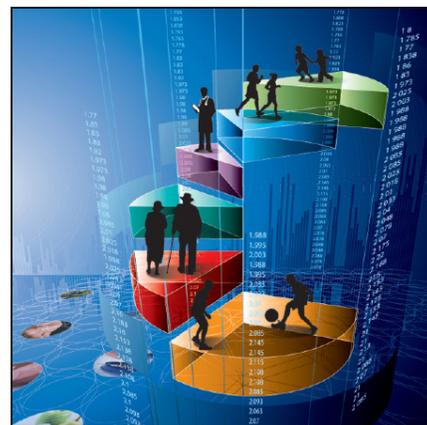


Health Reports

Positional accuracy of geocoding from residential postal codes versus full street addresses

by Saeeda Khan, Lauren Pinault, Michael Tjepkema and Russell Wilkins

Release date: February 21, 2018



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

© Minister of Industry, 2018

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.

Positional accuracy of geocoding from residential postal codes versus full street addresses

by Saeeda Khan, Lauren Pinault, Michael Tjepkema and Russell Wilkins

Abstract

Background: Postal codes are often the only geographic identifier available for assigning contextual or environmental information to a study population. This analysis assesses the influence of three factors—delivery mode type (mode of postal delivery), representative point type (source of latitude-longitude coordinates), and community size—on the accuracy of postal code spatial assignment.

Data and methods: PCCF+ (Postal Code Conversion File Plus) was used to assign delivery mode type, representative point type and community size to each individual in the 2011 Census of Canada. A sample (n = 1,004) was randomly selected with a minimum of 90 observations for each category of those three factors. Based on the address information of individuals in the sample, measures of positional accuracy for geocoding from residential postal codes (PCCF+) versus reference locations as determined by full street addresses (Google Maps) were calculated using a geographic information system. Accuracy was measured as the distance that the geocoded position differed from the full street address.

Results: Positional accuracy was related primarily to mode of postal delivery. Rural and mixed (partly urban, partly rural) modes had much higher geocoding error than did urban modes. Rural and small-town Canada and latitude and longitude based on dissemination area centroids had low accuracy, largely because of their close relationship to rural and mixed modes of delivery.

Interpretation: The accuracy of geocoding from postal codes can vary. Geocoding imprecision may result in misclassification, depending on the spatial resolution of the environmental or contextual measures. The spatial resolution required for a study helps to identify subpopulations that should be excluded because of inadequate positional accuracy.

Keywords: Community size, delivery mode type, misclassification, representative point type, street address

Geocoding based on a full street address is a highly accurate way of assigning geographic coordinates to an individual's residential location. However, address information is usually not available. As well, for many types of population health research, such precision is not necessary.

The positional accuracy required for a study depends on its context. To assign characteristics such as neighbourhood socio-economic variables, or to obtain census denominator data for a geographically defined at-risk population, reasonable accuracy is usually sufficient. By contrast, for some environmental health research, such as assigning walkability scores to urban residents, highly accurate spatial resolution may be required, since error of even a few hundred metres can result in substantial exposure misclassification.¹⁻⁵

In Canada, postal codes are often the only geographic identifier available for assigning contextual or environmental information to survey or administrative data. The Postal Code Conversion File (PCCF)⁶ and the Postal Code Conversion File Plus (PCCF+)⁷ provide the means to geocode datasets using only postal codes. To date, hundreds of studies have been published using these tools.⁸

The primary function of PCCF tools is to assign the full hierarchy of census geography (dissemination block, dissemination area, census tract, census metropolitan area or census agglomeration, census subdivision, census division, and province).⁹ A secondary function is assignment of latitude and longitude coordinates based on blockface, dissemination block, or dissemination area centroids.

This study compares the positional accuracy of geocoding using postal codes versus roof-top centroids derived from full street addresses (“reference locations”). The analysis is based on self-reported address data from the 2011 Census of Population. All reported postal codes were processed through PCCF+ Version 6C to obtain three variables that influence the accuracy of geocoding from postal codes: delivery mode type (DMT), representative point type, and community size. Positional accuracy is measured by the distance by which the geocoded position differs from the latitude/longitude of the reference location.

