

Health Reports

Accuracy of matching residential postal codes to census geography

by Lauren Pinault, Saeeda Khan and Michael Tjepkema

Release date: June 17, 2020



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

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2020

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Accuracy of matching residential postal codes to census geography

by Lauren Pinault, Saeeda Khan and Michael Tjepkema

Abstract

Background: Postal codes are often the only geographic identifier available to match subjects in a health dataset to census geography. This paper describes the characteristics of postal codes reported by the Canadian population on the census and, as an indicator of geocoding accuracy, the proportion that are linked to a single dissemination area (DA).

Data and methods: Postal codes reported on the 2016 Census questionnaire were matched to a combination of the Postal Code Conversion File (PCCF) and the Postal Code Conversion File Plus (PCCF+ version 7B) (reference date November 2018) to calculate population-weighted counts and the number of matches to DAs by province or territory, delivery mode type (DMT), population centre or rural area size, and census metropolitan area. The number of single matches to census tracts (CTs), census subdivisions (CSDs) and census divisions (CDs) was also calculated.

Results: In Canada, 72.6% of the population reported postal codes that matched to a single DA. This proportion was higher in urban cores (87.1%) and among postal codes for an urban street address (DMT=A) (85.3%) or apartment building (DMT=B) (95.3%), and was lower in rural areas (26.2% to 38.1%) and among rural postal codes (13.9%). In comparison, 89.3% and 95.4% of the population reported postal codes matching to a single CSD or CD, respectively, while 92.1% of the population that live within CT boundaries were matched to a single CT.

Interpretation: Matching postal codes to census geography is relatively accurate and frequently one to one in urban centres. In rural areas and for some types of postal code DMTs, alternative approaches to using the PCCF and PCCF+ for attaching census geography might be explored.

Keywords: postal code, Postal Code Conversion File Plus (PCCF+), delivery mode type, geocoding, geography

DOI: <https://www.doi.org/10.25318/82-003-x202000300001-eng>

In the majority of Canadian health survey or administrative datasets, the geographic information available to researchers is limited to a residential postal code rather than a full street address, often for confidentiality reasons. Postal codes are six-character alphanumeric codes created by Canada Post Corporation to sort and deliver mail.¹ Since postal codes do not always reflect well-defined, discrete and homogeneous spatial units, there is some uncertainty in using them to assign geographic attributes, such as exposures to environmental hazards. Postal code geography also does not always correspond with census geography, since many postal codes cross census boundaries.² Therefore, postal codes may be linked to one or more units of census geography.

Postal codes are used by a broad range of health research disciplines to assign contextual covariates to subjects. Neighbourhood estimates of socioeconomic deprivation and other demographic variables have been used to predict behavioural problems in children³ and higher rates of depression in urban areas.⁴ The Public Health Agency of Canada has created area-based indicators to capture health status and health determinants to highlight regional health inequalities in Canada.⁵ In the absence of individual-level socioeconomic data in a health dataset, researchers have also used area-based estimates as a proxy for personal income to adjust for model confounding. For example, neighbourhood income was used to adjust models of diabetes prevalence among immigrants to Ontario,⁶ in a study of patients with rheumatoid arthritis who received specialized services,⁷ and in a study of people who sought treatment for HIV/AIDS in British Columbia.⁸ Postal codes have been used

alongside inconsistent street address records in a coroners' dataset to improve the geolocation of drug overdose sites, therefore allowing researchers to evaluate the efficacy of a supervised injection site in Vancouver.⁹

In environmental health research, geography based on postal code is used in geographic information systems to link survey and administrative health datasets to environmental data from a range of sources. For example, for the purpose of cross-sectional analysis, obesity and diabetes data from the Canadian Community Health Survey have been attached to area-based population counts and residential density from the census; availability of walkable destinations from government datasets; and indicators of land-use mix, street connectivity, and the density of fast-food restaurants and grocery stores from proprietary datasets.^{10,11} In longitudinal mortality studies, large population-based health cohorts have been assigned estimates of ambient air pollution, greenness and blue space based on coordinates derived from postal codes reported on tax files.¹²⁻¹⁵

Statistics Canada produces two products that attach six-character postal codes to units of census geography: (1) the Postal Code Conversion File (PCCF),² which provides all possible matches between postal codes and census geography, and (2) the Postal Code Conversion File Plus (PCCF+),¹⁶ which uses both population weighting and random allocation to assign specific census geography to individuals in a dataset. Essentially, the PCCF+ uses population weighting to inform the selection of one of the matches provided on the PCCF file for each record. A secondary function of the PCCF+ is to assign a representative point of latitude and longitude to each individual; the

What is already known on this subject?

- Postal codes are often the only geographic identifier available to assign census geography to subjects in a health dataset.
- The Postal Code Conversion File Plus (PCCF+) provides a means to match postal codes to units of census geography, using population weights.
- Positional accuracy (latitude/longitude coordinates) in the PCCF+ are most accurate for urban apartment buildings and street addresses.

What does this study add?

- This study uses postal codes reported on the 2016 Census of Population to determine how frequently postal codes are matched to one unit of census geography in the PCCF+, as a measure of accuracy.
- In Canada, 72.6% of the population matched to a single dissemination area (DA), 89.3% to a single census subdivision, and 95.4% to a single census tract.
- Accuracy (match to single DA) was higher in urban apartment buildings and street addresses, and was lower in rural areas.

positional accuracy of this assignment across Canada has been assessed in a previous paper.¹⁷

Although hundreds of health studies have used the PCCF or PCCF+ for geocoding,¹⁸ the accuracy of assigning census geography based on postal codes across all of Canada remains poorly studied. One such study in Nova Scotia estimated positional error in assigning census geography by comparing the PCCF+ with the Nova Scotia Civic Address File. The study found lower rates of error in urban areas than in rural areas, although some misclassification occurred in all areas of the province.¹⁹

This study describes the characteristics of residential postal codes of the Canadian population using the 2016 Census and determines how frequently these postal codes are matched to one or more dissemination areas (DAs), a unit of census geography. DAs were assessed because they are small, relatively stable and discrete areas that represent approximately 400 to 700 people.²⁰ DAs are the smallest geographic unit for which census data are disseminated,²⁰ and they are the unit most frequently used in health research to attach neighbourhood measures or environmental exposures based on geography.^{3,6} Since a geocoding error could be introduced if a postal code matches to more than one DA, the number of possible matches to DAs was used here as an indicator of geocoding accuracy. For comparison, the percentage of the population matched to single census tracts (CTs), census subdivisions (CSDs) and census divisions (CDs) was also determined.

