Received: 2 December 2022 | Accepted: 1 May 2023

DOI: 10.1111/cag.12867

RESEARCH ARTICLE / ARTICLE SCIENTIFIQUE

Canadian Geographies / Géographies canadiennes

Best practices for measuring community resources across Canada: A comparison of coding classifications

Marisa Young[1] | Sean Leipe[2] | Diana Singh[1]

[1]Department of Sociology, McMaster University, Hamilton, Ontario, Canada

[2]McMaster University, Ontario, Canada

Correspondence to / Adresse de correspondance
Marisa Young, Department of Sociology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4M4.
Email/Courriel: myoung@mcmaster.ca

Funding information
Ontario Ministry of Research, Innovation and Science, Grant/Award Number: ER16-12-113; Social Sciences and Humanities Research Council of Canada, Grant/Award Number: 430-2017-00292; Canadian Institutes of Health Research, Grant/Award Number: MOP-102730

Abstract

Social scientists, geographers, criminologists, and health scientists are often tasked with finding data to best capture the impact of “community context” on individual outcomes, including residential services, physical resources, and social institutions. One outlet for such data in Canada is Digital Map Technologies Inc. (DMTI) Spatial, which offers a national repository of over one million businesses and recreational points of interest. The database is generated through CanMap Streetfiles, which includes geocodes of each point's precise location. These data are available to researchers from their university data library and Esri Canada, but primarily available to private sector and government markets. That said, the goal of the current paper is to encourage researchers to access this rich yet under-utilized data source. Each service, business, or resource in the DMTI Spatial database is assigned to a respective category using Standard Industrial Classification codes and North American Industrial Classification System codes. It is not clear, however, which is the more reliable coding criteria. We provide an overview of our review of DMTI Spatial data and take-away suggestions for using this valuable resource for future research on meso-level residential markers.

KEYWORDS

community data, DMTI Spatial data, North American Industrial Classification System codes, Standard Industrial Classification codes

RÉSUMÉ

Les chercheurs en sciences sociales et en santé, les géographes et les criminologues sont souvent chargés de trouver des données permettant bien de saisir les effets de milieux sur les caractéristiques individuelles. Au Canada, l'un des débouchés pour ces données est Digital Map Technologies Inc. (DMTI) Spatial, qui offre un référentiel national de plus d'un million d'entreprises et de points d'intérêt. La base de données est générée par CanMap Streetfiles, qui comprend les géocodes de l'emplacement précis de chaque lieu. Ces données sont disponibles pour les chercheurs à travers leur université et Esri Canada, mais elles sont principalement destinées aux marchés du secteur privé et du gouvernement. L'objectif du présent article est d'encourager les chercheurs à utiliser à cette source de données riche mais sous-utilisée. Chaque service, entreprise ou ressource

de la base de données DMTI Spatial est assigné à une catégorie respective à l'aide des codes de la Classification Industrielle Standard et du système de Classification Industrielle Nord-Américain. Il n'est toutefois pas évident de dire quel est le critère de codification le plus fiable. Nous présentons une vue d'ensemble de notre examen des

données de DMTI Spatial et proposons des suggestions pour l'utilisation de cette ressource précieuse dans le cadre de futures recherches sur les marqueurs résidentiels.

MOTS CLÉS
données spatiales DMTI, données communautaires, codes de la classification industrielle standard, codes de la classification industrielle nord-américaine

Key messages

• The goal of this paper is to outline existing data source(s) and measures from DMTI Spatial that might help capture meso-level residential institutions.
• We recommend "best practices" for using DMTI Spatial data in researchers' own work to capture neighbourhood resources/amenities, or the social infrastructure of the community using either Standard Industrial Classification codes or North American Industrial Classification System codes.
• We conclude that Standard Industrial Classification codes in DMTI Spatial enhanced points of interest data are more complete—and more accurate—than North American Industrial Classification System codes.

INTRODUCTION

The questions social scientists, geographers, criminologists, and many health scientists ask are inherently multilevel in their form. We are curious about how individuals are impacted by their surrounding context, whether this context refers to the institutions with which we engage daily, or the cultural or social context that influences patterns of behaviour and inequality (Mills, 1959). A prominent area of sociological, criminological, geographical, and health science research questions the impact of residential communities on individual-level outcomes (Duncan & Kawachi, 2018 Price et al., 2012; Sampson, 2012a; van der Werf et al., 2019; Walton-Roberts et al., 2019).

This research engages community/neighbourhood effects and meso-level residential institutions literature to understand patterns of social interaction, cohesion, and capital; individual stressors, health, and well-being; criminal behaviours; child development; resettlement patterns; household behaviours; and chronic diseases (Oliver et al., 2007; O'Campo et al., 2015; Price et al., 2012; Sampson, 2012a; van der Werf et al., 2019; Walton-Roberts et al., 2019; Webb et al., 2017; Young & Wheaton, 2013). To appropriately address research questions of this nature, quality data that measure residential or area-level characteristics of "context" are necessary. In Canada, there are several country-wide outlets for these sorts of data. These include volunteer geographic information (VGI) data sources, such as OpenStreetMap (OSM), Citizen Science Portal, GeoGratis, Green Maps, and WikiMapia, for example. Some of the latter resources reference geographical markers and land characteristics only. OSM offers information on roads, buildings, and other geographical features. Our approach is contributory since we offer another key VGI outlet that is comprehensive, convenient to access and download, and comprises information on meso-level residential institutions, such as education and recreation resources, fire and police services, religious organizations, and food stores—Digital Map Technologies Inc. (DMTI) Spatial. Compared to other VGI sources, DMTI Spatial data measures are cited less in research, and, amongst those who have used this rich data source, many rely on indicators of food stores by region only (e.g., Doggett et al., 2021; Taylor et al., 2020; Vallée et al., 2020; Woudsma & Jakubicek, 2020).

The goal of the paper is to outline existing data source(s) and measures from DMTI Spatial that might help capture meso-level residential institutions. DMTI Spatial has been collecting and mapping data in Canada for over 20 years; these data are available to researchers employed at select Canadian universities, as well as government or non-profit organizations. Based on website searches, we identified the universities as having access to DMTI Spatial data, including: University of Toronto, McMaster University, University of British Columbia, University of Alberta, University of Calgary, University of Lethbridge, Carleton University, University of Waterloo, Simon Fraser University, University of Ottawa, McGill University, Concordia University, and University of New Brunswick. This information was available through the institutions' respective data websites. For example, the search for DMTI data access on the University of Toronto Maps and Data Website returned the following:

https://mdl.library.utoronto.ca/onesearch/DMTI. Furthermore, there are several outlets beyond the academic community that provide access to DMTI data, such as Esri Canada.