Delivery mode type (DMT)

DMT is a feature of postal codes that indicates the mode of mail delivery. DMTs can be grouped into three major categories: urban (A, B, E, G, M), rural (W), and mixed (H, J, K, T, X). Urban DMTs may be residential (A or B) or commercial and institutional (E, G, and M). Retired postal codes (Z) were excluded from the study, as were postal codes with a DMT of J (general delivery at an urban post office) or X (mobile routes in industrial areas), as they were rare in the census.

Urban DMTs. DMT A is for ordinary urban households with letter carrier delivery to the household or to a community mailbox. DMT B is for large urban apartment buildings; DMT E, for urban businesses; DMT G, for large urban commercial establishments and institutions with letter carrier delivery; and DMT M, for large urban commercial establishments and institutions that retrieve their mail at a lock box or bag at an urban

Table 1
Sample size and population represented, by delivery mode type, representative point type, and community size, Canada, 2011

	Sample		Unique matches	2011 Census population [†]	
	Number	%			%
Total	1,004	100.0	399		100.0
Delivery mode type (DMT)					
A (ordinary urban households)	408	40.6	182		69.6
B (large urban apartment buildings)	100	10.0	94		11.1
E, G, M (urban business and institutional)	90	9.0	82		0.7
H, K, T (mixed urban and rural)	180	17.9	37		2.1
W (rural)	226	22.5	4		13.9
Representative point type					
Blockface	474	47.2	289		75.7
Dissemination block	101	10.1	51		4.1
Dissemination area	429	42.7	59		18.1
Community size					
1,500,000 or more	204	20.3	98		34.8
500,000 to 1,499,999	150	14.9	86		18.5
100,000 to 499,999	126	12.5	63		17.9
10,000 to 99,999	259	25.8	106		12.1
Fewer than 10,000	265	26.4	46		15.3

[†] detail does not sum to 100% because of exclusion of 723,960 addresses with missing postal codes and 1,820 with postal codes having DMT of J (general delivery at urban post office) or X (mobile route in industrial area)

Source: 2011 Census of Canada.

post office. Most urban postal codes are linked to blockface or dissemination block representative points (centroids), usually with a single link, but sometimes with up to four.

Rural DMTs. Postal codes with “0” in the second position (DMT W) designate rural post offices, which may serve a large area encompassing multiple dissemination areas and census subdivisions (townships or villages), and all types of users, including apartment buildings, businesses, and institutions. DMT W also covers service to remote areas, including service via air stage offices,¹⁰ which were excluded from the sample. Most rural postal codes are linked to a large number of representative points (centroids of multiple dissemination areas).

Mixed DMTs. DMT H, K, and T are mixed (partly urban, partly rural). DMT H and T are for rural and suburban routes beginning at urban post offices but extending into rural areas. DMT of K (small post office boxes) may serve urban or rural clients, but at an urban post office. Most mixed mode postal codes are linked to multiple dissemination area centroids.

Representative point type

Representative point type refers to the source of the latitude and longitude coordinates assigned—in order of decreasing precision: blockface, dissemination block, dissemination area, or census subdivision. Census subdivision representative points are assigned as an interim measure pending further postal code address information, and occur rarely; postal codes with that representative point type were excluded.

Community size

Community size indicates the 2011 Census population: 1,500,000 or more (Toronto, Montreal and Vancouver census metropolitan areas—CMAs); 500,000 to 1,499,999 (Ottawa-Gatineau, Edmonton, Calgary, Quebec, Winnipeg and Hamilton CMAs); 100,000 to 499,999 (remaining 18 CMAs and 7 largest census agglomerations—CAs); 10,000 to 99,999 (remaining CAs); and less than 10,000 (rural and small-town Canada—all areas except CMAs or CAs).