Methods

How the Postal Code Conversion File Plus geocodes to dissemination areas

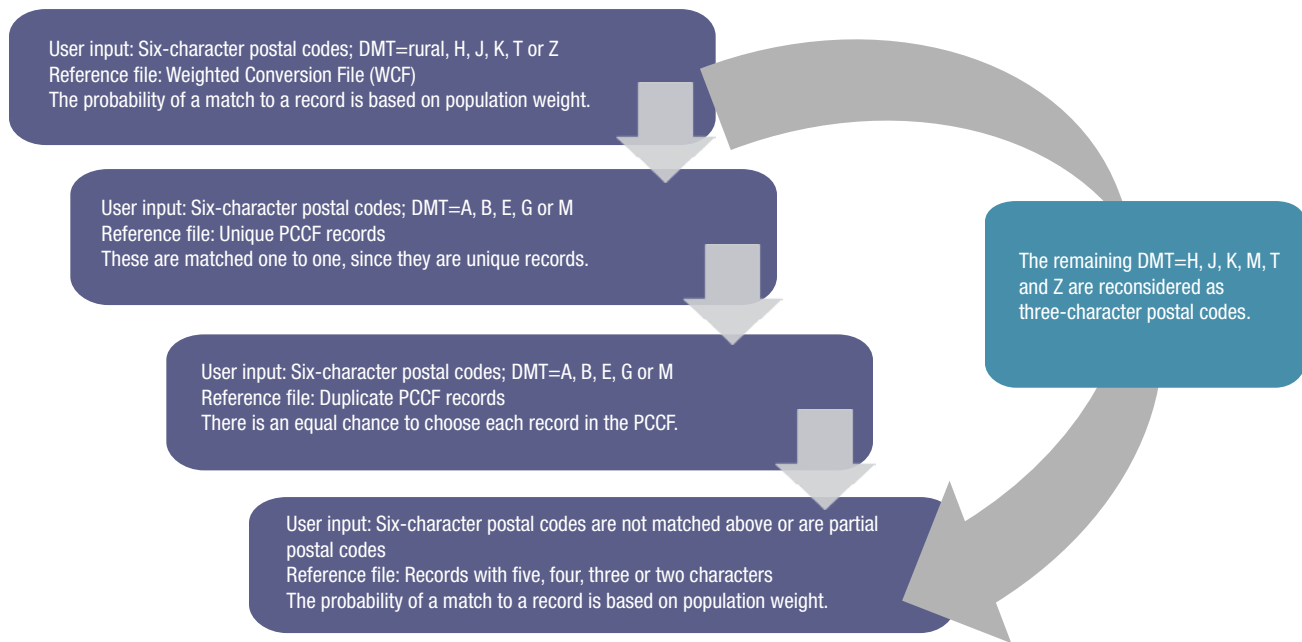
The PCCF includes all possible matches of postal codes to census geography (block face, dissemination block and DA). The PCCF+ program selects one of these matches for each postal code in a user-supplied health dataset using a stepwise process (Figure 1). First, rural postal codes (denoted by a second digit of “0”) and those with delivery mode types (DMTs) that indicate rural route delivery (DMT=H), general delivery (DMT=J), suburban service (DMT=T), post office boxes (DMT=K) or retired routes (DMT=Z) are matched to a subset of the weighted conversion file (WCF) based on probabilities from census population weights. The WCF is a dataset that combines postal codes from the PCCF with population counts from the census at the DA level. It is used to generate population weights for each combination of postal code and DA. Although postal codes in rural areas tend to have lower

geocoding accuracy because of their larger spatial extent (i.e., residences for a given postal code are spread over a larger area), the population weighting technique allows an individual-level dataset to be matched to census geography in a way that is consistent with the spatial distribution of the overall population.

PCCF records can include postal codes matched to one (unique) or many (duplicate) records of census geography. During the second step, the remaining postal codes that correspond with unique PCCF records (mostly urban postal codes) are matched to census geography. Third, the remaining postal codes that are matched to duplicate PCCF records are randomly matched with equal probability to all possible PCCF records. However, any remaining unmatched postal codes that have a DMT of rural, H, J, K or T are flagged during this step to be processed based on the first three characters of the postal code, called forward sortation areas (FSAs), in the final steps. In these final steps, the PCCF+ attempts to impute census geography based on partial postal codes (using the first five, four, three or two characters iteratively) using census population weights (Figure 1). Partial postal codes are considered where provided, or when full postal codes do not match to the PCCF.

Compared with the PCCF+ stepwise process, the original PCCF single-link indicator (SLI) selects one record for each postal code—generally the one with the largest number of dwellings.² Using the PCCF SLI, individuals in a health dataset would be matched only to DAs that are represented by the SLI. This results in clusters of high population areas surrounded by many DAs that are not matched to postal codes (Figure 2a). On the other hand, the PCCF+ allocates the rural population to DAs based on the population weights and allows users to ensure that the overall health dataset they are using is spatially distributed in a way that more closely resembles the actual population distribution (Figure 2b).

Figure 1
Flowchart of how the PCCF plus matches individuals in a user-supplied postal code dataset to subsets of PCCF records



Source: Postal Code Conversion File Plus version 7B (2019).

Descriptive statistics—methods

Postal codes for the entire population of Canada were obtained from the 2016 Census short-form questionnaire. All postal codes for the population of Canada were processed with PCCF+ version 7B to obtain the following variables: province or territory, DMT, population centre or rural area type, and census metropolitan area (CMA). PCCF+ version 7B uses postal code records as of November 2018.¹⁶

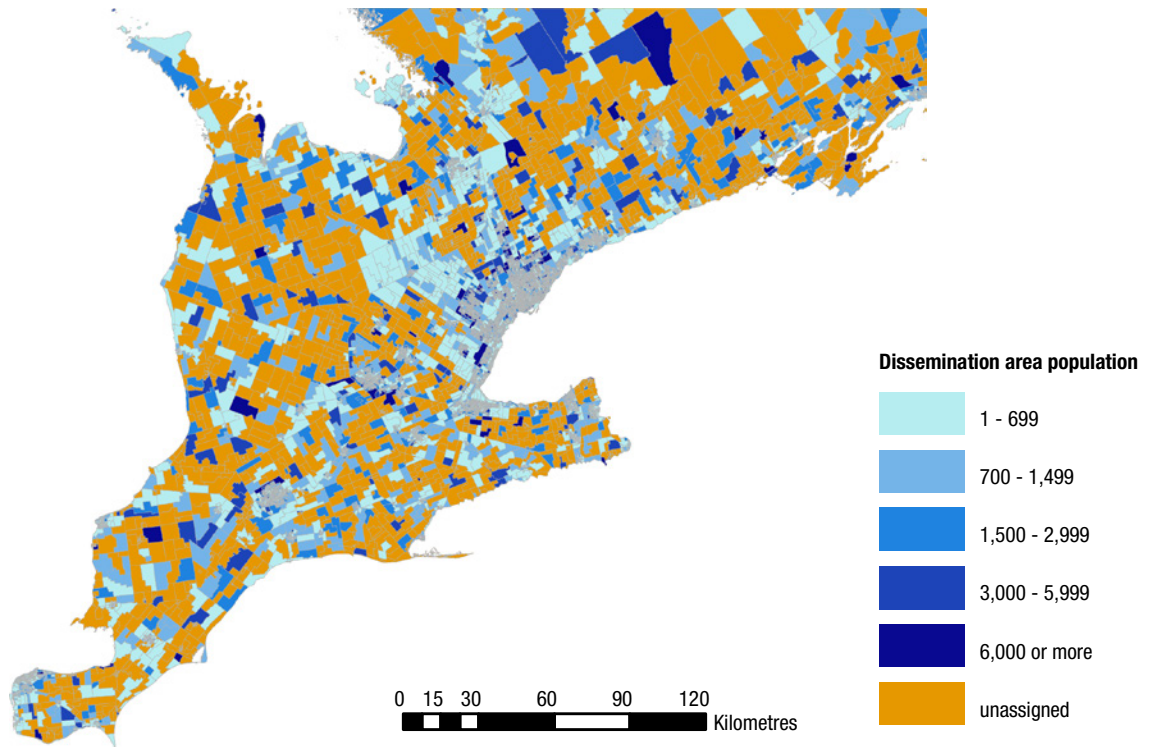
DMT indicates the method of mail delivery attributed to the postal code. It is useful to consider DMT in geocoding accuracy since previous studies have indicated that accuracy varies substantially by DMT category.¹⁷ The largest groupings of DMT categories are A (urban street address), B (urban apartment building) and rural postal code (no letter, but assigned a DMT of “W” by the PCCF+). DMTs E, G and M indicate business addresses or large volume receivers, and DMT K indicates post office boxes.¹ DMT Z indicates a postal code that has since been retired by Canada Post Corporation.¹ These were included since postal codes that appear

