Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

A pan-Canadian measure of active living environments using open data

by Thomas Herrmann, William Gleckner, Rania A. Wasfi,
Benoît Thierry, Yan Kestens and Nancy A. Ross

Release date: May 15, 2019

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

A pan-Canadian measure of active living environments using open data

by Thomas Herrmann, William Gleckner, Rania A. Wasfi, Benoît Thierry, Yan Kestens and Nancy A. Ross

Abstract

Background: Neighbourhood environments that support active living, such as walking or cycling for transportation, may decrease the burden of chronic conditions related to sedentary behaviour. Many measures exist to summarize features of communities that support active living, but few are pan-Canadian and none use open data sources that can be widely shared. This study reports the development and validation of a novel set of indicators of active living environments using open data that can be linked to national health surveys and can be used by local, regional or national governments for public health surveillance.

Data and methods: A Geographic Information System (GIS) was used to calculate a variety of measures of the connectivity, density and proximity to destinations for 56,589 dissemination areas (DAs) across Canada (2016 data). Pearson correlation coefficients were calculated to assess the association between each measure and the rates of walking to work and taking active transportation to work (a combination of walking, cycling and using public transportation) from census data. The active living environment measures selected for the final database were used to classify the DAs by the favourability of their active living environment into groups by k-medians clustering.

Results: All measures were correlated with walking-to-work and active-transportation-to-work rates at the DA level, whether they were derived using proprietary or open data sources. Coverage of open data was consistent across Canadian regions. Three measures were selected for the Canadian Active Living Environments (Can-ALE) dataset based on the correlation analysis, but also on the principles of suitability for a variety of community sizes and openly available data: (1) three-way intersection density of roads and footpaths derived from OpenStreetMap (OSM), (2) weighted dwelling density derived from Statistics Canada dwelling counts and (3) points of interest derived from OSM. A measure of access to public transportation was added for the subset of DAs in larger urban areas and was strongly related to active-transportation-to-work rates. Active-transportation-to-work rates were graded, in steps, by the five Can-ALE groups derived from the cluster analysis, although walking-to-work rates exceeded the national average only in the most favourable active living environments.

Interpretation: Open data may be used to derive measures that characterize the active living environments of Canadian communities.

Keywords: active living environments, active transportation, walking to work, open data, public health surveillance
DOI: https://www.doi.org/10.25318/82-003-x201900500002-eng

Modifiable elements of neighbourhood environments (e.g., number of sidewalks, proximity to commercial services, population density) can increase rates of active transportation (walking and cycling for the purpose of transportation, and using public transportation).[1-5] Public health and urban planning researchers often measure three characteristics of communities that support active travel: higher street connectivity (e.g., intersection density, route directness), higher density (e.g., population density, dwelling density), and greater numbers and diversity of nearby destinations.[4,6,7] Canadian research suggests that exposure to these favourable “active living environments” is associated with more optimal markers of health, including more optimal systolic blood pressure,[8] decreased obesity, overweight and diabetes prevalence,[9] and improved body mass trajectories among men.[10]

A national metric of active living environments is desirable to facilitate the direct comparison of communities, national surveillance of population health and data linkage with existing Canadian national health surveys (e.g., Canadian Community Health Survey, Canadian Health Measures Survey) and investigator-led cohort studies. Currently, very few pan-Canadian measures of the active living environment exist and those that do are not readily accessible or free to use.[11]

The purpose of this paper is to describe the development of the Canadian Active Living Environments (Can-ALE) dataset: a Canada-wide set of four individual and four summary measures that characterize the favourability of active living environments in Canadian communities at the dissemination-area (DA) level (Figure 1). This study reports on analyses which guided the selection of measures and derivation data sources for the dataset. The objective was to produce a national database entirely from open data and to evaluate the performance of open data compared with traditional or proprietary sources. A 2006 version and a 2016 version of Can-ALE are available to download online.