Sample selection

The analytical sample (n = 1,004) was randomly selected from the census dataset based on retaining a target of 100 observations for each DMT, representative point type, and community size category (Table 1). An effort was made to sample equally from each region (Atlantic, Ontario, Quebec, Prairies, British Columbia, and northern territories). When possible, the same observations were re-used across DMT, representative point type, and community size categories.

The study intentionally oversampled from rural and mixed mode postal codes, from less accurate dissemination area representative point type, and from the smallest census agglomerations and rural and small-town Canada (fewer than 10,000 population).

Geocoding from full street addresses

To determine reference locations, all street addresses in the sample were manually searched using Google Maps with satellite imagery. The latitude and longitude for the centroid of the building location were obtained. When possible, the building address was visually confirmed with Google Street View.

Addresses not located during manual geocoding because of poor satellite imagery or a combination of lack of street address and no Street View were excluded, and new addresses were resampled (n = 121). Records that required resampling tended to be in rural and remote areas or on Indian reserves.

If possible, reported postal codes were verified on the Canada Post website.¹¹ Except for mixed mode postal codes (DMT H or T), if the first three characters did not match the expected postal code, it was presumed to be erroneous and was excluded from the analytical dataset.

All addresses in the sample were matched with all possible postal code representative points from PCCF+. Most addresses had multiple possible representative points. Of 105 records with

very long distances (more than 50 km) between the PCCF+ representative point and the address location from Google Maps, 51 were excluded because the postal code or street address appeared to contain typographical errors.

Distance calculations

For each record in the sample, all possible representative points were extracted from PCCF+ 6C. All of them retained their weights, starting with those that were assigned population-based weights—rural postal codes (DMT W), mixed postal codes (DMT H, K, or T), urban postal codes (DMT A, B, E, G, or M) with matches to four or more dissemination areas, and incompletely enumerated Indian reserves. Postal codes with a match to a single representative point were assigned to that point with a weight of 1. Urban postal codes with matches to multiple representative points were assigned equal weights. This weighting strategy replicates the reference location assignment process in PCCF+.

The manually assigned reference locations based on full street addresses and the representative points generated by PCCF+ were mapped in Geographic

Information Systems (ArcGIS v.10, ESRI 2010). The Euclidean (straight-line) distance between each reference location and each possible representative point was measured.

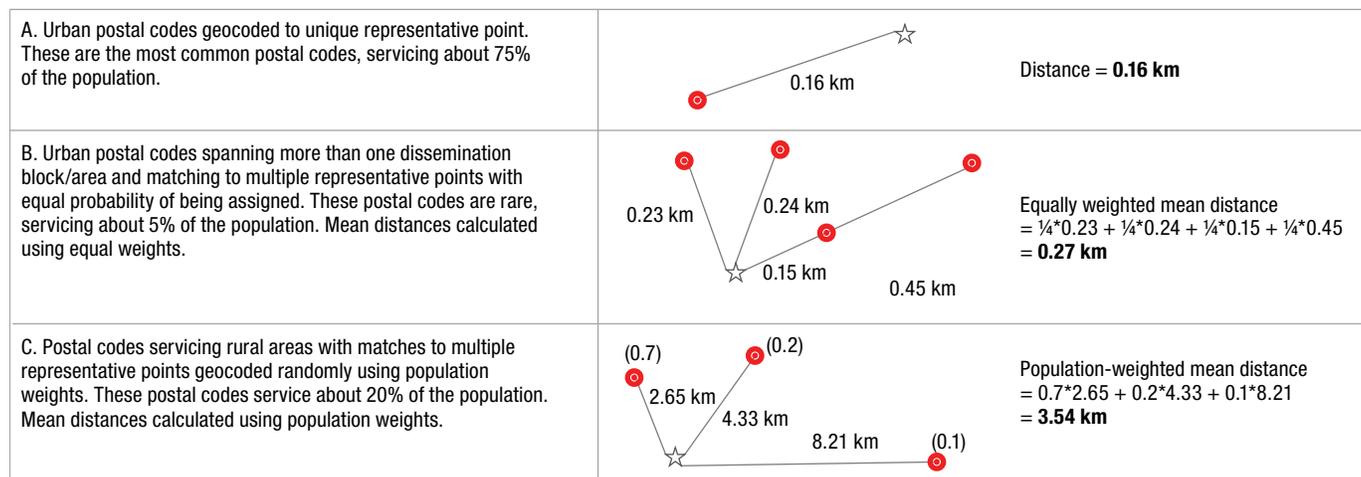
A weighted mean distance was calculated between the reference location and the PCCF+ representative points, using all the possible representative points for each postal code. For unique matches, only one representative point was possible (Figure 1A), but for urban records with multiple possible matches, all representative points had an equal probability of being assigned (Figure 1B). Figure 1C illustrates a rural case with three possible PCCF+ representative points for the postal code, with probabilities of assigning one of the three of 70%, 20%, and 10%. Those probability weights were used to calculate a weighted mean distance between the reference location and the three PCCF+ representative points. Using this process, one mean distance was calculated for each record.

