55-64	Fam.Income	Secondary	Trades	Families
Cases	0.36 **	0.75 **	0.66 **	0.50 **	-0.08 ns	0.25 **	-0.22 **	-0.25 **
Std.Cases	0.34 **	0.53 **	0.87 **	0.71 **	0.17 *	-0.09 ns	-0.42 **	-0.21 **
Seniors	0.35 **	0.49 **	0.60 **	0.31 **	-0.03 ns	-0.09 ns	-0.35 **	-0.47 **
Age.55-64	0.22 **	0.27 **	0.09 ns	0.50 **	0.29 **	-0.01 ns	-0.26 **	0.15 ns
Fam.Income	-0.06 ns	0.00 ns	-0.08 ns	0.10 ns	0.50 **	-0.68 **	-0.33 **	0.56 **
Secondary	0.09 ns	-0.12 ns	-0.17 *	0.04 ns	-0.42 **	0.75 **	0.25 **	-0.09 ns
Trades	-0.03 ns	-0.11 ns	-0.18 *	0.02 ns	0.02 ns	0.17 *	0.42 **	0.04 ns
Families	-0.13 ns	-0.19 *	-0.42 **	-0.05 ns	0.35 **	0.04 ns	0.24 **	0.74 **

Diagonal: univariate Moran's I. Upper half: Pearson's correlation . Lower half: bivariate Moran's I.

The correlation analysis also identifies additional traits consistent with this result. Significant and negative spatial cross-correlation between *Seniors* and *Families* indicates a highly clustered if not dichotomous spatial pattern, where neighbourhoods dominated by younger families alternate with neighbourhoods mostly occupied by seniors. The high correlation of *Income* with *Families* and with *Education* suggests that income exerts a strong but indirect influence on the spatial clustering, so that the observed residential pattern appears associated with economic factors, in addition to demographic ones. Nonetheless, age is confirmed as the variable that exhibits the most distinct spatial pattern.

Following these results, we analyze the response of the spatial autocorrelation index to variations in the spatial contiguity parameters for the variables *Cases*, *Standardized Cases*, and *Seniors*, as summarized in Figure 2. For all the

 ISBN: 978-1-61208-469-5

parameter combinations, *Seniors* exhibits the highest spatial autocorrelation values, whereas the values of *Cases* are constantly significant but relatively low. Age- and sex-standardized *Cases* exhibits greater spatial autocorrelation than the non-standardized variable. The number of nearest neighbours (*k*) has a greater impact on the spatial autocorrelation value, whereas the distance metric only provides minor adjustments. Since the analysis shows that the spatial autocorrelation is best expressed by relatively small neighbourhoods, defined by low *k* values, distance decay weighting will not be discussed.

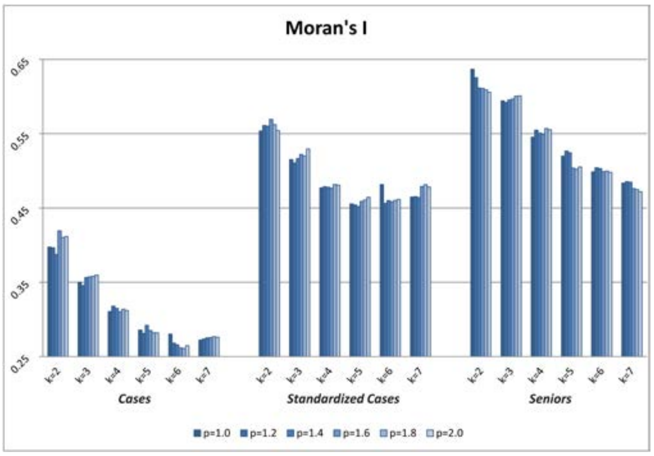

Figure 2. Moran's I as function of distance threshold and distance metric.

The spatial autocorrelation of *Cases* declines steeply as the number of nearest neighbours increases, and only parameter combinations of $k = 2$ and p values between 1.6 and 2 effectively capture the spatial autocorrelation in the variable. These combinations define very small neighbourhoods, as distance is short and metrics close to the straight line produce the shortest distance between centroids. For *Standardized Cases*, the trend is similar to *Cases*, and the most visible effect of the standardization is the increased value of the spatial autocorrelation. The spatial autocorrelation of *Seniors* exhibits less variation than *Cases* in response to variations in both parameters.

