

# An Analysis of Spatial Clustering of Stroke Types, In-hospital Mortality, and Reported Risk Factors in Alberta, Canada, Using Geographic Information Systems

Susan van Rheenen, Timothy W.J. Watson, Shelley Alexander, Michael D. Hill

**ABSTRACT: Background:** Despite advances in the quality and delivery of stroke care, regional disparities in stroke incidence and outcome persist. Spatial analysis using geographic information systems (GIS) can assist in identifying high-risk populations and regional differences in efficacy of stroke care. The aim of this study was to identify and locate geographic clusters of high or low rates of stroke, risk factors, and in-hospital mortality across a provincial health care network in Alberta, Canada. **Methods:** This study employed a spatial epidemiological approach using population-based hospital administrative data. Getis-Ord  $G_i^*$  and Spatial Scan statistics were used to identify and locate statistically significant “hot” and “cold” spots of stroke occurrence by type, risk factors, and in-hospital mortality. **Results:** Marked regional variations were found. East central Alberta was a significant hot spot for ischemic stroke (relative risk [RR] 1.43,  $p < 0.001$ ), transient ischemic attack (RR 2.25,  $p < 0.05$ ), and in-hospital mortality (RR 1.50,  $p < 0.05$ ). Hot spots of intracerebral hemorrhage (RR 1.80,  $p < 0.05$ ) and subarachnoid hemorrhage (RR 1.64,  $p < 0.05$ ) were identified in a major urban centre. Unexpectedly, stroke risk factor hot spots (RR 2.58,  $p < 0.001$ ) were not spatially associated (did not overlap) with hot spots of ischemic stroke, transient ischemic attack, or in-hospital mortality. **Conclusions:** Integration of health care administrative data sets with geographic information systems contributes valuable information by identifying the existence and location of regional disparities in the spatial distribution of stroke occurrence and outcomes. Findings from this study raise important questions regarding why regional differences exist and how disparities might be mitigated.

**RÉSUMÉ: Analyse de l'agrégation spatiale de types d'accident vasculaire cérébral, de mortalité hospitalière et de facteurs de risque rapportés en Alberta, au Canada, au moyen de la méthode GIS. Contexte:** Malgré les progrès réalisés dans le domaine de la qualité et de la prestation des soins aux patients atteints d'un accident vasculaire cérébral (AVC), des inégalités régionales persistent quant à l'incidence de l'AVC et à son issue clinique. Une analyse spatiale au moyen de systèmes d'information géographique (GIS) peut aider à identifier des populations à haut risque et des différences régionales dans l'efficacité des soins prodigués aux patients atteints d'un AVC. Le but de cette étude était d'identifier et de localiser des agrégats géographiques de taux élevés ou bas d'AVC, de facteurs de risque et de mortalité hospitalière au sein d'un réseau provincial de soins de santé en Alberta, au Canada. **Méthode:** Nous avons utilisé une démarche épidémiologique spatiale appliquée à des données administratives hospitalières de la population. Les méthodes statistiques Getis-Ord  $G_i^*$  et de scan spatial ont été utilisées pour identifier et localiser des points chauds et des points froids d'AVC par type, facteurs de risque et mortalité hospitalière. **Résultats:** Nous avons constaté qu'il existe des variations régionales importantes. La partie est de la région centrale de l'Alberta était un point chaud important en ce qui concerne l'AVC ischémique (risque relatif [RR] 1,43 ;  $p < 0,001$ ), l'ischémie cérébrale transitoire (RR 2,25 ;  $p < 0,05$ ) et la mortalité hospitalière (RR 1,50 ;  $p < 0,05$ ). Des points chauds d'hémorragie intracérébrale (RR 1,80 ;  $p < 0,05$ ) et d'hémorragie sous-arachnoïdienne (RR 1,64 <0,05) ont été identifiés dans un centre urbain important. À notre grande surprise, les points chauds pour les facteurs de risque de l'AVC (RR 2,58 ;  $p < 0,001$ ) n'étaient pas localisés aux mêmes endroits (ne coïncidaient pas) que les points chauds de l'AVC ischémique, de l'ischémie cérébrale transitoire ou de la mortalité hospitalière. **Conclusions:** L'intégration de données administratives de soins de santé à des systèmes d'information géographique fournit des informations précieuses en démontrant l'existence et en localisant des disparités régionales dans la distribution spatiale de la fréquence et de l'issue clinique de l'AVC. Nos constatations soulèvent d'importantes questions concernant les causes de ces différences régionales et comment elles pourraient être atténuées.

**Keywords:** Cluster analysis, GIS, hot spot analysis, spatial epidemiology, stroke, stroke mortality

doi:10.1017/cjn.2015.241

Can J Neurol Sci. 2015; 42: 299-309

From the Faculty of Medicine, Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada (SvR), Department of Clinical Neurosciences, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada (TWJW), Department of Geography, University of Calgary, Calgary, Alberta, Canada (SA), Department of Clinical Neurosciences, Community Health Sciences, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada (MDH)

RECEIVED JULY 1, 2014. FINAL REVISIONS SUBMITTED FEBRUARY 26, 2015.

Correspondence to: Susan van Rheenen, Faculty of Medicine, Department of Community Health Sciences, University of Calgary, TRW Building, 3rd Floor, 3280 Hospital Drive NW, Calgary, Alberta, Canada T2N 4Z6 Email: susan.vanrheenen@ucalgary.ca

Alberta is a Canadian province with diverse topography spanning 640,000 square kilometres and a population of 3,645,000.<sup>1</sup> Population density varies greatly. Two-thirds of the population resides in two major urban centres and the remainder in rural and smaller urban centres.<sup>2</sup> Stroke expertise is concentrated in the two comprehensive stroke centres (CSC) located in Edmonton and Calgary. There are 14 primary stroke centres (PSC) with thrombolysis protocols and telestroke access in place<sup>3</sup> strategically located throughout the province (Figure 1). In 2011, there were 6951 incident stroke cases<sup>4</sup> and approximately 36,000 stroke survivors in Alberta (Michael D. Hill, unpublished data, 2013).