on the 2016 Census may have been retired since that date.

Population centre and rural area types are discrete geographic units that characterize the urban–rural continuum using a combination of population size, contiguity and density.²⁰ Core populations are defined as at least 50,000 people (CMA) or 10,000 people (census agglomeration [CA]). A secondary core is similar to a core in that it comprises at least 10,000 people, but it differs in that it was once the core of a CA that later merged with a growing CMA. Fringe areas include population centres within a CMA or CA with fewer than 10,000 people. Otherwise, the area is classified as rural.²⁰ CMAs are defined as having a total population of at least 100,000 people, of which 50,000 people must live in the core, while CAs must have a total population of at least 10,000 people.²⁰ Because of sample sizes, some analyses combined rural areas inside CMAs or CAs with rural areas outside CMAs or CAs into “rural areas,” and population centres outside CMAs and CAs with secondary cores as “secondary population centres.”

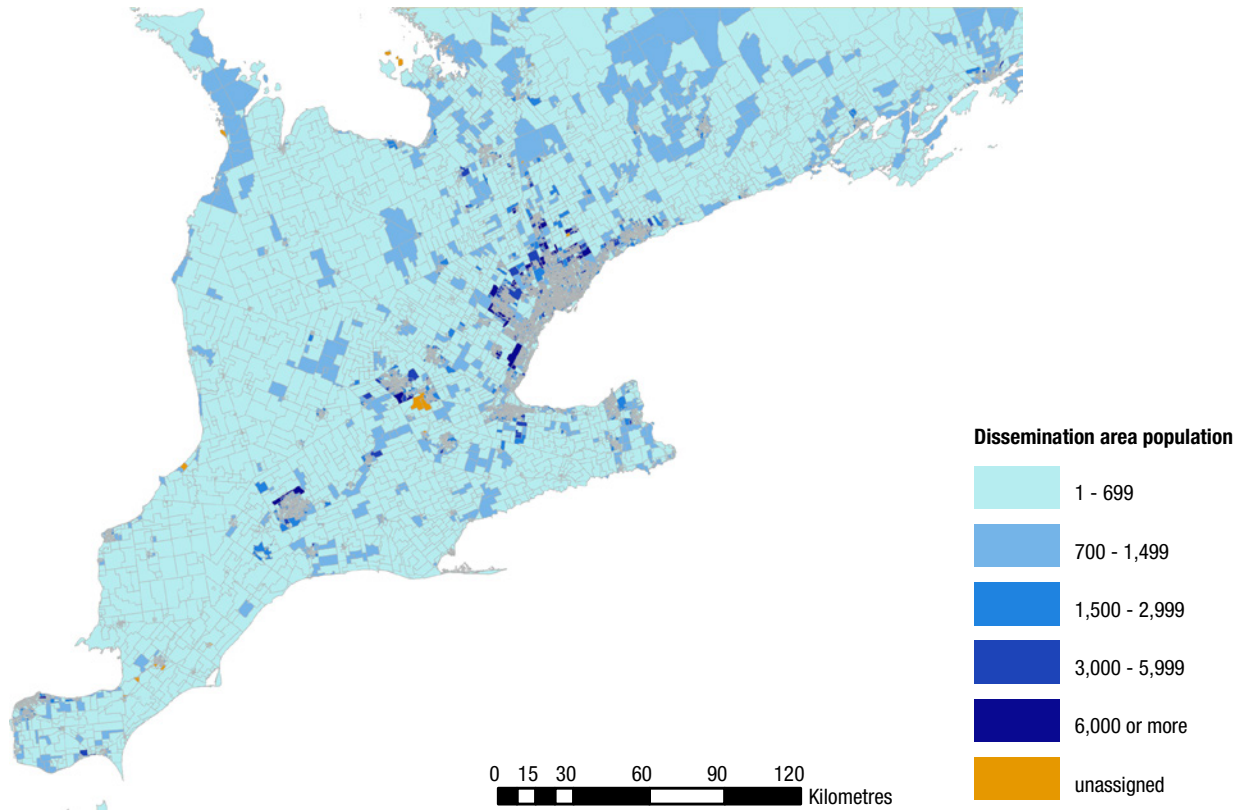
Since postal codes can be assigned to more than one DA, the number of possible matches between each postal code and DA (i.e., records in the PCCF) was calculated separately using the original PCCF. For example, if a postal code was matched to three possible DAs in the PCCF, a count of three was recorded. These counts were merged with the 2016 Census data to obtain population counts for each postal code–DA match. The percentage of the population with postal codes that link to one DA was used as an indicator of greater geocoding accuracy. Population counts and the percentages of the population linked to one or multiple DAs were calculated by province or territory, DMT, and population centre or rural area type. The percentage of the population with a postal code linked to a single CSD, CD or CT was also calculated for comparison. CDs are a subdivision of provinces and territories that are a set of neighbouring municipalities, often with common services.²⁰ CSDs are a subdivision of CDs and generally representative of single municipalities or areas treated as municipal equivalents for statistical purposes (e.g., Indian reserves).²⁰ CTs

Figure 2a
Postal Code Conversion File single-link indicator



Source: 2016 Census of Canada; Postal Code Conversion File (2018); Postal Code Conversion File Plus version 7B (2019).

Figure 2b
Postal Code Conversion File Plus



Source: 2016 Census of Canada; Postal Code Conversion File (2018); Postal Code Conversion File Plus version 7B (2019).