Based on the calculated distance (or mean distance) between the reference location and the location geocoded from the postal code, measures of error, such as median distance and interquartile range (IQR), were calculated by DMT, representative point type, and com-

munity size. In addition, the percentages of the sample geocoded to within various distances (0.5, 1, 5, and 10 km) from the reference location were calculated.

Of the 4,401 matches generated after merging with the postal code representative points, outliers with improbably long distances (more 50 km) were reviewed manually (178 matches, representing 105 records). A large portion were deemed plausible, as they were in suburban or rural areas, and were geocoded using population-weighted allocation (79 matches representing 54 sample records). They were retained because residents of some remote areas may travel long distances to retrieve their mail. Furthermore, the low population weight prevented outliers from skewing the mean distance calculation. Three geographic matches in the north (two in Dawson City; one in Yellowknife) that represented extremely long distances (up to 1,068 km) were deemed improbable and removed. The remaining long distances were typographical errors in census-reported postal codes or errors in geocoding the street address and were removed (99 matches, representing 51 records). The final study sample was n = 1,004 addresses.

Figure 1
Sample calculations for mean distance between respondent house location (from satellite imagery) and postal code representative points (from PCCF+)



● PCCF+ representative point
 ☆ respondent house location
 ——— Euclidean distance
 () DA population weight

Results

Table 2 shows the median (and IQR) distance between the reference locations and the population-weighted mean PCCF+ representative points, by DMT, representative point type, and community size. Urban postal codes had greater accuracy and lower variability than did rural and mixed postal codes. The best accuracy was for large apartment buildings (DMT B) with a median distance of 110 m (IQR 60 m to 200 m), followed by ordinary households (DMT A) with a median of distance 160 m (IQR 80 m to 320 m). Commercial and institutional services (DMT E, G, or M) had relatively good accuracy, with a median distance of 210 m, but more variability (IQR 80 m to 1.25 km). The greatest median distances were for mixed mode services (rural services from urban post offices,

DMT H, K, or T): 5.87 km (IQR 3.81 km to 8.75 km) and rural services (DMT W): 5.20 km (IQR 2.93 km to 9.24 km).

The finer the scale of the data source for representative point types, the better the accuracy. The median distance for blockfaces was 130 m (IQR 60 m to 230 m), compared with 330 m (IQR 160 m to 750 m) for dissemination blocks and 5.53 km (IQR 2.89 km to 8.66 km) for dissemination areas, 95% of which had mixed and rural postal codes (data not shown).

Accuracy generally improved with increasing community size. Rural and small-town areas had the longest median distance: 5.60 km (IQR 2.42 km to 9.48 km). Median distances for other community size groups varied from 160 m (IQR 80 m to 830 m) in the largest urban size group to 330 m (IQR 130 m to 3.93 km) in the smallest.