Major advances have been made to the quality and delivery of specialized stroke care, but regional disparities in stroke incidence and mortality persist.<sup>5,6</sup> Documenting existing trends and identifying the reason(s) for these variances is fundamental to the design of rational health care planning, distribution of resources, and program development.<sup>7</sup> Although regional disparities in stroke exist in Alberta,<sup>3</sup> an analysis of the spatial distribution of stroke types and mortality has not previously been undertaken, and the extent to which regional differences in risk factors for stroke are associated with any disparities has not been fully considered. How much geographic variation exists, where it exists, what populations are affected, and whether preventable regional

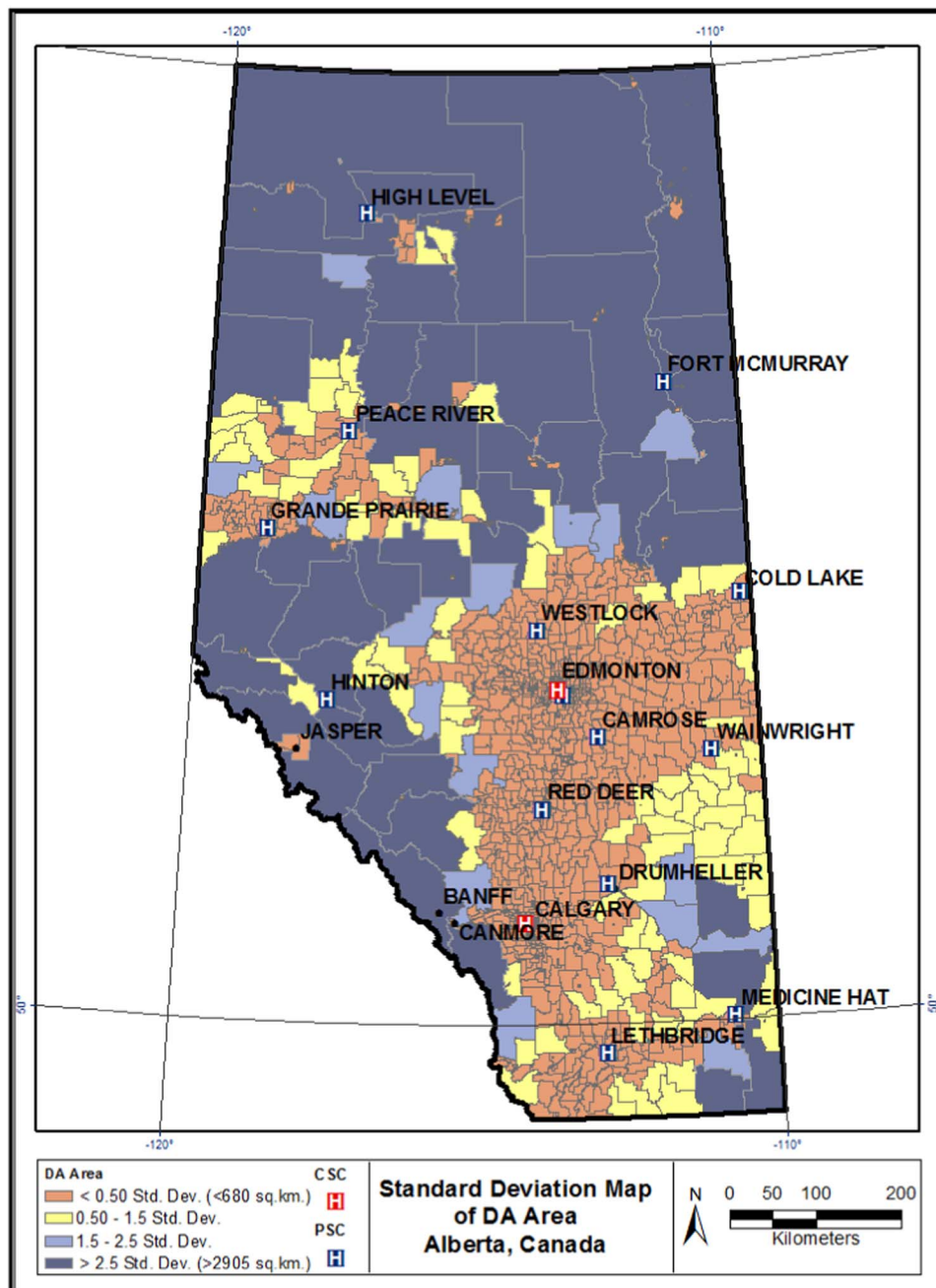


Figure 1: Standard deviation map of dissemination area size (km<sup>2</sup>).

disparities have been mitigated with targeted enhancements to stroke services are not known.

The aim of this investigation was to identify statistically significant geographic variances in the occurrence of major stroke types, in-hospital mortality (henceforth known as mortality) and known stroke risk factors using geographic information system (GIS)-based methods and administrative data sets in the province of Alberta.

## METHODS

### Study Design

This study employed a spatial epidemiological approach using population-based administrative data, collected over six years, integrating GIS. GIS has been used increasingly as an evidence-based tool in epidemiological research,<sup>8</sup> and in stroke research specifically.<sup>9,10</sup> The study was approved by the Conjoint Health Research Ethics Board at the University of Calgary.

### Data

Alberta, like other provinces in Canada, has a provincially administered government funded universal health care insurance system covering approximately 97% of permanent residents. Exceptions include Registered First Nations persons, prison inmates, and members of the military and the Royal Canadian Mounted Police, who have similar coverage federally.<sup>11</sup> All persons diagnosed with stroke who accessed the health care system in Alberta are included in the study database. Alberta inpatient discharge abstract data (Canadian Institute for Health Information [CIHI]-Discharge Abstract Database [DAD])<sup>12</sup> and ambulatory care data (Ambulatory Care Classification System/National Ambulatory Care Reporting System)<sup>13</sup> that includes emergency department visits for fiscal years 2002-2003 to 2007-2008, inclusive, were linked and all personal identifiers removed. Subjects selected for inclusion had an International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)<sup>14</sup> diagnosis of I63 (cerebral infarction), H34 (retinal arterial occlusion), I64 (stroke not specified as hemorrhage or infarction), I60 (subarachnoid hemorrhage [SAH]), I61 (intracerebral hemorrhage [ICH]), or G45 (transient ischemic attack [TIA]). I64 was coded with I63 and H34 as ischemic stroke based on the assumption that 87% of strokes are ischemic and 13% are hemorrhagic.<sup>15</sup> Patients younger than age 20-years were excluded from the analysis as mechanisms for stroke in children differ from those in adults. Patients without a health care number and/or valid Alberta postal code were also excluded.

A new stroke event was identified as follows. If multiple visits for the same patient occurred within 48 hours, the most serious event was selected (ICH > SAH > ischemic > TIA). If multiple visits within 48 hours were of the same stroke type, the first event was selected.

Known risk factors included in the analysis had an ICD-10 code of E10 (Insulin Dependent Diabetes Mellitus [DM]), E11-14 (Non-Insulin Dependent DM), I48 (atrial fibrillation), I10-I15 (hypertension), and E780, E781, E784, E785, E788, and E789 (hypercholesterolemia). These risk factors were identified from diagnosis codes reported in CIHI-DAD data (to 25 diagnostic occurrences).

Shapefiles for the spatial analysis including health care facilities, cities, and dissemination area (DA) and Alberta boundaries were obtained from Spatial and Numeric Data Services, University of Calgary.<sup>16</sup>

### Analysis

Administrative data were linked with Statistics Canada 2006 census population data and age-standardized rates, using 10-year age categories, were calculated using the direct method to preserve consistency for comparisons across geographic areas.<sup>17,18</sup> Ten-year age categories were chosen because they capture differences in age distribution<sup>19</sup> and are comparable to other epidemiological studies.<sup>20</sup> No adjustment was made for sex to avoid spreading the data too thinly and because age generally has a much stronger impact on stroke incidence and mortality.<sup>19</sup> The date of designation of a health care facility as PSC was not considered in this analysis. STATA/IC 12.1<sup>21</sup> software was used for the nonspatial portion of the analysis.