There is a fee for access, and most participating universities host these data through the Scholars GeoPortal online geospatial library service. Despite its wide availability, few researchers across universities and outside the academic community have used these data in analyzing residential effects. These data present several methodological challenges in terms of assessing validity and reliability. We recommend "best practices" for using DMTI Spatial data in researchers' own work to capture neighbourhood resources/amenities or the social infrastructure of the community (Klinenberg, 2018).

This paper addresses the following research questions: (1) Are there inconsistencies between Standard Industrial Classification (SIC) codes and North American Industrial Classification System (NAICS) code counts? We use data from 2011 and 2019 to address this research question. (2) Which are more accurately representative—SIC or NAICS codes? We conclude the following: SIC codes in DMTI Spatial enhanced points of interest (DMTI EPOI) data are more complete than NAICS codes; and SIC codes in DMTI EPOI data are more accurately representative than NAICS codes. We recognize that the data and overview we provide does not definitively answer these questions; they shed light on the necessary precautions when using these data. This is not to undermine the integrity of the DMTI Spatial data, but rather to offer a guide to reviewing and accessing these data for geographical and social science research.

LITERATURE REVIEW

DMTI Spatial datasets: Description and prior works

The EPOI dataset is a comprehensive Canadian national database compiled by DMTI Spatial, containing over 1,000,000 business and recreational points of interest with regular updates (DMTI Spatial, n.d.-a). These points of interest contain SIC codes and NAICS codes used to place them within categories, along with other identifying features such as street address, website, and contact information. DMTI Spatial provides additional database products targeted towards specific resource classes such as education and food distribution points, which will be referred to in this paper as DMTI-R or "refined" datasets. These datasets are available to researchers through services such as the Scholars GeoPortal online data service, which is a project of the Ontario Council of University Libraries that provides access to a wealth of commerciallylicensed geospatial data. The EPOIs have been used for current sociological research on land use patterns, accessibility to health- related resources, correlations of recreation facility access to substance abuse behaviours, identification of food deserts, and more (Doggett et al., 2021; Vallée et al., 2020; Woudsma & Jakubicek, 2020).

Clary and Kestens (2013) used field validation to evaluate the ability of the 2010 DMTI EPOI dataset to accurately capture food outlets in Montreal, Canada, using SIC codes in DMTI EPOI data and additional information such as business names. The dataset was determined to have a "moderate" capacity to detect points of interest present in the field, while a novel modified accuracy measure that tolerated mismatches in business names or slight imprecisions in location found the representativity of the EPOI database to be 77%.

Daepp and Black (2017) evaluated the validity of five sources of food outlet data, including the 2013 EPOIs, and assessed the effects of dataset error on food environment measures. Although all datasets scored relatively poorly and the SIC/NAICS codes from the EPOIs were found to be inadequate for detailed resource classification, there were still high correlations between proximity and density measures and those created from ground-truthed data.

Taylor et al. (2020) compared food stores selected using NAICS codes from the 2015 EPOIs to an official "gold standard" dataset compiled by the Government of Newfoundland and Labrador using various accuracy metrics. The study found that food stores from the EPOI dataset had fair to moderate overall agreement with the provincial ground-truthed validation data. Grocery stores had the greatest sensitivity, positive predictive values, and concordance with the provincial dataset, while greater misclassifications or geocoding errors were found with retailers in rural areas and gas stations.

SIC and NAICS coding systems

SIC codes are four-digit numerical codes that were established in 1987 by the United States Government to classify domestic production. SIC codes in DMTI EPOI data are also used within Canada to classify establishments, companies, and enterprises according to their primary activity (Government of Canada, 2016).

The NAICS is the current standard used by Canadian federal statistical agencies; it replaced the SIC system throughout North America in 1997 (Pierce & Schott, 2012). These codes cover 20 primary classes of industries including manufacturing, retail trade, educational services, health care, and food services (Government of Canada, 2016). The NAICS system provides more specific industry definitions than the SIC codes in DMTI EPOI data, with some activities reclassified between the two (Pierce & Schott, 2012).

Several works have used SIC and NAICS codes for purposes such as identifying available healthy food options, community resources, and industrial sector-wide changes, along with evaluating their accuracies compared to ground-truthed field datasets (Boone-Heinonen et al., 2013; Franco et al., 2008; Kile & Phillips, 2009; Pierce & Schott, 2012; Powell et al., 2011; Townley et al., 2018). Fewer studies have compared the accuracy of results between the two classification systems across one or more industrial sectors, while most have focused on food outlets.

Kile and Phillips (2009) directly assessed the ability of SIC, NAICS, and Global Industry Classification System codes to identify different classes of high-tech firms compared to classification via direct analysis of the firms' actual business activities. While the study found large variations in the effectiveness of all industry classification codes across different high-tech sectors, there was no across-the-board improvement in classification results when using NAICS codes instead of SIC codes in DMTI EPOI data. Errors of omission were generally larger than errors of commission, implying that many high-tech firms were incorrectly not included within the NAICS or SIC code datasets under consideration.

Powell et al. (2011) compared urban retail food outlets identified using primary SIC and NAICS codes with field-observed data and calculated various measures of accuracy. The authors found moderate agreement between field validation and secondary datasets based on SIC and NAICS codes, although exact classifications for individual sub-classes ranged from fair to poor. Food outlets were consistently undercounted in the secondary datasets when compared to field observations, which is consistent with studies using similar data.

The association between meso-level regional institutions and individual-level outcomes

Our paper aims to support more accessible and comprehensive data for social scientists, geographers, criminologists, and health researchers wanting to address the impact of residential region on individual-level outcomes. Past research on area-level effects draws on a variety of data sources to capture residential variations in institutions, community disorder or disadvantage, and resources. The most popular data sources used to capture these variations include census data and systematic social observation tools (SSO)—at least in the social sciences, health sciences, and criminology (Cottagiri et al., 2021; Duncan & Kawachi, 2018; Sampson, 2012b; Sampson et al., 1997). These sources are commonly used to capture "neighbourhood disadvantage" and rely on indicators such as aggregated social/demographic composition of the area (i.e., census data), or markers of disorder or disarray in communities (i.e., SSO tools, which are "social systematic observation tools", see Sampson et al. (2002), for a review; Young and Singh, 2022). Yet these approaches and data sources neglect to consider the available "resources and amenities" in geographical propinquity. The latter reflects what Eric Klinenberg (2018, p. 5) refers to as the "social infrastructure"—defined as "the physical places and organizations that shape the way people interact."

In Canada, there are several outlets of area-level data available to researchers beyond census data and SSO tools that might capture these alternate community-based criteria, like social infrastructure and local environment. These are designed through VGI, including sources like OSM, Citizen Science Portal, GeoGratis, Green Maps, Walk Score, Community Atlas, and WikiMapia, for example. OSM is the most similar to DMTI Spatial data in terms of data comprised by the two sources, and the geographic span of information across time and place. There are, however, several key differences, based on three criteria.