Table 2
Median distance and interquartile range between location geocoded from full street address and location geocoded from postal code, by delivery mode type, representative point type, and community size, Canada, 2011

	Kilometers		
	Median distance	Interquartile range	
		from	to
Total	0.46	0.12	4.89
Delivery mode type			
A (ordinary urban households)	0.16	0.08	0.32
B (large urban apartment buildings)	0.11	0.06	0.20
E, G, M (urban business and institutional)	0.21	0.08	1.25
H, K, T (mixed urban and rural)	5.87	3.81	8.75
W (rural)	5.20	2.93	9.24
Representative point type			
Blockface	0.13	0.06	0.23
Dissemination block	0.33	0.16	0.75
Dissemination area	5.53	2.89	8.66
Community size			
1,500,000 or more	0.16	0.08	0.83
500,000 to 1,499,999	0.22	0.09	0.52
100,000 to 499,999	0.20	0.08	2.53
10,000 to 99,999	0.33	0.13	3.93
Fewer than 10,000	5.60	2.42	9.48

Source: 2011 Census of Canada.

Table 3
Percentage geocoded from postal code to within various distances of location geocoded from full street address, by delivery mode type, representative point type, and community size, Canada, 2011

	Distance			
	500 m or less	1 km or less	5 km or less	10 km or less
Delivery mode type				
A (ordinary urban households)	87.3	92.2	99.3	100.0
B (large urban apartment buildings)	92.0	97.0	100.0	100.0
E, G, M (urban business and institutional)	65.6	74.4	87.8	96.7
H, K, T (mixed urban and rural)	1.1	3.3	38.9	82.2
W (rural)	0.0	0.4	47.3	78.3
Representative point type				
Blockface	91.4	94.7	98.1	99.8
Dissemination block	65.4	78.2	98.0	99.0
Dissemination area	2.3	4.4	45.9	80.9
Community size				
1,500,000 or more	72.1	77.5	92.2	99.5
500,000 to 1,499,999	74.0	80.0	88.7	98.0
100,000 to 499,999	64.3	68.3	84.1	98.4
10,000 to 99,999 (any CMA/CA with fewer than 100,000)	56.4	61.0	82.2	94.2
Fewer than 10,000 (any non-CMA/CA)	9.1	9.4	45.7	76.2

CMA = Census Metropolitan Area

CA = Census Agglomeration

Source: 2011 Census of Canada.

Table 3 shows the percentage of the study sample geocoded by PCCF+ to within various distances from the full street address. Of the sample with urban residential postal codes, 87% of ordinary households (DMT A), 92% of large apartment buildings (DMT B), and 66% of businesses or institutions (DMT E, G, M) were geocoded to within 500 m of their full street address. By contrast, only 47% with rural postal codes (DMT W) and 39% with mixed urban and rural postal codes (DMT H, K, T) were geocoded to within 5 km of their residence.

By representative point type, 91% of records that geocoded to a blockface and 65% that geocoded to a dissemination block were within 500 m of their full street address. Only 46% of records with dissemination area representative points were geocoded to within 5 km of their residence.

The percentages of the sample geocoded to within 500 m of their full street address were 56% for communities of 10,000 to 99,999 population and more than 70% for communities of at least 500,000. Just 46% of the sample in rural areas and small towns were geocoded to within 5 km of their residence.

Discussion

This study examines the positional accuracy of geocoding based on PCCF+ compared with geocoding from full street addresses.

The results support the literature on postal code geocoding—urban areas have much less error than rural areas.^{12,13} The analysis further demonstrates the importance of DMT to geocoding accuracy. In addition to previously documented problems geocoding rural regions,^{12,14} this study reveals high levels of error associated with mixed DMTs. Such DMTs include rural and suburban routes from urban post offices and general delivery and post office box services at urban post offices, which serve many rural residents.