DA was selected as the geographic unit of analysis. DAs are small, relatively stable geographic units with a population of between 400 and 700 persons,<sup>22</sup> designed to be socioeconomically relatively homogeneous. Postal codes were used to identify rural and urban geographic areas. Depending on the definition, Canada's rural population may vary between 22% and 38% of the total population (23% for postal codes).<sup>23</sup> Postal codes were converted to latitude and longitude coordinates using the 2006 and 2011 Postal Code Conversion Files,<sup>24</sup> and data were imported into ArcGIS ArcMap 10.<sup>25</sup> Statistics Canada 2006 Census data were suppressed for 15 of 3757 DAs with stroke events because of reserve land, unreliable census data, or small population numbers (<100 persons for six-character postal codes<sup>26</sup>).

Individual stroke events were aggregated according to DA. Annualized mean age-standardized stroke and mortality rates were calculated for each DA and used for the cluster analysis. DAs with suppressed data were assigned a rate of "0" for the cluster analysis if no strokes were reported during the six-year time span.

Data were projected in NAD83, 10-degree Transverse Mercator.<sup>27</sup> Two widely used spatial statistical methods, the Getis-Ord  $G_i^*$  statistic in ArcGIS<sup>28</sup> and Spatial Scan statistic in SaTScan<sup>29</sup> were employed to identify "hot spots" and "cold spots" of stroke occurrence and mortality. High variances in rates can occur when using geographic units with small at-risk populations. For this reason, raw age-standardized rates using the direct method were smoothed using the spatial empirical Bayes smoothing technique (second-order queen weights) in GeoDa<sup>30</sup> for the Getis-Ord  $G_i^*$  analysis. Age-standardized rates using the indirect method were used for the Spatial Scan statistical analysis. Statistical significance for hot spots was set as DAs with a  $G_i^*$  Z-score of >1.96, and Spatial Scan statistic relative risk (RR) >1.0 and  $p < 0.05$ . Statistical significance for cold spots was defined as DAs with a  $G_i^*$  Z-score <-1.96, and a Spatial Scan statistic RR < 1.0 and  $p < 0.05$ . A total of 999 Monte Carlo simulations were used to obtain  $p$  values for the Spatial Scan statistic. Proportions of stroke events with known risk factors for stroke were calculated for each DA and comparisons were made (1) within and external to the cluster boundary and (3) between hot and cold clusters.

*Getis-Ord  $G_i^*$  statistic.* Distance band was selected as the method for conceptualization of spatial relationships because of the large variation in DA area.<sup>31</sup> A spatial weights matrix file was created using (1) a fixed distance band calculation, based on DAs with an area <2.58 SD (Figure 1), which reflected maximum spatial autocorrelation,<sup>32</sup> and (2) a minimum number of eight neighbours to ensure a reliable Z-score.<sup>33</sup>

**Table 1: Study population characteristics**

Fiscal year	2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008
Stroke total (n)	3673	3678	3724	3619	3665	3527
Age, mean (years)	72.8	72.6	72.4	72.0	72.3	71.8
Female (%)	52.6	49.5	49.7	49.3	49.3	50.2
Comorbidities (% of stroke events)						
Insulin-dependent DM	1.4	1.4	1.4	0.8	0.8	1.0
Non-insulin-dependent DM	19.6	18.8	17.6	18.0	20.8	21.8
Atrial fibrillation	16.4	16.6	15.9	16.6	17.4	16.9
Hypertension	54.3	56.0	53.4	53.7	56.7	58.5
Hypercholesterolemia	16.8	18.8	17.2	16.9	18.7	20.7
Stroke type (%)						
Ischemic	61.7	61.0	60.5	62.3	62.4	64.2
Transient ischemic attack	24.8	26.2	26.3	24.1	23.7	23.4
Intracerebral hemorrhage	9.2	7.9	8.0	8.4	8.2	6.9
Subarachnoid hemorrhage	4.3	4.9	5.2	5.2	5.7	5.5
Urban* (%)	72.5	74.0	73.9	75.0	73.8	74.9
Ambulance (%)	59.0	60.9	64.2	65.4	99.9	100.0
Transport destination (%)						
Comprehensive stroke centre <sup>†</sup>	42.4	46.9	46.8	49.7	49.1	47.2
Urban	49.4	54.9	52.8	56.1	55.5	53.7
Rural	23.9	24.0	29.7	30.7	30.8	27.8
Primary stroke centre <sup>‡</sup>	16.8	16.3	17.1	15.2	18.5	22.3
Urban	17.7	16.6	17.2	15.2	19.0	21.6
Rural	14.3	15.5	16.7	15.0	17.1	23.9
Other	40.8	36.8	36.1	35.1	32.4	30.5
Urban	32.9	28.5	30.0	28.7	25.5	24.7
Rural	61.8	60.5	53.6	54.3	52.1	48.3
Disposition (%)						
Transfer						
Inpatient	17.8	18.9	19.5	18.0	15.0	15.7
Long-term care	8.1	8.1	7.9	7.0	12.1	9.9
Other	1.8	1.1	0.7	1.1	1.4	1.2
Discharge						
Home	48.6	48.9	47.3	48.2	44.8	48.4
Home with support	10.0	9.5	10.4	12.1	12.9	11.7
Against medical advice	0.3	0.4	0.4	0.5	0.7	0.7
In-hospital mortality	13.4	13.1	13.8	13.1	13.1	12.4

\*Postal code definition.