Table 1
Percentage of the Canadian population by postal code reported on the 2016 Census, postal code characteristics and the percentage of the population matched to each delivery mode type

	Population of Canada [†] number	Percentage of the Canadian population	Percentage of the Canadian population by delivery mode type											
			A	B	E	G	H	J	K	M	T	Rural	Z	Missing
														percent
Canada	35,151,700	100.00	70.59	9.78	0.13	0.21	1.19	0.00	0.35	0.02	0.59	16.74	0.28	0.11
Province or territory														
Newfoundland and Labrador	519,700	1.48	51.83	0.73	0.02	0.08	1.45	0.00	0.61	0.01	0.20	44.68	0.08	0.31
Prince Edward Island	142,900	0.41	44.96	1.66	0.00	0.01	1.55	0.00	0.55	0.01	0.00	50.85	0.00	0.41
Nova Scotia	923,600	2.63	59.17	6.70	0.12	0.11	2.54	0.00	0.41	0.00	0.00	30.68	0.08	0.18
New Brunswick	747,100	2.13	96.12	2.40	0.17	0.11	0.00	0.02	0.16	0.00	0.01	0.00	0.76	0.24
Quebec	8,164,400	23.23	75.34	4.51	0.08	0.24	0.41	0.00	0.06	0.01	1.14	18.13	0.06	0.02
Ontario	13,448,500	38.26	68.21	14.48	0.16	0.23	1.96	0.00	0.14	0.01	0.75	13.74	0.20	0.12
Manitoba	1,278,400	3.64	62.11	6.86	0.03	0.35	1.43	0.01	0.77	0.03	0.03	27.89	0.32	0.17
Saskatchewan	1,098,500	3.12	52.56	3.36	0.02	0.20	1.54	0.00	1.10	0.09	0.14	40.31	0.39	0.29
Alberta	4,067,000	11.57	74.49	7.39	0.16	0.09	1.20	0.01	1.17	0.04	0.09	14.63	0.53	0.20
British Columbia	4,648,100	13.22	74.25	13.10	0.18	0.16	0.12	0.01	0.36	0.01	0.18	10.93	0.67	0.04
Yukon	35,900	0.10	65.67	2.55	0.01	0.26	0.10	0.00	8.17	0.00	0.00	22.84	0.03	0.39
Northwest Territories	41,800	0.12	34.42	4.65	0.23	3.97	0.00	0.00	4.58	0.04	0.00	52.01	0.09	0.00
Nunavut	35,900	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
Population centre or rural area type														
Core	24,401,500	69.42	84.53	13.65	0.16	0.23	0.14	0.00	0.07	0.01	0.19	0.70	0.27	0.05
Fringe	1,102,800	3.14	61.72	1.75	0.03	0.16	0.41	0.01	0.61	0.00	0.18	34.91	0.15	0.07
Rural area inside a CMA or CA	2,543,100	7.23	46.77	0.84	0.01	0.03	7.21	0.00	1.03	0.05	4.27	38.57	0.85	0.38
Population centre outside a CMA or CA	1,814,700	5.16	38.00	0.92	0.08	0.15	0.75	0.02	1.27	0.06	0.13	58.31	0.07	0.25
Rural area outside a CMA or CA	4,069,800	11.58	13.12	0.08	0.00	0.01	4.37	0.01	1.16	0.00	1.14	79.66	0.18	0.27
Secondary core	1,204,700	3.43	90.66	3.95	0.11	0.13	0.63	0.00	0.20	0.01	0.22	3.84	0.16	0.07
Missing	15,100	0.04	0.00	0.00	30.23	54.07	0.03	0.10	2.16	7.28	0.09	5.94	0.03	0.07

[†] rounded to the nearest hundred for confidentiality

Notes: Delivery mode types (DMT): A - urban street address, B - urban apartment building, E - business building, G - large volume receiver, H - rural route, J - general delivery, K - post office box, M - large volume receiver (post office box), T - suburban service, Rural - rural postal code, Z - retired postal code and Missing. CMA refers to census metropolitan area and CA refers to census agglomeration.

Source: 2016 Census of Canada; Postal Code Conversion File Plus version 7B (2019).

are only located within CMAs or CAs and are small, stable areas, usually with a population of fewer than 10,000 people.²⁰

Results

The majority of Canadians were serviced by a DMT of type A (urban street address, 70.6%), B (urban building address, 9.8%) or rural postal code (16.7%); however these proportions varied among the provinces and territories (Table 1). In Quebec, Ontario, Alberta and British Columbia, proportions were similar to national estimates. However, rural postal codes were more common in the Atlantic provinces (except New Brunswick, where rural postal codes have been discontinued by Canada Post Corporation), Manitoba, Saskatchewan and the three territories. DMT also varied substantially among population centre and rural area types. Urban street address (DMT=A) was the most common in a CMA or CA core or fringe, as well as in a secondary

core. However, in population centres and rural areas outside a CMA or CA, rural postal codes were the most common.

In Canada, 72.6% of the population's postal codes were matched to a single DA, and 9.8% were matched to two DAs (Table 2). However, these proportions varied by province and territory, DMT, and population centre or rural area type. Among provinces and territories, similarly high proportions of the population (68.0% to 79.5%) were matched to a single DA in Quebec, Ontario, Manitoba, Saskatchewan, Alberta and British Columbia. However, in Prince Edward Island, 48.0% of the population was matched to five or more DAs, and in Nunavut, 71.9% of the population was matched to two or three DAs. By DMT, individuals with urban street addresses (DMT=A) and those residing in apartment buildings (DMT=B) were matched to a single DA in 85.3% and 95.3% of cases, respectively. However, those with rural postal codes were most commonly

matched to five or more DAs (62.9%), and rarely matched to a single DA (13.9%). By population centre or rural area type, individuals living in a CMA or CA core or in a secondary core were matched to a single DA the majority of the time (from 74.9% to 87.1%). In contrast, individuals living in rural areas and population centres outside CMAs and CAs were more commonly matched to multiple DAs.

Table 3 presents the proportion of people matched to a single DA by combined characteristics. In general, individuals with urban street addresses (DMT=A) were matched to a single DA most of the time (72.6% to 98.4%), except in fringe areas of New Brunswick and Manitoba, secondary population centres in Saskatchewan and Alberta, and rural areas of the Prairies and British Columbia. People living in apartment buildings (DMT=B) had the highest accuracy (matching to a single DA 84.0% to 100.0% of the time), except in

Table 2
Percentage of the Canadian population matched to one dissemination area or more than one dissemination area in the Postal Code Conversion File Plus, by postal code characteristics

