Demographic Documents

Comparison of Place of Residence between the T1 Family File and the Census: Evaluation using record linkage

by Julien Bérard-Chagnon

Release date: September 26, 2017

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.

Standard table symbols

The following symbols are used in Statistics Canada publications:

.   not available for any reference period
..   not available for a specific reference period
...   not applicable
0   true zero or a value rounded to zero
0s   value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
p   preliminary
r   revised
x   suppressed to meet the confidentiality requirements of the *Statistics Act*
E   use with caution
F   too unreliable to be published
*   significantly different from reference category (p < 0.05)

COMPARISON OF PLACE OF RESIDENCE BETWEEN THE T1 FAMILY FILE AND THE CENSUS: EVALUATION USING RECORD LINKAGE

by Julien Bérard-Chagnon

TABLE OF CONTENTS

Tables

Figures

Charts

Maps

Acknowledgments

List of acronyms

ACS: American Community Survey

CA: Census agglomeration

CCB: Canada Child Benefit

CCTB: Canada Child Tax Benefit

CD: Census division

CMA: Census metropolitan area

CRA: Canada Revenue Agency

CSD: Census subdivision

DEM: Demography Division

FSA: Forward sortation area

HSMD: Household Survey Methods Division

IMDB: Longitudinal Immigration Database

ISD: Income Statistics Division

LAD: Longitudinal Administrative Databank

NHS: National Household Survey

NLSCY: National Longitudinal Survey of Children and Youth

PCCF: Postal Code Conversion File

SSMD: Social Survey Methods Division

T1FF: T1 Family File

Highlights

- Consistency of place of residence between tax data and census data is relatively high:
    - The province or territory of residence is the same in the tax data and in the census for more than 99% of the persons matched.
- However, the level of consistency decreases for smaller geographic levels:
    - The postal code is the same in the tax data and in the census for 92.9% of the persons matched.
- Consistency of postal code also varies from one region to another:
    - In census metropolitan areas, it tends to be lower in central areas and higher in outlying areas.
- Several characteristics are closely associated with lower consistency of postal code:
    - People who live in a collective dwelling, people who reported having moved in the 2011 National Household Survey, young adults, very elderly people, and people whose tax return was prepared by a third party (according to box 490 of the tax form), are especially likely not to have their postal code correspond with their tax-data postal code.
- The difference between the reference date for tax data and that for the census is likely to affect consistency of postal code.

Introduction

Home is where the heart is.
   - Pliny the Elder

Tax data are being used increasingly to measure and analyze the population and its characteristics. These data serve not only to support statistical processes such as surveys, but also to develop statistical and analytical products. For example, tax data are used to produce estimates of internal migration, which, in turn, are used calculate official population estimates. Tax data are also used to construct two databases commonly used by analysts and researchers, the Longitudinal Administrative Databank (LAD) and the Longitudinal Immigration Database (IMDB).

While this approach considerably reduces data production costs and Canadians' response burden, it also poses a number of challenges. One of the main challenges concerns differences in how place of residence is defined. Various statistical programs, including the census and population estimates and projections, use the concept of usual place of residence, whereas tax data give the mailing address of taxfilers.

The concept of place of residence is fundamental to the study of demography. Most of the statistical indicators used to shed light on key socioeconomic issues rely on the ability of data sources to put people in the "right place". A number of studies address the problems of determining the place of residence with certainty for particular segments of the population, including children in joint custody, residents of collective dwellings, interprovincial workers, and students who alternate between their place of study and the family home (National Research Council 2006; Laporte et al. 2012; Turcotte 2013). Conceptual differences in the place of residence can substantially affect the comparability of files and, as a result, impact the consistency of statistics and the interpretation of results.

The use of tax data to study the population is likely to continue increasing in the future given the expanded use of record linkages and the higher volume of statistical information made available to Statistics Canada.[1] At the same time, social trajectories are becoming more complex, which is making it more difficult to determine the place of residence for a number of demographic groups. In these circumstances, comparing tax data with other sources of demographic data is increasingly relevant. Using record linkage, this study examines the effect of differences in how place of residence is defined in tax data and census data.

The main objectives of this study are to:

- compare the concept of residence in tax data and in census data;
- examine consistency of place of residence between tax data and census data for a few geographic levels of interest;
- identify the main characteristics associated with consistency of place of residence.

---

1.   These activities are one of the priorities identified in Statistics Canada's Corporate Business Plan (Statistics Canada 2015a).

Chapter 1. Concepts of place of residence

While it may appear simple at first glance, the concept of place of residence is highly complex. Although most of the population is able to determine its place of residence with a high degree of certainty, this information is more difficult to establish for persons who have more than one residence. For example, some students and children in joint custody have more than one home and regularly alternate between them. Homeless persons by definition have no place of residence. Some individuals may see their place of residence not as the place where they spend the most time, but rather as the place with which they maintain the strongest economic or social ties. This can apply to workers who, for professional reasons, reside elsewhere than in the family home for much of the year.

Conceptual differences in how various sources of demographic data treat the place of residence were brought to the fore in the United States in the context of creating the American Community Survey (ACS), which replaced the long form of the U.S. census in the mid-2000s. The U.S. census uses the usual place of residence for enumerating the population. The ACS adds the condition that the person must live in that place for at least two months for it to be considered the usual place of residence. This additional condition can affect the comparability of the data calculated through the ACS, especially for highly mobile populations (Scardamalia 2014). While the U.S. Census Bureau noted these conceptual differences (U.S. Census Bureau 2004), the committee of experts, responsible for studying the rules of residence for the U.S. census, recommended that a question on usual place of residence be included in the ACS. This would allow for further exploration of the effect of conceptual differences between that survey and the census (National Research Council 2006: 265).

In Canada, a number of studies have compared tax data with census data at aggregated levels. They showed, among other things, that the coverage provided by tax data could vary, sometimes considerably, according to the characteristics of taxfilers (Bérard Chagnon 2008; He and Michalowski 2005). A comparison of several data sources used for the study of mobility also highlighted the probable effect of different definitions of migration on the numbers of migrants calculated on the basis of these sources (Vamderkamp and Grant 1988).

However, to our knowledge, very few Canadian studies have dealt directly with the issue of the consistency of the place of residence between these two data sources. Using record linkage, Bérard-Chagnon and Brennan (2014) found that more than 85% of linked persons had the same postal code in data from the Canada Child Tax Benefit (CCTB)[2] program and the census data. That study also highlighted the much lower level of consistency for CCTB recipients who had moved during the year, as well as a lag of a few months in updating postal codes for the CCTB. In general, that analysis suggests that there could be sizable differences between tax data and census data regarding place of residence.

1.1.   Concept of usual place of residence in the census

There is no standard definition of place of residence. This concept changes from one source to another according to the main use of the data. These conceptual differences can mean that a given individual is not listed as having the same place of residence from one source to another. This is especially likely to be the case for individuals whose place of residence is more difficult to determine.

The section of the 2011 Census questionnaire on place of residence is shown in the figure on the next page.

---

2.   This federal program, managed by the Canada Revenue Agency (CRA), is designed to provide financial assistance to families with children under 18 years of age through the monthly distribution of allowances. This program is now named Canada Child Benefit (CCB). Statistics Canada uses these data to calculate preliminary estimates of interprovincial migration.

Figure 1
Section on usual place of residence in the 2A questionnaire for the 2011 Census

Source: Statistics Canada, http://www23.statcan.gc.ca/imdb-bmdi/instrument/3901_Q1_V4-eng.pdf
(accessed November 3, 2015).

Canadian censuses use a *de jure* concept of residence.[3] Consequently, the address refers to the usual place of residence, which is defined as the dwelling in Canada in which a person lives most of the time.[4] This approach is necessary for the proper planning of community services, such as schools and public transit, for the allocation of funds to the different levels of government, and for the redistribution of electoral districts.

However, the concept of residence is not clear for all situations in life. The census form includes rules for cases in which the usual place of residence is harder to determine.[5] For example, students who come back to live with their parents during school breaks must be enumerated in the family home, even if they spend much of the year elsewhere. These specifications make it clear that determining the place of residence can be complex for some groups. In fact, the problems encountered in accurately determining the place of residence are one of the main sources of coverage error in censuses (Statistics Canada 2015b).

---

3. A *de jure* census (*de jure*: according to law) refers to a census in which residence is determined on the basis of an official definition. This approach contrasts with the *de facto* method, which consists in counting people where they are located on the day of the census. The *de jure* method has been used for Canadian censuses since 1871.
4. The detailed definition of the concept of usual place of residence in the 2011 Census is available at http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/pop126-eng.cfm (accessed November 5, 2015).
5. The electronic questionnaire for the 2011 Census included a series of questions designed to ensure that the population was counted in the right place. A question was also asked in the event that the respondent was not sure whether a member of the household should be included in that household. A follow-up with the household was then conducted to determine whether the person should be included. It is reasonable to believe that this approach yielded a better enumeration of the population.

1.2. Concept of mailing address in tax data

The figure below shows the sections of the T1 tax return dealing with the place of residence. In the tax data, taxfilers must report three things in connection with their place of residence: their mailing address, their province or territory of residence on December 31 of the taxation year, and their province or territory of residence at the time of filing if it is different from the one in the mailing address. The mailing address is required so that the Canada Revenue Agency (CRA) can contact the taxfiler. Accordingly, the address is not intended to determine where the taxfiler lives most of the time, but rather the place at which to reach him or her efficiently. Taxfilers can even give a mailing address that is not their place of residence. This could, for example, be the case with a young adult who has just left the parental home and for whom the family home continues to be the anchor for managing his or her tax returns.

Also, some situations specific to taxation can affect taxfilers' place of residence. Some taxfilers have their tax return prepared by an accountant or a family member, who might give his or her own mailing address instead of the taxfiler's. Similarly, the method of creating the T1 Family File (T1FF), the fiscal file used in this study, puts CCTB children in the home of the taxfiler who is receiving the benefits,[6] which is not necessarily the child's actual place of residence.

Figure 2
Sections on the taxfiler's address in the 2014 tax return

Source: Canada Revenue Agency, T1 form for the 2014 taxation year.

---

6.   The methodology used to create the T1FF is summarized in "Chapter 2. Data used" on page 10.

Chapter 2. Data used

A number of studies have compared tax data with census data. However, since these studies use mostly aggregated totals, many factors, including the coverage of the files, can interact together to explain the differences observed between the different sources. The innovative aspect of the present study is that it uses record linkage between the 2010 T1FF tax data, sent by taxfilers in the spring of 2011, and the 2011 Census to directly compare individuals' place of residence in the two sources. Instead of the T1 tax data, the T1FF was used for the linkage for two reasons. First, the more extensive coverage of the entire population in the T1FF provides a more complete picture of consistency with respect to place of residence. Second, different researchers and analysts already use the T1FF file extensively, especially for calculating official estimates of internal migration, constructing the LAD and the IMDB, and establishing economic dependency profiles at very fine geographic levels. Data from the 2011 National Household Survey were then added using existing linkage keys for this database and the census to take advantage of the characteristics of the population available in the NHS.

This chapter describes the data used in the present study. It begins with a brief introduction to the files that were linked. Next, it describes the linkage method used. It ends by providing the definition of place of residence followed in this study and describing the characteristics examined.

2.1.   T1 Family File

The Income Statistics Division (ISD) has produced the T1FF annually at Statistics Canada since 1982, for the purpose of recreating Canada's population and family universe. The ISD constructs the file by taking data on individuals who completed a T1 return of income for the reference year and combining those data with information on non-filing spouses and data on children from the CCTB, vital statistics, and a historical file. It then groups individuals into census families by means of a complex methodology.[7] This approach serves to create an annual file that covers approximately 95% of the Canadian population and contains basic tax and demographic information on the Canadian population (Bérard-Chagnon 2008).

2.2.   Census and National Household Survey

Statistics Canada conducts the census of population in May every five years to develop a statistical portrait of Canada. The data are used not only to meet the requirements of various laws, but also to support decision making in a wide variety of very different fields, ranging from the planning of community services to the conduct of market studies. The short form, which collects basic demographic and linguistic data, is sent to all Canadian households. In 2011, the census long form was replaced by the NHS. This voluntary survey collected more detailed socioeconomic data, such as level of schooling, employment status, and place of birth. The NHS was administered to roughly 30% of households, targeted only private households, and obtained an unweighted response rate of 68.6%.

---

7.   For more information on how the T1FF is constructed, see Lessard (2011).

2.3.   Record linkage

The study used proven techniques to link the T1FF data and the census data deterministically in five successive waves. The linkage was based on name, date of birth, sex, and family information.[8] Note that the study did not use geographic information as a linkage key so as not to bias the matching.[9] Table 1 shows the five waves and the numbers matched in each.