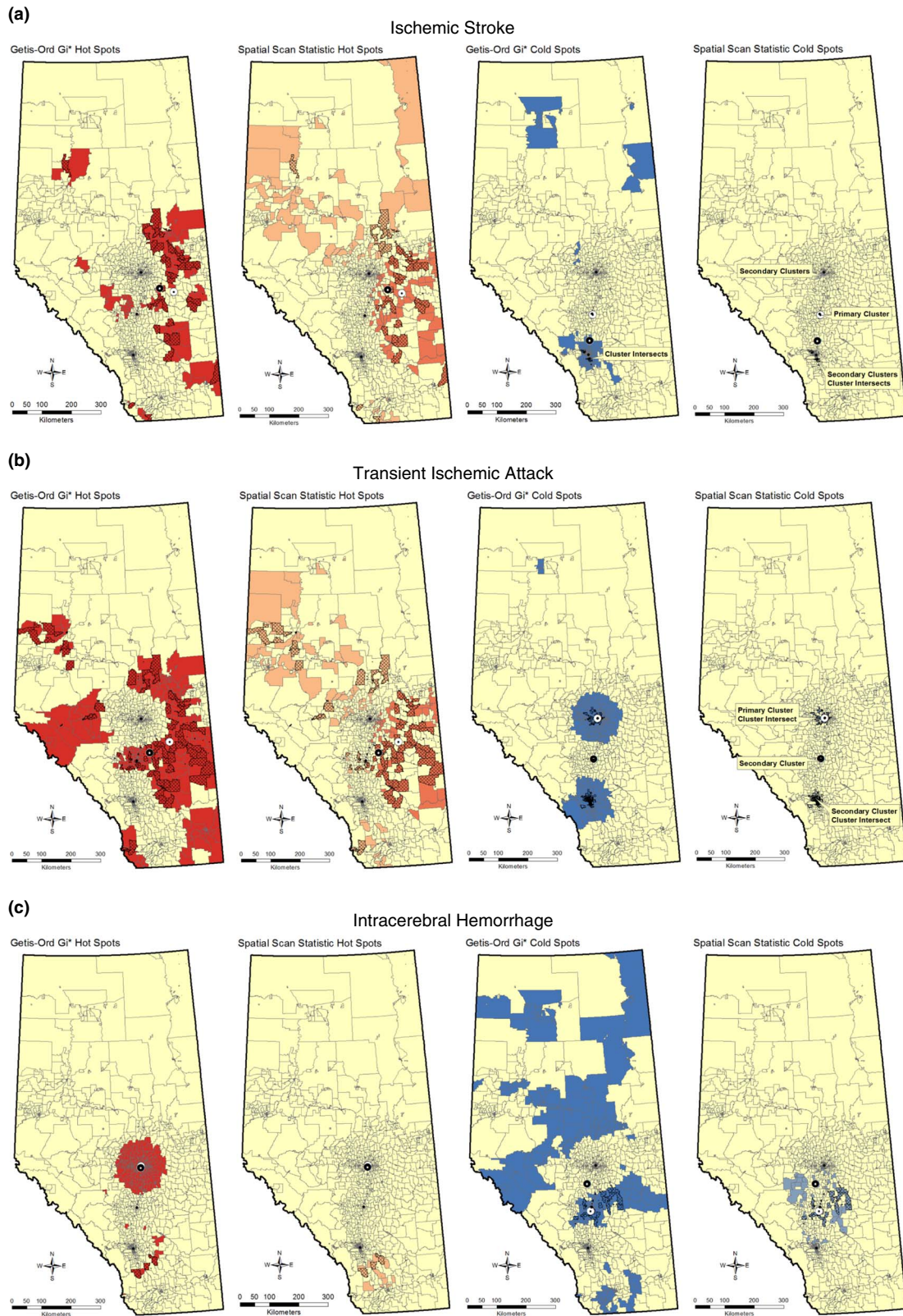
<sup>†</sup>Comprehensive stroke centre: computed tomography scan, tissue plasminogen activator, stroke team, neurosurgical/neuro-interventional expertise, acute stroke unit.

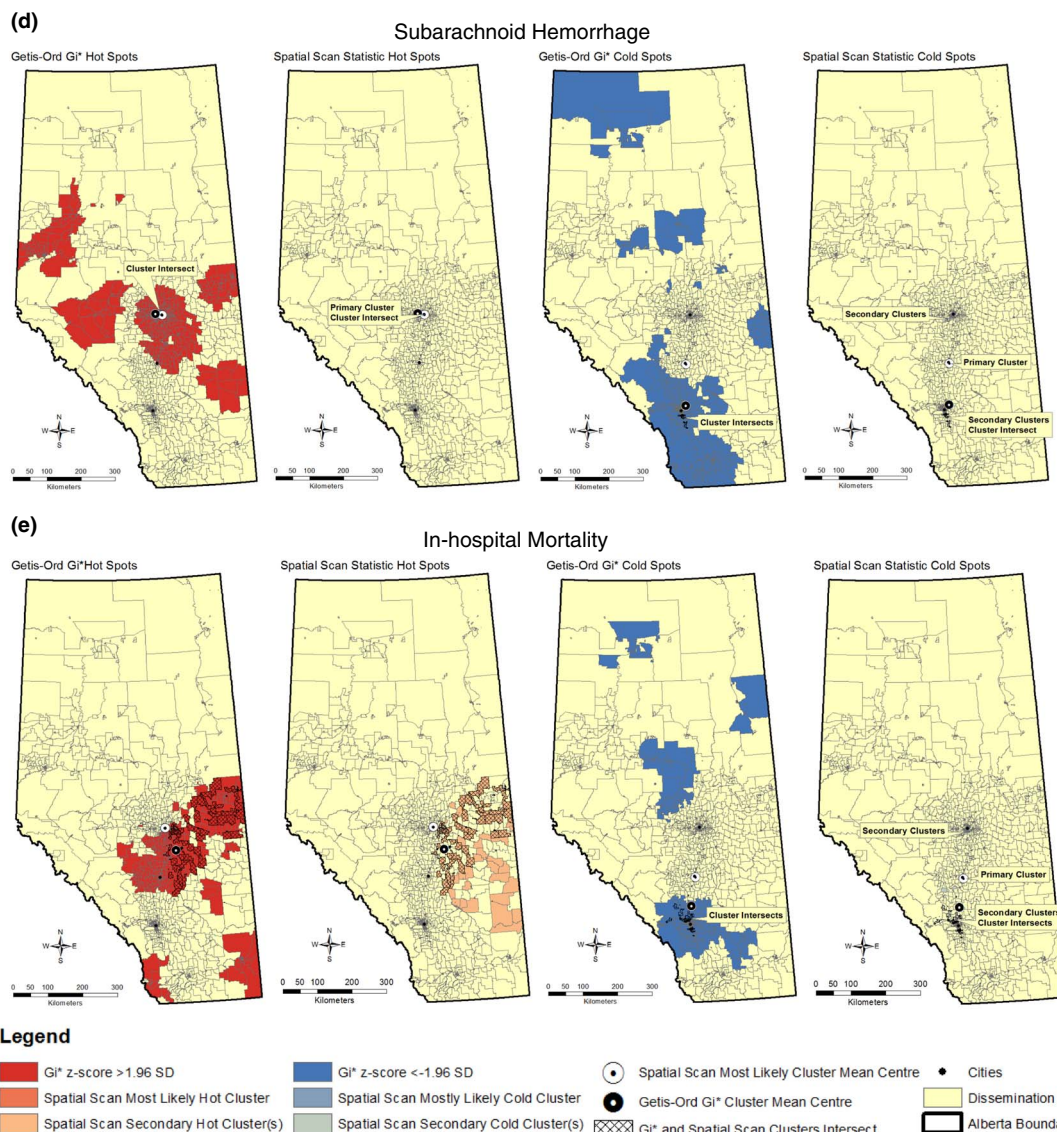
<sup>‡</sup>Primary stroke centre: computed tomography scan, tissue plasminogen activator, stroke expertise onsite or available by telehealth, acute stroke care provision.

*Spatial Scan statistic.* A purely spatial, discrete Poisson model was used for this analysis. Clusters of high or low rates were scanned for simultaneously using a circular scanning window set to a maximum size of 40% based on the population of the largest DA<sup>34</sup> and a size considered sufficiently large to ensure that potential small and large clusters could be included.<sup>29</sup>

## RESULTS

Ischemic stroke and TIA accounted for 86.8% of all strokes. Stroke types were evenly divided amongst men and women except for SAH, which occurred more often in women (62%). SAH also occurred at a younger age (56.7 years vs ischemic [73.2 years] and ICH [70.7 years]).