	Population of Canada ^a number	Percentage of the population with postal codes, by number of dissemination area links					
		1	2	3	4	5 or more	Missing
		percent					
Canada	35,151,700	72.56	9.77	3.03	1.98	12.54	0.12
Province or territory							
Newfoundland and Labrador	519,700	62.30	12.09	4.97	4.95	15.38	0.31
Prince Edward Island	142,900	44.01	3.82	1.30	2.45	48.01	0.41
Nova Scotia	923,600	57.07	8.35	3.76	2.25	28.39	0.18
New Brunswick	747,100	85.74	8.52	2.87	1.36	1.27	0.24
Quebec	8,164,400	71.94	9.43	3.47	2.64	12.50	0.02
Ontario	13,448,500	72.21	9.47	2.07	0.94	15.18	0.12
Manitoba	1,278,400	71.32	9.89	4.54	3.40	10.67	0.19
Saskatchewan	1,098,500	67.95	12.48	5.53	4.61	9.12	0.31
Alberta	4,067,000	72.35	11.58	3.33	2.47	10.02	0.24
British Columbia	4,648,100	79.47	9.24	3.17	2.10	5.99	0.04
Yukon	35,900	95.70	2.61	0.02	0.00	1.29	0.39
Northwest Territories	41,800	69.84	13.94	5.97	2.63	7.55	0.06
Nunavut	35,900	28.15	31.30	40.55	0.00	0.00	0.00
Delivery mode type							
A - urban street address	24,812,400	85.32	11.24	2.09	0.68	0.68	0.00
B - urban apartment building	3,438,200	95.25	2.38	1.28	0.86	0.23	0.00
E - business building	46,100	90.64	4.96	1.75	1.79	0.86	0.00
G - large volume receiver	72,400	95.80	2.56	0.98	0.54	0.11	0.00
H - rural route	420,000	6.80	4.11	3.29	5.64	80.16	0.00
J - general delivery	1,400	43.29	27.69	1.54	27.48	0.00	0.00
K - post office box	124,200	63.67	9.19	5.27	4.50	17.37	0.00
M - large volume receiver (post office box)	5,400	53.43	29.97	10.62	0.81	4.37	0.80
T - suburban service	208,400	6.74	8.14	9.47	10.47	65.18	0.00
Rural - rural postal code	5,884,000	13.86	8.28	7.57	7.32	62.94	0.04
Z - retired postal code	100,000	10.94	24.47	15.49	14.96	34.14	0.00
Missing	39,300	0.00	0.00	0.00	0.35	0.17	99.48
Population centre or rural area type							
Core	24,401,500	87.12	9.04	1.76	0.73	1.30	0.05
Fringe	1,102,800	52.31	11.84	3.44	2.31	30.04	0.07
Rural area inside a CMA or CA	2,543,100	38.10	11.30	5.10	4.44	40.63	0.43
Population centre outside a CMA or CA	1,814,700	39.84	9.08	4.55	5.17	41.09	0.27
Rural area outside a CMA or CA	4,069,800	26.15	11.15	8.60	6.73	47.09	0.29
Secondary core	1,204,700	74.90	15.84	2.93	1.00	5.26	0.07
Missing	15,100	82.27	7.95	1.77	1.66	6.36	0.00

^a rounded to the nearest hundred for confidentiality

Notes: CMA refers to census metropolitan area and CA refers to census agglomeration.

Source: 2016 Census of Canada; Postal Code Conversion File Plus version 7B (2019).

rural Alberta and secondary population centres in Manitoba. Individuals with rural postal codes were generally less frequently matched to a single DA, but results varied by province and territory as well as by population size. The proportions of people matched to a single DA varied substantially for all of the other DMTs examined. Note that not all provinces and territories have all DMTs and population centre or rural area types.

Table 4 provides population estimates, the proportion of the population matched

to the three most common DMTs and the proportion of the population matched to a single DA, for all CMAs in Canada. For most CMAs, the vast majority of the population was linked to urban street addresses (DMT=A) and to a single DA. In Ontario, the CMAs of Kingston, Peterborough and Greater Sudbury had a higher proportion of the population with rural postal codes (20.5% to 24.6%) and a lower overall proportion of the population linked to a single DA (61.9% to 70.2%). For Barrie, Ontario, only 52.5%

of the population was linked to a single DA despite 80.7% of the population having an urban street address (DMT=A).

To place the previous results into context, the percentages of the population linked to a single CSD, CD or CT by different characteristics are provided in Table 5. In general, most of the population (89.3%) was matched to a single CSD, with a notably lower percentage in Prince Edward Island (47.5%) and Nova Scotia (75.9%), rural areas (51.1% to 68.3%), and population centres outside a CMA or CA (67.9%). The proportion of the population matched to a single CSD was extremely high for DMTs A, B, E, G and M (greater than 99%), but was considerably lower for DMTs of types H and rural (43.5% and 44.1%). Results for CDs followed similar patterns as those for CSDs; however, the percentage match to a single CD was higher than for a CSD. Almost 26.7 million Canadians were matched to a CT, of whom 92.1% were matched to a single CT record. The percentage of postal codes matched to a single CT was lower for DMTs of types H, T, rural and Z, as well as for all areas outside a CMA or CA.

Discussion

This study uses postal codes that match to a single DA as an indicator of greater accuracy in assigning DA-level neighbourhood variables. Overall, 72.6% of the population of Canada was matched to a single DA, and this proportion was higher (87.1%) when restricted to Canadians residing within the urban core of a CMA or CA. Individuals with a DMT of urban street address (DMT=A) or apartment building (DMT=B) were also most commonly linked to a single DA. These estimates were provided by province and territory and by specific CMA and CA to inform studies of provincial and regional health datasets. For larger units of census geography, the percentages of the population with a postal code that linked to a single unit were 89.3% for CSDs, 95.4% for CDs and 92.1% for CTs.

Table 3
Percentage of the Canadian population with a postal code that links to one dissemination area, by delivery mode type, population centre type, and province or territory

Delivery mode type	Population of Canada within a delivery mode type and population centre type combination number	Percentage of people with postal codes linked to one dissemination area, by province or territory												
		N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
A - Urban street address														
Core	17,963,500	88.4	92.2	86.5	92.9	88.2	85.3	91.0	92.7	85.6	88.7	97.6	77.6	...
Fringe	530,400	98.4	...	84.0	67.2	79.8	76.6	67.6	90.7	81.5	78.7
Rural area	1,274,900	80.0	88.0	72.6	84.0	82.4	73.2	68.6	67.6	54.7	58.6	88.8	87.3	...
Secondary core	1,400,000	83.1	...	90.0	84.7	82.9	75.7	91.8	46.9	69.4	86.4
B - Urban apartment building														
Core	3,176,400	91.2	100.0	99.0	97.9	96.4	95.2	92.2	94.7	93.6	96.1	95.7	100.0	...
Fringe	19,000	100.0	100.0	100.0	99.2	100.0	...	100.0	96.5
Rural area	18,400	...	100.0	100.0	98.6	94.3	100.0	86.9	84.0	59.0	85.8	...	100.0	...
Secondary core	61,000	100.0	...	100.0	85.5	99.1	97.3	72.5	92.2	91.2	95.8
E, G, K or M														
Core	98,600	83.4	88.2	77.6	100.0	85.8	91.0	98.2	85.1	69.0	83.2	100.0	61.8	...
Fringe	5,000	4.9	97.3	82.8	45.8	...	100.0	100.0	97.0
Rural area	52,100	90.4	8.5	8.7	45.6	82.0	64.4	91.5	45.2	84.5	25.0	81.5	0.0	...
Secondary core	25,100	99.3	...	100.0	68.2	97.0	94.4	100.0	100.0	32.8	93.8
H or T														
Core	5,300	100.0	0.0	0.0	...	8.9	1.3	...	15.6	13.2	10.2
Fringe	2,000	28.1	100.0	...	100.0	0.0
Rural area	35,100	4.1	0.0	0.0	24.1	5.2	2.1	22.7	30.1	24.6	21.2
Secondary core	200	0.0	...	1.3	0.0	19.1
Rural - Rural postal code														
Core	4,400	41.1	21.7	0.1	...	0.0	0.5	0.0
Fringe	16,200	0.0	0.0	0.0	...	0.0	1.2	1.2	13.1	24.4	8.6
Rural area	599,400	38.6	1.5	3.3	...	10.7	4.3	31.2	42.2	15.1	11.5	100.0	55.3	54.3
Secondary core	195,400	13.9	...	0.2	...	3.9	5.1	21.4	27.1	22.6	55.5	...	79.9	0.0