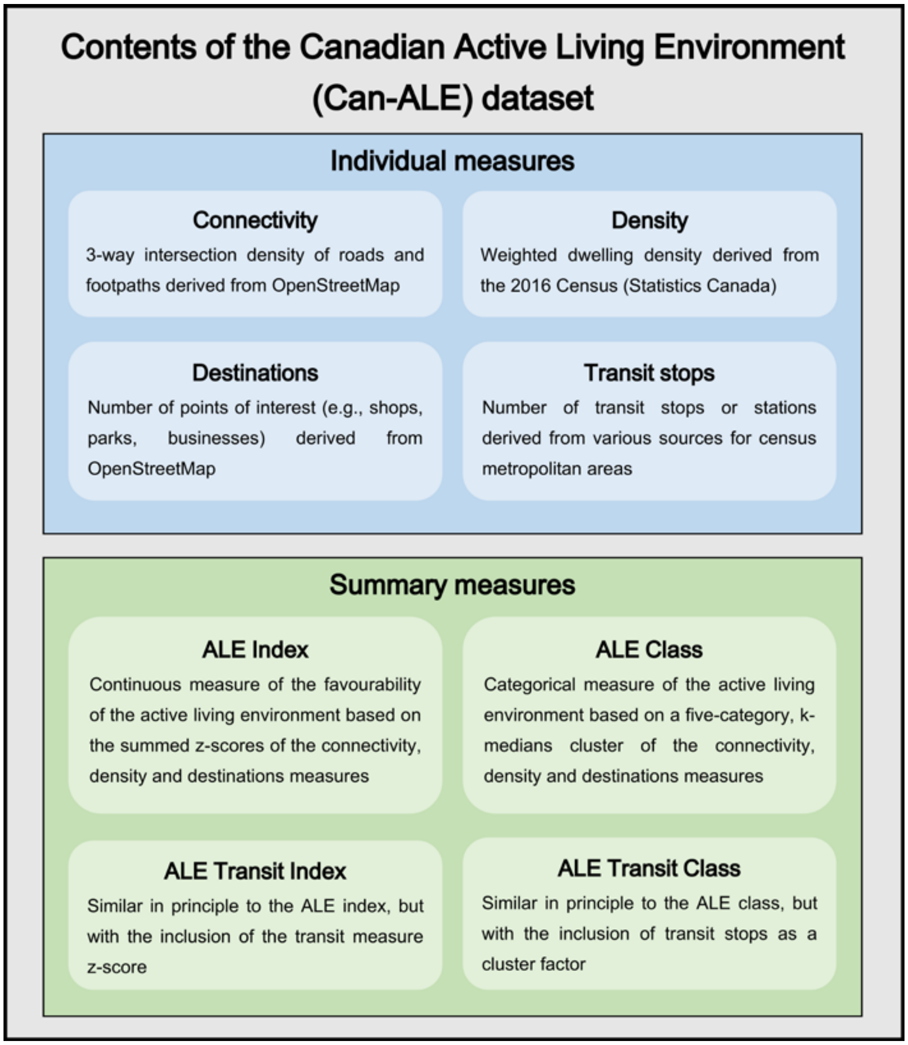
Data and methods

Study design

Multiple candidate measures of active living environments were created to represent connectivity, density and access to destinations for 56,589 DAs (2016 data) in Canada using a Geographic Information System (GIS). The candidate measures were derived using ArcMap v10.5, and the measures included in the final Can-ALE dataset were derived using PostGIS v2.3.3, a GIS extension for the PostgreSQL object-relational database

Authors: Thomas Herrmann (thomas.herrmann@mcgill.ca), William Gleckner, Rania A. Wasfi and Nancy A. Ross are with the Department of Geography at McGill University. Rania A. Wasfi, Benoît Thierry and Yan Kestens are with the Département de médecine sociale et préventive at the Centre de recherche du Centre hospitalier de l’Université de Montréal (CRCHUM) and the École de santé publique at the Université de Montréal (ESPUM).

Figure 1
Individual measures and summary measures included in the 2016 Canadian Active Living Environments (Can-ALE) dataset



management system (coding available from authors upon request). DAs are geographic units defined by Statistics Canada. They are the smallest standard geographic area for which complete census data are disseminated across Canada and have populations of 400 to 700 people.[12] A correlation analysis was used to inform decisions on the measures used in the final pan-Canadian dataset. These decisions were guided by principles according to which the selected measures were (1) associated with walking rates, (2) suitable to a variety of built environments (e.g., urban, suburban, rural), and (3) openly available to researchers and the public health community. A k-medians cluster analysis—an iterative process used to assign observations (in this case DAs) to a group with the closest within-group median values and maximal between-group median values—was conducted to guide the creation of a categorical variable characterizing the favourability of the active living environment of Canadian communities for both research and public health surveillance.

Unit of analysis

The principal unit of analysis was a circular, one-kilometre buffer around the centroid of the DA. Previous studies have shown differences in associations between measures and walking behaviour or health outcomes according to the type of geographic unit used to derive them.[13-15] As a result, buffer choices are often debated by researchers. To determine the most appropriate buffer shape and size for the national dataset, 12 active-living measures were derived using 4 types of buffers, which varied according to shape (circular versus street-network-based) and radius (500 metres versus 1 kilometre) for a subset of DAs on the island of Montréal. Using walking-for-transportation rates derived from the 2013 Montréal Origin-Destination Survey, 10 of the 12 measures had the highest correlations with walking rates when derived using the circular, one-kilometre buffer. In general, measures derived from the 500-metre network buffers were the least associated with walking rates (results available from the authors). Network buffers, by design, are a type of connectivity measure[16,17] and their use essentially causes connectivity to be counted twice in summary measures. The strength of the associations with walking rates and the confounding influence of network buffers led to the adoption of one-kilometre, circular buffers. Circular buffers are also favourable for computing resources relative to network buffers.

Data

Active living environment measures

The active living environment measures were derived from three primary sources: Statistics Canada (2016 Road Network File, population and dwelling counts), 2017 DMTI Spatial (Enhanced Points of Interest [EPOI] file) and OpenStreetMap (OSM) (road and footpath network, points of interest [POIs]). The 2016 Statistics Canada Road Network File is a digital representation of Canada's national road network.[18] The 2017 DMTI EPOI file contains approximately 1.7 million geocoded records of businesses and institutions (e.g., schools, hospitals) and other selected features.[19] DMTI Spatial is a proprietary data source available to Canadian university researchers through special research licensing and is the mainstay of destination-based measures. OSM is a digital world map predicated on open data licensing and user-generated data.[20] OSM contains spatial data for transportation infrastructure (e.g., streets, footpaths, rail stations), natural and topological features, POIs (e.g., shops, schools), and administrative boundaries. While most researchers in Canada use proprietary datasets to derive active living environment measures, OSM data have been validated with government and proprietary datasets and used for research in Canada[21] and other Western countries.[22,23] OSM data were obtained from the Geofabrik data extract server in October 2017.[24]

Two connectivity measures (three-way intersection density and four-way intersection density) were calculated three times, each time using a different methodological approach or derivation source. First, intersection density was derived using road intersections from the 2016 Statistics Canada Road Network File. Second, intersection density was calculated using OSM road features, which are identical, in principle, to the Statistics Canada file. Third, intersection density was calculated from OSM intersections of roads, footpaths and recreational trails. Limited-access roads (e.g., highways and freeways) were removed from each of the road files before calculating intersection density.