**Figure 2:** Hot and cold clusters of major stroke types and in-hospital mortality identified using Getis-Ord  $G_i^*$  and Spatial Scan statistics.

From 2002 to 2007, ICH rates decreased from 9.2% to 6.9% and SAH rates increased from 4.3% to 5.5% (Table 1). Over this time span, the proportion of the study population (by DA) with reported non-insulin-dependent DM, hypertension, and hypercholesterolemia increased. There was a trend to increased use of emergency medical services for transport, and a greater proportion of patients were transported to either a PSC or CSC by 2007. Mortality decreased from 13.4% in 2002 to 12.5% in 2007.

Median DA study population size was 410 persons (interquartile range 345, 525). Of 5357 DAs in Alberta, 3757 DAs reported  $\geq 1$  stroke event. A near-equal proportion of urban and rural DAs had rates of zero for ischemic stroke (~11%), ICH (~71%), SAH (~75%), and mortality (~57%). TIA was reported more frequently in rural DAs as having a rate of zero (48% vs 31%).

Statistically significant spatial clusters of stroke types and mortality are presented in Figure 2. Spatial mean centre, the point constructed from the average x and y coordinates for DA

centroids<sup>35</sup> of the  $G_i^*$  and Spatial Scan statistic “most-likely” clusters are plotted. Characteristics of confirmed clusters using the two methods are summarized in Table 2.

A noncontiguous hot spot of ischemic stroke was located diagonally across the centre of the province in rural DAs northeast of Westlock extending southward to the east of Calgary (Figure 2A). TIA hot spots exhibited a similar spatial pattern (Figure 2B). Contrary to what would be expected from epidemiological data, ischemic stroke and TIA hot spots had DAs with lower proportions of all reported risk factors excluding insulin-dependent DM as compared with DAs external to cluster boundaries (Table 3). DAs within ischemic stroke and TIA cold spots had the highest levels of reported risk factors (Table 3).

ICH and SAH hot spots were identified in the northernmost of the two major urban centres (Figure 2C, D). DAs within the ICH and SAH hot spots had higher proportions of strokes with all reported risk factors excluding insulin-dependent DM as compared with DAs outside cluster boundaries (Table 3). ICH and

**Table 2: Dissemination area characteristics of Getis-Ord  $G_i^*$  and spatial scan statistic cluster intersects**

Cluster intersect	DA (n)	Distance* (km)	Urban (%)	CSC (mean %)	PSC (mean %)	Other (mean %)	Primary location
Ischemic							
High	70	49.5	31.4	20.9	21.7	57.4	Noncontiguous band extending northeast of Westlock to east of Calgary
Low	127	89.2	97.6	80.7	1.7	17.6	Northwest and south Calgary
TIA							
High	294	77.4	59.9	23.2	28.2	48.5	Noncontiguous band extending from northeast of Westlock, south to Red Deer and Drumheller; Peace River region
Low	2445	138.8	98.4	70.0	4.9	25.1	Edmonton and Calgary and surrounding areas
ICH							
High	816	5.8	98.0	58.8	8.9	32.3	Edmonton; south of Calgary
Low	201	93.5	85.6	15.7	57.0	27.3	Noncontiguous band south of Camrose to south southwest of Red Deer
SAH							
High	737	20.4	98.2	58.2	9.1	32.8	Edmonton
Low	131	122.0	93.9	73.4	1.2	25.4	Noncontiguous band from south Calgary to High River
Mortality							
High	585	70.8	87.0	54.0	16.2	29.9	Noncontiguous band from Cold Lake to southeast of Red Deer
Low	335	84.0	95.8	82.5	1.3	16.1	North and south Calgary; Okotoks

\*Spatial mean distance: Euclidean or straight-line distance between  $G_i^*$  and spatial scan most-likely cluster spatial mean centres.

SAH were the two stroke types with higher proportions of reported hypertension within hot spot boundaries.

A noncontiguous mortality hot spot was identified in east central Alberta in a similar geographic distribution to the hot spot for ischemic stroke (Figure 2E). DAs within the mortality hot spot had lower proportions of reported non-insulin-dependent DM and hypertension compared with DAs external to the cluster boundary. DAs within the mortality cold spot had higher proportions of atrial fibrillation, hypertension, and hypercholesterolemia compared with DAs external to the cluster boundary (Table 3). Within the mortality hot spot, CSC was less frequently the transport destination and non-stroke centre was more frequently the transport destination (eTable 1, eTable 2).

A cluster analysis of risk factors identified hot spots in the two large urban centres with CSCs (Calgary: RR 2.59,  $p < 0.001$ ; Edmonton: RR 2.43,  $p < 0.001$ ) and one of the urban primary stroke centres (Medicine Hat:  $p < 0.05$ ) (Figure 3). There was an inverse relationship between risk factor hot spots and hot spots for ischemic stroke, TIA, and mortality (Figure 3, Table 3).

## DISCUSSION

Significant hot and cold spots of stroke types and mortality were identified (Figure 2). There was overlap of clusters across statistical methods and, in relation to the geographical expanse of the province, spatial mean centres of hot spots calculated for each statistic were not far apart in distance (Table 2).

### Ischemic Stroke and TIA

DAs within ischemic stroke and TIA hot spots had lower proportions of all reported risk factors excluding insulin-dependent DM compared with DAs outside of the hot spot and DAs within cold spots (Table 3). This is contrary to the established link

between stroke risk factors and stroke occurrence.<sup>36</sup> The most likely explanation is a lower incidence of recognition and reporting of stroke risk factors in these clusters with high rates of ischemic stroke and TIA. Furthermore, the high rates of reported stroke risk factors in cold spots for ischemic stroke and TIA suggests a strong and potentially causal link between recognition and reporting of stroke risk factors and occurrence of ischemic stroke and TIA. These findings serve to emphasize the importance of primary prevention. Clusters with lower proportions of reported risk factors may be indicative of populations with decreased access to primary care and/or that are less amenable to seeking and receiving preventive care. Forty percent of TIAs have magnetic resonance imaging evidence of infarction,<sup>37</sup> and differences in access to magnetic resonance imaging and extent of investigation as well as urban-rural differences in the coding accuracy of TIA versus ischemic stroke<sup>38</sup> may partially explain the urban bias for TIA cold spots located in Edmonton and Calgary (Figure 2B).

### ICH and SAH

ICH and SAH were the two major stroke types in which clusters of high rates had higher reported risk factors. ICH and SAH hot spots had DA populations with higher proportions of all reported risk factors excluding IDDM compared with DAs outside of cluster boundaries (Table 3). Underdiagnosis or subtherapeutic treatment of hypertension in this region may be important factors in addition to risk factors unaccounted for in the analysis. Data relating to incidence of smoking, a major risk factor for SAH, were not available.

### Mortality

DAs within the mortality hot spot had lower proportions of reported non-insulin-dependent DM, atrial fibrillation, and hypertension and higher proportions of hypercholesterolemia (Table 3).