The first wave consists of individuals for whom an exact match on sex, place of birth, and name is established for more than one family member. The majority of the linked individuals are from the first wave. Subsequently, in the second wave, the study uses SAS's COMPGED function to relax linkage rules for the name.[10] Only individuals for whom a match was established for more than one family member are retained. The third wave uses *Mix Match*[11] to establish a match on the given name. The rules in the fourth wave establish a match for families in

Table 1

Description of linkage waves and number linked (number and percent) by wave

Linkage wave	Persons linked	
	number	percent
1) Exact match on sex, date of birth, and name for more than one family member	12,697,283	68.2
2) Relaxed match on name	542,169	2.9
3) Match on given name using *Mix Match*	361,126	1.9
4) Match for at least three family members	2,543,298	13.7
5) Unattached persons added	2,656,730	14.3
Duplicates removed	-177,125	-1.0
Total	**18,623,481**	**100.0**

Source: House, Georgina. 2014. *The Linking Process,* working document, Household Survey Methods Division, Statistics Canada, 1 page.

which the sex and date of birth correspond for at least three members. Lastly, the fifth, and final, wave includes the matches made in the previous waves for unattached persons.

The matched file contains a total of 18,623,481 individuals, for a linkage rate of 57.0% of the population in the 2010 T1FF. This linkage rate is similar to the rates obtained for other linkages between tax data and census data when one excludes links created from geographical information. A comparison with another linkage between the 2010 T1FF and the 2011 Census, also conducted at Statistics Canada, showed a correspondence of approximately 97% in the matches created by the two linkages. Most of the unsuccessful matches were due to links created for twins where the linkage created a match with a different twin. This situation has a negligible effect, since it is quite reasonable to assume that twins live in the same place.

The matched T1FF–census file includes 3,991,431 observations after the NHS data are added.

---

8.   Georgina House performed this linkage. For more information on the linkage techniques used, see House (2014).

9.   Since the purpose of this study is to compare the place of residence in two sources, the use of geographic information to construct the linkage would bias the comparison: the persons matched would then be much more likely to have the same geographic information.

10.   This function calculates a generalized measurement of the Levenshtein distance between two character strings in order to establish a match. This distance consists in the number of characters that must be added, deleted or replaced to go from one character string to another.

11.   *Mix Match* is a generalized record linkage environment that, among other things, lends itself to simultaneous comparison of character strings. For example, this program can be used to match given names, such as John and Jonathan.

Table 2 gives an overview of the quality of the match, showing the distributions for the T1FF, the census and the matched file on a few basic characteristics.

In general, the distributions of the characteristics studied here and observed in the matched file are, on the whole, faithful to those in the T1FF and the census. The matched file includes a slightly higher proportion of imputed children and spouses than the 2010 T1FF. This is because the linkage technique uses the relationships between family members to maximize the number of matched records. For the same reason, the matched file also includes a higher proportion of married couples and a lower proportion of persons aged 65 and older.

Of course, when interpreting the results of the linkage and making inferences to the population as a whole, one must keep in mind that the population in the matched file depends not only on the linkage rate but also on the presence of individuals in each of the two source files. Several studies have drawn attention to census undercoverage (Statistics Canada 2015b) and the incompleteness of tax data (Bérard-Chagnon 2008; Aydemir and Robinson 2006). In both cases, males in their twenties and immigrants are especially likely to not be enumerated or to not have completed a tax return. Therefore, the results presented in this study reflect only the situation of matched individuals and not that of the Canadian population as a whole. One should exercise great caution when making an inference to the entire population.

Table 2

Distribution data from the 2010 T1 Family File, the 2011 Census and the matched file (percent) for selected characteristics

Characteristic	2010 T1FF	2011 Census	Linkage
	percent		
Type of record in the T1FF			
Living taxfiler	78.0	…	75.7
Imputed child / spouse	21.8	…	24.1
Other[1]	0.3	…	0.2
Age group			
0 to 17 years	21.4	20.5	25.3
18 to 24 years	8.9	9.1	8.5
25 to 39 years	19.8	19.2	20.3
40 to 64 years	35.5	35.6	35.7
65 years and older	14.6	14.5	10.3
Missing[2]	0.0	1.2	0.0
Composition of census family			
Married couple	58.5	…	64.2
Common-law couple	13.9	…	13.9
Lone-parent family	10.8	…	7.7
Person not in family	16.9	…	14.3

1. This group consists primarily of living taxfilers associated with deceased taxfilers.
2. The census matched file is the Census Response Database, which contains the information directly provided by respondents before the imputation stages. For this reason, it may contain respondents whose age and date of birth are missing.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

2.4.   Definition of place of residence for the study

In this study, place of residence is defined according to the postal code. This decision is based on both the fact that a postal code is a very detailed geographic unit that approximates an individual's actual place of residence, primarily in urban areas where most of the population resides,[12] and on the fact that there are problems involved in processing individuals' complete addresses. A list of postal codes from the Postal Code Conversion Files (PCCF) were used to clean up postal codes in the matched file. According to this criterion, the postal codes of 95.9% of matched individuals are valid and could, therefore, be used in this study.[13] There is a match on place of residence when the postal code (or any other geographic level examined) in the T1FF is identical to the one in the census.

---
12. For more information on the definition of postal codes, see Statistics Canada (2015c).
13. See "Table A.1" in the Appendix showing the proportion of individuals whose postal codes are valid according to selected characteristics.

2.4.1.   Difference in reference dates

An important point to consider is that there is a time lag between the reference date for T1FF tax data and the census reference date. Whereas Census Day in 2011 was May 10,[14] tax returns for the year 2010 contain taxfilers' information dating from March or April 2011.[15] Furthermore, the geographic information that the CRA sent to Statistics Canada for purposes of constructing the T1FF was the most current information available to the CRA on December 31, 2011. That difference is likely to account for some of the inconsistencies observed in this report between the postal codes in the tax data and those in the census. For example, individuals who moved between Census Day and December 31, 2011, would naturally have a different place of residence in the two sources if they reported their move to the CRA, and if the CRA updated their mailing address in its databanks.

2.5.   Characteristics used

When the information available in the T1FF, census and NHS files is combined, the result is a very rich database. Three criteria guided the search for characteristics likely to be associated with consistency of place of residence.[16]

First, a number of individuals are in a situation in which their place of residence is more difficult to determine. Essentially, this can result from two factors: foremost, for some subpopulations, the place of residence is especially difficult to identify, both in the tax data and in the census. This is the case, for example, with persons who have a thirty party prepare their tax return. Additionally, some individuals maintain economic and social ties with more than one place of residence. This can mean that their mailing address differs from their usual place of residence. To target the characteristics associated with these situations, both the literature[17] and experts at Statistics Canada were consulted. This criterion includes interprovincial workers, couples not living in the same household, and children in joint custody.

Second, Canadians are highly mobile. The NHS data show that more than one Canadian in ten changed his or her usual place of residence between 2010 and 2011. It is important to examine the effect of mobility on the consistency of one's place of residence, for two reasons. While most of these moves are over a fairly short distance, they may nevertheless reflect a major transition in the mover's life that involves creating and maintaining major ties with more than one place of residence for a given time. Given the difference between the reference dates for the two files examined in this study, and the extent to which the tax data are able to adequately capture moves, mobility is likely to be a major correlate of consistency of place of residence.[18] In addition to the NHS question that deals explicitly with mobility, this group of characteristics includes age, immigrant status, and being a renter.[19]

Third, the methodology for constructing the T1FF and the linkage, as well as the choice of the postal code to approximate the place of residence, may also affect the consistency of the place of residence. Therefore, a few methodological variables, such as the wave in which the individual was added to the linkage, were also targeted for this study.

Examining these characteristics will make it possible to both identify the demographic groups for which the consistency of the postal code is lower and to better understand the mechanisms that underlie this situation.

---

14.   Except for the early enumeration, which took place between February and April 2011 in remote and northern areas and on Indian reserves in northern Canada (Statistics Canada 2012).
15.   Late filers comprise a sizable group. However, this group is not included in the T1FF or, as a result, in the matched file used in this analysis.
16.   These criteria were established to structure the analysis. Since they are not mutually exclusive, some characteristics could meet more than one criterion. The complete list of characteristics used in this analysis is available in "Table A.2" in the Appendix.
17.   Studies conducted on this subject include National Research Council (2006), Laporte et al. (2013) and Turcotte (2013).
18.   The characteristics associated with mobility are therefore likely to be related to those associated with maintaining ties with more than one place of residence, the latter being the first group of characteristics identified for this study. However, considering the prevalence of migration and the highly probable effect of the difference in reference dates, these characteristics were included in another group.
19.   Sources for the characteristics identified by means of this criterion include Turcotte and Vézina (2008), Dion and Coulombe (2008), and Finnie (2000).

Chapter 3. Consistency of place of residence by geographic level

This chapter examines the consistency of place of residence in the T1FF and the census for different geographic levels of interest, namely postal code, forward sortation area (FSA),[20] census subdivision (CSD),[21] census division (CD),[22] census metropolitan area (CMA),[23] and province or territory. These geographic levels are obtained by geocoding postal codes by means of the PCCF for the T1FF and by going directly into the census databases.

The following chart illustrates the consistency between the T1FF and the census for different geographic levels.

Chart 1
Geographic consistency rates (percent) by geographic level
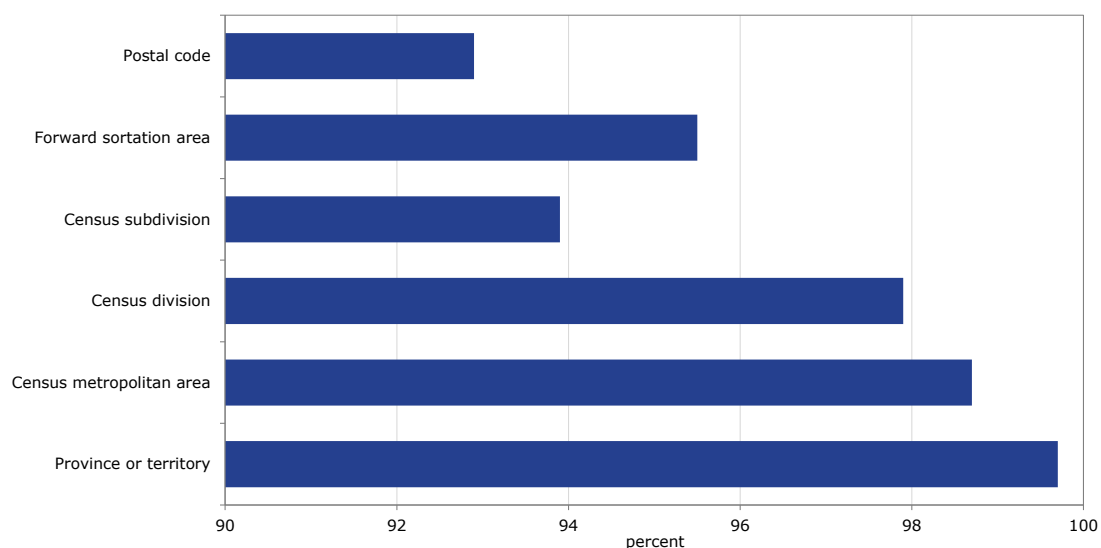
Note: In the interest of comparability, the province for the T1FF is obtained by using the postal code and not the code of the province reported by the taxfiler.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

In general, the consistency of the place of residence between the tax data and the census data is relatively high: 92.9% of matched individuals have the same postal code in the T1FF and the census. The rate climbs to 95.5% for FSAs and to 93.9% for CSDs. The slightly lower consistency for CSDs than for FSAs may be explained by the fact that, while FSAs are a very fine geographic level in highly urban areas, they can extend over much larger areas than CSDs in the rest of Canada. Consistency is even greater for more aggregated geographic levels, approaching 98% for CDs and exceeding 99% at the provincial/ territorial level. This means that practically all matched individuals indicated the same province or territory of residence in the two databases examined in this study.

---

20. A forward sortation area (FSA) corresponds to the first three characters of the postal code. In May 2011, Canada had 1,638 FSAs.
21. "Census subdivision (CSD)" is a generic term that designates municipalities (as defined by provincial/territorial laws) or areas deemed to be municipal equivalents for statistical reporting purposes (e.g., Indian reserves, Indian settlements and unorganized territories).
22. A census division (CD) is a group of neighbouring municipalities joined together for the purposes of regional planning and managing common services (such as police or ambulance services). These groupings are established under laws in effect in certain provinces. For example, a CD might correspond to a county, a municipalité régionale de comté, or a regional district. In other provinces and in the territories, where legislation does not provide for such areas, Statistics Canada defines equivalent areas for statistical reporting purposes in cooperation with these provinces and territories. An example of a CD is La Vallée-du-Richelieu, which includes several municipalities, including Beloeil and Chambly.
23. A census metropolitan area (CMA) is an area consisting of one or more neighbouring municipalities situated around a core. A CMA must have a total population of at least 100,000, of which 50,000 or more live in the core. Montréal and its suburbs are an example of a CMA.