Four density measures were derived, and they varied in terms of underlying data (population count versus dwelling count) and calculation method (gross density versus weighted density). Gross population density and gross dwelling density were derived by dividing the population or dwelling count by the area of the DA. Weighted population density and weighted dwelling density were calculated by aggregating the population or dwelling count of each DA within the buffer and dividing by the area of the buffer. If a DA fell entirely within the buffer, its entire population and its dwellings were added to the count for the buffer. If a DA fell partially within the buffer, the population or dwelling counts of the DA were adjusted according to the proportion of the DA within the buffer. For example, if a DA with 1,000 inhabitants was only 25% within the buffer area, the buffer was assigned 250 of its inhabitants. Major coastal water bodies and DAs with no data were excluded from this calculation. The full and adjusted values were summed to determine an approximate population and dwelling count for the buffer. Population and dwelling counts were obtained from the 2016 Census conducted by Statistics Canada.

The POI measures differed only by underlying data source. The DMTI EPOI dataset contains the geographic coordinates of businesses and government institutions in Canada, as well as a few additional special POIs (e.g., provincial and national parks, border crossings). OSM POIs consist of a wide variety of mapped features (e.g., schools, shops, parks, benches, ATMs, soccer fields) obtained from administrative data sources and user contributions. Both measures were calculated by counting the number of POIs within the one-kilometre circular buffer. Additionally, a public transportation stop measure was derived for DAs within census metropolitan areas (CMAs)—urban areas with a population of 100,000 or more. This measure reflects the number of transit stops or stations (e.g., bus stops, light rail stations, subway stations) within the DA buffer. The coordinates of each transit stop were obtained in December 2017 from datasets on municipal or transit agency websites that conform to the General Transit Feed Specification (GTFS) data standard. Areas outside of CMAs were omitted because of inconsistent availability of GTFS data. The transit stop measure was derived for 35,338 DAs (97.1% of DAs within CMAs). Spatial data on transit stop locations were not found for some smaller CMAs: Belleville, Ontario; Peterborough, Ontario; Saguenay, Quebec; and Trois-Rivières, Quebec.

Active transportation rates

Two active transportation rates were calculated from Question 43 a) of the 2016 Census long-form questionnaire.[25] This question asks members of the labour force aged 15 or older with a fixed workplace address how they get to work. These rates were calculated by aggregating the number of pedestrians, cyclists and public transportation users for all DAs intersecting the circular, one-kilometre buffer of the DA centroid. If a DA fell only partially within the buffer, a smaller number of commuters proportional to the area of the DA within the buffer was aggregated. For example, if a DA was 25% within the buffer area and reported 40 pedestrian commuters, it was estimated that only

What is already known on this subject?

- Active living environments characterized by higher street connectivity, residential density and land use mix or proximity to destinations are associated with higher rates of walking and cycling for transportation.
- Exposure to these environments is associated with increased physical activity levels and more optimal markers of health.

What does this study add?

- This study describes the development of a free-to-use, pan-Canadian database of active living environment measures—the first of its kind in Canada—available for linkage to national datasets or investigator-driven cohorts.
- Measures derived from open data in Canada are similarly associated with walking-to-work and active-transportation-to-work rates as those derived from traditional or proprietary sources, or more strongly associated.
- Active-transportation-to-work rates are graded by active living environment clusters, while walking-to-work rates exceed the national average only in the most favourable active living environments.

10 pedestrian commuters lived within the DA buffer. The walking-to-work rate reflects the proportion of this population that reports walking as their primary mode of transportation to work. The active-transportation-to-work rate reflects the proportion of the same population that walks, cycles or uses public transportation to get to work. Public transportation use was included as active transportation, as public transportation has been shown to generate physical activity via walking to and from transit stops.[26-28]

Statistical methods

Pearson correlation coefficients were calculated to assess the association between walking-to-work rates, active-transportation-to-work rates and the 13 active living environment candidate measures (Table 1). To assess whether there was a regional bias in the coverage of user-contributed OSM data, the proportion of OSM POIs to DMTI POIs in five geographical regions (Atlantic, Quebec, Ontario, Prairies and British Columbia) was calculated. It is important to note that OSM and DMTI POI datasets contain different types of records. The DMTI EPOI file mainly consists of records of businesses, while OSM POIs contain certain commercial businesses (e.g., grocery stores, restaurants, clothing stores) as well as features of public spaces and streets (e.g., benches, picnic tables, tennis courts, food stalls, ATMs, postal boxes). Accordingly, the number of points in each dataset cannot be directly compared, and instead, the proportion of OSM to DMTI points was examined for different regions of the country. The strength of the correla-