**Table 3: High and low stroke and in-hospital mortality cluster intersects and Dissemination Area proportion of stroke events with risk factor(s).**

% of Stroke Events With Risk Factor(s) by DA (SD)	High		Low	
	Cluster	Not Cluster	Cluster	Not Cluster
<b>Ischemic</b>				
IDDM*	1.9 (13.5)	1.1 (10.4)	1.1 (10.3)	1.1 (10.6)
NIDDM†	17.1 (37.7)	19.5 (39.6)	16.6 (37.2)	19.5 (39.6)
Atrial Fibrillation‡	11.1 (31.4)	16.9 (37.5)	19.8 (39.8)	16.6 (37.2)
Hypertension§	44.3 (49.7)	55.9 (49.7)	63.3 (48.2)	55.2 (49.7)
Hypercholesterolemia <sup>¶</sup>	12.4 (33.0)	18.4 (38.8)	19.2 (39.4)	18.1 (38.5)
<b>Transient Ischemic Attack</b>				
IDDM	1.5 (12.1)	1.1 (10.3)	1.0 (9.6)	1.3 (11.4)
NIDDM	14.7 (35.4)	20.2 (40.2)	20.7 (40.5)	18.2 (38.6)
Atrial Fibrillation	11.7 (32.1)	17.5 (38.0)	19.2 (39.4)	14.2 (34.9)
Hypertension	42.8 (49.5)	57.6 (49.4)	62.1 (48.5)	49.1 (50.0)
Hypercholesterolemia	11.3 (31.7)	19.3 (39.5)	23.5 (42.4)	13.0 (33.7)
<b>Intracerebral Hemorrhage</b>				
IDDM	1.0 (9.5)	1.2 (10.8)	1.7 (12.9)	1.1 (10.3)
NIDDM	21.6 (41.2)	19.0 (39.2)	16.0 (36.7)	19.7 (39.8)
Atrial Fibrillation	20.7 (40.5)	15.8 (36.5)	15.3 (36.0)	16.7 (37.3)
Hypertension	61.2 (48.7)	54.2 (49.8)	48.1 (50.0)	56.0 (49.6)
Hypercholesterolemia	28.4 (45.1)	16.0 (36.6)	11.8 (32.2)	18.7 (39.0)
<b>Subarachnoid Hemorrhage</b>				
IDDM	0.9 (9.3)	1.2 (10.8)	0.8 (9.1)	1.1 (10.6)
NIDDM	21.7 (41.2)	19.0 (39.2)	16.4 (37.0)	19.5 (39.6)
Atrial Fibrillation	20.7 (40.5)	15.9 (36.5)	19.2 (39.4)	16.6 (37.2)
Hypertension	61.9 (48.6)	54.2 (49.8)	56.8 (49.5)	55.4 (49.7)
Hypercholesterolemia	29.0 (45.2)	16.1 (36.8)	16.9 (37.5)	18.2 (38.6)
<b>In-hospital Mortality</b>				
IDDM	1.1 (10.3)	1.1 (10.6)	1.1 (10.2)	1.1 (10.6)
NIDDM	18.5 (38.8)	19.6 (39.7)	16.7 (37.3)	19.6 (39.7)
Atrial Fibrillation	16.6 (37.2)	16.7 (37.3)	19.4 (39.5)	16.5 (37.1)
Hypertension	51.9 (50.0)	56.2 (49.6)	61.4 (48.7)	55.0 (49.8)
Hypercholesterolemia	20.8 (40.6)	17.6 (38.1)	18.3 (38.7)	18.1 (38.5)

\*IDDM - ICD-10 Code E10.

†NIDDM – ICD-10 Codes E11-E14.

‡Atrial Fibrillation – ICD-10 Code I48.

§Hypertension – ICD-10 Codes I10-I15.

¶Hypercholesterolemia – ICD-10 Codes E780, E781, E784, E785, E788, E789.

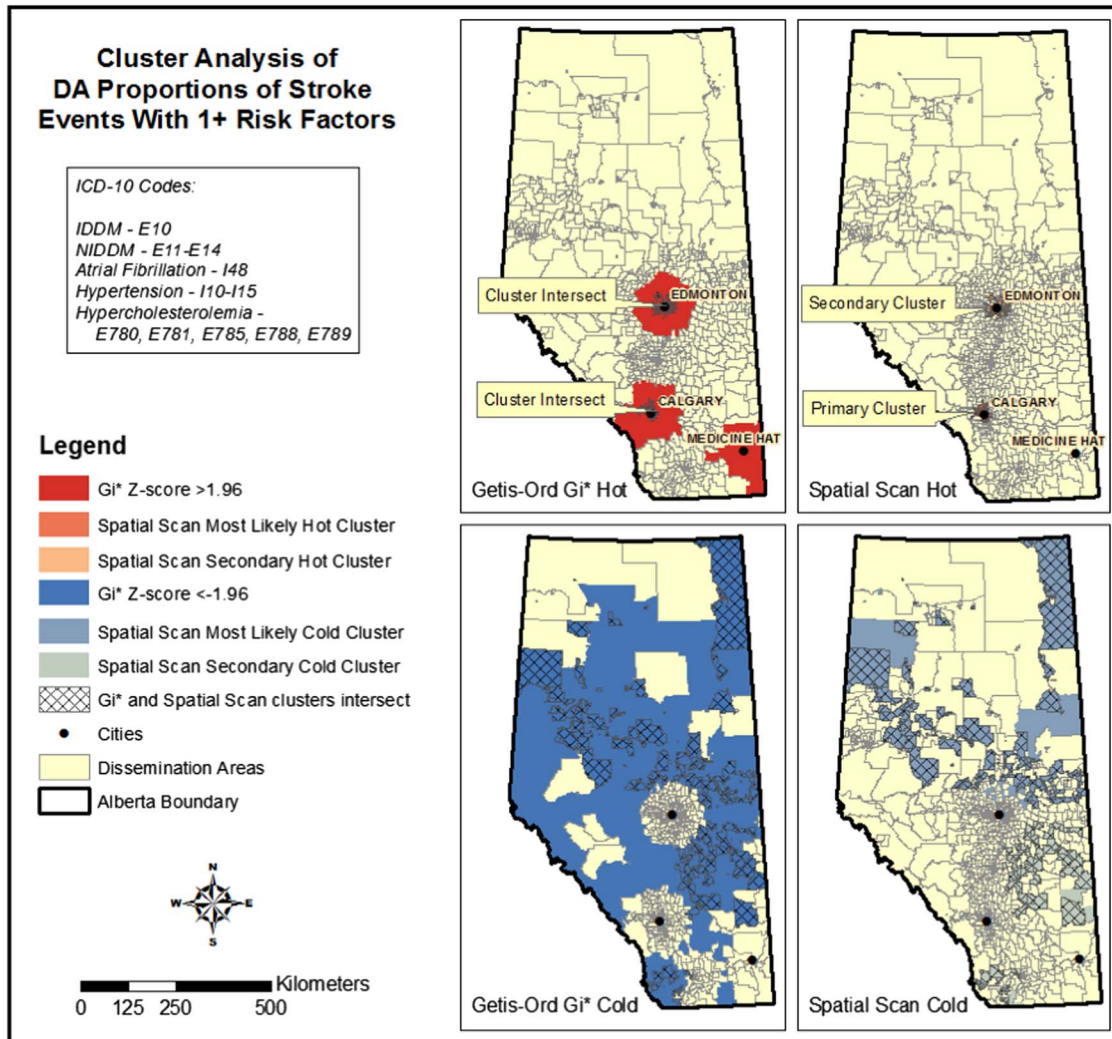
There was a higher occurrence of ischemic stroke, ICH, and SAH within the Spatial Scan “most-likely” high clusters (eTable 1). This finding raises questions regarding access to and efficacy of primary and secondary prevention programs and in-hospital stroke care in this region. Transport destination may have contributed to this finding as mortality hot spots had lower proportions of events transported to a CSC and a higher proportion of events transported to a non-designated stroke centre (eTable 2). Time from symptom onset to recanalization<sup>39-41</sup> and specialized stroke care<sup>42</sup> are crucial factors in morbidity and mortality after stroke. Prolonged transport times, suboptimal transport destination,<sup>6,43</sup> quality of care issues (i.e.