Chart 2

Consistency of postal code (percent), by census province or territory

percent
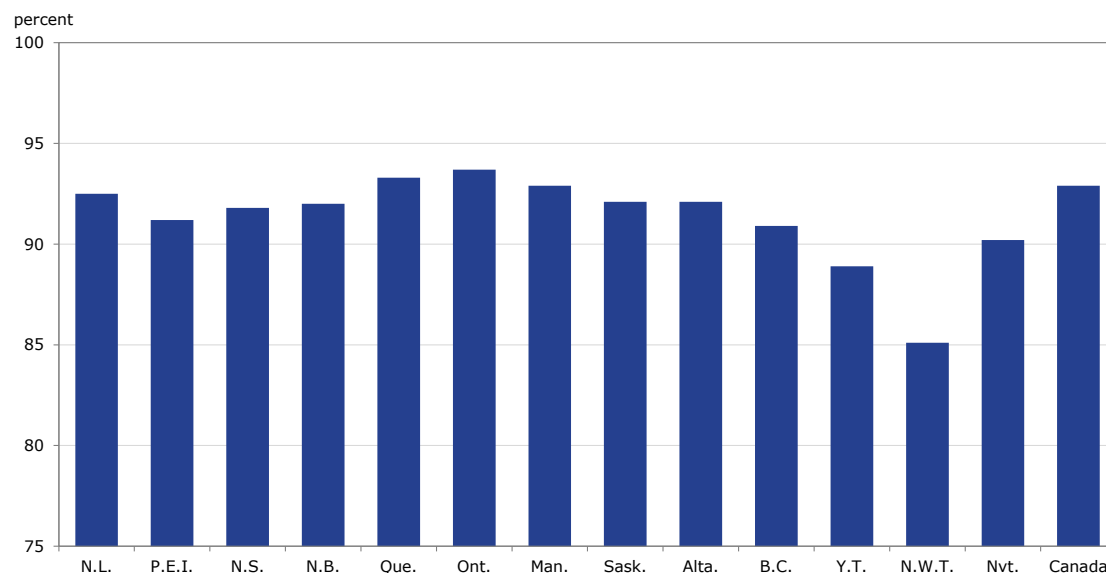
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

These results suggest, then, that, while the place of residence recorded in the tax data may differ from the one reported in the census, the difference is evident mainly at very fine geographic levels. However, the fairly high levels of consistency seen here should not obscure two points. First, although 92.9% of matched individuals have the same postal code in the two sources, the corollary is that the postal codes do not match for 7.1% of the individuals in the linkage. If the matching results can be inferred to the population as a whole, nearly 2.5 million persons would not have indicated the same postal code in the census and the tax data. This number is equivalent to the population of the Vancouver CMA, the third-largest in Canada.

Second, there are sizable regional variations as to the consistency of the place of residence. The chart above shows these variations for the provinces and territories.

The consistency of the postal code is generally lower in the territories. It reaches a low of 85.1% in the Northwest Territories. This means that, for that territory, the postal codes of nearly 15 matched individuals in 100 are not the same in the census and the T1FF. In addition to being much less populous than the provinces, all three territories have a population that is generally more mobile[24] (Willbond 2014). In contrast, postal code consistency exceeds 93% in Ontario and Quebec.

---

24.   The links between consistency of postal code and mobility will be explored in "Chapter 4.   Consistency of postal code by individuals' characteristics" on page 24.

Map 1
Consistency of postal code (percent) by census division from census data
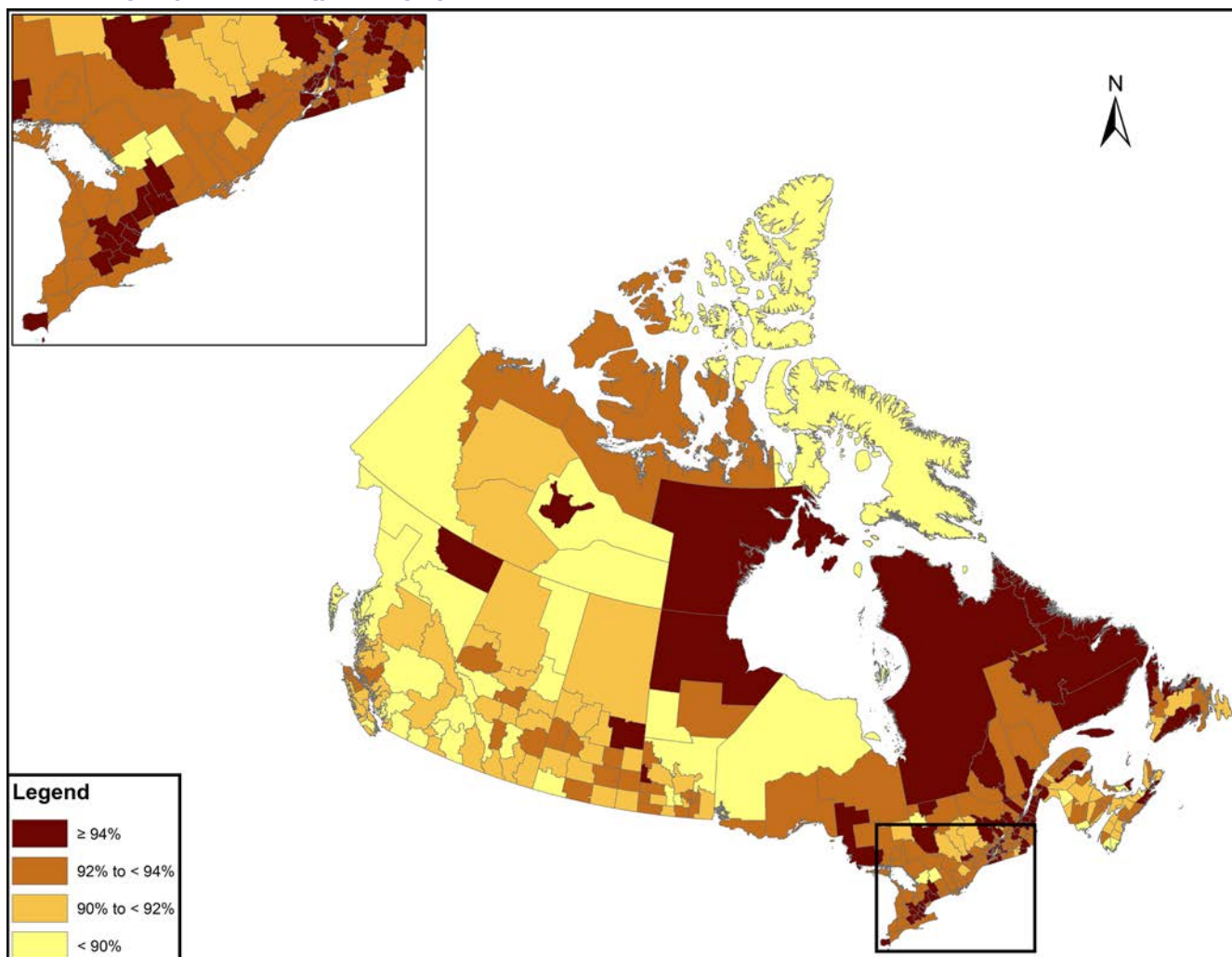
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Regional variations in consistency can also be seen at finer geographic levels. This is illustrated in Map 1 (above), which shows the rates of consistency of postal code by census division.

Consistency of postal code ranges from 81.9% (Region 6 [N.W.T.]) to 96.9% (La Haute-Côte-Nord [Que.]). A total of 5 of the 293 CDs have a consistency rate below 85%, while 24 CDs have a rate exceeding 95%. Some CDs in rural areas tend to have lower consistency of postal code. In particular, the three CDs that show the lowest postal code consistency are Region 6 (N.W.T.), Stikine (B.C.), and Division No. 19 (Man.). In contrast, the CDs in some more urbanized areas generally have consistency levels exceeding 90%. Of course, some CDs with very low populations also have very high consistency levels that result from the effects associated with their small numbers of residents.

An examination of postal code consistency at the FSA level reveals a divide between metropolitan cores and outlying areas. Maps 2, 3 and 4 highlight this situation for Canada's three most populous CMAs: Montréal, Toronto and Vancouver. For these three CMAs, consistency levels for linked individuals who live in the central FSAs tend to be below 90% and even reach 85%. Conversely, for the population living in a number of outlying FSAs, consistency levels exceed 95%. These differences may result from two phenomena. First, urban FSAs are generally much smaller in area than rural FSAs. Consequently, moving is very likely to involve a change of FSA. In the suburbs and in more rural areas, moving a short distance might not involve a change of FSA. Second, the characteristics of the population of central districts differ from those of the population of peripheral and rural areas: individuals living in highly urban settings tend to be younger and more mobile, and these two characteristics are closely associated with lower postal code consistency.[25]

The main consequence of the differences in consistency levels is that the population of some FSAs varies considerably depending on the source examined. While the T1FF population ranges between 96% and 112% of the census population for 95% of FSAs, it is sometimes much higher or lower than the census population. Maps 5, 6 and 7 illustrate this situation. They show the ratio between the T1FF population and the census population for the Montréal, Toronto and Vancouver CMAs. In all three cases, the T1FF figure is very close to the census figure in the suburbs. However, in some central areas, the T1FF has much larger numbers than the census. In the Montréal CMA, the H3A and H3B FSAs are noteworthy. These two areas are located in downtown Montréal and have many office towers and financial institutions. Some tax filers might have a postal code that falls within these areas on their tax return because they have the return prepared by an accountant.[26] The H3B FSA also includes the headquarters of the *Curateur public du Québec*, which administers the tax files of some filers who are under curatorship. As a result, the T1FF contains more than 6,000 persons who reported a postal code in this FSA, whereas in the census this was the usual place of residence of barely more than 100 people.

The M5H, M5G and M5C FSAs, located in the Toronto CMA, as well as the V6C and V6E FSAs in Vancouver, also have a much larger population in the T1FF than in the census. The T1FF includes nearly three times as many individuals under the M5G FSA and more than four times the number of individuals within the M5H FSA. These FSAs are also located in downtown Toronto or Vancouver in districts that include a number of office towers.

Some sources yield contrasting numbers for some FSAs outside these CMAs as well. The FSAs of V8W (Victoria) and T5J (Edmonton) have a substantially higher population in the T1FF than in the census. Once again, both these FSAs are in downtown neighbourhoods. This confirms that the match between the numbers in the T1FF and those in the census is weaker in the downtowns of large cities.

---

25. The links between these characteristics and consistency of postal code are examined in "Chapter 4. Consistency of postal code by individuals' characteristics" on page 24.
26. The links between consistency of postal code and having one's tax return prepared by a third party are examined in "Chapter 4. Consistency of postal code by individuals' characteristics" on page 24.

Map 2
Consistency of postal code (percent), by census forward sortation area, Montréal CMA

Legend

■	≥ 95%
■	90% to < 95%
■	85% to < 90%
■	< 85%
■	Not available

Note: Forward sortation areas with fewer than 100 persons matched are not represented.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Map 3
Consistency of postal code (percent), by census forward sortation area, Toronto CMA

Legend

■	≥ 95%
■	90% to < 95%
■	85% to < 90%
■	< 85%
■	Not available

Note: Forward sortation areas with fewer than 100 persons matched are not represented.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Map 4

Consistency of postal code (percent), by census forward sortation area, Vancouver CMA

Legend

- ≥ 95%
- 90% to < 95%
- 85% to < 90%
- < 85%
- Not available

Note: Forward sortation areas with fewer than 100 persons matched are not represented.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Map 5
Ratio of the population of T1FF to census, by census forward sortation area, Montréal CMA

Legend

■	≥ 1,15
■	1,05 to < 1,15
■	1,00 to < 1,05
■	0,95 to < 1,00
■	< 0,95
■	Not available

Note: Forward sortation areas with fewer than 100 persons matched are not represented.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Map 6
Ratio of the population of T1FF to census, by census forward sortation area, Toronto CMA

Legend

■	≥ 1,15
■	1,05 to < 1,15
■	1,00 to < 1,05
■	0,95 to < 1,00
■	< 0,95
■	Not available

Note: Forward sortation areas with fewer than 100 persons matched are not represented.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Map 7

Ratio of the population of T1FF to census, by census forward sortation area, Vancouver CMA

Legend

- ≥ 1,15
- 1,05 to < 1,15
- 1,00 to < 1,05
- 0,95 to < 1,00
- < 0,95
- Not available

Note: Forward sortation areas with fewer than 100 persons matched are not represented.
Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Chapter 4. Consistency of postal code by individuals' characteristics

The results examined in "Chapter 3. Consistency of place of residence by geographic level" on page 14 revealed differences in the place of residence identified in tax data and the census. In some instances, they also showed substantial variations in the consistency of place of residence for different geographic levels. This chapter continues the evaluation, presenting levels of postal code consistency according to various characteristics of individuals. It gives the results for characteristics associated with more complex situations related to the usual place of residence, for characteristics relating to high mobility, and for methodological characteristics.