Table 1
Summary of rates and measures used in correlation analysis

Measure or unit	Definition or method	Source
Walking-to-work rates	The proportion of employed persons (ages 15 and older) who primarily commute by walking	2016 Census (Statistics Canada)
Active-transportation-to-work rates	The proportion of employed persons (ages 15 and older) with a fixed workplace address who primarily commute by walking, cycling or using public transportation	2016 Census (Statistics Canada)
Three-way intersection density	The number of ≥ three-way intersections on roads per square kilometre, excluding road segments identified as limited-access highways	Road Network File (Statistics Canada, 2016)
Three-way intersection density	The number of ≥ three-way intersections on roads per square kilometre, excluding roads classified as motorways (highways, freeways) or slip roads (e.g., highway entrance and exit ramps)	Road features (OpenStreetMap, downloaded 2017)
Three-way intersection and footpath density	The number of ≥ three-way intersections on roads, footpaths and trails per square kilometre, excluding roads classified as motorways (highways, freeways) or slip roads (e.g., highway entrance and exit ramps)	Road features (OpenStreetMap, downloaded 2017)
Four-way intersection density	The number of ≥ four-way intersections on roads per square kilometre, excluding road segments identified as limited-access highways	Road Network File (Statistics Canada, 2016)
Four-way intersection density	The number of ≥ four-way intersections on roads per square kilometre, excluding roads classified as motorways (highways, freeways) or slip roads (e.g., highway entrance and exit ramps)	Road features (OpenStreetMap, downloaded 2017)
Four-way intersection and footpath density	The number of ≥ four-way intersections on roads, footpaths and trails per square kilometre, excluding roads classified as motorways (highways, freeways) or slip roads (e.g., highway entrance and exit ramps)	Road features (OpenStreetMap, downloaded 2017)
Gross population density	Population per square kilometre of the dissemination area (DA)	2016 Census (Statistics Canada)
Weighted population density	Population per square kilometre of the DA buffer, aggregated from each DA within the polygonal buffer and weighted according to the proportion of the DA within the buffer	2016 Census (Statistics Canada)
Gross dwelling density	Dwellings per square kilometre of the DA	2016 Census (Statistics Canada)
Weighted dwelling density	Dwellings per square kilometre of the DA buffer, aggregated from each DA within the polygonal buffer and weighted according to the proportion of the DA within the buffer	2016 Census (Statistics Canada)
DMTI points of interest	The number of points of interest (e.g., businesses, schools, hospitals)	Enhanced Points of Interest (DMTI Spatial, 2017)
OpenStreetMap points of interest	The number of points of interest (e.g., shops, schools, parks)	Points of interest (OpenStreetMap, downloaded 2017)
Transit stops	The number of public transportation stops or stations (e.g., bus stops, light rail stations, subway stations)	Various municipal and transit agency websites

tions in this analysis and the spatial and regional distribution of the input datasets were the primary findings that informed the selection of Can-ALE measures.

K-medians clustering was performed to classify Canadian DAs into five categories that characterize the favourability of the active living environment. K-medians clustering is a partition clustering method where the user specifies the number of clusters (k), then an iterative process is used to assign observations (in this case DAs) to a group with the closest median values. The cluster analysis was based on the three pan-Canadian DA-level measures selected for Can-ALE: (1) three-way intersection density of roads and footpaths derived from OSM, (2) weighted dwelling density derived from Statistics Canada dwelling counts, and (3) POIs derived from OSM. The k-medians clustering method was used, as the right skewness of the active living environment measures made the k-means approach unsuitable. A descriptive analysis compared walking-to-work rates; active-transportation-to-work rates; and average connectivity, density and POI values by cluster group.

Results

Correlation analysis

Most measures were modestly (R = 0.20 to 0.39) to moderately (R = 0.40 to 0.59) correlated with walking-to-work rates and moderately to strongly (R ≥ 0.60) correlated with active-transportation-to-work rates (Table 2), regardless of derivation source. Of the connectivity measures, the four-way intersection density of roads and footpaths derived from OSM was the most strongly associated with walking-to-work rates (R = 0.42). The three-way intersection density of roads and footpaths derived from OSM was one of two connectivity measures that were most strongly associated with active-transportation-to-work rates (R = 0.68). However, four-way intersections were sparsely distributed throughout Canada (especially in rural and certain suburban areas) and this made the four-way intersection measure less desirable than the three-way intersection measure for a pan-Canadian dataset. For instance, 55.4% of rural DA buffers in Canada (i.e., those outside CMAs or census areas) had no four-way intersections, whereas only 33.8% of rural DA buffers had no three-way intersections.

Dwelling density measures were more strongly associated with walking-to-work rates than population density measures, and the weighted density derivation method was more strongly associated with both walking-to-work and active-transportation-to-work rates. Gross population density (R = 0.23) and gross dwelling density (R = 0.29) were both modestly associated with walking-to-work rates, while both weighted population density (R = 0.82) and weighted dwelling density (R = 0.82) were strongly associated with active-transportation-to-work rates.

Both of the DMTI (R = 0.55) and OSM (R = 0.53) POI measures were similarly associated with walking-to-work rates, and the OSM POIs were somewhat more strongly associated with active-transportation-to-work rates (R = 0.66) than the DMTI POIs

Table 2
Summary statistics and correlations between active living environment measures and walking-to-work and active-transportation-to-work rates for dissemination area buffers in Canada (n = 54,678)

Measure	Correlation with walking-to-work rates	Correlation with active-transportation-to-work rates	Summary statistics: Number of dissemination areas	Mean	Minimum	Maximum
Three-way intersection density (Statistics Canada)	**0.28**	**0.65**	56,589	30.1	0.0	180.3
Three-way intersection density (OSM)	**0.25**	**0.63**	56,589	30.3	0.0	168.5
Three-way intersection and footpath density (OSM)	**0.39**	**0.68**	56,589	48.8	0.0	1,081.8
Four-way intersection density (Statistics Canada)	**0.40**	**0.68**	56,589	9.1	0.0	99.9
Four-way intersection density (OSM)	**0.37**	**0.67**	56,589	9.3	0.0	89.7
Four-way intersection and footpath density (OSM)	**0.42**	**0.64**	56,589	16.0	0.0	581.6
Gross population density	**0.23**	**0.53**	54,800	3 426.1	0.0	482,859.2
Gross dwelling density	**0.30**	**0.54**	54,800	1 489.3	0.0	245,618.9
Weighted population density	**0.29**	**0.82**	56,089	2 213.2	0.0	25,530.9
Weighted dwelling density	**0.40**	**0.82**	56,089	941.7	0.0	14,747.6
DMTI points of interest (DMTI POIs)	**0.55**	**0.57**	56,589	191.2	0.0	12,396.0
OpenStreetMap points of interest (OSM POIs)	**0.53**	**0.66**	56,589	53.7	0.0	2,119.0
Transit stops*	**0.45**	**0.70**	35,338	46.8	0.0	252.0
ALE Index	**0.47**	**0.78**	56,089	0.1	-2.1	43.2

* Transit stop correlation coefficients were derived for the 35,338 dissemination areas (DAs) in census metropolitan areas.