The regression summarized in Table 2 is based on the spatial contiguity parameters derived from the auxiliary process. It models the association between the standardized dependent variable and the pool of socioeconomic variables. Significant explanatory variables are: low education attainments, family median income, and, -with negative coefficient- family status and technical education, the latter possibly a proxy for young, low income groups. The model does not include the variable that is most highly correlated with the dependent, i.e., *Seniors*, due to the confounding effect of the indirect standardization; hence, the goodness of fit is relatively low, e.g., the value of Anselin's [1] pseudo-R^2 is 0.19.

TABLE II. SPATIAL AUTOREGRESSIVE MODEL

Stdandardized Cases (k = 3, p = 2)

	β value	Std.Error	t value	Pr(>|t|)
Intercept	14.03	3.77	3.73	0.00
Secondary	0.03	0.00	6.01	0.00
Family median income	0.14	0.03	5.28	0.00
Trades	-0.02	0.01	-4.04	0.00
Families	-0.01	0.00	-3.40	0.00

Log. likelihoo	Pseudo R^2	Res. Std. Error	Rho	Residual Moran
-679.70	0.19	2.92	0.32	-0.03

In its linear specification, the model exhibits significant residual spatial autocorrelation; hence, a spatially autoregressive specification is presented, which exhibits non-significant residual spatial autocorrelation and a significant autoregressive coefficient, rho. The calibration of the spatial contiguity parameters on the auxiliary process, *Seniors*, is a way of representing this variable in the model. Arguably, this method enhances the significance of the rho parameter, attaining a more effective reduction of the residual spatial autocorrelation and more reliable regression estimates.

V. DISCUSSION

The use of an auxiliary process as an alternative to the use of the primary process for the definition of the spatial contiguity parameters leads to an improved neighbourhood definition, enhancing the reliability of the spatial regression model estimates. In the application discussed here, the resulting neighbourhoods are larger; moreover, the analytical results are less sensitive to variations in the neighbourhood size, defined by the contiguity parameters. These two results are important and related, both deriving from the choice of a particular auxiliary process. The tiny neighbourhoods defined by the primary process model clusters of *Cases*, which are isolated, as shown by their low spatial autocorrelation. Conversely, the larger neighbourhoods calibrated on the auxiliary process effectively model the distribution of the highest-risk population, *Seniors*. Therefore, modelling *Cases* based on the spatial distribution of *Seniors* provides not only a more reliable, but also a more interpretable model. Overall, the greater analytical stability obtained through of the use of the auxiliary process is an important result, which can potentially reduce the impact of the modifiable aereal unit problem (MAUP) [41], [32]. As a future research direction, the method discussed here shall be tested on different spatial units, e.g., communities or dissemination areas, where previous analyses [6] implemented directly on the primary process have suggested a large impact of the MAUP.

The distribution of seniors in Calgary has been studied within several disciplines, and it is conceptually understood as influenced by economic cycles and age of community, among other factors [36]. As an additional line of enquiry, the spatial structure of the process shall be analyzed in light of that literature. Conversely, other crucial aspects, such as range of spatial interaction, shall be confirmed by qualitative analyses [23], [37]. An integration of these two lines of enquiry is expected to substantially improve understanding and representation of the spatial pattern of the high-risk population in its interaction with the incidence of cardiac disease [33].

For the application presented here, the *k* parameter chosen through the auxiliary process is only marginally larger than the one selected for the primary process, and the difference between the resulting neighbourhoods is larger in conjunction with the distance metric. Hence, the impact on the reliability of the regression estimates may be moderate if measured simply by variance indicators; however, the spatial contiguity parameters also affect model inference, impacting model selection procedures. Comparing several regression specifications, the exclusion of *Seniors* as a predictor is constantly accompanied by increased residual spatial autocorrelation, suggesting that *Seniors* is the process associated with the observed spatial autocorrelation.

One important extension of the current analysis will be its application on acute coronary syndrome (ACS). While the proposed method presents many advantages with respect to modelling seniors, it shifts the analytical focus away from younger adults, where the prevalence is very low (0.01%), and its spatial modelling remains challenging.

A number of health conditions are associated with specific demographic segments, and in all those cases, the use of an appropriate auxiliary process can improve the analytical results. Testing is underway on ACS, congenital birth defects and child obesity. Further lines of enquiry shall include analyses of different contiguity configurations, such as threshold distance vs. nearest neighbours, distance metrics beyond the Minkowski range, and assessment of distance decay functions on larger neighbourhoods.