Methodology notes

This section incorporates the results from the National Household Survey (NHS), a voluntary survey that replaced the census long-form questionnaire in 2011. Three aspects of the NHS are relevant here: it was administered to roughly 30% of households; it did not cover collective dwellings; and its unweighted response rate was 68.6%. In comparison, the census was sent to all Canadian households, including collective dwellings, and covered more than 97% of the Canadian population. However, although these methodological differences affect postal code consistency, the effect is fairly small. The overall postal code consistency rate based on census data was 92.9%, and the rate based on the NHS was 94.1% the NHS. This difference is cut by half when the census figures are limited to private households (93.4%).

Given the study's objectives and the linkage rate, the results drawn from the NHS presented are not weighted. It was considered more appropriate for each person included in the matched file to have the same weight for purposes of comparing postal codes.

To make the results easier to interpret, the study set the threshold for lower postal code consistency at 90%. Results for the characteristics not included here are provided in «Table A.3» in the Appendix.

4.1.   Consistency of postal code: Results for characteristics associated with the concept of place of residence
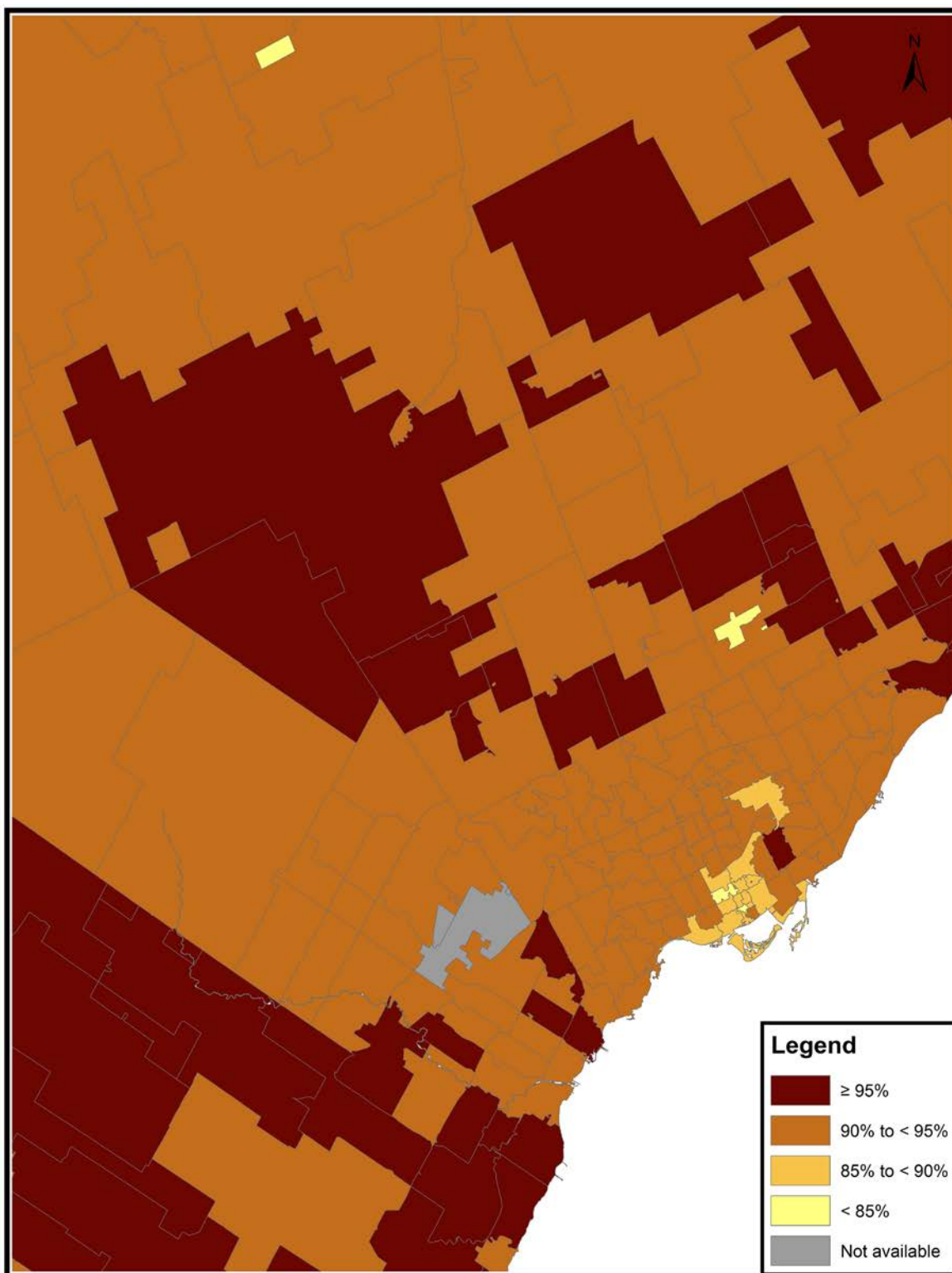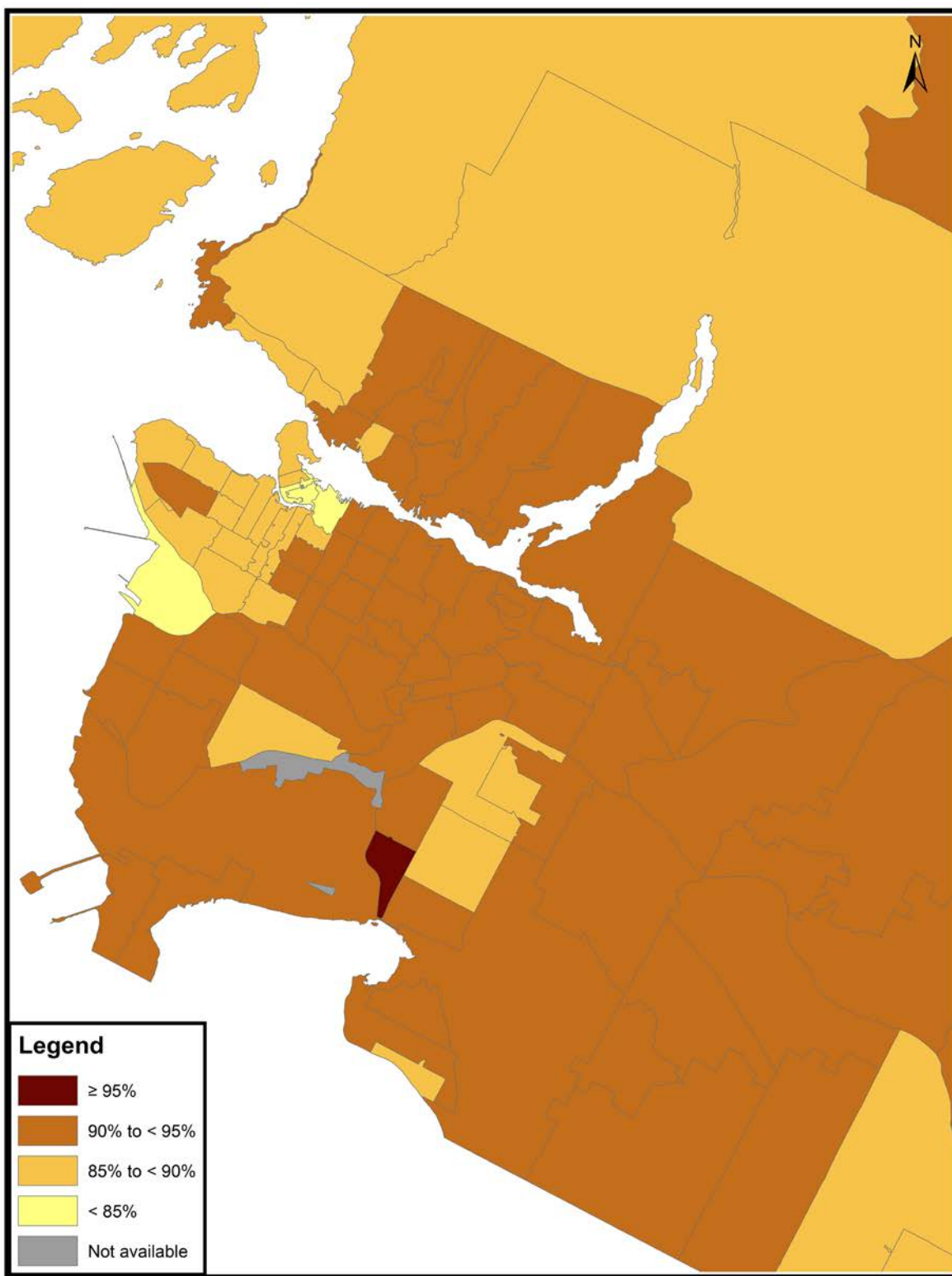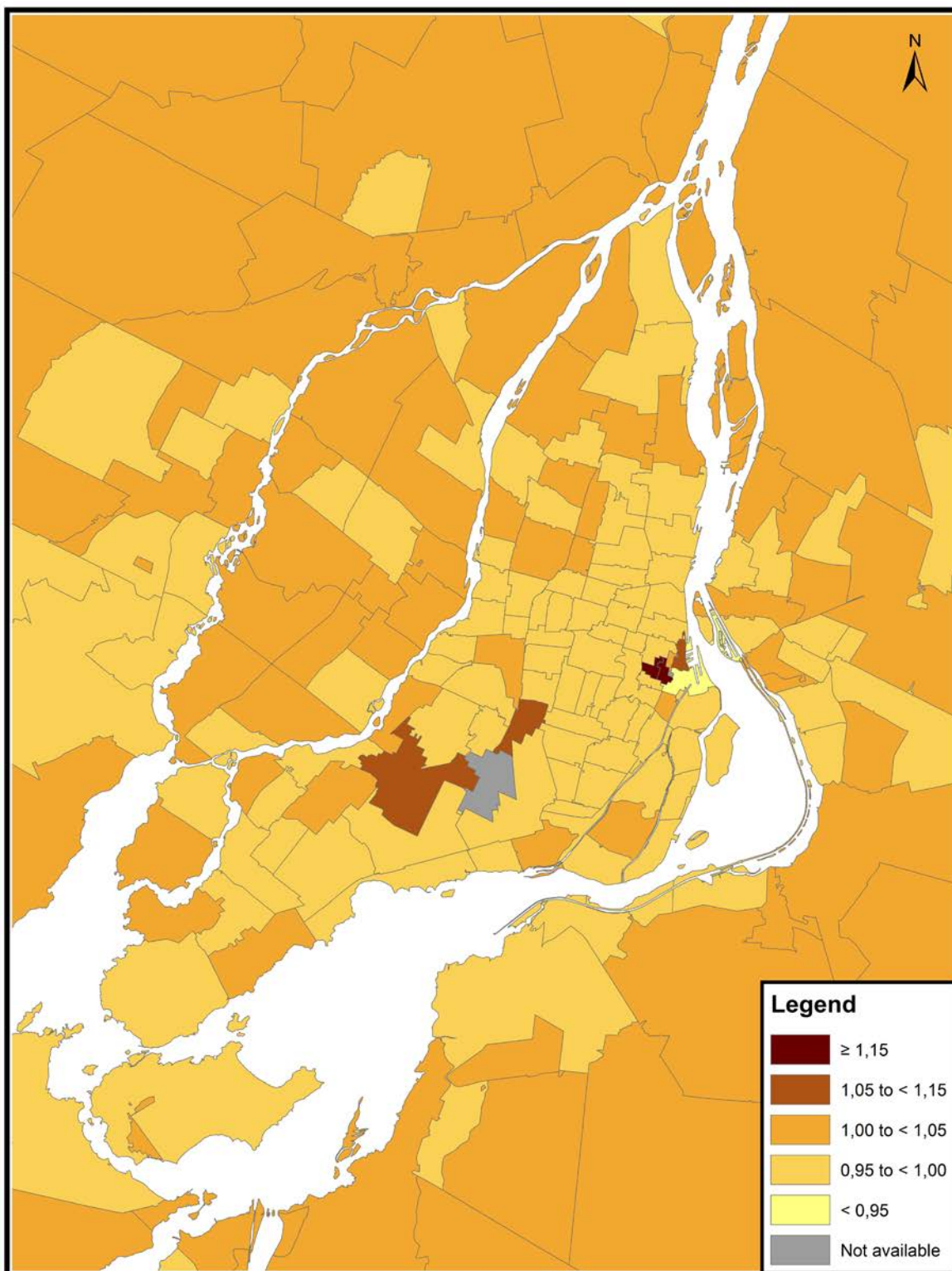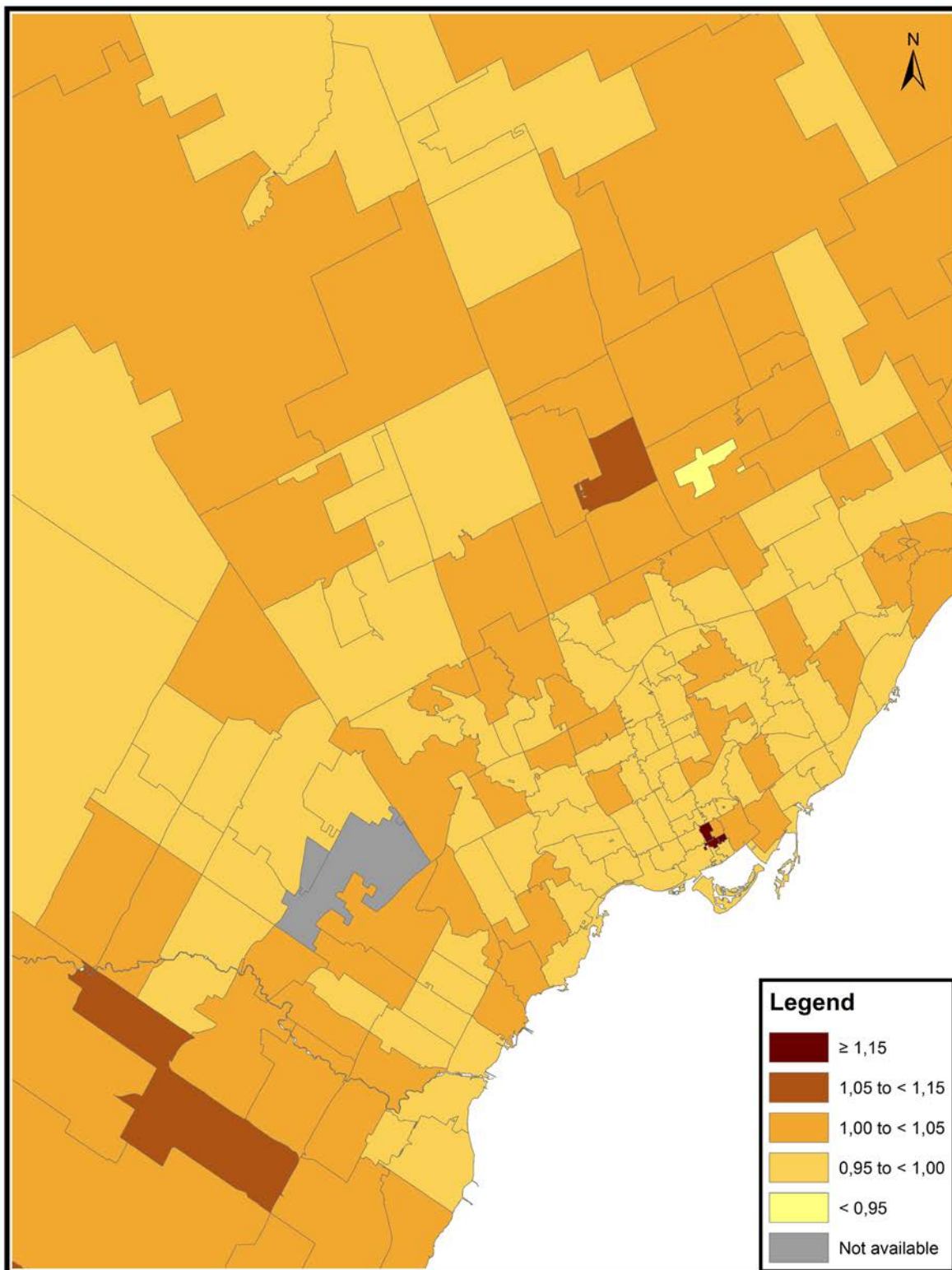
Table 3 (on the next page) shows the characteristics related to the concept of place of residence associated with lower postal code consistency.