Notes: Walk-to-work and active-transportation-to-work rates are based on Question 43 a) of the 2016 Census, which was asked to employed members of the labour force aged 15 and older with a fixed workplace address ("How did this person **usually** get to work?").

Sources: OpenStreetMap; Statistics Canada Road Network File; Statistics Canada, 2016 Census; DMTI Spatial, Inc.; Barrie Transit; Brantford Open Data Portal; Calgary Transit; Edmonton Transit System; Strathcona County Transit; St. Albert Transit; Greater Sudbury Transit; Guelph Transit; Halifax Transit; Hamilton Street Railway; Burlington Transit; BC Transit; City of Kingston; Grand River Transit; Lethbridge Open Data Portal; London Transit Commission; Codiac Transpo; Agence métropolitaine de transport; Société de transport de Montréal; Société de transport de Laval; Réseau de transport de Longueuil; Durham Region Transit; OC Transpo; Société de transport de l'Outaouais; Réseau de transport de la Capitale; Société de transport de Lévis (open data site); City of Regina; The City of Saint John Open Data Catalogue; City of Saskatoon; Société de transport de Sherbrooke; Regional Municipality of Niagara; Metrobus Transit (St. John's, NL); Thunder Bay Transit; Toronto Transit Commission; GoTransit; MiWay; Brampton Transit; York Region; Oakville Transit; TransLink; City of Windsor; and Winnipeg Transit.

(R = 0.57). A summary measure of the favourability of the active living environment measures or "ALE Index" (for definition, see Figure 1) was moderately associated with walking-to-work rates (R = 0.47) and strongly associated with active-transportation-to-work rates (R = 0.78). Within the subset of urban DAs (n = 35,319), the transit stop measure was correlated with both walking-to-work (R = 0.45) and active-transportation-to-work rates (R = 0.70). OSM data coverage varied minimally across five different regions of Canada (Atlantic, Quebec, Ontario, Prairies and British Columbia). The proportion of OSM POIs to DMTI POIs nationally is 17.9%, which varies regionally from a low of 14.1% in Quebec to a high of 20.7% in Ontario.

Cluster analysis

Five cluster groups had the lowest variation of median values for each component measure within the cluster groups and the most variation across groups. The cluster groups were ordered according to the favourability of the active living environment: DAs in Group 1 represent the least favourable active living environments and those in Group 5 represent the most favourable active living environments in Canada. A cluster group was not assigned to the 500 DAs where dwelling density could not be derived (i.e., areas where Statistics Canada does not disseminate data on dwelling counts).

The majority of Canadian DAs (64.3%) were in the lowest two cluster groups (groups 1 and 2). This reflects the right skewness of the active living environment measures on which the cluster groups are based. Rural areas and the outermost areas of cities and towns tended to be in Group 1. DAs in Group 5 tended to be in the central business districts of Canada's largest cities. Groups 2 to 4 tended to be located in residential neighbourhoods of urban areas (Figure 2).

Active-transportation-to-work rates were positively graded by the cluster group and increased by similar magnitude by cluster group (Table 3). In contrast, the walking-to-work rates were similar and below the national average (6.0%) for groups 1 to 3 (Group 1: 5.3%, Group 2: 4.8%, Group 3: 5.3%).