Even in urban areas, it is important to distinguish between ordinary residential and commercial or institutional delivery. Large commercial and institutional estab-

ishments with more than one physical location (campuses, branches, outlets) may have all their mail addressed to a single postal code (DMT M, with lock box or bag pick up at a post office) and make deliveries from that location to the final destinations. It is also possible that

some of the reported census postal codes with a DMT of E, G, or M pertained to place of work rather than residence. Either problem may have contributed to the greater distances observed for urban business and institutional postal codes (DMT E, G, M), compared with urban residential postal codes (DMT A, B).

Community size and representative point type influenced positional accuracy, largely because of their relationship to the mode of mail delivery (DMT). Most residents of rural and small-town Canada use rural or mixed mode postal codes, and in PCCF+, such postal codes are almost exclusively linked to dissemination area centroids.

Despite the risk of positional inaccuracy and population misclassification associated with various DMT categories, removal of a specific DMT or community size group might introduce selection bias.

The positional accuracy required for a study depends on the spatial resolution of the environmental or contextual measures of interest.^{2,3} A mismatch between the measure and the positional accuracy of geocoding from postal codes alone can result in misclassification. A comparison of 1996-Census-assigned enumeration area income quintiles with those assigned by an older version of PCCF+ showed that misclassification ranged from 3% to 10% in urban areas, but from 39% to 50% in rural areas.¹⁴ In urban areas, misclassification tended to result in assignment to the next-higher or -lower income quintile; in rural areas, 15% to 20% of misclassification was off by two or more quintiles.

The trend in air pollution studies is toward finer-scale data. For instance, satellite-derived ambient fine particulate matter (PM_{2.5}) estimates were developed at 10-km²-resolution,¹⁵ but have been refined to 1-km²-resolution.¹⁶ Finer spatial resolution is possible in urban population centres, but may be problematic in rural areas, where the accuracy of geocoding from postal codes is much reduced.

The demand for urban form data (for example, walkability, density, urban sprawl, and greenness) has increased—

for instance, in studies of the impact of the built environment on physical activity, overweight, and obesity.¹⁷⁻²⁰ Some of these impacts may be examined at coarser scales, such as census metropolitan area or dissemination area,^{18,21} but others rely on environmental data specific to a residential block.²⁰ The degree of spatial resolution in urban form studies typically varies between 500 m (a 5-minute walk) and 1 km. Based on the present analysis, such studies should exclude rural (DMT W) and mixed (DMT H, J, K or T) delivery modes, but could include urban residential delivery modes (DMT A and B) in smaller communities.

Most Canadians have a single postal code that corresponds to their residence. Others, particularly in rural areas, may have an additional postal code corresponding to a post office box or rural route from an urban post office. Postal codes that correspond to post office boxes and rural routes from urban post offices are common in population-based files; however, geocoding from such “mixed” mode postal codes has the same level of accuracy as geocoding from rural postal codes.

Some postal codes with a DMT of E, G, or M may correspond to legitimate places of residence (for instance, nursing homes, university residences, prisons, apartments in a mainly business building), but the majority are commercial or institutional establishments. Their inclusion in administrative records probably reflects reporting place of work rather than residence. PCCF+ flags such records to draw attention to the possible non-residential nature of those postal codes.

Strengths and limitations

A strength of this study was the ability to randomly sample address data from the Census of Population, the most complete survey of residential postal codes in Canada.²² This made it possible to compile an analytical file that included different DMTs, representative point types, and community sizes across the country. As well, the study benefitted

What is already known on this subject?

- In Canada, postal codes are often the only geographic identifier available for assigning contextual or environmental information to a study population.
- The Postal Code Conversion File (PCCF) and Postal Code Conversion File Plus (PCCF+) provide the means to geocode datasets using only postal codes.
- The literature on geocoding based on postal codes shows much less error in urban areas than in rural areas.

What does this study add?