... not applicable

Notes: Delivery mode type: E - business building; G - large volume receiver; K - post office box; M - large volume receiver (post office box); H - rural route; T - suburban service.

Source: 2016 Census of Canada; Postal Code Conversion File Plus version 7B (2019).

In general, the finding that geocoding in urban areas is more accurate is consistent with the literature. Several other studies have examined point accuracy in geocoding (i.e., latitude and longitude), finding that postal codes can provide relatively accurate point estimates in urban centres, but are less accurate in rural regions.^{17,21-23} For example, a study of patients undergoing cardiac catheterization in Calgary found that postal code coordinates were within a city block of the residence 87.9% of the time.²¹ In a previous study of the accuracy of geocoding latitude and longitude using the PCCF+, median distances between the representative latitude and longitude (i.e., in the PCCF+) and the true location of the residence were small (0.16 km to 0.33 km) in communities of at least 10,000 people, but quite large in rural areas (5.60 km).¹⁷

In addition to urban and rural variables, users can consider the DMT

variable to enhance their knowledge of postal code geocoding accuracy. In a previous study, postal codes with DMTs A and B had the shortest distances between the representative latitude and longitude and the true location of the residence, whereas rural postal codes had some of the longest distances.¹⁷ Other DMTs had varying degrees of accuracy, but the current study indicates that these DMTs are used by about 2.8% of the Canadian population. Accuracy in this group is heterogeneous because it is dependent on the postal code route. For example, in a rural route (DMT=H), the route may begin at a post office and then follow a nonlinear path to a rural area, cutting across several DAs in the process. DMT is provided as an output variable by the PCCF+, and the program also flags some DMTs (i.e., postal codes associated with businesses) as possibly non-residential in the problem output file. The PCCF+ also pulls some of the DMTs that are

more likely to link to multiple DAs and attaches contextual data to a less precise three-character postal code (FSA) to reflect this uncertainty.

The PCCF+ is particularly useful in assigning census geography (e.g., DAs) for studies that attach neighbourhood-level covariates to a health dataset. DAs are frequently used because they are the smallest geographic unit of publicly available census data from Statistics Canada. Additionally, most neighbourhood-level socioeconomic indicators are produced at the DA level, such as the Canadian Marginalization Index (CAN-Marg)²⁴ and the Quebec index of material and social deprivation.²⁵ In environmental health research, several environmental indicators are also produced at the DA level, such as the Canadian Active Living Environments Database (Can-ALE).^{26,27} Most studies using these data focus on urban population centres, where DA-level geocoding

is most accurate. In addition, many contextual covariates are spatially autocorrelated, meaning that adjacent neighbourhoods (i.e., DAs) might have similar values of covariates attributed to them. For example, ethnic concentration might be similar across adjacent DAs in a given urban centre. In these cases, if a postal code is randomly assigned to an adjacent DA, the neighbourhood estimate for ethnic concentration is still likely to be similar. Therefore, issues may arise for studies that include either more rural areas, or a particularly heterogeneous neighbourhood covariate across the landscape.

For primarily urban health datasets (e.g., city-level cohorts), DA assignment may therefore be considered reasonably accurate. However, in national cohorts, or cohorts that are primarily rural, DA assignment may present a problem. In these cases, if a population dataset is sufficiently large and the research question requires that the overall population distribution be only generally representative, inaccurate DA assignment might be mitigated by the average assignment of DAs across the landscape based on population weighting. However, in cases where a research question requires that the DA correspond with specific members of a cohort, such as when comparing individual characteristics with neighbourhood characteristics, or where a cohort is smaller in size, then this inaccuracy in geocoding might present a serious limitation. In these cases, a recommended alternative approach might be to consolidate DAs to a larger unit, such as a CSD or CD, to improve accuracy in matching postal codes to census geography.

The PCCF+ can help users identify different sources of possible error in an input health dataset by providing some diagnostic functionality. Missing or invalid postal codes in the source file will be flagged if they are not present in the PCCF. In these cases, users can manually check the postal codes, address ranges and street spelling using the Canada Post Corporation Internet lookup tool, or verify these postal codes against other sources (e.g., with members of a particular community). In its output, the

Table 4**Percentage of the population in census metropolitan areas by delivery mode type and percentage of the population with postal codes linked to a single dissemination area**

Census metropolitan areas and census agglomerations	Population (published, [†] enumerated) number	Percentage of the population by delivery mode type				Percentage of the population linked to a single DA
		A	B	Rural	Other	
Not in a CMA or CA	...	20.7	0.3	73.0	5.8	30.3
Census agglomeration	...	78.4	3.7	12.7	5.1	72.9
Census metropolitan areas						
St. John's	205,955	93.0	1.4	3.7	1.6	84.1
Halifax	403,390	80.4	14.7	4.3	0.6	78.6
Moncton	144,810	93.2	4.5	0.0	2.2	88.6
Saint John	126,202	95.7	3.7	0.0	0.6	88.9
Saguenay	160,980	82.3	1.4	10.0	6.2	78.7
Québec	800,296	84.1	6.7	7.8	1.4	81.7
Sherbrooke	212,105	85.1	3.2	8.6	3.1	76.3
Trois-Rivières	156,042	91.0	2.2	5.5	1.2	88.1
Montréal	4,098,927	89.6	6.6	2.5	1.2	84.1
Ottawa–Gatineau	1,323,783	80.1	9.9	8.0	2.0	76.3
Kingston	161,175	64.0	9.5	20.5	6.1	70.2
Peterborough	121,721	61.9	6.0	24.6	7.2	66.4
Oshawa	379,848	87.2	7.5	2.7	2.4	76.5
Toronto	5,928,040	74.1	23.5	1.3	1.1	83.8
Hamilton	747,545	76.2	13.4	8.5	1.9	78.4
St. Catharines–Niagara	406,074	69.3	7.3	19.4	4.0	73.9
Kitchener–Cambridge–Waterloo	523,894	83.2	10.8	4.5	1.5	82.8
Brantford	134,203	73.1	4.9	14.3	7.6	72.4
Guelp	151,984	75.0	11.5	8.1	5.5	73.3
London	494,069	73.1	16.8	7.4	2.6	81.9
Windsor	329,144	80.9	5.7	10.6	2.9	78.7
Barrie	197,059	80.7	4.0	12.7	2.6	52.5
Greater Sudbury	164,689	64.9	7.7	23.3	4.2	61.9
Thunder Bay	121,621	91.8	3.9	3.9	0.3	85.2
Winnipeg	778,489	82.8	10.5	5.1	1.5	86.3
Regina	236,481	86.5	4.5	5.4	2.4	86.9
Saskatoon	295,095	76.5	6.9	13.5	3.1	80.4
Calgary	1,392,609	88.7	9.1	1.1	1.1	84.5
Edmonton	1,321,426	84.5	10.4	3.2	1.6	82.1
Kelowna	194,882	84.8	12.1	2.9	0.2	82.0
Abbotsford–Mission	180,518	86.9	12.2	0.0	0.8	73.5
Vancouver	2,463,431	78.7	19.4	0.2	1.7	89.8
Victoria	367,770	90.6	8.9	0.0	0.5	88.5