An examination of the data in this table reveals several situations where conceptual differences between the postal code in the census and in the tax files could affect postal code consistency. First, consistency is very low (32.8%) for linked persons who had their tax return prepared by a third party (according to box 490). This is likely because tax preparers tend to put their own mailing address on the tax return, rather than that of the taxfiler. These people are usually very elderly: approximately 30% of persons who had their return prepared by a third party (according to box 490) were aged 80 and older, compared with less than 3% of the linked population.

Matched persons who live in a collective dwelling[27] also have a much lower consistency rate than the rest of the population (31.4%). The majority of people residing in collective dwellings are very elderly individuals who are living in a residence that offers specialized care and services. Such individuals constitute a segment of the population for which it is often harder to determine the place of residence with certainty (National Research Council 2006). Also, because of their age, they may be more likely to have their tax return prepared either by someone close to them or by a professional.

Interprovincial workers—workers whose job is in a province other than their province of residence[28] (89.4%)—and individuals who do not live with their spouse (89.0%) have slightly lower consistency levels. In both cases, these persons are especially likely to maintain ties with more than one place of residence, for members of the first group because of their distance from their workplace and for members of the second group because of their connection with more than one home.

---

27. A collective dwelling is an establishment used for commercial, institutional or communal purposes, such as a hotel, a hospital or a work camp.
28. Persons living in the Ottawa — Gatineau CMA or in the census agglomerations (CAs) of Campbellton, Hawkesbury and Lloydminster were excluded from this analysis because their metropolitan area overlaps two provinces.

Table 3

Distribution in the matched file and consistency rates (percent) between the T1FF and census postal codes, by characteristic associated with the concept of place of residence

Characteristic	Distribution in matched file	Consistency rate
	percent	
Overall	100.0	92.9
Return prepared by third party (box 490)[1]		
No	99.4	93.2
Yes	0.6	32.8
Living in a collective dwelling		
Not living in a collective dwelling	99.1	93.4
Living in a collective dwelling	0.9	31.4
Province or territory of work[2]		
Same province or territory	99.4	94.5
Other province or territory	0.6	89.4
Presence of spouse in household		
Spouse present	99.5	95.0
Spouse absent	0.5	89.0
Household on-reserve		
No	98.9	94.2
Yes	1.1	87.6
Aboriginal identity		
No Aboriginal identity	96.3	94.2
First Nations	2.2	89.0
Métis	1.2	91.8
Inuit	0.2	93.9
More than one Aboriginal identity	0.1	92.6

1. This variable is taken from box 490 on the T1 tax return. This box provides information on the preparer of the return only if a fee was charged. Note that not all preparers fill in this box even when they are charging a fee. Therefore, this variable does not cover everyone who had his or her return prepared by a third party.

2. Persons living in the Ottawa — Gatineau CMA or in the census agglomerations (CA) of Campbellton, Hawkesbury and Lloydminster were excluded from this analysis because their metropolitan area overlaps two provinces.

Sources: Statistics Canada, 2011 Census, 2011 National Household Survey, and 2010 T1 Family File.

The rate postal code consistency is also lower for persons living on an Indian reserve (87.6%) and members of First Nations (89.0%). Reserves are not only mostly located in rural areas but they are also places with very different social and political dynamics than the rest of the country.

4.1.1. Children in joint custody

Children in joint custody regularly alternate between their parents' homes. In the census, they must be enumerated in the home in which they spend the most time. If they divide their time equally between their two parents, they must be enumerated in the home they are in on Census Day. However, in the tax data, a child is generally included with the taxfiler who is receiving the tax benefits associated with that child, and the taxfiler's mailing address is not always identical with the child's usual place of residence. As a result, this group is especially likely to have a different place of residence in the two sources.

Individuals in this group are more difficult to identify based on the data available in the linkage. We have therefore tried to identify them indirectly using census information on the individual's status within his or her family. This information serves to identify children who live in a lone-parent family or who are the biological or adopted child of only one of the two members of the couple. Such children are especially likely to be in joint custody; based on the definition used here, one child in five aged 0 to 17 is in this situation.[29]

Linked children who are identified as being likely to be in joint custody have a postal code consistency rate of 86.3%, compared with 94.0% for other children; this represents a gap of nearly 8 percentage points. Of course, this gap is quite probably underestimated, since not all children identified by this approach are necessarily in joint custody. On this subject, the NLSCY data show that 27.9% of children aged 0 to 9 in a custody arrangement were in joint custody. In any case, the more complex situation with respect to determining the place of residence of children in joint custody appears to have a sizable impact on the consistency of the postal code.

---

29. The proportion obtained here of children likely to be living in a custody situation is similar to the one seen in the National Longitudinal Survey of Children and Youth (NLSCY). The data on early childhood development cohorts in Cycle 5 (2002/2003) to Cycle 7 (2006/2007) show that 15.4% of Canadian children aged 0 to 9 were living in a custody situation, a proportion similar to the one identified using the present definition for the same age group (15.7%).

4.2.   Consistency of postal code: Results for characteristics associated with mobility

Mobility is a factor closely associated with the consistency of the postal code. The chart below shows consistency rates based on migrant status and type of migration.

From an examination of this chart, two main findings emerge. First, the fact of having moved during the past year, either to a location within the same municipality or to elsewhere in Canada, is closely related to postal code consistency. Linked individuals who moved have a consistency level of less than 85%, compared with 95.4% for those who remained at the same address. Second, the consistency level varies only slightly according to the type of migration. This suggests that it is the fact of having moved, not the type of migration, which is associated with the consistency of the postal code.

Chart 3

Consistency rates (percent) between the T1FF and census postal codes, by migrant status and type of migration
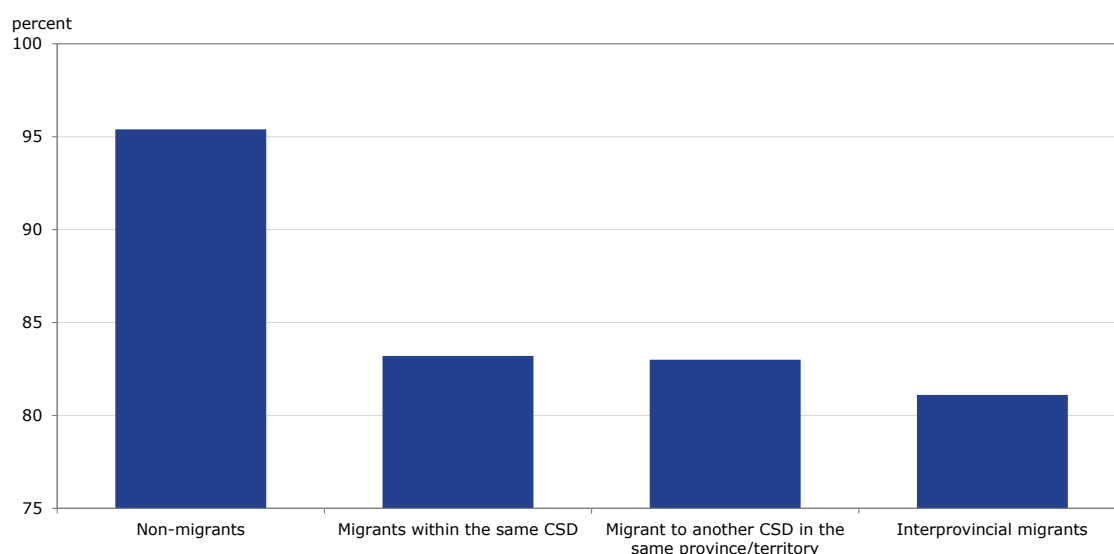
Sources: Statistics Canada, 2011 Census, 2011 National Household Survey, and 2010 T1 Family File.

While these results may be explained by the difference between the reference dates for the two files,[30] it is also possible that some of the gap may be due to migratory movements that are not adequately captured by the tax data. As mentioned earlier, highly mobile individuals such as some young adults might give the mailing address for the parental home as long as their place-of-residence situation is fairly unstable. On this subject, Morissette and Bérard-Chagnon (2014) showed that for measuring interprovincial migration, the differences between tax data and NHS data can be quite substantial. Since migration is essentially a change in the place of residence, these results tend to reaffirm the limitations of tax data for identifying the usual place of residence.

Table 4 on the next page shows the differences in postal code consistency for a few characteristics related to mobility. These data serve to identify particular segments of the population that have lower postal code consistency, largely because of their greater mobility.