VI. CONCLUSION

A multivariate regression model estimates the association between cardiac catheterization and socioeconomic factors. In the presence of spatial dependencies, the use of a spatially autoregressive model increases the reliability of the model estimates. Such reliability can be further improved by an appropriate definition of the spatial contiguity parameters. Of the catheterizations recorded over five years in the study area, almost half are performed on individuals aged 65 years or older. This association is well known and understood, and it suggests a strong association between seniors' residential location and the spatial pattern of catheterization cases. Therefore, *Seniors* is identified as an auxiliary process for the calibration of the spatial contiguity parameters of the model. The method enhances the reliability of the regression estimates and the model selection, rendering the auxiliary-based model more efficient, interpretable, and stable over variations in the supporting spatial units.

ACKNOWLEDGMENT

I acknowledge the Natural Science and Engineering Research Council of Canada and the GEOIDE Network of Centers of Excellence for funding this research. I also thank APPROACH initiative researchers for data and support. I am grateful to Olesya Elikan and all the students who helped with bibliographic research, database maintenance, and visualization.

REFERENCES

[1] L. Anselin, SpaceStat tutorial. Regional Research Institute, West Virginia University, Morgantown, West Virginia, 1993.

[2] L. Anselin, Spatial Econometrics: Methods and Models. Kluwer Academic Publisher, New York, 1988.

[3] L. Anselin, A. K. Bera, R. Florax and M. J. Yoon, "Simple diagnostic tests for spatial dependence", Regional Science and Urban Economics 26, 1996, 77-104.

[4] T. Augustin, T. A. Glass, B. D. James and B. S. Schwartz, "Neighbourhood psychosocial hazards and cardiovascular disease: The Baltimore Memory Study". American Journal of Public Health 98, 2008, 1664–1670.

[5] A. J. Bagnall, S. G. Goodman, K. A. A. Fox, R. T. Yan, J. M. Gore, A. N. Cheema, T. Huynh, D. Chauret, D. H. Fitchett, A. Langer and A. T. Yan, "Influence of Age on Use of Cardiac Catheterization and Associated Outcomes in Patients With Non-ST-Elevation Acute Coronary Syndromes". American Journal of Cardiology 103, 2009, 1530-1536.

[6] S. Bertazzon, S. Olson and M. Knudtson, "A spatial analysis of the demographic and socioeconomic variables associated with cardiovascular disease in Calgary (Canada)". Applied Spatial Analysis and Policy 3, 2010, 1-23.

[7] B. Chaix, M. Rosvall and J. Merlo, "Neighborhood socioeconomic deprivation and residential instability: Effects on incidence of ischemic heart disease and survival after myocardial infarction". Epidemiology 18, 2007, 104–111.

[8] N. Cressie, Statistics for Spatial Data. Wiley, New York, 1993.

[9] M. Cutchin, "The need for the ''new health geography'' in epidemiologic studies of environment and health". Health & Place 13, 2007, 725–742.

[10] A. V. Diez Roux, S. Merkin, D. Arnett, L. Chambless, M. Massing, F. Nieto, P. Sorlie, M. Szklo, H. Tyroler and R. Watson, "Neighborhood of residence and incidence of coronary heart disease". New England Journal of Medicine 345, 2001 99-106.

[11] S. Dray, "A New Perspective about Moran's Coefficient: Spatial Autocorrelation as a Linear Regression Problem". Geographical Analysis 43, 2001, 127–141

[12] T. Dummer, "Health geography: supporting public health policy and planning". Canadian Medical Association Journal 178, 2008, 1177–1180.

[13] P. Elliott and D. Wartenberg, "Spatial epidemiology: Current approaches and future challenges". Environmental Health Perspectives 12, 2004, 998–1006.

[14] A. Fotheringham, "Spatial structure and distance-decay parameters". Annals of the Association of American Geographers 71, 1981, 425-436.

[15] A. Fotheringham, C. Brundson and M. Charlton, "Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis". Environment & Planning A 30, 1998, 1905–1927.

[16] A. Gatrell, G. Lancaster, A. Chapple, S. Horsley and M. Smith, "Variations in use of tertiary cardiac services in part of North-West England". Health & Place. 8, 2002, 147-153.

[17] A. Gatrell, J. Popay and C. Thomas, 2004. "Mapping the determinants of health inequalities in social space: can Bourdieu help us?" Health & Place. 10, 2004, 245-257.

[18] Y. Gerber, S. Jacobsen, R. Frye, S. Weston, J. Killian and V. Roger, "Secular Trends in Deaths From Cardiovascular Diseases: A 25-Year Community Study". Circulation 113, 2006, 2285-2292.

 ISBN: 978-1-61208-469-5

[19] A. Getis, "A history of the concept of spatial autocorrelation: A geographer's perspective". Geographical Analysis 40, 2008, 297–309.

[20] A. Getis and J. Aldstadt, "Constructing the Spatial Weights Matrix Using a Local Statistic". Geographical Analysis 36, 2004, 90-104.