First, OSM is an open-source and community-driven mapping project. That is, public participation contributes to the collection and editing of map data (OpenStreetMap, n.d.). Alternatively, DMTI Spatial data are from both public and private sources, including government institutes, field surveys, and aerial imagery (DMTI Spatial, n.d.-a).

Second, the two outlets also differ based on their data structure. For example, OSM has a geospatial structure where each reference point is represented by a set of coordinates and attributes. This structure is beneficial in looking at data layers, like land use and roads, for example (Mapbox, n.d.). DMTI Spatial data are not structured in that way. Instead, markers indicate the presence of a building, location, or institution, for example, but layered data are not provided. For that reason, DMTI Spatial data might be best used for extracting data on community resources (DMTI Spatial, n.d.-a).

Third, the two outlets differ in terms of licensing. OSM data are open access and therefore accessible to all to use and share. There are, however, some expectations for users to provide attribution and information about changes made to the data (OpenStreetMap, n.d.). Alternatively, DMTI Spatial is privately owned and the data are protected by copyright. Anyone wanting to use these data must purchase a licence (DMTI Spatial, n.d.-b).

There have been multiple studies comparing OSM and DMTI Spatial data on road networks (e.g., Zhang & Malczewski, 2017) and land use (e.g., Vaz & Arsanjani, 2015). That said, no studies to our knowledge have compared the two sources in terms of meso-level residential institutional data points. In fact, some researchers use data from both sources to test their research questions (Su et al., 2015; Vallée et al., 2020). The goal of this paper, however, is not to compare OSM and DMTI Spatial data, but to provide additional information about—and successful approaches to using—DMTI Spatial data if researchers choose to engage with this rich source.

DATA AND METHODS

DMTI Spatial data on a variety of resources/amenities per geographical location across years can be accessed through Scholars Geoportal (also see https://www.dmtispatial.com/). Datasets are separated into two categories: (1) composite DMTI EPOI, which is a national database of over one million business and recreational points of interest across Canada, and (2) refined datasets (DMTI-R), which represent only specific categories of resources such as education points, food distribution points, and health and emergency services.

The DMTI Spatial files, including the composite EPOI and DMTI-R files, are generated as part of the CanMap Streetfiles product dataset (digital map data), which includes geocodes of each point's precise location. All DMTI Spatial datasets were matched to the Canadian Census Divisions (CD) polygon shapefile, which was downloaded from Statistics Canada's open data portal. These files were then loaded into the ArcMap 10.6 GIS software package for processing. Both the CD shape file and DMTI Spatial datasets were reconfigured to the same coordinate system. In the cases presented, each service, business, or resource in the DMTI Spatial datasets is assigned to a respective category using both SIC and NAICS codes. In most cases, we compare the two sets of codes for select amenities across two time points, 2011 and 2019.

How does DMTI Spatial collect their data?

DMTI Spatial collects data through a variety of sources. One source is aerial imagery that captures high-resolution images of the earth's surface. This imagery is used to create maps and other location-based data products. Another source is from field surveyors who collect location-based data, including measures on property boundaries for road networks. Remote sensing technology is used to collect data on the earth's surface, including data from satellites and drones and other remote sensing tools. Data also come from all three levels of government, including municipal, provincial, and federal agencies. Finally, commercial outlets provide DMTI Spatial partners with data on location-based services, including retail, recreation, education, and real estate. The DMTI Spatial data discussed in this paper likely come from the latter two sources— government and commercial sources (Canadian GIS, n.d.; DMTI Spatial, n.d.-b).

DMTI Spatial measures of resources/amenities

We included compiled data on four main amenities: education resources, food stores, police stations, and fire stations. These resources were selected because they were the cleanest—in terms of coding—from the DMTI Spatial data. Recreation facilities are also considered, secondarily. There are fewer inconsistencies for the recreation measures from the data. We provide recommendations, but do not include evidence from these data to demonstrate our arguments. Note, healthcare data is also available through DMTI Spatial. However, from spot checks our team did not find these data to be reliable.

With regard to education resources, education-specific SIC codes in DMTI EPOI data were used to extract points from the EPOI dataset, which was compared to a more specifically targeted DMTI-R dataset. The DMTI EPOI education point layer is a curated dataset that includes the point locations of elementary schools, high schools, colleges, cégeps, and universities. Our selection of SIC codes in DMTI EPOI data representing education resources is found in Table 1. The DMTI-R dataset included points with several additional primary SIC codes in DMTI EPOI data such as SIC 8322 (individual and family services) and SIC 7991 (physical fitness facilities).

For the purposes of comparison, data on food stores came from both the EPOI and DMTI-R datasets in 2019. Analysis in 2011 relied only on the EPOI datasets as comparable DMTI-R datasets could not be located through Scholars GeoPortal or elsewhere. The DMTI-R food distribution dataset includes the locations of food distribution depots across Canada with a broad range of SIC codes in DMTI EPOI data. Our selection of SIC codes in DMTI EPOI data representing food stores (Table 1) was extracted from the EPOI datasets, while the DMTI-R dataset included the additional codes of SIC 5411 (candy, nut, and confectionery stores) and SIC 5499 (food stores).

Counts for all police and security services in 2019 were extracted from both the DMTI EPOI and DMTI-R datasets for the purposes of comparison, while analysis in 2011 relied on the EPOIs only due to data availability. The DMTI-R health and emergency services dataset indicates emergency service points in Canada including, but not limited to, healthcare facilities, ambulance stations, police service locations, fire stations, and veterinarian locations. Resources from both datasets with a primary SIC code of 9221 (police protection) were extracted and included in our analyses.

Counts for all fire protection services in 2019 were extracted from both the DMTI EPOI and DMTI-R health and emergency services datasets for the purposes of comparison, while analysis in 2011 relied on the EPOIs only due to data availability. Resources from both datasets with a primary SIC code of 9224 (fire protection) were extracted and included in our analyses.

Recreation establishments were tallied from DMTI EPOI datasets including all business and organizations, such as sport and amusement parks, beaches, swimming pools, riding academies and schools, carnival operation, exposition operation, horse shows, and picnic grounds. These were selected using SIC 7999 (amusement and recreation services, not elsewhere classified). To the best of our knowledge, there are no DMTI-R datasets for recreation establishments for the years addressed in this study (2011 and 2019).

TABLE 1 SIC codes in DMTI EPOI data used to classify specific types of resources using the DMTI-EPOI datasets.