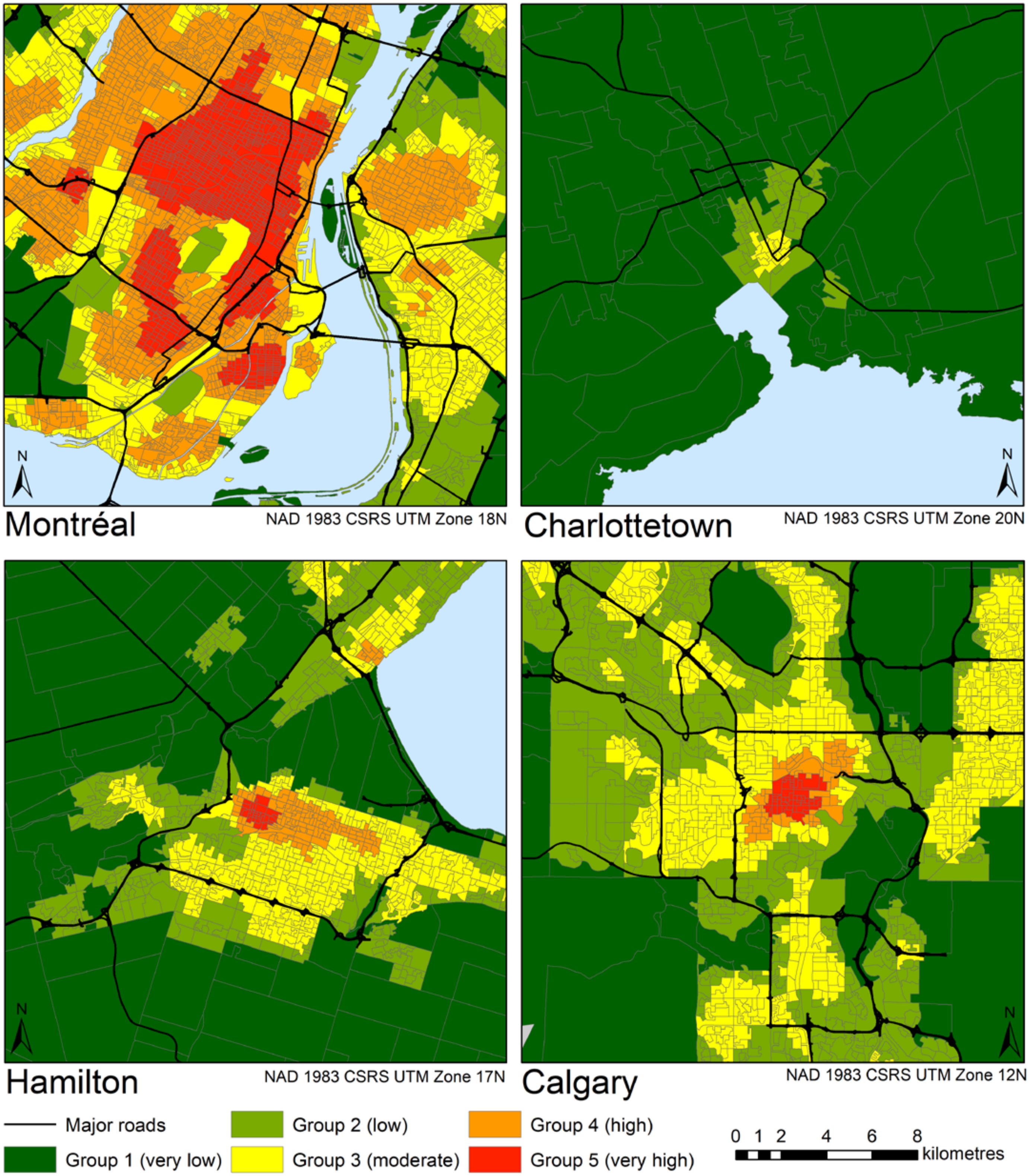
Discussion

This study informed the development of the first pan-Canadian database of active living environment measures derived entirely from non-proprietary sources. Database content was guided by principles according to which the measures selected needed to be associated with walking rates, suitable to a variety of built environments (e.g., urban, suburban, rural), and openly available to researchers and the public health community. Four individual measures (three fully pan-Canadian; transit stop proximity was available only for a subset of DAs in urban areas) were ultimately selected for the database, which also contains composite measures. Three-way intersection density using the OSM road and footpath features was selected based on its higher correlation with active-transportation-to-work rates, relative to the measures with roads alone. Three-way intersection density was deemed more appropriate for the national database, as four-way intersections are concentrated in cities with a grid-like street pattern and are less common in rural and certain suburban areas of Canada, which may lead to the connectivity in these communities being underestimated. Weighted dwelling density was selected because of its higher association with walking-to-work rates relative to population density measures and the much higher association between the weighted density methods and active-transportation-to-work rates. OSM POIs were selected because of their stronger association with active-transportation-to-work rates relative to DMTI POIs and good evidence of regional similarities in feature coverage. Additionally, the transit stop measure was selected and derived for DAs within CMAs and was strongly associated with both walking-to-work and active-transportation-to-work rates.

The Canadian active living environment measures derived from open data are similarly associated with active transportation behaviour as measures derived from traditional sources, or more strongly associated. For the connectivity measures, this may be attributable to the presence of footpath features in OSM, a feature that is not well-mapped in other Canadian datasets (Statistics Canada, DMTI Streetfiles). In 2010, even before the road network was considered complete in Germany, Zielstra and Zipf noted that OSM was more optimized for pedestrian routing than an authoritative GPS dataset in medium- and large-sized German cities.[22] Today, the OSM road network features are considered complete in Canada and most Western countries, and the coverage of footpaths and trails continues to grow.[29] Another advantage of open datasets, such as OSM, is the emerging international comparison work that is facilitated by using replicable methods and data that are available internationally. For example, these methods have been replicated to produce a comparable active living environment measure for Wales.[30]

No research has been conducted to assess the completeness of OSM POI data in Canada or elsewhere. Unlike the DMTI EPOI dataset, which primarily contains records of businesses, OSM POIs contain many features for which no publicly available, exhaustive dataset exists (e.g., beaches, playgrounds, fountains). However, research on OSM's POI features suggests that the features in OSM are spatially and descriptively accurate and that the number of POIs is growing exponentially.[23,31,32] Furthermore, the small-scale, on-street features mapped by OSM are conceptually appealing given their similarity to micro-scale features of active living environments (e.g., benches, park features), which are associated with increased physical activity, but are traditionally difficult to map without conducting field audits.[33] Although the incompleteness of OSM data may present a barrier to research for those studying particular attribute types (e.g., liquor stores, transit

Figure 2
Distribution of cluster groups in four Canadian urban areas



stops),[34] the OSM POI measure is distributed similarly to the DMTI POIs by region, and both measures are similarly associated with walking-to-work and active-transportation-to-work rates.

Active-transportation-to-work rates are graded positively by the five cluster groupings identified in the k-medians cluster analysis. This contrasts with evidence that only the most favourable active living environments (groups 4 and 5, which represent the communities of only 13.6% of Canadians) can support walking and cycling commute rates above the national average. Previous research found that public transportation users can meet recommended physical activity levels regardless of the active living friendliness of their neighbourhood.[26] As higher walking-to-work rates are concentrated in cluster groups 4 and 5, and as less than 2% of Canadians cycle to work,[35] the positive grading of active-transportation-to-work rates is largely explained by increased public transit use by cluster group. Together, these findings suggest that public transportation use may be a practical alternative to increasing physical activity levels in places where walking to work is impractical.

Cluster groups with more favourable active living environments (groups 4 and 5) had fewer DAs than those with less favourable living environments (groups 1 and 2). Because of a very high dwelling density and a high concentration of destinations in a few Canadian neighbourhoods—particularly the urban cores of major cities—measures of active living environments at the national level are highly right-skewed. Researchers have stratified communities in the past using equal interval categories (e.g., quartiles or quintiles) and have found significant relationships between these active living features and health outcomes for the highest quartile or quintile.[9,36] Future research may find more robust results if categories that reflect the right skewness of active living environment measures are used.