---

30.   It should be kept in mind that the postal code in the census dates from May 10, 2011 while the one in the tax data dates from December 2011.

An examination of the consistency of the postal code by age reveals two things. First, the consistency rate declines moderately starting in the early twenties, falls below 90% between ages 25 and 29 and then begins rising at around age 30. This period in the life cycle is generally marked by several moves, motivated by leaving the parental home, attending a postsecondary educational institution, entering the labour market or purchasing a first property. In fact, the propensity to move peaks at between ages 25 and 29 (Dion and Coulombe 2008). Young adults who move are especially likely to maintain ties with another residence because of the often temporary nature of their moves. Furthermore, the process of leaving the parental home is becoming more complex, largely because of the democratization and lengthening of postsecondary education. It appears that increasingly, young adults' accession to residential independence involves periods of dependence and semi-autonomy (Billette et al. 2006), during which they might continue to put the mailing address of the family home on their tax return while being enumerated in their "independent" home.

Second, postal code consistency drops considerably among very elderly persons, falling below the 80% threshold starting at age 80 and even reaching close to 60% for persons aged 90 and older. This drop is due to the combined effect of the above-mentioned lower consistency rate for very elderly people who live in a collective dwelling and the fact that very elderly persons are much more likely to

Table 4

Distribution in the matched file and rates of consistency (percent) between the T1 Family File and census postal codes, by characteristic associated with mobility

Characteristic	Distribution in matched file	Consistency rate
	percent	
General	100.0	92.9
Age group		
0 to 9 years	14.1	91.5
10 to 19 years	14.0	93.6
20 to 24 years	5.8	91.0
25 to 29 years	6.0	89.2
30 to 39 years	14.5	92.0
40 to 49 years	16.5	94.7
50 to 59 years	14.2	95.4
60 to 69 years	8.6	94.9
70 to 79 years	4.2	93.1
80 years and older	2.3	79.9
Immigrant status and period of immigration		
Individual born in Canada	79.2	94.2
Immigrant	20.4	93.8
Landed between 2006 and 2011	3.4	88.7
Landed between 2001 and 2005	3.4	92.9
Landed between 1996 and 2000	2.5	94.3
Landed prior to 1996	11.1	95.6
Non-permanent resident	0.5	86.9
Total household income		
Under $25,000	7.6	89.9
Between $25,000 and $ 74,999	32.9	92.8
Between $75,000 and $149,999	41.1	95.2
$150,000 and over	18.4	95.6
Renter status		
Owner	81.2	95.7
Renter	18.2	87.1
Band housing	0.6	87.8

Sources: Statistics Canada, 2011 Census, 2011 National Household Survey, and 2010 T1 Family File.

live in a collective dwelling than the rest of the population. When only persons who live in a private dwelling are considered, postal code consistency for very elderly persons remains around 90%. Very elderly persons are also more likely to have their tax return prepared by an accountant or a family member, which is another factor closely associated with lower postal code consistency.

While the consistency level for matched immigrants is similar to that for the Canadian-born population, it follows a gradient clearly defined by the period in which the immigrant has landed. For recent immigrants who landed between 2006 and 2011, the consistency level is 88.7%, and it gradually increases for earlier cohorts of immigrants. This is mainly because recent immigrants are more mobile, seemingly because the process of settling in a new country is a gradual process (Houle 2007; Okonny-Myers 2010). Moreover, the study of Houle (2007) revealed that more than one recent immigrant in ten had moved to another CSD in the 24 months after settling in Canada. Non-permanent residents[31] also had a consistency level below 90%. This is due to both a greater propensity to migrate and the fact than non-permanent residents are more likely to have been matched in the fifth linkage wave.[32]

The level of consistency is also correlated with household income. Persons with a total household income below $25,000 are less likely to have indicated the same postal code in the census and the tax data. Clearly, this characteristic is related to renter status, which is also linked with consistency. Renters and individuals who live in band housing[33] have a consistency level below 90%. The links between being a renter and the greater propensity to move are obvious.

---

31. A non-permanent resident is an individual who has received a temporary resident permit, such as for work or study.

32. Nearly one-third of non-permanent residents were added to the linkage in the final wave, compared with 12% for Canadian-born persons. Section 4.3 will explore the relationship between the linkage wave and the consistency of the postal code.

33. For historical and statutory reasons, shelter occupancy on reserves does not lend itself to the usual classification by standard tenure categories. Therefore, a special category, band housing, was created starting with the 1991 Census.

4.3.   Consistency of postal code: Results for methodological characteristics

The table below shows two methodological characteristics that are correlated with the consistency of the postal code.

As can be seen from the data in this table, the methods used to construct the T1FF and the linkage are likely to affect postal code consistency.

The level of consistency is lower for persons matched in the fifth and final linkage wave (84.8%). Because of the approach used to construct the linkage, the links created in the last waves are thought to be slightly less robust, and therefore the chances of having a false match[34] are greater. Of course, postal codes are less likely to correspond for a false match. Also, persons matched in the final wave have different characteristics from those matched in the previous waves, especially as regards their propensity to move and their living in a collective dwelling. As seen above, these situations can play a role in explaining the lower consistency for persons matched in the fifth wave.

Although living taxfilers who are linked with a deceased taxfiler constitute only a small proportion of taxfilers, they have a lower level of postal code consistency (87.1%). These taxfilers tend to be considerably older, and nearly 80% of them were matched in the final wave. These factors might explain the lower consistency for this group.

Table 5
Distribution in the matched file and rates of consistency (percent) between the T1FF and census postal codes, by two methodological characteristics

Characteristic	Distribution in matched file	Consistency rate
	percent	
Overall	100.0	92.9
Linkage wave		
1) Exact match	69.0	94.6
2) More relaxed constraint on name	2.9	93.3
3) *Mix Match*	2.0	93.1
4) At least three family members with match on sex and date of birth	12.3	92.3
5) Unattached persons added	13.8	84.8
Type of record in T1FF		
Living taxfiler	75.7	93.1
Imputed child/spouse	24.1	92.2
Living taxfiler linked with a deceased taxfiler	0.2	87.1

Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

---

34.   A "false match" is a match between two different persons, a match which theoretically should not have been created.

Chapter 5. Determinants of postal code consistency

"Chapter 4. Consistency of postal code by individuals' characteristics" on page 24 highlighted several characteristics associated with lower postal code consistency. Those characteristics indicate that different segments of the population appear to be especially likely not to have the same place of residence in the census and in the T1FF tax data. This chapter pursues that analysis, identifying the main factors that are correlated with the consistency of the postal code using a multivariate logistic regression model.

> **Methodology notes**
>
> The matched file includes nearly 4 million records for NHS respondents. Because of the very large number of observations, the usual thresholds of statistical significance are less appropriate (Lin et al. 2013; Sullivan and Feinn 2012).
>
> We therefore drew partially on Ferguson (2009) in determining the characteristics for which the association with postal code consistency is statistically significant. The results shown here are considered statistically significant if the 95% confidence interval for the odds ratio is either totally greater than 1.50 or totally less than 0.67. Although Ferguson proposes a threshold of 2.00 (or 0.50) for determining that an odds ratio is statistically significant in the social sciences, we have opted for a slightly less restrictive threshold. This decision, which is subjective, is based in part on the assumption that the noise caused by the difference between the reference dates for the two sources examined could weaken the relationships between the characteristics and postal code consistency. Also, we added a second criterion, namely that the chi-square statistic must be greater than 100 (which corresponds to a p-value considerably less than 1%). This criterion was chosen so as to include a measure that is well tested in the social sciences, the p-value, while establishing a subjective threshold which, according to our experience, properly takes account of the very large number of observations.
>
> This combined approach makes it possible to target characteristics of interest in that not only are they closely related to the consistency of the postal code but the association is also fairly strong. For more information on the main limitations in using p-values for determining statistical significance thresholds, see for example Wasserstein and Lazar (2016). Although only the results that are statistically significant according to the criteria established here are reported, «Table A.4» on page 39 in the Appendix includes the results for all the characteristics included in the model.
>
> The NHS was administered only to the population living in private households. Therefore it is not possible to include the effect of living in a collective dwelling in the multidimensional analysis.
>
> The variable constructed to identify children potentially living in joint custody was omitted from the analyses in this chapter because of the more approximate nature of its construction.
>
> Lastly, the statistical significance thresholds used here do not refer to the relationships that unite the characteristics in the context of a superpopulation. Instead, they serve as tools for evaluating differences in postal code consistency among various demographic groups. Accordingly, the results represent only the linkage data and not the Canadian population as a whole. Any inference to the Canadian population should be made with great caution.

The table on the right shows the main results of this modelling.

The data in this table show, firstly, that even when controls for the effect of different characteristics were applied, a number of them remain strongly associated with the level of postal code consistency. Also, while the regression model confirms several results described in "Chapter 4. Consistency of postal code by individuals' characteristics", it also brings out links that were less obvious at the descriptive stage.

The fact of having moved during the year preceding the NHS remains highly correlated with postal code consistency. Individuals who moved still exhibit considerably lower consistency than those who did not. This applies to movers regardless of the type of move they made, and could be due to the difference in the reference dates for the two files.

Despite controlling for the effect of mobility, children, young adults and individuals who rent their dwelling still exhibit lower consistency. This suggests that mobility does not totally explain these groups' lower consistency levels, which were identified earlier.

While the effect of age on consistency continues to be statistically significant for young adults according to our criteria, the effect seen in the descriptive stage for elderly persons aged 80 and older tends to fade in light of the possible effect of other factors. This appears to result mainly from the exclusion of persons living in collective dwellings for the multivariate model because the NHS data were used. The descriptive analysis showed that the consistency level for very elderly persons living in private dwellings remains high.

Taxfilers who had their return prepared by a third party according to box 490 continue to have a lower consistency level, even when the effect of age is taken into account.

Table 6

Odds ratios and 95% confidence intervals for characteristics significantly associated with postal code consistency

Characteristic	Odds ratio
Migrant status	
Non-migrant	Reference
Migrant within the same CSD	0.39 [0.39 to 0.40]
Migrant to another CSD in the same province/territory	0.34 [0.33 to 0.35]
Interprovincial migrant	0.36 [0.35 to 0.38]
Aboriginal identity	
No identity	Reference
Inuit	2.19 [1.92 to 2.50]
Age group	
0 to 9 years	0.51 [0.49 to 0.53]
10 to 19 years	0.58 [0.56 to 0.61]
20 to 29 years	0.60 [0.59 to 0.62]
40 to 49 years	Reference
Census metropolitan area of residence	
Toronto	Reference
British Columbia (non-CMA)	0.58 [0.56 to 0.59]
Yukon	0.47 [0.42 to 0.54]
Northwest Territories	0.33 [0.29 to 0.37]
Nunavut	0.23 [0.19 to 0.28]
Renter status	
Owner	Reference
Renter	0.46 [0.46 to 0.47]
Household size	
1 person	2.11 [2.04 to 2.18]
4 persons	Reference
Status of person within census family	
Married spouse or common-law partner	Reference
Child	2.23 [2.15 to 2.31]
Lone-parent family status	
Not a lone-parent family	Reference
Lone-parent family	0.65 [0.64 to 0.66]
Type of record in the T1FF	
Living taxfiler	Reference
Imputed child/spouse	0.57 [0.55 to 0.59]
Linkage wave	
1) Exact match	Reference
5) Unattached persons added	0.31 [0.30 to 0.31]
Return prepared by a third party (box 490)	
No	Reference
Yes	0.05 [0.05 to 0.06]
Number of observations	3,991,353
Cox and Snell R-squared (percent)	4.4

Notes: Only categories that are statistically significant based on the established thresholds are shown here. "Table A.4" on page 39 in the Appendix includes results for all the characteristics included in the model.

Sources: Statistics Canada, 2011 Census, 2011 National Household Survey, and 2010 T1 Family File.

The Inuit, who exhibited a slightly lower level of consistency than that of the population with no aboriginal identity, now have a significantly higher level. This population's lower consistency level would therefore seem to be attributable to its specific characteristics, such as its younger age structure and its geographic location in northern Canada. Since these characteristics are also related to lower consistency, taking them into account in the model serves to isolate the effect of having reported an Inuit identity in the NHS.

The results of the regression analysis shed more light on the correlations between some family characteristics and the consistency of the postal code. When other factors are taken into account, children in a census family[35] are more likely to have the same postal code in the two sources. Also, the level of postal code consistency is lower for persons living in a lone-parent situation. This could reflect the sometimes ambiguous situation of children in joint custody with regard to their place of residence.