percentage receiving tissue plasminogen activator), or other unmeasured differences in stroke systems of care could account for mortality hot spots. Stroke severity, an important risk factor for mortality after stroke, could not be assessed because of a lack of clinical data.

### Risk Factors

Hot spots of risk factors were identified and located in Calgary, Edmonton, and Medicine Hat (Figure 3). These cities have well-established stroke prevention programs in place. This supports findings stemming from the cluster analyses of stroke types and





**Figure 3:** A cluster analysis of DA proportions of stroke events with one or more reported risk factors.

mortality suggesting that identification and reporting of risk factors is positively associated with significantly lower ischemic stroke and TIA occurrence and mortality. Spatial patterns may be related to differing risk distributions, regional differences in the coding of risk factors, or the type and quality of care that is associated with having a reported risk factor. Risk factors appeared to be less important in the spatial distribution of ICH and SAH.

This study highlights the power of using GIS to understand statistically significant regional variability in stroke occurrence, risk factors, and outcomes. These types of analyses can be particularly useful in identifying gaps in service, in planning and evaluation of health systems and health care delivery, and in (re)allocation of resources.<sup>7</sup> Strengths of this research study include that spatial analyses were performed by stroke type and that clusters were confirmed using two statistical methods. A broad regional analysis was conducted, but the geographic unit of analysis was sufficiently small to permit a more refined delineation of cluster boundaries and a later, more detailed examination of explanatory variables including local population characteristics and geography using linked census data and shape files. A polygon analysis removed spatial autocorrelation of events because of

underlying population density. The small numbers problem resulted in greater variation of rates in DAs with smaller populations, but this was mitigated by a spatial smoothing technique and, in areas of overlap, by the use of the Spatial Scan statistic. Findings were validated in that ischemic stroke and TIA had similar geographical distributions and that risk factor hot spots were located in regions with well-established stroke prevention programs and acute care service provision in place.

Potential limitations of this study include those associated with the use of administrative data: (1) lack of clinical detail, (2) regional differences with ICD-10 coding of stroke type, (3) regional differences in reporting of risk factors,<sup>44</sup> (4) inaccurate documentation of postal codes, and (5) errors associated with the collection of Statistics Canada census data including under- or overcoverage of specific populations.<sup>45</sup> Although the consistency for diagnosis typing for the most responsible diagnosis in CIHI-DAD data is high, the consistency for diagnosis typing of pre- and post-admit comorbidities is lower.<sup>45</sup> In Alberta, coding of atrial fibrillation, DM, and hypertension was found to have a higher degree of accuracy compared to coding of hyperlipidemia.<sup>38</sup> An important potential limitation is that the

postal code documented on the clinical health record may not accurately represent location of residence, location of stroke, or risk exposure. Residential mobility was not taken into account. Limitations associated with the spatial data analysis include: (1) position error, (2) DAs limited capacity to reveal what is happening in rural areas, (3) the modifiable areal unit problem (rate changes with change in geographic units), (4) the small numbers problem (greater variability in rate estimates), and (5) ecological and atomistic fallacy. It remains a useful but high-level tool for evaluation of disease incidence and outcomes.

## SUMMARY

Using whole population administrative data for a large geographic region of Canada integrated with GIS, this study identified significant regional variations in ischemic stroke, TIA, ICH, SAH and in-hospital mortality. Risk factor hot spots were identified in regions with significantly low ischemic, TIA, and in-hospital mortality rates. ICH and SAH did not have this same inverse association. Further studies using logistic and spatial regression techniques investigating associations between stroke and mortality hot spots that incorporate novel socio-demographic, environmental, and geographic (e.g. proximity to specialized care) predictors will inform public health policy with respect to the targeting of risk reduction strategies; enhancements to the provision of preventive stroke care, pre- and in-hospital (hyper)acute stroke care, and the future location of stroke centres.

## ACKNOWLEDGEMENTS AND FUNDING

We gratefully acknowledge support by Peter Peller, Head, Spatial Data and Numeric Services, University of Calgary and Vicki Stagg, Statistical Programmer, University of Calgary. This study was partly funded by the Heart & Stroke Foundation and the Hotchkiss Brain Institute.

## DISCLOSURES

SvR, TWJW, and MDH have nothing to disclose. SA is an employee at the University of Calgary.

## STATEMENT OF AUTHORSHIP

SvR wrote the primary manuscript, designed the study, obtained data, and conducted the analysis. TWJW, SA, and MDH provided help with elements of the study and provided critical revisions to the final manuscript.

## SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/cjn.2015.241>

## REFERENCES

1. Statistics Canada. Population and dwelling counts, for Canada, provinces and territories, 2011 and 2006 censuses. <http://www.statcan.gc.ca/tables-tableaux/sum-som/101/cst01/agrc42j-eng.htm>. Accessed June 2013.
2. Statistics Canada. Population, urban and rural, by province and territory. <http://www.statcan.gc.ca/tables-tableaux/sum-som/101/cst01/demo62a-eng.htm>. Accessed June 2013.
3. Jeerakathil T, Burrige D, Thompson G, Fang S, Hill MD. The Alberta Provincial Stroke Strategy: Improving Stroke Care Across Alberta. Interim Evaluation Report. Edmonton: Alberta Provincial Stroke Strategy. December 2010.
4. Alberta Health and Wellness. Interactive Health Data Application (IHDA). [http://www.ahw.gov.ab.ca/IHDA\\_Retrieval/](http://www.ahw.gov.ab.ca/IHDA_Retrieval/). Accessed October 2014.
5. Sarti C, Rastenyte D, Zygimantas C, Tuomilehto J. International trends in mortality from stroke, 1968 to 1994. *Stroke*. 2000; 31:1588-601.
6. Prabhakaran S, Ward E, John S, et al. Transfer delay is a major factor limiting the use of intra-arterial treatment in acute ischemic stroke. *Stroke*. 2011;42:1626-30.
7. Wennberg J, Gittelsohn A. Small area variations in health care delivery. *Science*. 1973;182:1102-8.
8. Boulos MN. Towards evidence-based, GIS-driven national spatial health information infrastructure, and surveillance services in the United Kingdom. *Int J Health Geogr*. 2004;3:1-50.
9. Pedigo A, Aldrich T, Odoi A. Neighborhood disparities in stroke and myocardial infarction mortality: a GIS and spatial scan statistics approach. *BMC Public Health*. 2011;11:644.
10. Han D, Carrow SS, Rogerson PA, Munschauer FE. Geographical variation of cerebrovascular disease in New York State: the correlation with income. *Int J Health Geogr*. 2005;4:25.
11. Health Canada. Canada Health Act Annual Report 2007-2008. Introduction. <http://www.hc-sc.gc.ca>. Accessed June 2013.
12. Canadian Institute for Health Information (CIHI). Discharge Abstract Database (DAD) Metadata. [http://www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/acute+care/dad\\_metadata](http://www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/acute+care/dad_metadata). Accessed June 2013.
13. Canadian Institute for Health Information (CIHI). National Ambulatory Care Reporting System (NACRS). [http://www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/emergency+care/NACRS\\_METADATA](http://www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/emergency+care/NACRS_METADATA). Accessed June 2013.
14. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th revision. <http://apps.who.int/classifications/icd10/browse/2010/en>. Accessed December 2013.
15. Centers for Disease Control and Prevention. Stroke facts. <http://www.cdc.gov/stroke/facts.htm>. Accessed October 2014.
16. University of Calgary. Spatial and Numeric Data Services (SANDS). <http://library.ucalgary.ca/sands>. Accessed June 2013.
17. Lai PC, So FM, Chan KW. Spatial epidemiological approaches in disease mapping and analysis. Boca Raton, FL: CRC Press; 2009.
18. Beyer KMM, Tiwari C, Rushton G. Five essential properties of disease maps. *Annals AAG*. 2012;102:1067-75.
19. Buescher PA. Age-adjusted death rates. Statistical primer. State Center for Health Statistics. Raleigh, NC: North Carolina Department of Health and Human Services August 1998: 1-11.
20. Klein RJ, Schoenborn CA. Age adjustment using the 2000 Projected U.S. population. Healthy People 2010 statistical notes. Hyattsville, MD: National Center for Health Statistics. 2001;20:1-10.
21. StataCorp LP. Stata/IC 12.1 for Windows. College Station, TX: 1985-2011 (rev. 20 March 2013).
22. Statistics Canada Census Dictionary. Dissemination area (DA). <http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo021-eng.cfm>. Accessed January 2014.
23. du Plessis V, Beshiri R, Bollman RD, Clemenson H. Definitions of "rural." Statistics Canada, Ottawa, ON; 2002.
24. Statistics Canada. Postal Code Conversion File (PCCF), reference guide 2006, 2011. <http://www.statcan.gc.ca/bsolc/olc-cel/olc-el?catno=92F0153G&CHROPG=1&lang=eng>. Accessed June 2013.
25. Environmental Systems Research Institute. ArcMAP 10.0. ArcGIS Desktop 10. Redlands, CA: ESRI Inc.; 1999-2010.
26. Statistics Canada. Data quality and confidentiality standards and guidelines (public). <http://www12.statcan.gc.ca/census-recensement/2011/ref/DQ-QD/conf-eng.cfm>. Accessed September 2013.
27. Environmental Systems Research Institute. Transverse mercator. ArcGIS Resource Center Desktop 10 website. [http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Transverse\\_Mercator/003r00000046000000/](http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Transverse_Mercator/003r00000046000000/). Accessed June 2013.
28. Environmental Systems Research Institute. How hot spot analysis (Getis-Ord Gi\*) works. ArcGIS Resource Center Desktop 10. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/005p00000011000000>. Accessed June 2013.
29. Kulldorff M. SaTScan user guide for version 9.0. 2010. <http://www.satscan.org/>. Accessed June 2013.

30. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geograph Anal.* 2006;38:5-22.
31. Environmental Systems Research Institute. Selecting a conceptualization of spatial relationships: best practices. ArcGIS Resource Center. Desktop 10.0. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p00000005000000>. Accessed September 2013.
32. Environmental Systems Research Institute. Selecting a fixed distance band value. ArcGIS Resource Center Desktop 10. [http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Modeling\\_spatial\\_relationships/](http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Modeling_spatial_relationships/). Accessed June 2013.
33. Environmental Systems Research Institute. How Hot Spot Analysis (Getis-Ord  $G_i^*$ ) works. ArcGIS Resource Center Desktop 10. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p0000000110000000>. Accessed June 2013.
34. Walsh SJ, DeChello LM. Geographical variation in mortality from systemic lupus erythematosus in the United States. *Lupus.* 2001;10:637-46.
35. Environmental Systems Research Institute. Mean center (spatial statistics). ArcGIS Resource Center Desktop 10 website. <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p0000000180000000>. Accessed June 2013.
36. Lackland DT, Roccella EJ, Deutsch AF, et al. Factors influencing the decline in stroke mortality. a statement from the American Heart Association/American Stroke Association. *Stroke* December 2013. <http://stroke.ahajournals.org/content/early/2013/12/05/01.str.0000437068.30550.cf>. Accessed January 2014.
37. Kidwell CS, Alger JR, Di Salle F, et al. Diffusion MRI in patients with transient ischemic attacks. *Stroke.* 1999;30:1174-80.
38. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using International Classification of Diseases, Revisions 9 and 10. *Stroke.* 2005;36:1776-81.
39. National Institute for Neurological Disorders and Stroke r-tPA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* 1995;333:1158-8.
40. Hacke W, Kaste M, Bluhmki E, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med.* 2008;359:1317-29.
41. Mazighi M, Serfaty JM, Labreuche J, et al. Comparison of intravenous alteplase with a combined intravenous-endovascular approach in patients with stroke and confirmed arterial occlusion (RECANALISE STUDY): a prospective cohort study. *Lancet Neurol.* 2009;8:802-9.
42. Stroke Unit Trialists' Collaboration. Collaborative systematic review of the randomised trials of organised inpatient (stroke unit) care after stroke. *BMJ.* 1997;314:1151-9.
43. Schwamm LH, Pancioli A, Acker JE, et al. Recommendations for the establishment of stroke systems of care. *Circulation.* 2005;111:1078-91.
44. Canadian Institute for Health Information. CIHI Data Quality Study of the 2009 – 2010 Discharge Abstract Database. Revised June 2012. Accessed February 2015. [https://secure.cihi.ca/free\\_products/Reabstraction\\_june19revised\\_09\\_10\\_en.pdf](https://secure.cihi.ca/free_products/Reabstraction_june19revised_09_10_en.pdf).
45. Statistics Canada. Population coverage error. [http://www12.statcan.gc.ca/census-recensement/2006/ref/tp-guides/tp/coverage-couverture/cov-couv\\_p03-eng.cfm](http://www12.statcan.gc.ca/census-recensement/2006/ref/tp-guides/tp/coverage-couverture/cov-couv_p03-eng.cfm). Accessed July 2013.