Resource type	SIC codes in DMTI EPOI data selected	Code definition
Education	8211	Elementary and secondary schools
	8221	Colleges and universities
	8222	Junior colleges and technical institutes
	8244	Business and secretarial schools
	8299	Schools and educational services, not elsewhere classified
Food stores	5141	Groceries general line
	5411	Grocery stores
	5421	Meat and fish markets

5431 Fruit and vegetable markets

5451 Dairy products stores

5461 Retail bakeries

Fire stations 9224 Fire protection

Police stations 9221 Police protection

Recreation 7999 Amusement and recreation services, not elsewhere classified

Analyses

We took several analytical approaches to answer our research questions. First, we examined inconsistencies between SIC and NAICS code counts using data from 2011 and 2019 from Canada, broadly, and then focused specifically on food stores to demonstrate these inconsistencies. Second, we examined data from two case study regions—Calgary and Hamilton—to demonstrate the accuracy and representativeness of DMTI Spatial data codes using 2019 data only. These sites were chosen based on their comparability in size, and as a reflection of distance across Canada. We provide these two case studies as evidence for our argument. We conducted spot checks with counts of education and food store resources from 2019, and compared these to 2019 Google Street View (GSV) data on those same locations. For education spot checks we use the DMTI EPOI education point layer; for food store spot checks we use the DMTI‐R food distribution dataset.
Third, we demonstrated the key differences between classes of DMTI Spatial datasets by showing comparisons in SIC code counts across resources for Calgary and Hamilton for 2019. Additionally, we undertook more detailed comparisons for counts of food stores across Canada, generally—and Hamilton, specifically—during this time period.
Note, our team took caution to correct for potential errors in DMTI Spatial's data collection and coding processes. For example, we deleted duplicated resource tags. After initial spot checks, several instances were found where duplicate points representing the same resource existed in the same location in space. To resolve this, a new field was created using each point's latitude and longitude coordinates. All datasets were then aggregated based on this field so that each specific latitude and longitudinal location only contained one point. Once each specific resource was obtained (through any noted DMTI Spatial dataset), and a latitude and longitudinal point was identified, the team used a spatial join to obtain a new shapefile containing counts of each resource falling within each CD. Duplicate cases were then deleted. Next, the team standardized all resources/amenities by population (resource/per 100,000 population). Population values were derived from Statistics Canada for 2011 and 2016. Finally, resources/amenity counts for each category were converted to z‐scores (mean = 0; std = 1) for the purposes of comparison within and across time points.

RESULTS

We broadly found that there are inconsistencies between SIC and NAICS codes using DMTI EPOI data from 2011 and 2019. Specifically, we demonstrated that SIC codes in DMTI EPOI data were more complete. For example, using 2011 data for the entirety of Canada, 433,035 records were lacking a primary NAICS code (28.7%), while only 9,784 records lacked a primary SIC code (< 0.01%). Of the records without a primary NAICS value, 424,340 (98%) had at least one primary SIC code value.

The 2019 data presented a similar picture. Among all DMTI EPOI data points, 608,294 records did not have a primary NAICS code (34.6%), while only 30,788 records were lacking a primary SIC code (0.02%). In other words, of these records without a primary NAICS value, 578,529 (95%) had a primary SIC code.
To help confirm these results, we focused on one DMTI data resource, food stores, using EPOI selections and a DMTI‐R food distribution dataset. Using the 2011 EPOIs we found 30,310 total points according to SIC codes in the DMTI EPOI data. Amongst these, 7,673 points (25.3%) are missing a NAICS code. For food stores within the 2019 EPOI dataset, the results are similar: for the 34,816 total observations with primary SIC codes in the DMTI EPOI data, 10,819 (31%) points are missing a NAICS code.
These results suggest that when conducting analysis based on NAICS codes within the EPOI datasets, there is a much larger potential for errors of omission compared to selecting resources using SIC codes in DMTI EPOI data. The missing NAICS codes are based on the top two categories assigned to these outlets. Beyond that, the references become unclear and diffuse to the point that we could not justifiably combine those records into a designated category.
We examined 2019 data from the two case study regions, Calgary and Hamilton, to demonstrate the validity of SIC and NAICS codes. Our research team conducted spot checks for five locations for education and food stores across the two regions. Tables 2 and 3 present these results for the Hamilton and Calgary areas, respectively. We identified several inconsistencies in our spot checks from the 2019 DMTI Spatial data and GSV images. For example, in the Hamilton spot checks for education, three of the five locations did not have NAICS codes. Two of the

TABLE 2 Spot checks for Hamilton, selected resources for 2019 from Scholars GeoPortal vs. GSV.

Resource and Data Link

Education Food (1) Yorkview School Mac's Grocery Store Address 86 Cameron Rd. N. 36 York Rd. SIC Code 8211 5411 NAICS Code 0000 (No Code) 44512 GSV Check X In Dundas, not Hamilton X Circle K (2) Sacred Heart Separate School The Candy Shop Address 5 Hamilton Ave. 314 Barton Street E. SIC Code 8211 5441 NAICS Code 0000 (No Code) 445292 GSV Check √ X Coffee shop - “Stir it Up” (3) Woodview Delta Section 23 Almanar Food Market Address 1284 Main Street E. 388 Concession St. SIC Code 8211 5411 NAICS Code 61111 44512 GSV Check √ X Empty parking lot (4) Grade Learning – Stoney Creek Hamilton Korean Food Market Address 800 Queenston Road 16 Parkdale Ave N SIC Code 8211 5411 NAICS Code 61111 44511 GSV Check X X Husky (5) St. Luke Separate School Metro Grocery Store Address 345 Albright Rd. 1900 King Street E SIC Code 8211 5411 NAICS Code 0000 (No Code) 44511 GSV Check √ √

TABLE 3 Spot checks for Calgary, selected resource for 2019 from Scholars GeoPortal vs. GSV.

Resource and Data Link

Education Food

(1) River Valley School Foothills Foods

Address 3127 Bowwood Dr. NW 3716 61 Ave. SE

SIC Code 8211 5411

NAICS Code 0000 (No Code) 44512

GSV Check X Unclear X Unclear

(2) Rundle School Ocean Edge

Address 4120 Rundlehorn Dr. NE 4101 19 St. NE

SIC Code 8211 5441

NAICS Code 0000 (No Code) 0000

GSV Check √ X Unclear

(3) Mother Mary Greene School Bow Food Convenient Store Address 115 Edenwold Dr. NW 7930 Bowness Rd. NW SIC Code 8211 5411

NAICS Code 0000 44512

GSV Check √ X Buckley Drug Store

(4) Learning Experience Calgary Co-operative Association LTD. Address 17107 James McKevitt Rd. SW 2220 68 St. NE

SIC Code 8211 5461

NAICS Code 61111 0000

GSV Check √ √

(5) University of Calgary Herbal Magic Weight Loss and Nutrition Centres Address 2500 University Dr. NW 14 Richard Way SW