It is worth noting that the descriptive analysis revealed that persons living alone were only slightly less likely to have the same postal code in the census and in the T1FF than were four-person households.[36] However, this situation changes when the other characteristics are taken into account, with the result that living alone is now associated with a higher level of consistency. This result is mainly due to the fact that persons living alone are not only more mobile but are also more likely to have been matched in the final linkage wave. Controlling for these characteristics brings out the real links between postal code consistency and household size.

Some regional effects are associated with postal code consistency. Persons who live in the territories or in non-CMA areas in British Columbia have lower consistency than those who live in the Toronto CMA.

The type of record in the T1FF and the linkage wave continue to be correlated with the consistency of the postal code even when controlling for other factors. Imputed children and spouses exhibit significantly lower consistency than living taxfilers. The geographic information for imputed persons is obtained indirectly from that of the taxfiler to whom they are attached.[37] This information is therefore slightly more likely not to reflect the imputed person's real place of residence, such as in the case of a child in joint custody. Since the majority of imputed persons are children, controlling for the effect of age yields a better picture of the association between the consistency of the postal code and the type of record. For example, taxfilers 18 years of age have a consistency rate of 95.1%, compared with 91.2% for imputed persons of the same age.

Individuals matched in the fifth linkage wave continue to exhibit much lower consistency than those matched in the initial wave. These results suggest that this situation is not solely due to the particular characteristics of the persons matched in this wave. Consequently, it is possible that the greater propensity to have a false match in this wave is a factor that helps to explain this situation.

Finally, while the regression analysis shed light on a number of factors associated with postal code consistency, the relatively low level of the adjustment statistic in the Cox and Snell model[38] (4.4%) indicates that these characteristics explain only a small part of the differences in postal code consistency levels. Other factors could be associated with place of residence consistency. The difference in the reference dates of the two files might also contribute significantly to variations in consistency levels. However, it is not possible at this time to adjust the data to take account of this difference.

---

35. A census family is a married or common-law couple (with or without children) or a lone-parent family.
36. Persons living alone have a postal code consistency level of 92.6%, compared with 95.0% for persons living in a four-person household.
37. As noted earlier, these individuals are added to taxfilers by means of data from the CCTB, vital statistics or a historical file in the case of children or a tax return in the case of spouses.
38. This adjustment measure is designed to reproduce the R^2 statistic of the linear regression in a logistic regression context using likelihood functions.

Conclusion

Tax data are used increasingly to measure the population and its characteristics as well as to shed light on different demographic and social issues. This trend raises a number of challenges, including differences in the concept of place of residence. While the census enumerates the population at its usual place of residence, tax data collect taxfilers' mailing address. This study examines the issue using a record linkage between data from the 2011 Census, the NHS and the 2010 T1FF.

First, this study found that 92.9% of the matched individuals had the same postal code in the census and in the T1FF tax data. This proportion climbs to more than 99% when more aggregated geographic levels are considered. However, consistency is lower, sometimes substantially, for some segments of the population. Persons who live in a collective dwelling, migrants, young adults, very elderly people and persons who had their tax return prepared by a third party (according to box 490 of the tax return form) are especially likely to have lower consistency. In general, these include demographic groups for which it is more difficult to determine the place of residence. These individuals seem to be more likely to report a different place of residence in the census and in the tax data, largely because of the different definitions of residence in the two sources. The variety of characteristics associated with lower consistency is also indicative of the diversity of the life paths that can lead to ambiguity when it comes to determining the usual place of residence and the mailing address reported on the tax return.

While the results of this study are enlightening, they also raise several questions. A number of taxfilers have their return prepared by an accountant or someone close to them, which can definitely influence the consistency of the place of residence. If the preparer did not fill in box 490, it is impossible to determine whether the taxfiler completed the return himself or herself.

The difference between the reference dates for the two sources may explain much of the variations in consistency for particular groups, especially those whose characteristics are associated with mobility. However, the effect of this difference cannot be isolated at this time. Furthermore, several characteristics likely to be associated with lower consistency of the place of residence cannot be observed directly through the data used in this study. These characteristics include postsecondary students who alternate between their place of study and the family home, children in joint custody, the shadow population[39] and individuals who have just separated or divorced.

One of the main uses of tax data in demography is for estimating internal migration. This demographic phenomenon is defined as a change in the usual place of residence, and therefore a rigorous measurement of migration depends on a good measurement of the place of residence. The consistency levels calculated in this study, especially for persons who moved according to the NHS, point to various limitations in the tax data for the purposes of measuring migration. The tax data appear to be somewhat less appropriate for measuring migration of highly mobile groups, such as young adults. Therefore, it would be very worthwhile to determine the extent to which tax data lend themselves to calculating precise estimates of migration by examining the consistency of migration data using the record linkage used here.

A final point is that more generally, issues regarding the place of residence reflect the growing diversity of social trajectories. Although joint custody of children and the increase in the number of people residing in collective dwellings as a result of population aging tend to be relatively recent social phenomena, they are likely to be enduring features of Canadian society in the 21st century. In conjunction with the growing use of administrative data in demography, these changes will likely lead researchers to conduct more data comparison studies such as the one described here. Such studies will not only provide a better understanding of the possible uses of a growing number of files available, but also a better grasp of the demographic dynamics that characterize the Canadian population.

---

39. This population is defined broadly as consisting of persons living or working temporarily in an area but maintaining a permanent residence elsewhere (Aylward 2006).

Bibliography

AYDEMIR, Abdurrahman and Chris ROBINSON. 2006. "Return and Onward Migration Among Working Age Men", *Analytical Studies Branch Research Paper Series,* no. 273, Statistics Canada Catalogue no. 11F0019M, 49 pages.

AYLWARD, Mary Joan. 2006. *Shadow Populations in Northern Alberta,* Northern Alberta Development Council, July 2006, 53 pages.

BÉRARD-CHAGNON, Julien. 2008. *Analyse de la couverture du fichier T1FF de 2005,* working document, Demography Division, Statistics Canada, June 2008, 60 pages.

BÉRARD-CHAGNON, Julien and James BRENNAN. 2014. "Évaluation du fichier de la Prestation fiscale canadienne pour enfants à l'aide d'un couplage de données", *Acte de conférence du Symposium 2010 de méthodologie de Statistique Canada,* Ottawa, 6 pages.

BILLETTE, Jean-Michel, Céline LE BOURDAIS and Benoît LAPLANTE. 2006. "Le contexte de l'indépendance résidentielle au Canada", *Cahiers québécois de démographie,* volume 35, no. 1. spring 2006, pages 83 to 121.

DION, Patrice and Simon COULOMBE. 2008. "Portrait of the mobility of Canadians in 2006: Trajectories and characteristics of migrants", *Report on the Demographic Situation in Canada: 2005 and 2006,* Statistics Canada Catalogue no. 91-209-X, pages 78 to 108.

FERGUSON, Christopher J. 2009. "An Effect Size Primer: A Guide for Clinicians and Researchers", *American Psychological Association,* volume 40, no. 5, pages 532 to 538.

FINNIE, Ross. 2000. "Who Moves? A Panel Logit Model Analysis of Inter-Provincial Migration in Canada", *Analytical Studies Branch Research Paper Series,* no. 142, Statistics Canada Catalogue no. 11F0019M, 33 pages.

HE, Jiaosheng and Margaret MICHALOWSKI. 2005. "Research on Modifications to the Method of Preliminary Estimates of Interprovincial Migration", *Demographic Documents,* no. 7, Statistics Canada Catalogue no. 91F0015M, 64 pages.

HOULE, René. 2007. "Secondary Migration of New Immigrants to Canada", *Our diverse cities,* no. 3, summer 2007, pages 16 to 24.

HOUSE, Georgina. 2014. *The Linking Process,* working document, Household Survey Methods Division, Statistics Canada, 1 page.

LAPORTE Christine, Yuqian LU and Grand SCHELLENBERG. 2012. "Inter-provincial Employees in Alberta", *Analytical Studies Branch Research Paper Series,* no. 350, Statistics Canada Catalogue no. 11F0019M, 57 pages.

LESSARD, Martin. 2011. *Création du T1FF,* presentation given during Demography Division lunch-time seminars, Statistics Canada, Ottawa, March 2011.

LIN, Mingfeng, Henry C. LUCAS and Galit SHMUELI Jr. 2013. "Too Big to Fail: Large Samples and the p-Value Problem", *Information Systems Research,* pages 1 to 12.

MORISSETTE, Denis and Julien BÉRARD-CHAGNON. 2014. *Evaluation of the 2011 Errors of Closure for Canada, Provinces and Territories,* working document, Demography Division, Statistics Canada, July 2014, 65 pages.

NATIONAL RESEARCH COUNCIL. 2006. *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census,* Panel on Residence Rules in the Decennial Census, Daniel L. CORK and Paul R. VOSS, editors, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC, The National Academies Press, 355 pages.

OKONNY-MYERS, Ima. 2010. *The Interprovincial Mobility of Immigrants in Canada,* Citizenship and Immigration Canada, Ci4-47, 29 pages.

SCARDAMALIA, Robert L. 2014. *Aging in America*, County and City Extra Series*,* Bernan Press, 416 pages.

STATISTICS CANADA. 2012. *Overview of the Census,* 2011 Census, Catalogue no. 98-302-X, 49 pages.

STATISTICS CANADA. 2015a. *Corporate Business Plan – Statistics Canada – 2015/2016 to 2017/2018,* http://www.statcan.gc.ca/eng/about/bp, accessed December 14, 2015.

STATISTICS CANADA. 2015b. *Census Technical Report: Coverage,* 2011 Census, Catalogue no. 98-303-X, 152 pages.

STATISTICS CANADA. 2015c. *Postal CodeOM Conversion File (PCCF), Reference Guide,* April 2015, http://www.statcan.gc.ca/pub/92-154-g/92-154-g2015001-eng.htm, accessed November 6, 2015.

SULLIVAN, Gail M. and Richard FEINN. 2012. "Using Effect Size-or Why the P Value Is Not Enough", *Journal of Graduate Medical Education,* September 2012, pages 279 to 282.

TURCOTTE, Martin and Mireille VÉZINA. 2010. "Migration from central to surrounding municipalities in Toronto, Montréal and Vancouver", *Canadian Social Trends,* winter 2010, no. 90, Statistics Canada Catalogue no. 11-008-X, 24 pages.

TURCOTTE, Martin. 2013. "Living apart together", *Insights on Canadian Society,* March 5, 2013, Statistics Canada Catalogue no. 75-006-X, 11 pages.

U.S. CENSUS BUREAU. 2004. *Meeting the 21st Century Demographic Data Needs-Implementing the American Community Survey, Report 4: Comparing General Demographic and Housing Characteristics with Census 2000,* May 2004, 87 pages.

VAMDERKAMP, John and Kenneth E. GRANT. 1988. "Canadian Internal Migration Statistics: Some Comparisons and Evaluations", *Canadian Journal of Regional Science,* volume 11, no. 1, spring 1988, pages 9 to 32.

WASSERSTEIN, Ronald L. and Nicole A. LAZAR. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose", *The American Statistician,* volume 70, no. 2, pages 129 to 133.

WILLBOND, Stephanie. 2014. "Migration: Interprovincial, 2011/2012", *Report on the Demographic Situation in Canada,* Statistics Canada Catalogue no. 91-209-X.

Appendix tables

Table A.1

Proportion of individuals in the linkage whose postal codes in the T1FF and the census are valid (percent), by selected characteristics

Characteristic	Valid postal codes
	percent
General	**95.9**
Province or territory	
Newfoundland and Labrador	97.7
Prince Edward Island	96.3
Nova Scotia	95.6
New Brunswick	95.3
Quebec	94.1
Ontario	96.3
Manitoba	95.7
Saskatchewan	95.3
Alberta	97.1
British Columbia	96.9
Yukon	92.6
Northwest Territories	96.9
Nunavut	92.6
Age group	
0 to 9 years	96.1
10 to 19 years	96.2
20 to 29 years	96.6
30 to 39 years	96.6
40 to 49 years	96.4
50 to 59 years	96.1
60 to 69 years	95.3
70 to 79 years	93.0
80 years and older	87.1
Linkage wave	
1) Exact match	98.4
2) More relaxed constraint on name	95.9
3) *Mix Match*	99.0
4) At least three family members with match on sex and date of birth	86.4
5) Unattached persons added	92.6
Type of record in the T1FF	
Living taxfiler	95.8
Imputed child/spouse	96.1
Living taxfiler linked with a deceased taxfiler	91.7

Sources: Statistics Canada, 2011 Census and 2010 T1 Family File.