[21] A. Getis and K. Ord, "The analysis of spatial association by use of distance statistics". Geographical Analysis 24, 1992, 189–206.

[22] W. A.Ghali and M. L. Knudtson, "Overview of the Alberta provincial project for outcome assessment in coronary heart disease". Canadian Journal of Cardiology 16, 2000, 1225–1230.

[23] A. Grimes, D. C. Maré and M. Morten, "Defining Areas and Linking Geographical Data: an Example from New Zealand". Population, Space and Place 16, 2010, 165–170.

[24] R. Haining, The Special Nature of Spatial Data. The SAGE handbook of spatial analysis.In: A.S. Fotheringham and P. A. Rogerson (Eds.) SAGE, London, 2009.

[25] E. J. Holowaty, T. A. Norwood, S. Wanigaratne, J. J. Abellan and L. Beale, "Feasibility and utility of mapping disease risk at the neighbourhood level within a Canadian public health unit: an ecological study". International Journal of Health Geographics 9,21, 2010.

[26] A. Iftimi, F. Montes, A. M. Santiyán and F. Martínez-Ruiz, 'Space–time airborne disease mapping applied to detect specific behaviour of varicella in Valencia, Spain', *Spatial and Spatio-temporal Epidemiology*, 14–15, 2015, 33-44.

[27] V. Jürgens, S. Ess, M. Schwenkglenks, T. Cerny, and P. Vounatsou, 'Using lung cancer mortality to indirectly approximate smoking patterns in space', *Spatial and Spatio-temporal Epidemiology,* 14–15, 2015, 23-31.

[28] G. A. Kaplan and J. E. Keil, "Socioeconomic factors and cardiovascular disease: a review of the literature". Circulation 88, 1993, 1973–1998.

[29] E. G. Lakatta, Age-associated cardiovascular changes in health: impact on cardiovascular disease in older persons". Heart Failure Reviews 7, 2002, 29-49.

[30] D. Manuel, M. Leung and K. Nguyen, "Burden of cardiovascular disease in Canada". Canadian Journal of Cardiology 19, 2003, 997–1004.

[31] J. McKay and G. A. Mensah, The Atlas of Heart Disease and Stroke. Geneva: World Health Organization, 2005.

[32] M.-P. Parenteau and M. C. Sawada, "The modifiable areal unit problem (MAUP) in the relationship between exposure to NO2 and respiratory health". International Journal of Health Geographics, 2011, 10:58.

[33] B. Preston and M. W. Wilson, 'Practicing GIS as Mixed Method: Affordances and Limitations in an Urban Gardening Study', *Annals of the Association of American Geographers,* 104 (3), 2014, 510-29.

[34] S. Preston, P. Heuveline and M. Guillot, Demography: Measuring and Modeling Population Processes. Blackwell Publishing, Oxford, 2000.

[35] A. Rattner and B. A. Portnov, "Distance decay function in criminal behaviour: a case of Israel". Annals of Regional Science 41, 2007, 673–688.

[36] B. Sandalack and A. Nicolai, The Calgary Project: Urban form/urban life. University of Calgary Press, Calgary, 2006.

[37] S. M. Santos, D. Chor and G. L. Werneck, "Demarcation of local neighborhoods to study relations between contextual factors and health". International Journal of Health Geographics 9, 34, 2010.

[38] R. Shahid, S. Bertazzon, M. Knudtson and W. Ghali, "Comparison of distance measures in spatial analytical modeling for health service planning". BMC Health Services Research 9:200, 200.

[39] Statistics Canada, 2002. Cartographic Boundary Files: 2001 Census. Reference Guide. Catalogue no. 92F0171GIE. Minister of Industry, Ottawa.

[40] Statistics Canada. The Canadian Population in 2011: Population Counts and Growth. http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-310-x/98-310-x2011001-eng.cfm. Last accesses on 3/15/2016.

[41] D. J. Unwin, "GIS, spatial analysis and spatial statistics". Progress in Human Geography 20, 1996, 540-441.

[42] L. Waller and C. Gotway, Applied Spatial Statistics for Public Health Data.: John Wiley & Sons, Hobeken, 2004.

[43] R. Wilkins and S. Khan, 2010. PCCF + Version 5G User's Guide. Catalogue no. 82F0086-XDB. Health Statistics Division, Statistics Canada, Ottawa. http://data.library.utoronto.ca/datapub/codebooks/cstdli/pccf_health/pccf5h/MSWORD.PCCF5H.pdf. Last accessed on 3/15/2016.

 ISBN: 978-1-61208-469-5