SIC Code N/A 5499

NAICS Code N/A 446191

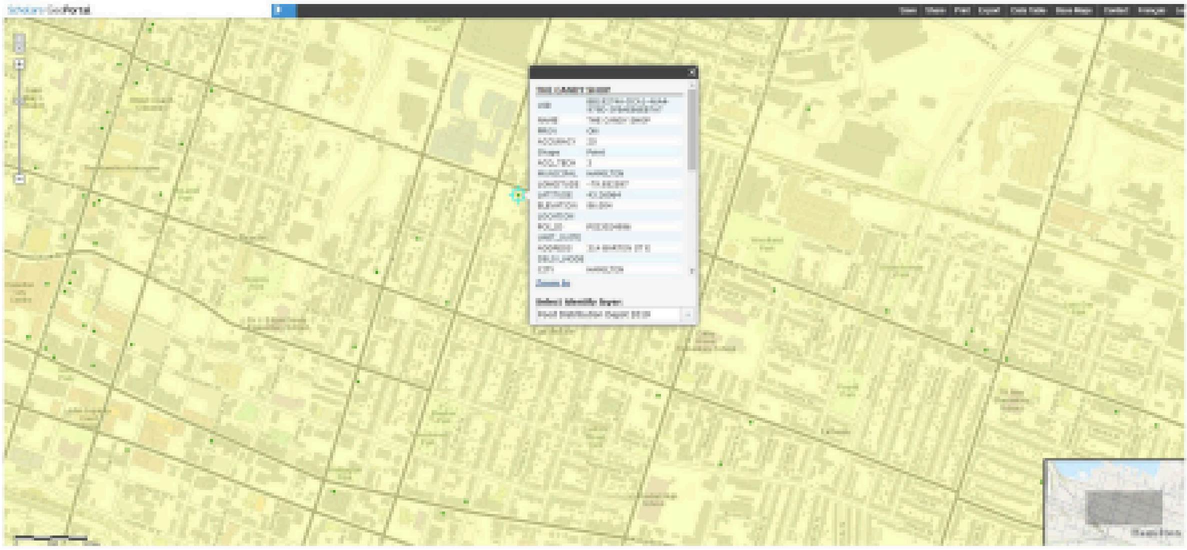
GSV Check √ X Unclear; Pita Pit

cases were coded incorrectly through DMTI, compared to GSV—where either the address was wrong (Yorkview School) or the point of interest could not be located (Grade Learning, Stoney Creek).

The same patterns were noted for our Hamilton spot checks for food stores: four of the five locations were not verified by GSV. These inconsistencies ranged from minor within-resource-category changes (e.g., the Mac's Grocery Store had been renamed to Circle K) to more impactful changes that would impact any analysis conducted using specific SIC or NAICS codes (e.g., the Hamilton Korean Food Market is now a Husky Gas Bar; the Almanar Food Market could not be found on GSV; and the Candy Shop is now a coffee shop—see Figure 1). It is important to note that these resources are either in a completely different resource category than their previous state or no longer exist, which underscores the importance of spot checking data.

The spot checks for Calgary education resources were somewhat more precise. We used the University of Calgary as one of our focused criteria. Noteworthy here is the number of resources embedded within the University area that are coded separately. For example, we have highlighted the University Care Centre and the University Elementary Schools (see Table 3 and Figure 2). These results suggest that there are multiple education resources attached to one location (i.e., the University of Calgary), which might distort coding efforts depending on the researcher's approach. Results from the Calgary food store spot checks were similar to those in Hamilton: four of the five locations could not be clearly confirmed on GSV, and only three of these locations had both SIC and NAICS codes.

(a)



(b)