- Using self-reported address data from the 2011 Census of Population, this study compares the positional accuracy of geocoding based on postal codes versus full street addresses.
- All postal codes in the census were processed through PCCF+ Version 6C; positional accuracy was measured by the distance by which the geocoded position differed from the street address.
- Positional accuracy based on postal codes is primarily dependent on delivery mode type.
- The best accuracy was for large urban apartment buildings, followed by ordinary urban households.
- Accuracy generally improved with increasing community size.
- Rural and small-town postal codes, along with post office boxes, had the highest levels of error.

from the availability of Google mapping software to help manually geocode addresses using satellite photography and confirm the addresses using Google Street View. Ortho-rectified aerial photography and satellite imagery are among the most accurate methods for geocoding residences.^{1,3,23}

Two types of error arise when using self-reported address data, particularly postal codes, for spatial analysis: reporting errors and geocoding or positional/geographic assignment errors. Although this study focused on the latter, attempts were made to mitigate census reporting errors, particularly in the postal code. Errors in the first three digits (forward sortation area) generate larger spatial errors than do those in the last three digits (local delivery unit). An evaluation of the impact of address error on public health surveillance in Montreal found that of the 10% of records in the dataset with errors in the address, almost 80% were the result of inaccurate postal codes; 20% contained errors in the forward sortation area, 60% contained errors in the local delivery unit, and 20% contained errors in both.²⁴ For the present study, when possible, postal codes were verified on the Canada Post website¹¹; census postal codes (other than for DMT H, K or T) whose forward sortation

area did not match the forward sortation area generated from the full address were excluded.

This study's reference location used Google Maps satellite imagery to assign latitude and longitude to roof-top centroids based on full street addresses. However, a degree of error is associated with this method.^{25,26} Furthermore, the quality of the satellite images varied and may have contributed to error in assigning the home centroid. This was mitigated by excluding all addresses for which a centroid could not be selected for the house or a small cluster of adjacent buildings (for example, farmhouse and adjacent barns). The resampling that resulted from poor-quality satellite imagery and lack of Google Street View may have introduced bias, because these areas are more remote and disproportionately impact certain groups, such as residents of the far North or Indian Reserves.

As indicated by the wide IQR for DMT A (80 m to 320 m), urban DMTs can, themselves, be heterogeneous, probably reflecting the density of dwellings and period of construction (post-World War II suburban development versus pre-war downtowns). Further analysis could explore the implications of such characteristics on the accuracy of geocoding from postal codes.

Conclusion

Based on address information from a random sample of census respondents, this study assesses positional accuracy of geocoding from postal codes versus full street addresses. Positional accuracy was related to mode of delivery (urban, rural, or mixed), source of latitude and longitude information (blockface, dissemination block, or dissemination area centroids), and community size. Differences across community size groups and representative point types were related to mode of mail delivery.

The results highlight the impact of delivery mode type on positional accuracy. Both rural and mixed (partly urban, partly rural) postal codes had much higher geocoding error than did urban postal codes. These findings demonstrate the importance of understanding how the positional accuracy of geocoding from postal codes differs depending on the nature of the postal codes. Because such differences can substantially alter the effect of environmental and contextual variables, first stratifying by delivery mode type or community size is recommended. The spatial resolution required by the environmental or contextual measures of interest can help analysts to identify subpopulations that should be excluded from a study. ■