Table A.2
Characteristics used in this analysis to examine the consistency of the place of residence

Characteristics associated with concept of place of residence	Characteristics associated with mobility	Methodological characteristics
Tax return prepared by third party (box 490)	Mobility one year ago	Linkage wave
Aboriginal identity	Immigrant status	Type of record in T1FF
Living in a collective dwelling	Immigration period	Rural postal code
Province/territory of work	Total household income	-
Metropolitan influenced zone	Highest educational attainment	-
Primary household maintainer	Renter status	-
School attendance	Marital status	-
Presence of spouse in household	Industrial sector	-
Household on-reserve	Census family size	-
Lone-parent family	Household size	-
Children in joint custody[1]	Family status	-
-	Age	-

1. Derived characteristic based on those available in the census.

Table A.3

Distribution in the matched file and consistency rate (percent) between the T1FF and census postal codes for characteristics not included in the analysis

Characteristic	Distribution in matched file	Consistency rate
	percent	
General	100.0	92.9
Metropolitan influenced zone		
CA or CMA	86.0	94.2
Strong MIZ	3.9	94.6
Moderate MIZ	5.2	94.0
Weak MIZ	4.2	92.8
No MIZ	0.5	92.3
Territories outside CAs	0.2	91.0
Primary household maintainer		
No	62.7	94.1
Yes	37.3	94.1
School attendance		
Did not attend school	81.0	94.6
Attended school	19.0	93.8
Lone-parent family status		
Not a lone-parent family	89.7	95.0
Lone-parent family	10.3	90.3
Highest educational attainment		
No secondary school diploma	15.8	94.0
Secondary school diploma	34.2	94.6
College diploma	19.7	94.7
University diploma	30.2	94.2
Marital status		
Single	40.5	93.2
Married or common-law	51.7	94.9
Separated or divorced	5.0	92.5
Widow/widower	2.8	93.9
Size of census family		
Unattached person	10.4	90.9
2 persons	22.8	94.0
3 persons	21.5	93.7
4 persons	30.0	95.2
5 persons	11.3	95.0
6 or more persons	4.1	93.8
Rural postal code		
No	86.7	92.8
Yes	13.3	93.5
Industry Sector (North American Industry Classification System)		
Agriculture, forestry, fishing and hunting	1.4	94.3
Mining, quarrying, and oil and gas extraction	1.0	93.4
Utilities	0.7	95.3
Construction	4.5	93.9
Manufacturing	6.8	95.2
Wholesale trade	3.3	94.8
Retail trade	8.7	94.3
Transportation and warehousing	3.2	94.6
Information and cultural industries	1.9	94.3
Finance and insurance	3.7	94.7
Real estate and rental and leasing	1.3	93.6
Professional, scientific and technical services	6.1	94.0
Management of companies and enterprises	0.1	93.7
Administrative and support, waste management and remediation services	2.9	93.4
Educational services	6.5	95.0
Health care and social assistance	8.5	94.3
Arts, entertainment and education	1.7	94.4
Accommodation and food services	4.4	93.1
Other services (except public administration)	3.2	93.8
Public administration	6.2	94.9
Did not work in 2010 or 2011	23.8	94.6
Household size		
1 person	7.4	92.6
2 persons	21.9	93.7
3 persons	21.1	93.5
4 persons	29.5	95.0
5 persons	12.7	94.7
6 or more persons	7.3	93.5
Status of person within census family		
Married spouse or common-law partner	51.5	95.0
Lone parent	3.7	91.8
Child	34.5	94.0

Sources: Statistics Canada, 2011 Census, 2011 National Household Survey, and 2010 T1 Family File.

Table A.4

Logistic regression coefficients, standard errors and chi-squared statistics for postal code consistency by selected characteristics

Characteristic	Coefficient	Standard error	Chi-squared statistic
Intercept	3.78	0.01	62,761.6
Migrant status			
Non-migrant		Reference	
Migrant within the same CSD	-0.94	0.01	20,717.4
Migrant in another CSD in the same province/territory	-1.08	0.01	14,642.9
Interprovincial migrant	-1.01	0.02	3,191.6
Aboriginal identity			
No identity		Reference	
First Nations	-0.10	0.02	38.4
Métis	-0.12	0.02	43.4
Inuit	0.78	0.07	136.0
More than one aboriginal identity	0.01	0.06	0.0
Province or territory of work[1]			
Same province or territory		Reference	
Other province or territory	-0.44	0.03	202.0
Immigrant status and period of immigration			
Individual born in Canada		Reference	
Immigrant landed between 2006 and 2011	-0.26	0.01	649.1
Immigrant landed between 2001 and 2005	-0.16	0.01	194.6
Immigrant landed between 1996 and 2000	-0.07	0.01	23.7
Immigrant landed before 1996	0.00	0.01	0.0
Non-permanent resident	0.04	0.02	2.8
Metropolitan influenced zone (MIZ)[2]			
CA or CMA		Reference	
Strong MIZ	-0.16	0.01	128.0
Moderate MIZ	-0.12	0.01	94.9
Weak MIZ	-0.01	0.01	58.3
No MIZ	-0.03	0.03	0.7
Territories (outside CAs)	0.65	0.08	71.0
Primary household maintainer			
No		Reference	
Yes	0.05	0.01	69.8
Marital status			
Single	0.19	0.02	132.9
Married or common-law		Reference	
Separated or divorced	0.09	0.02	20.7
Widow/widower	0.25	0.02	120.1
Highest educational attainment			
No secondary school diploma	-0.01	0.01	3.0
Secondary school diploma		Reference	
College diploma	0.00	0.01	0.2
University diploma	-0.06	0.01	69.8
School attendance			
No		Reference	
Yes	-0.05	0.01	36.0
Presence of spouse in household			
Spouse present		Reference	
Spouse absent	-0.39	0.04	107.7
Age group			
0 to 9 years	-0.67	0.02	1,047.2
10 to 19 years	-0.54	0.02	886.7
20 to 29 years	-0.50	0.01	2,946.3
30 to 39 years	-0.30	0.01	1,395.9
40 to 49 years		Reference	
50 to 59 years	0.17	0.01	336.4
60 to 69 years	0.13	0.01	118.0
70 to 79 years	0.13	0.02	62.6
80 years and older	0.01	0.02	0.1

See notes at the end of the table on

Table A.4
Logistic regression coefficients, standard errors and chi-squared statistics for postal code consistency by selected characteristics - continued

Characteristic	Coefficient	Standard error	Chi-squared statistic
Industry Sector (North American Industry Classification System)			
Agriculture, forestry, fishing and hunting	-0.19	0.02	59.7
Mining, quarrying, and oil and gas extraction	-0.09	0.03	10.7
Utilities	-0.04	0.03	1.4
Construction	-0.10	0.02	37.5
Manufacturing		Reference	
Wholesale trade	-0.06	0.02	9.6
Retail trade	-0.05	0.01	13.1
Transportation and warehousing	-0.07	0.02	16.3
Information and cultural industries	-0.01	0.02	0.1
Finance and insurance	-0.07	0.02	17.8
Real estate and rental and leasing	-0.24	0.02	106.1
Professional, scientific and technical services	-0.12	0.01	70.6
Management of companies and enterprises	-0.27	0.09	9.8
Administrative and support, waste management and remediation services	-0.08	0.02	22.8
Educational services	0.03	0.02	4.8
Health care and social assistance	-0.06	0.01	16.1
Arts, entertainment and recreation	-0.07	0.02	9.2
Accommodation and food services	-0.10	0.02	41.9
Other services (except public administration)	-0.12	0.02	47.2
Public administration	0.00	0.02	0.0
Did not work in 2010 or 2011	-0.02	0.02	1.6
Renter status			
Owner		Reference	
Renter	-0.77	0.01	20,449.8
Band housing	-0.20	0.03	43.7
Household size			
1 person	0.75	0.02	1,950.8
2 persons	0.18	0.01	158.0
3 persons	0.08	0.01	33.8
4 persons		Reference	
5 persons	-0.02	0.01	3.1
6 or more persons	-0.15	0.01	112.0
Status of person within census family			
Married spouse or common-law partner		Reference	
Lone parent	0.30	0.02	208.5
Child	0.80	0.02	2,056.8
Lone-parent family status			
Not a lone-parent family		Reference	
Lone-parent family	-0.43	0.01	2,071.1

Table A.4

Logistic regression coefficients, standard errors and chi-squared statistics for postal code consistency by selected characteristics - continued

Characteristic	Coefficient	Standard error	Chi-squared statistic
Census metropolitan area of residence			
St. John's	-0.38	0.03	153.5
Newfoundland and Labrador non-CMA	-0.26	0.03	94.0
Prince Edward Island	-0.43	0.04	138.9
Halifax	-0.09	0.02	15.3
Nova Scotia non-CMA	-0.44	0.02	492.5
Moncton	-0.16	0.03	23.4
Saint John	-0.12	0.04	10.5
New Brunswick non-CMA	-0.29	0.02	214.0
Saguenay	-0.11	0.03	12.4
Québec	-0.03	0.02	4.9
Sherbrooke	-0.22	0.03	77.4
Trois-Rivières	-0.02	0.03	0.3
Montréal	0.03	0.01	14.6
Ottawa — Gatineau	0.16	0.05	10.6
Quebec non-CMA	-0.02	0.01	2.2
Kingston	-0.18	0.03	29.8
Peterborough	-0.13	0.04	9.1
Oshawa	0.14	0.03	30.4
Toronto		Reference	
Hamilton	0.09	0.02	26.8
St. Catharines — Niagara	0.00	0.02	0.0
Kitchener — Cambridge — Waterloo	0.11	0.02	32.0
Brantford	-0.06	0.04	2.8
Guelph	0.06	0.04	2.5
London	0.01	0.02	0.3
Windsor	0.08	0.03	9.2
Barrie	-0.18	0.03	36.4
Greater Sudbury	-0.02	0.03	0.5
Thunder Bay	-0.01	0.04	0.1
Ontario non-CMA	-0.12	0.01	100.1
Winnipeg	0.04	0.02	6.9
Manitoba non-CMA	-0.29	0.02	184.3
Regina	-0.05	0.03	2.9
Saskatoon	-0.25	0.02	102.9
Saskatchewan non-CMA	-0.28	0.02	181.2
Calgary	-0.14	0.01	128.5
Edmonton	-0.16	0.01	152.2
Alberta non-CMA	-0.41	0.01	952.0
Kelowna	-0.42	0.03	228.8
Abbotsford — Mission	-0.42	0.03	211.5
Vancouver	-0.27	0.01	865.5
Victoria	-0.30	0.02	229.5
British Columbia non-CMA	-0.55	0.01	2,112.5
Yukon	-0.75	0.06	139.8
Northwest Territories	-1.12	0.06	359.5
Nunavut	-1.47	0.11	189.7
Household on-reserve			
No		Reference	
Yes	-0.45	0.03	280.6

Table A.4

Logistic regression coefficients, standard errors and chi-squared statistics for postal code consistency by selected characteristics - end

Characteristic	Coefficient	Standard error	Chi-squared statistic
Type of record in T1FF			
Living taxfiler		Reference	
Imputed child/spouse	-0.57	0.02	998.8
Other[3]	-0.10	0.05	4.3
Linkage wave			
1) Exact match		Reference	
2) More relaxed constraint on name	-0.14	0.01	117.5
3) *Mix Match*	-0.14	0.01	84.9
4) At least three family members with match on sex and date of birth	-0.11	0.01	208.2
5) Unattached persons added	-1.18	0.01	15,817.5
Return prepared by third party (box 490)[4]			
No		Reference	
Yes	-2.94	0.02	24,041.2
Sex			
Female		Reference	
Male	-0.01	0.00	5.0
Total household income			
Less than $25,000	-0.17	0.01	389.5
Between $25,000 and $74,999	-0.08	0.01	205.9
Between $75,000 and $149,999		Reference	
$150,000 and over	-0.08	0.01	128.8
Rural postal code			
No		Reference	
Yes	0.21	0.01	507.0
Size of census family			
2 persons	-0.22	0.01	236.6
3 persons	-0.21	0.01	237.1
4 persons		Reference	
5 persons	0.02	0.01	1.7
6 or more persons	0.05	0.02	7.8
Number of observations			3,991,353
Cox and Snell R-squared (percent)			4.4

1. Persons living in the Ottawa — Gatineau CMA or in the census agglomerations (CAs) of Campbellton, Hawkesbury and Lloydminster were excluded from this analysis because their metropolitan area overlaps two provinces.
2. A MIZ is a category assigned to a municipality not included in either a CMA or a CA depending on the percentage of its resident employed labour force that commute to work in the core of any CMA or CA.
3. This group consists mostly of living taxfilers associated with deceased taxfilers.
4. This variable is obtained from box 490 on the T1 tax return. Box 490 provides information only on the preparer of the return if a fee was charged. Consequently, this variable does not cover all persons who had their return prepared by a third party.

Sources: Statistics Canada, 2011 Census, 2011 National Household Survey, and 2010 T1 Family File