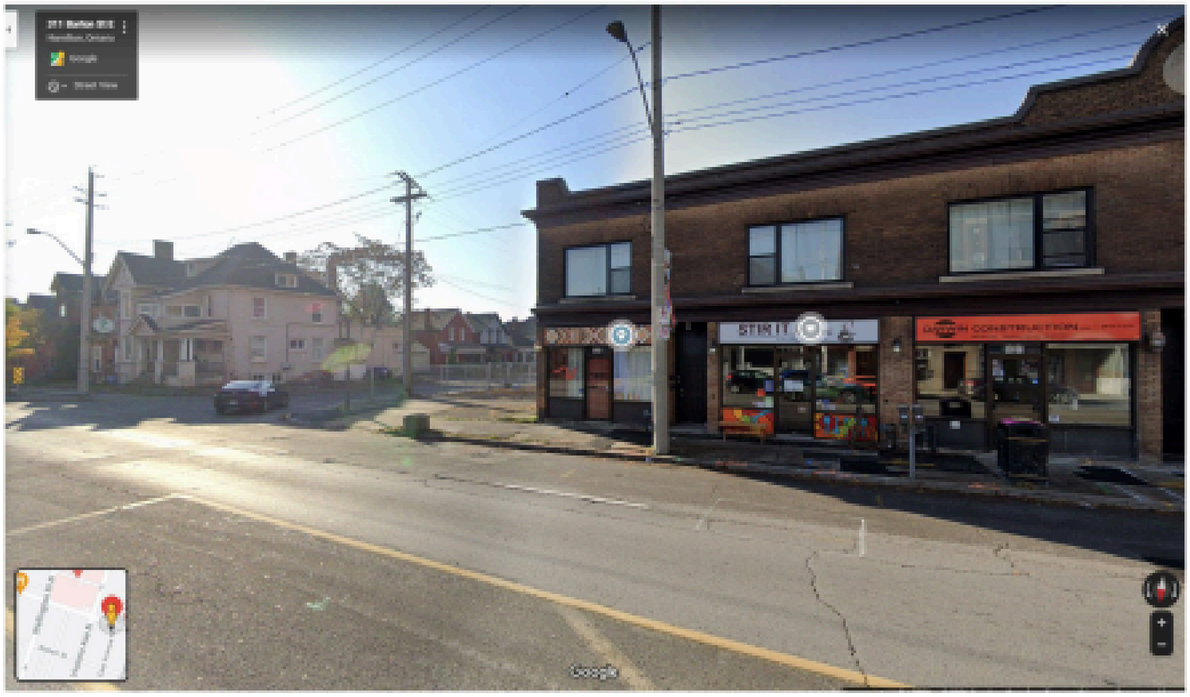


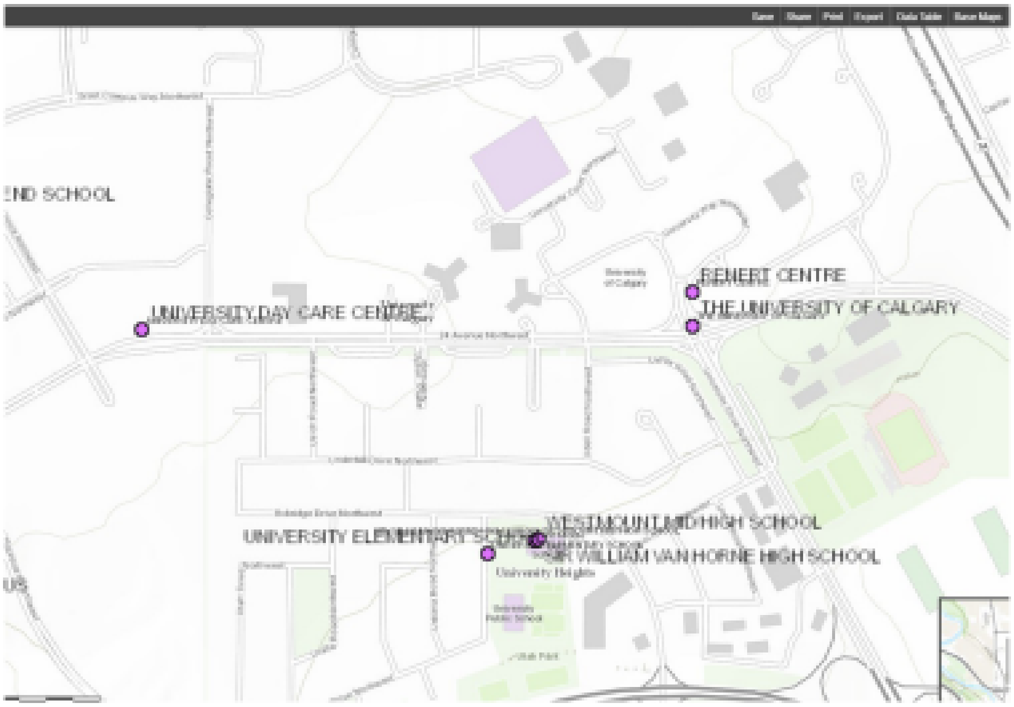
FIGURE 1 Spot check example for Hamilton, selected resources for 2019 from Scholars GeoPortal vs. GSV. Notes: Example of “The Candy Shop” in Hamilton (see Table 2) at 314 Barton Street E. (Scholars GeoPortal), which is currently a coffee shop named “Stir it Up” (GSV).

These results suggest that there are complexities to the coding of residential resources. For example, there were certain education and food store locations that were either miscoded or do not contain valid NAICS codes. Instead of checking and cross-referencing each potential code, we offer a standardized approach to ensure efficacy in data coding and analyses. According to our counts, the EPOIs included 34,816 total points according to selected SIC codes in DMTI EPOI data, while the DMTI-R food distribution data returned 40,725 total points (unmodified). When we excluded SIC 5441 (candy, nut, and confectionery stores) and 5499 (misc. food stores) from the DMTI-R, we recorded 33,288 total points. These differences might be limited but should be noted. For example, when we focused on Hamilton, we could see a difference of 81 available food stores when including SICs 5441 and 5499. If these data are used without discretionary criteria, it might overestimate the available resources to Hamilton residents. The same might apply to residents in other areas of Canada. To examine differences between resource counts using the EPOI and DMTI-R datasets, we specifically focused on food stores. However, the other resources considered in the paper present similar challenges.

(a)

UNIVERSITY OF CALGARY
2500 University Drive NW
Google

(b)



END SCHOOL
BENERT CENTRE
THE UNIVERSITY OF CALGARY
UNIVERSITY DAY CARE CENTRE
WESTMOUNT MID HIGH SCHOOL
UNIVERSITY ELEMENTARY SCHOOL
SIR WILLIAM VAN HORNE HIGH SCHOOL
University Heights

FIGURE 2 Spot check example for Calgary, selected resources for 2019 from Scholars GeoPortal vs. GSV. Notes: Example of “The University of Calgary” in Calgary (see Table 3) at 2500 University Ave. NW (GSV), and available through Scholars GeoPortal.

Education

The DMTI-R education point datasets have a much broader range of SIC codes than our selection from the DMTI EPOI data (Table 1), but a smaller number of resulting points. There appear to be several points of interest within this dataset that correctly correspond to educational resources, but are not assigned an education-related SIC code. Some specific examples are Carleton University (Ottawa), which is coded as SIC 5912 (drug stores and proprietary stores), and St Martin's Manor (Hamilton), which is coded as SIC 8322 (individual and family services). Both locations could be coded as education resources but are not included within the specific DMTI-R dataset. While all these locations could

TABLE 4 Comparison of 2019 SIC code counts across DMTI datasets.

Region	Calgary				Hamilton			
Resource								
Dataset	Educ	Food	Fire	Police	Educ	Food	Fire	Police
EPOI[a]	2373	1007	99	61	902	407	35	12
DMTI-R[b]	709	1257	70	60	288	488	33	12

[a]DMTI Enhanced Points of Interest

[b]DMTI Refined Dataset (individual products targeted towards specific resource classes).

reasonably be considered as education resources, they would be missed if only selecting data from the EPOIs through SIC codes in the DMTI EPOI data. These results are based on spot checks from the Greater Toronto and Hamilton Area (Table 4). The focus of these examples showcases the cleaning undertaken (and necessary) when using DMTI Spatial data. We hope that our examples from above shed light on the necessary precautions when using these data. This is not to cast doubt on the integrity of the data; rather it speaks to our contribution, namely the provision of a guide to reviewing and accessing these data for geographical and social science research.

When selecting education points from the EPOIs in Hamilton for 2019, entries for Young Drivers of Canada, CM Dancing, and A to G Piano Teaching were listed under SIC 8299 (schools and educational services, not elsewhere classified). While these do correspond to the formal definition of SIC 8299, it is not obvious whether they are appropriate for sociological analysis based on strictly educational resources. While SIC 8299 is also included within the 2019 DMTI-R education point dataset, none of the aforementioned resources are included which implies some level of manual selection in its development.

Police stations

Police stations were correctly identified by the primary SIC code 9221 in both the EPOI datasets and the DMTI-R health and emergency services point dataset. This primary SIC code was also attributed to private security or home safety type organizations/stores grouped in with the police stations. For example, in Ontario, the Ontario-Toronto Housing Security, Ontario Small Business and Consumer Service, and Ontario Securities Commission were included under the primary SIC code 9221.

Fire stations

There are major differences when registering fire stations, based on the EPOIs with primary SIC code 9224 and the DMTI health and emergency services point dataset. For EPOIs where the primary SIC code is 9224, it appears that many fire safety or protection organizations or stores are grouped in with the fire stations. For example, Fire Industry Repair and Maintenance Inc., Adler Insulation, and Safety Boss Inc. all fall under SIC 9224.