Limitations

One limitation of this study is the use of commuting data from the census to derive walk-to-work and active-transportation-to-work rates. Commuting data reflect only a portion of all active travel behaviour. Future researchers may wish to use measures of physical activity and health status from surveys, such as the Canadian Health Measures Survey, to analyze the association between these measures and other forms of active living. The use of aggregated transportation data for both analyses suggests an ecological association between active living environment characteristics and increased physical activity through transportation, but not necessarily a causal relationship.

Often, land use mix is included in active living environment measures instead of or in addition to destinations and POIs. While a lack of free land-use data prevented the calculation of land use mix for Can-ALE, destinations are often assessed alongside or instead of land use mix,[4,37,38] and estimations of land use mix using the conventional entropy formula have been criticized for overestimating the mix of uses in certain areas.[39]

The use of open data to derive the Can-ALE database makes these measures more accessible, but also presents a challenge for reproducibility. OSM has existed for only about a decade and is always changing, which limits the exact replication of research. With the ongoing additions of new POIs by users, it may be difficult to attribute changes in OSM to changes in the built environment or simply to improvements in the data quality and quantity. Until the POIs have reached saturation at the national level (i.e., the annual number of additions or edits to the dataset is less than 3% of the total number of points[40]), OSM POIs may not be a suitable source of data for studies that wish to examine changes in the built environment. However, the OSM road features are considered complete[29] and may be suitable for longitudinal research.

Conclusion

Addressing the prevalence of sedentary behaviour in Canada may require environmental interventions that support routine, daily physical activity. However, to date few national datasets exist to characterize active living environments and support research and public health surveillance. Can-ALE addresses this

Table 3
Descriptive and summary statistics of the five cluster groups

Cluster group	All groups	1 (very low)	2 (low)	3 (moderate)	4 (high)	5 (very high)
Number of dissemination areas (DAs)	**56,089**	20,722	15,327	13,077	4,541	2,422
Percentage of Canadian population (%)	**99.7**	33.0	29.1	24.0	8.7	4.9
Walk-to-work rate	**6.0**	5.2	4.8	5.3	8.2	17.9
Active-transportation-to-work rate (%)	**18.0**	7.2	13.9	23.2	42.7	63.2
Three-way intersection density of roads and footpaths						
Average	**49.2**	**7.4**	**49.0**	**75.1**	**100.3**	**173.5**
Minimum	0.0	0.0	5.8	17.3	31.1	44.5
Maximum	1,081.8	223.5	379.7	382.1	415.4	1,081.8
Weighted dwelling density						
Average	**941.7**	**71.1**	**661.4**	**1,272.3**	**2,548.5**	**5,366.5**
Minimum	0.0	0.0	339.2	933.2	1,840.2	3,610.3
Maximum	14,747.6	344.6	948.3	1,844.8	3,621.1	14,747.6
Points of interest						
Average	**54.1**	**5.2**	**37.9**	**66.1**	**120.7**	**386.3**
Minimum	0.0	0.0	0.0	4.0	9.0	30.0
Maximum	2,119.0	285.0	388.0	714.0	829.0	2,119.0

Notes: Values reflect the counts, percentages and means for DAs assigned to each cluster group. Walk-to-work and active-transportation-to-work rates are based on Question 43 a) of the 2016 Census, which was asked to employed members of the labour force aged 15 and older with a fixed workplace address ("How did this person **usually** get to work?"). The sample size and percentage of the Canadian population do not add to all of Canada, as 500 DAs were omitted because dwelling density was not derived.
Sources: OpenStreetMap; Statistics Canada, 2016 Census.

gap by providing a freely accessible dataset that is validated with national active-transportation data. The study draws attention to the methodological considerations of building measures for the national scale, and in particular, to the fact that three-way intersection density and dwelling density are suitable for national measurements. This study also expands the Canadian evidence base on the validity of using open data and suggests that certain well-mapped elements of OSM (e.g., footpaths, micro-scale features) may improve the predictive capacity of active living environment measures. As OSM reaches completion or saturation, open data may serve as an important derivation source for active living environment measures and support the creation of longitudinal datasets.

The findings also draw attention to the spatial distribution of active living environments and active travel behaviour in Canada. The non-normal distribution of active living environment measures and active travel in Canada suggests that the use of equal interval categories to stratify neighbourhoods may lead to the relationship between the built environment and health status being underestimated. In contrast, the strong grading of active transportation use by active living environment cluster group draws attention to the role public transportation may play in increasing physical activity among a larger group of Canadians. ■