... not applicable

[†] Statistics Canada, Focus on Geography Series, 2016 Census, 2017.

Notes: CMA refers to census metropolitan area and CA refers to census agglomeration. DMT refers to delivery mode type and DA refers to dissemination area.

Source: 2016 Census of Canada; Postal code Conversion File Plus version 7B (2019).

PCCF+ provides users with a quality indicator for PCCF records that were geocoded using the automated system.¹⁶ The PCCF+ also flags probable non-residential postal codes, that is, DMTs E, G and M. Since the PCCF+ uses population weighting in rural areas to attach census geography, it might be preferable to use the PCCF SLI to assign geography to these flagged postal codes. Canada Post Corporation can also retire postal codes and reinstate (or “rebirth”) them later in a different area. Therefore, users are encouraged to verify records where the

DMT=Z, indicating retired postal codes, and to understand that there may not be exact correspondence between the current location of a postal code and a historical one. This may be particularly relevant in longitudinal health studies. Historical DMTs are also provided as output and can be used to check the previous type of a postal code before it was retired and reinstated.

An additional consideration in using the PCCF+ to assign residential census geography is that the program relies on census population weighting to perform

Table 5
Percentage of the Canadian population matched to a single dissemination area, census subdivision, census division or census tract in the Postal Code Conversion File Plus, by postal code characteristics

	Population of Canada [‡] number	Percentage of the population with a postal code linked to a single DA, CSD or CD			Population within CTs [‡] number	Percentage of the population in a CT with a postal code linked to a single CT	
		DA	CSD	CD		percent	
		percent				percent	
Canada	35,151,700	72.56	89.27	95.43	26,685,200	92.13	
Province or territory							
Newfoundland and Labrador	519,700	62.30	81.76	97.44	205,400	97.14	
Prince Edward Island	142,900	44.01	47.51	81.46	0	...	
Nova Scotia	923,600	57.07	75.94	88.86	430,000	91.21	
New Brunswick	747,100	85.74	97.12	99.69	372,700	97.48	
Quebec	8,164,400	71.94	89.68	94.19	6,029,900	93.53	
Ontario	13,448,500	72.21	88.08	93.95	11,523,200	89.72	
Manitoba	1,278,400	71.32	88.29	95.88	788,800	92.86	
Saskatchewan	1,098,500	67.95	81.86	97.93	536,100	94.98	
Alberta	4,067,000	72.35	92.39	98.17	3,188,700	93.29	
British Columbia	4,648,100	79.47	94.53	99.47	3,610,400	95.18	
Yukon	35,900	95.70	98.70	100.00	0	...	
Northwest Territories	41,800	69.84	98.41	99.69	0	...	
Nunavut	35,900	28.15	100.00	100.00	0	...	
Delivery mode type							
A - urban street address	24,812,400	85.32	99.50	99.93	21,238,500	96.81	
B - urban apartment building	3,438,200	95.25	99.90	99.99	3,328,500	96.79	
E - business building	46,100	90.64	99.98	100.00	41,100	93.65	
G - large volume receiver	72,400	95.80	99.77	99.82	60,300	98.48	
H - rural route	420,000	6.80	43.48	88.01	149,100	19.68	
J - general delivery	1,400	43.29	96.29	100.00	700	76.85	
K - post office box	124,200	63.67	91.39	99.98	29,500	82.08	
M - large volume receiver (post office box)	5,400	53.43	99.89	100.00	2,300	60.33	
T - suburban service	208,400	6.74	58.71	95.66	128,100	26.72	
Rural - rural postal code	5,884,000	13.86	44.07	74.13	1,620,400	36.63	
Z - retired postal code	100,000	10.94	84.35	94.47	86,700	24.30	
Missing	39,300	0.00	88.18	100.00	200	3.59	
Population centre or rural area type							
Core	24,401,500	87.12	99.33	99.84	22,261,600	96.23	
Fringe	1,102,800	52.31	80.37	88.59	947,700	75.45	
Rural area inside a CMA or CA	2,543,100	38.10	68.34	84.35	1,827,500	60.40	
Population centre outside a CMA or CA	1,814,700	39.84	67.85	84.37	128,900	44.46	
Rural area outside a CMA or CA	4,069,800	26.15	51.06	81.50	308,500	49.56	
Secondary core	1,204,700	74.90	97.88	98.81	1,200,000	93.67	
Missing	15,100	82.27	99.93	100.00	11,100	92.83	

... not applicable

[‡] rounded to the nearest hundred for confidentiality

Notes: CMA refers to Census metropolitan areas; CA refers to Census agglomeration; DA refers to dissemination area; CSD refers to census subdivision; CD refers to census division and CT refers to census tract. Census tracts are only in urban areas.

Source: 2016 Census of Canada; Postal Code Conversion File Plus version 7B (2019).

random allocation of census geography. As a result, the program assumes that the geographic distribution of the health dataset is similar overall to that of the Canadian population. However, this may not always be the case. For example, researchers conducting a health study of an on-reserve Indigenous population may wish to manually select PCCF records that match to a reserve rather than to an adjacent town or city if it is known that the study respondents live on the reserve itself. Similarly, researchers studying the elderly population may wish

to select PCCF records that selectively include seniors' residences. Therefore, using the population distribution of Canada to conduct the random allocation of census geography, particularly in rural areas, presents an important limitation of the PCCF+ program for specialized health datasets where the distribution may be different from that of the general population.