These results suggest that points in the EPOI dataset might not be suitable for use when explicitly looking for fire stations or other emergency services. While all resources fall under the SIC 9224 umbrella of “fire protection services,” they do not necessarily correspond to the type of resource a researcher may be looking to assess using this code. Points in the DMTI-R (health and emergency services point) datasets where the primary SIC code is 9224 all appear to be legitimate fire stations. These findings support the conclusion that more specifically targeted DMTI-R datasets (e.g., health and emergency, food distribution, etc.) appear to be manually selected compared to the presence of resources according to SIC codes in DMTI EPOI data.

DISCUSSION

Our paper set out to address the challenges researchers face in finding or collecting data to capture the impact of residential services, physical resources, and social institutions on individual outcomes. We highlighted the DMTI Spatial national repository of over 1,000,000 Canadian businesses and recreational points of interest, which includes geocodes of each point's precise location. Each service, business, or resource in the DMTI Spatial database is assigned to a respective category using SIC and/or the more recently updated NAICS. It is not clear, however, which is the more reliable coding criteria for capturing specific resource classes for quantitative analysis. Furthermore, DMTI Spatial offers multiple databases of location points, but does not offer criteria for when to use which dataset. Finally, the data available for download require cleaning and standardization to increase their efficacy in predicting respective outcomes. We have addressed these points, and our research questions, through two case studies featuring two regions in Canada (Calgary and Hamilton). Using DMTI Spatial data from 2011 and 2019, these case studies enabled us to contribute knowledge that will facilitate the use of these data by other researchers.

Our research is timely, particularly with regard to research on mental health. As we enter the third year of the pandemic in Canada, psychological problems are escalating (Aknin et al., 2022). In fact, COVID-19 has created what many have called a mental health echo pandemic (Janson, 2020). While our data do not coincide with the timeline of the pandemic, they speak to potential residential stressors or resources that might impact residents' mental well-being and other individual-level outcomes. Our work has also contributed by profiling the use of DMTI Spatial data measures—most of which look at indicators of food stores by region (Doggett et al., 2021; Taylor et al., 2020; Vallée et al., 2020 Woudsma & Jakubicek, 2020). Our study's findings help to provide information and suggestions for employing DMTI Spatial data to better assess meso-level residential institutional impacts on individual-level outcomes. This is among the first attempt (to our knowledge) to document helpful guidance in using this rich context-level data source.

Based on our case studies, we conclude the following. First that SIC codes in DMTI EPOI data are more complete than NAICS codes, and second, that SIC codes in DMTI EPOI data are more accurately representative than NAICS codes. We discuss our results and limitations of our study in the following sections.

Consistencies between SIC and NAICS codes

Our findings, based on 2011 and 2019 data, suggest that there are inconsistencies between the accuracy and representativeness of the two coding schemas. Within the DMTI EPOI data, there are far more points that lack NAICS codes than lack SIC codes. Given this, SIC codes in DMTI EPOI data should be used to ensure completeness. This conclusion is supported by the number of cases we document across selected resources that are missing NAICS codes, but have a valid SIC code (e.g., food stores coded across Canada). Missing codes can cause serious data reliability issues, particularly as many of the researchers who use this rich data source rely solely on indicators of food stores by region (Doggett et al., 2021; Taylor et al., 2020; Vallée et al., 2020; Woudsma & Jakubicek, 2020). Their results might be dependent on which codes the authors use to determine food resources. Using food stores as an example from 2011 EPOIs, we found 30,310 total points according to SIC codes in DMTI EPOI data. Amongst these, 7,673 (25.3%) of these points are missing a NAICS code. Our findings are further underscored by recent publications using DMTI data measures. Taylor et al. (2020), for example, found poor agreement between food stores within the 2015 EPOI dataset and a ground-based provincial dataset. We conclude that the large percentage of false negatives was likely exacerbated by the use of NAICS codes to extract food stores from the EPOI data.

Accuracy and representativity of SIC versus NAICS

Our case study results for both regions of Canada highlighted that there are inconsistencies in the coding across some spot checks. For example, there were certain education and food store locations that were either miscoded or did not include valid NAICS codes. This was true for the Calgary and Hamilton food store spot checks, where four out of the five locations could not be clearly confirmed on GSV, and only three of the Calgary locations had both SIC and NAICS codes. The accuracy and representativity of SIC codes in DMTI EPOI data ranged from minor to impactful. For example, in some cases what would be coded as a food outlet no longer existed. We underscore that these findings provide a foundation for researchers to advocate for—and contribute to—regular updates to these data to ensure accuracy and representativity of measure counts for the future.

Key differences between DMTI datasets

Using 2019 data with SIC codes in DMTI EPOI data across Canada we found differences between some DMTI-R datasets and comparable SIC code selections from the EPOIs. However, the magnitude of these differences depends on the specific datasets explored. Education, police stations, and fire protection points offered stronger evidence of the differences between dataset codes. We concluded that the DMTI-R datasets may be more appropriate for analysis depending on the specific resource, but many of these are not available before 2015, which can complicate multi-year comparison studies. Given these disparities researchers should carefully review any differences in resource definitions before selecting which datasets and classification codes to use within a study, as these could impact the results of analyses.

CONCLUSIONS AND LIMITATIONS

We conclude with several observations for researchers about DMTI Spatial data. We found that the SIC codes are more complete than the NAICS codes. Second, we found that DMTI-R data are comprehensive and can help sociologists best capture residential-level resource availability. We recognize that these data are not available for every researcher, but for those who have access, we suggest using DMTI-R data sources. Despite the contributions of our study, our suggestions are limited by our selected years and locations of the DMTI Spatial data. We chose these specific criteria based on the broader impetus of this study, which is the Family-Friendly Community Resources for Better Balance, Health and Well-Being Study (Young and Singh, 2022), which set the parameters of the study reported on in this paper. We encourage future researchers to add to our recommendations by exploring other years and geographical parameters of DMTI Spatial data.

FUNDING INFORMATION

Ontario Ministry of Research and Innovation Early Research Award (Funding Reference Number: ER16-12-113), Social Science and Humanities Research Council (Funding Reference Number: 430-2017-00292), Canadian Institutes of Health Research (Funding Reference Number: MOP-102730) Dr. Marisa Young, PI.