References

- Mazumdar S, Rushton G, Smith BJ, et al. Geocoding accuracy and the recovery of relationships between environmental exposures and health. *International Journal of Health Geographics* 2008; 7: 13.
- DeLuca PF, Kanaroglou PS. Effects of alternative point pattern geocoding procedures on first and second order statistical measures. *Spatial Science* 2008; 53(1): 131-41.
- Bonner MR, Han D, Nie J, et al. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology* 2003; 14(4): 408-12.
- Burra T, Jerrett M, Burnett RT, Anderson M. Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario. *The Canadian Geographer* 2002; 46(2): 160-71.
- Guernsey JR, Dewar R, Weerasinghe S, et al. Incidence of cancer in Sydney and Cape Breton County, Nova Scotia 1979-1997. *Canadian Journal of Public Health* 2000; 91(4): 285-92.
- Statistics Canada. *Postal Code^{DM} Conversion File (PCCF), Reference Guide, 2016* (Catalogue 92-154-G) Ottawa: Statistics Canada, 2016.
- Statistics Canada. *Postal Code^{DM} Conversion File Plus (PCCF+) Version 6C, Reference Guide* (Catalogue 82-E0086-XDB) Ottawa: Minister of Industry, 2016.
- Peller P. *An Analysis of the Postal Code Conversion File's Use in Research*. Calgary, Alberta: University of Calgary, 2011.
- Statistics Canada. *Census Dictionary—Census Year, 2011* (Catalogue 98-301-X2011001). Available at: <http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/index-eng.cfm>
- Canada Post Corporation. Air stage offices list. Available at: <https://www.canadapost.ca/tools/pg/prices/RCRZ-e-AIR.pdf>. Accessed July 11, 2016.
- Canada Post Corporation. *Postal Code Lookup Tool*. Available at: www.canadapost.ca/cpo/mc/personal/postalcode/fpc.jsf
- 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.

Positional accuracy of geocoding from residential postal codes versus full street addresses • Methodological Insights

13. Ng E, Wilkins R, Perras A. How far is it to the nearest hospital? Calculating distances using the Statistics Canada Postal Code Conversion File. *Health Reports* 1993; 5(2): 179–88.
14. Wilkins R. Neighbourhood income quintiles derived from Canadian postal codes are apt to be misclassified in rural but not urban areas. Ottawa: Statistics Canada, 2004. Available at: [file=r3jc96paper.pdf](#) [available online at www.ResearchGate.net]
15. Crouse DL, Peters PA, Hystad P, et al. Ambient PM_{2.5}, O₃, and NO₂ exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environmental Health Perspectives* 2015; 123(11): 1180–6.
16. Pinault L, Tjepkema M, Crouse DL, et al. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian Community Health Survey cohort. *Environmental Health* 2016; 15(18): 1–15.
17. Poulidou T, Elliott SJ. Individual and socio-environmental determinants of overweight and obesity in urban Canada. *Health & Place* 2010; 16: 389–98.
18. Seliske L, Pickett W, Janssen I. Urban sprawl and its relationship with active transportation, physical activity and obesity in Canadian youth. *Health Reports* 2012; 23(2): 17–25.
19. de Sa E, Ardern CI. Associations between the built environment, total, recreation, and transit-related physical activity. *BMC Public Health* 2014; 14(693): 1–8.
20. Polsky JY, Moineddin R, Dunn JR, et al. Absolute and relative densities of fast-food versus other restaurants in relation to weight status: Does restaurant mix matter? *Preventive Medicine* 2016; 82: 28–34.
21. Winter M, Barnes R, Venners S, et al. Older adults' outdoor walking and the built environment: Does income matter? *BMC Public Health* 2015; 15(876): 1–8.
22. Mechanda K, Puderer, H. *How Postal Codes Map to Geographic Areas*. Geography Working Paper Series (Statistics Canada Catalogue 92F0138MIE – No. 001) Ottawa: Minister of Industry, 2007.
23. Bow CJD, Water 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: 1–9.
24. Zinszer K, Jauvin C, Verma A, et al. Residential address error in public health surveillance data: A description and analysis of the impact on geocoding. *Spatial and Spatio-temporal Epidemiology* 2010; 1: 163–8.
25. Ubukawa T. *An Evaluation of the Horizontal Positional Accuracy of Google and Bing Satellite Imagery and Three Road Data Sets Based on High Resolution Satellite Imagery*. New York, New York: Centre for International Earth Science Information Network, The Earth Institute at Columbia University, 2013: 1–16.
26. Potere D. Horizontal positional accuracy of Google Earth's high resolution imagery archive. *Sensors* 2009; 8: 7973–81.