References

1. Public Health Agency of Canada. *The Chief Public Health Officer's Report on the State of Public Health in Canada 2017 – Designing Healthy Living*. Ottawa: Public Health Agency of Canada, 2018.
2. Cervero R, Kockelman K. Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment* 1997; 2(3): 199-219.
3. Kaczynski AT. Neighborhood walkability perceptions: associations with amount of neighborhood-based physical activity by intensity and purpose. *Journal of Physical Activity and Health* 2010; 7(1): 3-10.
4. Glazier RH, Creatore MI, Weyman JT, et al. Density, Destinations or Both? A Comparison of Measures of Walkability in Relation to Transportation Behaviors, Obesity and Diabetes in Toronto, Canada. *PLOS ONE* 2014; 9(1): e85295.
5. Winters M, Brauer M, Setton EM, Teschke K. Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving. *Journal of Urban Health* 2010; 87(6): 969-93.
6. Booth G, Creatore M, Moineddin R, et al. Unwalkable neighborhoods, poverty, and the risk of diabetes among recent immigrants to Canada compared with long-term residents. *Diabetes Care* 2013; 36(2): 302-8.
7. Frank LD, Schmid TL, Sallis JF, et al.. Linking objectively measured physical activity with objectively measured urban form. *American Journal of Preventive Medicine* 2005; 28(2): 117-25.
8. Hajna S, Dasgupta K, Ross NA. Laboratory-Assessed Markers of Cardiometabolic Health and Associations with GIS-Based Measures of Active-Living Environments. *International Journal of Environmental Research and Public Health* 2018; 15(10): 2079.
9. Creatore MI, Glazier RH, Moineddin R, et al. Association of Neighborhood Walkability with Change in Overweight, Obesity, and Diabetes. *JAMA* 2016; 315(20): 2211-20.
10. Wasfi RA, Dasgupta K, Orpana H, Ross NA. Neighborhood Walkability and Body Mass Index Trajectories: Longitudinal Study of Canadians. *American Journal of Public Health* 2016; 106(5): 934-40.
11. Chiu M, Shah BR, Maclagan LC, et al. Walk Score® and the prevalence of utilitarian walking and obesity among Ontario adults: A cross-sectional study *Health Reports* 2015; 26(7): 3-10.
12. Statistics Canada. *Boundary Files, Reference Guide*, Census year 2016 (Catalogue 92-160-G) Ottawa: Statistics Canada, 2017.
13. James P, Berrigan D, Hart JE, et al. Effects of buffer size and shape on associations between the built environment and energy balance. *Health & Place* 2014; 27: 162-70.
14. Boone-Heinonen J, Popkin BM, Song Y, Gordon-Larsen P. What neighborhood area captures built environment features related to adolescent physical activity? *Health & Place* 2010; 16(6): 1280-6.
15. Grasser G. Walkability and Health. Development of GIS based indicators of walkability relevant to public health for surveillance and planning purposes in Graz [Dissertation]. Graz: Medical University of Graz; 2014.
16. Manaugh K, El-Geneidy A. Validating walkability indices: How do different households respond to the walkability of their neighborhood? *Transportation Research Part D: Transport and Environment* 2011; 16(4): 309-15.
17. Oliver LN, Schuurman N, Hall AW. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International Journal of Health Geographics* 2007; 6(1): 41.
18. Statistics Canada. *Road Network File, Reference Guide: Census Year 2016* (Catalogue 92-500-G) Ottawa: Statistics Canada, 2017.
19. DMTI Spatial Inc. CanMap® Content Suite User Manual v2017.3. Markham, ON: 2017.
20. Haklay M, Weber P. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing* 2008; 7(4): 12-8.
21. Zhang H, Malczewski J. Accuracy Evaluation of the Canadian OpenStreetMap Road Networks. *International Journal of Geospatial and Environmental Research* 2018; 5(2).
22. Zielstra D, Zipf A. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. 13th AGILE International Conference on Geographic Information Science; 2010; Guimarães, Portugal.
23. Girres JF, Touya G. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*. 2010; 14(4): 435-59.
24. Geofabrik GmbH. GEOFABRIK // Home [Internet]. Karlsruhe 2018. Available at: https://www.geofabrik.de/.
25. Statistics Canada. *Dictionary, Census of Population, 2016* (Catalogue 98-301-X) Ottawa, Statistics Canada, 2016.
26. Wasfi RA, Ross NA, El-Geneidy AM. Achieving recommended daily physical activity levels through commuting by public transportation: Unpacking individual and contextual influences. *Health & Place* 2013; 23: 18-25.

27. Langlois M, Wasfi RA, Ross NA, El-Geneidy A. Can transit-oriented developments help achieve the recommended weekly level of physical activity? *Journal of Transport & Health* 2016; 3(2): 181-90.

28. Besser LM, Dannenberg AL. Walking to public transit: steps to help meet physical activity recommendations. *American Journal of Preventive Medicine* 2005; 29(4): 273-80.

29. Barrington-Leigh C, Millard-Ball A. The world's user-generated road map is more than 80% complete. *PLOS ONE* 2017; 12(8): e0180698.

30. Fry R, Akbari A, Mah SM, Ross NA. Measuring Active Living Environments: An international comparison between Canada and Wales. *International Journal of Population Data Science* 2018; 3(2).

31. Goodchild MF, Li L. Assuring the quality of volunteered geographic information. *Spatial Statistics* 2012; 1: 110-20.

32. Vandecasteele A, Devillers R. Improving volunteered geographic data quality using semantic similarity measurements. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2013; 1(1): 143-8.

33. Cerin E, Nathan A, Van Cauwenberg J, et al. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity* 2017; 14(1): 15.

34. Bright J, De Sabbata S, Lee S, et al. OpenStreetMap data for alcohol research: Reliability assessment and quality indicators. *Health & Place* 2018; 50: 130-6.

35. Statistics Canada. *Commuting Destination (5), Main Mode of Commuting (10), Sex (3) and Age (5) for the Employed Labour Force Aged 15 Years and Over Having a Usual Place of Work, in Private Households of Canada, Provinces and Territories, Census Divisions and Census Subdivisions, 2016 Census - 25% Sample Data.* Data Tables, 2016 Census. Ottawa: Statistics Canada, 2017.

36. Hajna S, Ross NA, Joseph L, et al. Neighbourhood Walkability and Daily Steps in Adults with Type 2 Diabetes. *PLOS One*. 2016; 11(3): e0151544.

37. McCormack GR, Giles-Corti B, Bulsara M. The relationship between destination proximity, destination mix and physical activity behaviors. *Preventive Medicine*. 2008; 46(1): 33-40.

38. James P, Hart JE, Laden F. Neighborhood walkability and particulate air pollution in a nationwide cohort of women. *Environmental Research* 2015; 142: 703-11.

39. Manaugh K, Kreider T. What is mixed use? Presenting an interaction method for measuring land use mix. *Journal of Transport and Land Use* 2013; 6(1): 63-72.

40. Gröchenig S, Brunauer R, Rehrl K. Digging into the history of VGI data-sets: results from a worldwide study on OpenStreetMap mapping activity. *Journal of Location Based Services* 2014; 8(3): 198-210.