REFERENCES

Aknin, L. B., De Neve, J. E., Dunn, E. W., Fancourt, D. E., Goldberg, E., Helliwell, J. F., Jones, S. P., Karam, E., Layard, R., Lyubomirsky, S., Rzepa, A., Saxena, S., Thorton, E. M., VanderWeele, T. J., Whillans, A. V., Zaki, J., Karadag, O., & Ben Amor, Y. (2022). Mental health during the first year of the COVID-19 pandemic: A review and recommendations for moving forward. Perspectives on Psychological Science, 17(4), 915–936.
Boone-Heinonen, J., Diez-Roux, A. V., Goff, D. C., Loria, C. M., Kiefe, C. I., Popkin, B. M., & Gordon-Larsen, P. (2013). The neighborhood energy balance equation: Does neighborhood food retail environment+ physical activity environment= obesity? The CARDIA study. PloS One, 8(12), e85141. Clary, C. M., & Kestens, Y. (2013). Field validation of secondary data sources: A novel measure of representativity applied to a Canadian food outlet database. International Journal of Behavioral Nutrition and Physical Activity, 10, 1–9.
Cottagiri, S. A., De Groh, M., Srugo, S. A., Jiang, Y., Hamilton, H. A., Ross, N. A., & Villeneuve, P. J. (2021). Are school-based measures of walkability and greenness associated with modes of commuting to school? Findings from a student survey in Ontario, Canada. Canadian Journal of Public Health, 112, 331–341.
Daepp, M. I., & Black, J. (2017). Assessing the validity of commercial and municipal food environment data sets in Vancouver, Canada. Public Health Nutrition, 20(15), 2649–2659.
Doggett, A., Godin, K. M., Schell, O., Wong, S. L., Jiang, Y., & Leatherdale, S. T. (2021). Assessing the impact of sports and recreation facility density within school neighbourhoods on Canadian adolescents' substance use behaviours: Quasi-experimental evidence from the COMPASS study, 2015–2018. BMJ Open, 11(8), e046171.
Duncan, D. T., & Kawachi, I. (Eds.). (2018). Neighborhoods and health. Oxford University Press.
Franco, M., Roux, A. V. D., Glass, T. A., Caballero, B., & Brancati, F. L. (2008). Neighborhood characteristics and availability of healthy foods in Baltimore. American Journal of Preventive Medicine, 35(6), 561–567.
Government of Canada. (2016). Archived - Standard Industrial Classification - Establishments (SIC-E) 1980. Statistics Canada. https://www23.statcan.gc.ca/ imdb/p3VD.pl?Function=getVD&TVD=53446. Link no longer active; last accessed 5 July 2021.
Janson, K. (2020). COVID-19 and mental health: Heading off an echo pandemic. Canadian Mental Health Association.
Kile, C. O., & Phillips, M. E. (2009). Using industry classification codes to sample high-technology firms: Analysis and recommendations. Journal of Accounting, Auditing & Finance, 24(1), 35–58.
Klinenberg, E. (2018). Palaces for the people. Crown.
Mills, C. W. (1959). The sociological imagination. Oxford University Press.
O'Campo, P., Wheaton, B., Nisenbaum, R., Glazier, R. H., Dunn, J. R., & Chambers, C. (2015). The neighbourhood effects on health and well-being (NEHW) study. Health & Place, 31, 65–74.
Oliver, L. N., Dunn, J. R., Kohen, D. E., & Hertzman, C. (2007). Do neighbourhoods influence the readiness to learn of kindergarten children in Vancouver? A multilevel analysis of neighbourhood effects. Environment and Planning A, 39(4), 848–868.
Pierce, J. R., & Schott, P. K. (2012). A concordance between ten-digit US Harmonized System Codes and SIC/NAICS product classes and industries. Journal of Economic and Social Measurement, 37(1–2), 61–96.
Powell, T. C., Lovallo, D., & Fox, C. R. (2011). Behavioral strategy. Strategic Management Journal, 32(13), 1369–1386.
Price, K., Plante, C., Goudreau, S., Boldo, E. I. P., Perron, S., & Smargiassi, A. (2012). Risk of childhood asthma prevalence attributable to residential proximity to major roads in Montreal, Canada. Canadian Journal of Public Health, 103(2), 113–118.
Sampson R. J. (2012a). Great American city: Chicago and the enduring neighborhood effect. University of Chicago Press.
Sampson, R. J. (2012b). Moving and the neighborhood glass ceiling. Science, 337(6101), 1464–1465.
Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. Annual Review of Sociology, 28(1), 443–478.
Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. Science, 277(5328), 918–924.
Su, J. G., Apte, J. S., Lipsitt, J., Garcia-Gonzales, D. A., Beckerman, B. S., de Nazelle, A., Texcalac-Sangrador, J. L., & Jerrett, M. (2015). Populations potentially exposed to traffic-related air pollution in seven world cities. Environment International, 78, 82–89.
Taylor, N. G., Stymest, J., & Mah, C. L. (2020). Validation of a province-wide commercial food store dataset in a heterogeneous predominantly rural food environment. Public Health Nutrition, 23(11), 1889–1895.

Townley, G., Brusilovskiy, E., Snethen, G., & Salzer, M. S. (2018). Using geospatial research methods to examine resource accessibility and availability as it relates to community participation of individuals with serious mental illnesses. American Journal of Community Psychology, 61(1–2), 47–61. Vallée, J., Shareck, M., Le Roux, G., Kestens, Y., & Frohlich, K.L. (2020). Is accessibility in the eye of the beholder? Social inequalities in spatial accessibility to health-related resources in Montréal, Canada. Social Science & Medicine, 245, 112702.
van der Werf, P., Seabrook, J. A., & Gilliland, J. A. (2019). Food for naught: Using the theory of planned behaviour to better understand household food wasting behaviour. The Canadian Geographer/Le Géographe canadien, 63(3), 478–493.
Vaz, E., & Arsanjani, J. (2015). Crowdsourced mapping of land use in urban dense environments: An assessment of Toronto. The Canadian Geographer/Le Géographe canadien, 59(2), 246–255.
Walton-Roberts, M., Veronis, L., Wayland, S. V., Dam, H., & Cullen, B. (2019). Syrian refugee resettlement and the role of local immigration partnerships in Ontario, Canada. The Canadian Geographer/Le Géographe canadien, 63(3), 347–359.

Webb, S., Janus, M., Duku, E., Raos, R., Brownell, M., Forer, B., Guhn, M., Hertzman, C., Maggi, S., Roos, N., Sayers, M., & Muhajarine, N. (2017). Neighbourhood socioeconomic status indices and early childhood development. SSM-Population Health, 3, 48–56.
Woudsma, C., & Jakubicek, P. (2020). Logistics land use patterns in metropolitan Canada. Journal of Transport Geography, 88, 102381.
Young, M., & Singh, D. (2022). The Canadian family-friendly community resources study for better balance, health and well-being. Community, Work & Family, 1–27. https://doi.org/10.1080/13668803.2022.2151869
Young, M., & Wheaton, B. (2013). The impact of neighborhood composition on work-family conflict and distress. Journal of Health and Social Behavior, 54(4), 481–497.
Zhang, H., & Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap road networks. International Journal of Geospatial and Environmental Research, 5(2), Article 1.

How to cite this article: Young, M., S. Leipe, and D. Singh. (2024). Best practices for measuring community resources across Canada: A comparison of coding classifications. Canadian Geographies / Géographies canadiennes, 68, 115–128 https://doi.org/10.1111/cag.12867