Limitations

This study depended on postal codes captured by the 2016 Census of Population. The postal code variable is derived from three possible sources: the original Statistics Canada address files (i.e., the postal code used in the census mail-out), postal codes reported by respondents and postal codes mentioned on census form 3 (the personal questionnaire). Valid postal codes obtained from the address files used in mail-outs were used preferentially. In rare cases where a valid postal

code was not obtained from any of the three sources, it was imputed using a donor postal code. Another source of error is that respondents may not necessarily have reported the postal code at their place of residence, and instead may have reported a business postal code or post office box.²⁰ The extent to which this occurs in the census dataset is not presently known.

Matches between postal codes and DAs were derived from the PCCF, which is regularly updated with all new postal codes introduced by Canada Post Corporation.² Many of these records are manually edited and verified by Statistics Canada as they are added. However, any positional errors in the original PCCF file that have not been corrected during manual edits would also translate to errors in the PCCF+ processing, and

therefore affect DMT assignment and the DA–postal code counts.

Conclusions

This study highlights the importance of considering several geographic variables, including DMT, when determining the accuracy of matching postal codes to census geography. In general, the majority of Canadians are urban residents (69.4%) or have DMTs of type A or B (80.4%)—most of which were linked to a single DA (85.3% to 95.3%). Therefore, DA-level neighbourhood covariates can be relatively accurately assigned with postal codes in studies of urban Canada. A handful of specific CMAs, such as Barrie, Ontario, are less accurate.

However, in rural areas and areas with rural postal codes, where postal codes likely straddle the boundaries of several

DAs, postal codes were most frequently matched to multiple DAs. In these cases, the PCCF+ uses a random allocation algorithm to select a DA and, at a population level, assign the DA in a way that replicates the distribution of the overall population. However, for smaller (i.e., less representative) health datasets, or for datasets where it might be important to accurately match the neighbourhood to individual covariates, postal codes may not provide enough geographic detail for geocoding at the DA level. In these cases, either supplemental geographic information (e.g., address files) or consolidation to a larger census geographic unit (e.g., CTs) might help to mitigate geocoding error. ■

References

1. Canada Post. *Delivery Mode Data Technical Specifications*. Ottawa: Canada Post, 2014.
2. Statistics Canada. *Postal Code^{OM} Conversion File (PCCF), Reference Guide, 2018*. (Catalogue 92-154-G) Ottawa: Statistics Canada, 2018.
3. Boyle MH, Lipman EL. Do places matter? Socioeconomic disadvantage and behavioral problems of children in Canada. *Journal of Consulting and Clinical Psychology* 2002; 70(2): 378–89.
4. Matheson FI, Moineddin R, Dunn JR, et al. Urban neighborhoods, chronic stress, gender and depression. *Social Science and Medicine* 2006; 63: 2604–16.
5. Public Health Agency of Canada. *Pan-Canadian Health Inequalities Tool*. Ottawa: Public Health Agency of Canada, the Pan-Canadian Public Health Network, [Statistics Canada, and the Canadian Institute of Health Information](http://infobase.phac-aspc.gc.ca/health-inequalities/), 2017. Available online: <http://infobase.phac-aspc.gc.ca/health-inequalities/>
6. Creatore MI, Moineddin R, Booth G, et al. Age- and sex-related prevalence of diabetes mellitus among immigrants to Ontario, Canada. *Canadian Medical Association Journal* 2010; 182(8): 781–9.
7. Shipton D, Glazier RH, Guan J, et al. 2004. Effects of use of specialty services on disease-modifying antirheumatic drug use in the treatment of rheumatoid arthritis in an insured elderly population. *Medical Care* 2004; 42(9): 907–13.
8. Wood E, Montaner JSG, Tyndall MW, et al. Prevalence and correlates of untreated Human Immunodeficiency Virus Type 1 infection among persons who have died in the era of modern antiretroviral therapy. *The Journal of Infectious Diseases* 2003; 188: 1164–70.
9. Marshall BDL, Milloy M-J, Wood E, et al. Reduction in overdose mortality after the opening of North America's first medically supervised safer injecting facility: a retrospective population-based study. *Lancet* 2011; 377: 1429–37.
10. Poulidou T, Elliott SJ. Individual and socio-environmental determinants of overweight and obesity in Urban Canada. *Health & Place* 2010; 16: 389–98.
11. 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.
12. Pinault LL, Weichenthal S, Crouse D, et al. Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort. *Environmental Research* 2017; 159: 405–15.
13. Crouse D, Pinault L, Balram A, et al. Urban greenness and mortality in Canada's largest cities: a national cohort study. *The Lancet Planetary Health* 2017; 1(7): e289–e297.
14. Crouse D, Balram A, Hystad P, et al. Associations between living near water and risk of mortality among urban Canadians. *Environmental Health Perspectives* 2018; 126(7): 077008.
15. Christidis T, Erickson A, Pappin AJ, et al. The impact of low concentrations of PM air pollution on mortality in the Canadian Community Health Survey-Mortality cohort. *Environmental Health* 2019; *in press*.
16. Statistics Canada. *Postal Code^{OM} Conversion File Plus (PCCF+) Version 7B, Reference Guide*. Ottawa: Statistics Canada, 2019.
17. Khan S, Pinault L, Tjepkema M, et al. Positional accuracy of geocoding from residential postal codes versus full street addresses. *Health Reports* 2018; 29(2): 3–9.
18. Peller P. *An Analysis of the Postal Code Conversion File's Use in Research*. Calgary, Alberta: University of Calgary, 2011.
19. Terashima M, Kephart G. Misclassification errors from postal code-based geocoding to assign census geography in Nova Scotia, Canada. *Canadian Journal of Public Health* 2016; 107(4-5): e424–30.
20. Statistics Canada. *Dictionary, Census of Population, 2016*. (Catalogue 98-301-X2016001) Ottawa: Statistics Canada, 2018.
21. Bow CJD, Waters NM, Faris PD et al. Accuracy of city postal code coordinates as a proxy for location of residence. *International Journal of Health Geographics* 2004; 3: 5.

Accuracy of matching residential postal codes to census geography • Methodological Insights

22. Shah TI, Bell S, Wilson K. Geocoding for public health research: Empirical comparison of two geocoding services applied to Canadian cities. *The Canadian Geographer* 2014; 58(4): 400–17.
23. Healy MA, Gilliland JA. Quantifying the magnitude of environmental exposure misclassification when using imprecise address proxies in public health research. *Spatial and Spatio-temporal Epidemiology* 2012; 3: 55–67.
24. Matheson FI, Dunn JR, Smith KLW, et al. Development of the Canadian Marginalization Index: A new tool for the study of inequality. *Canadian Journal of Public Health* 2012; 103(S2): S12–6.
25. Pampalon R. *The Québec Index of Material and Social Deprivation: Methodological Follow-up, 1991 through 2006*. Institut National de Santé Publique du Québec, Gouvernement du Québec, 2011.
26. Herrmann T., Gleckner W., Wasfi R.A., et al. A pan-Canadian measure of active living environments using open data. *Health Reports* 2019; 30(5): 16-25.
27. Colley RC, Christidis T, Michaud I, et al. The association between walkable neighbourhoods and physical activity across the lifespan. *Health Reports* 2019; 30(9) : 3 - 13.