

Historical tax information in the Canadian Housing Survey database

The Canadian Housing Survey (CHS) is a new survey jointly developed by Canada Mortgage and Housing Corporation (CMHC) and Statistics Canada. The survey will run biennially until 2028 in every province and territory. The survey collects information on dwelling and neighbourhood satisfaction, on key housing indicators related to core housing need, as well as many other important dwelling and housing characteristics including households living in social and affordable housing.

This article describes two pre-linked CHS datasets that will help inform CMHC's policies on appropriate housing supports and contribute to the evaluation of housing outcomes. Administrative tax data, for the past 10 years, have been used to derive income and housing mobility information for CHS respondents and members of their household. The data are accessible from Statistics Canada Research Data Centres (RDCs).

The following sections describe what information is available in the pre-linked datasets, how they were constructed, known data limitations and what users should know when using this information.

What data are available

Information on respondent income and residential dwellings over time is captured in two sets of data pre-linked to the CHS which are available in the Research Data Centers (RDC): the historical income and historical address components.

The historical income component consists of annual income from 2008 to 2017 for all CHS respondents and members of their households. The main source of information is administrative tax data – specifically the T1 Family Files (T1FF) and T4 files. Together, these files describe respondent income (including wage and salary income for T1 non-filers), number of jobs, and family composition across time.

About CMHC Research Insights

Research insights are summaries of our research reports.

These insights:

- identify a housing research issue, gap or need
- provide an overview of the research project undertaken to address it
- present major findings of the research

The research presented in this series explore the areas of Housing Need, Housing Finance, Housing Supply and Outcomes of the National Housing Strategy.

Sign Up

Get the latest findings directly in your inbox



cmhc.ca/researchnewsletter

The historical address component captures the geographic locations of all CHS respondents and members of their households from 2008 to 2018. The main source of information is mailing addresses kept by Canada Revenue Agency for tax filers. Statistics Canada obtains this information from the annual administrative T1 Personal Master File (PMF).

The geographic descriptors assigned to each mailing address include:

- A sequential marker differentiating mailing addresses the respondents have used between the years 2008 to 2018; and
- Geographic boundaries describing street blocks¹ for these residences (i.e. Census dissemination area block and postal code).

How the datasets are created

The first step in creating the CHS historical income and address components involves linking CHS respondent information with tax data. The CHS roster file is statistically matched in the Social Data Linkage Environment (SDLE)² using variables which include full name, date of birth, sex, marital status, geographic location, and the phone number used in collection. Linking the CHS to the SDLE helps facilitate data extractions for this project from other Statistics Canada data holdings – specifically tax files.

Records from the annual T1FFs and T4s from 2008 to 2017 are extracted based on deterministic matches between the SDLE matched CHS roster file and tax data files. In cases where there are duplicate T1FF records for an individual, only one record is kept for each reference year. Similarly, duplicate T4 records for a given individual and job are removed.

¹ In rural areas, census dissemination area blocks and postal codes are likely to cover areas larger than a “street block”.

² Social Data Linkage Environment (SDLE) at Statistics Canada enables statistical linkage across multiple datasets in the social domain. SDLE aims to standardize processes, improve methods and enhance data quality.

Records for the historical address component are extracted following the same deterministic process. Geocoded mailing addresses³ from the T1 PMF are assigned to each respondent when the data are available. When these data are missing, other available information (e.g. survey response to when respondents moved from their previous dwelling) is used to impute a geographic location when possible. Reserve codes are assigned when no dwelling address information can be derived and for cases where the individual is 15 years old or younger since they will not have filed taxes.

Data limitations

Limitations of both the collected survey and administrative data can result in unmatched records. Over 80% of CHS respondents were found in the SDLE. This rate of matching is comparable with other surveys matched to the SDLE. The likelihood of a successful match is not equal for all respondents, for example:

- individuals with a complete and valid birthdate (year, month and day) are more likely to be matched;
- individuals born before 2010 are more likely to be matched; but
- individuals in the northern territories are less likely to be matched.

Furthermore, even when an individual can be matched to a record in the SDLE, they may not have worked or filed taxes for a specific year. Table 1 describes the CHS respondents with matching tax records by age range.

When considering matched records across time, there are fewer matched records in the earlier reference years for many reasons (e.g. higher presence of younger adults or recent immigrants). Table 2 describes the lower matching rates in 2008 compared with the 2017 tax year by selected characteristics.

The same factors that limit matching of historical income and CHS survey data also apply when assigning dwelling history. Additionally, the Dwelling Universe File used for this purpose, provides a snapshot of the address register at a point in time which can result in missing information for earlier years. However, the impact of these limitations is mitigated somewhat through imputation for survey identified

non-movers. Table 3 describes the CHS respondents with historical address information by adjusted age ranges. The DBUID identifies Census dissemination area blocks.

Also, geographic descriptors at the dissemination block level may also be more limited in certain subpopulations such as persons living in the northern territories or in rental households, as shown in Table 4.

User notes

Comparing historical and current income

The CHS database includes annual income data from two sources: the T1FF for years 2008 to 2017, and the Administration Personal Income Masterfile (APIM) for 2018. These two programs are not strictly comparable and should be used together with caution.

While the main income concepts between APIM and T1FF match well, there are processing differences for these two programs. For example, APIM enhances data from T1 records with information available from other tax slips. Individuals who did not file T1 forms may still have income on APIM which is imputed from other tax slips not used by the T1FF program. As a result, coverage of total income for a given year with APIM may be more complete than for T1FF. Additionally, the SDLE non-linkage issue is mitigated for the 2018 APIM income by imputing the income of unmatched persons.

On the other hand, T1FF program describes information at a census family level through family formation, which the APIM program does not capture.

Interpretation of mailing address

The dwelling history extracted from tax files is based on mailing addresses which may not match the addresses at which CHS respondents actually reside. To assess this potential discrepancy, the 2018 ARUID of the geocoded mailing addresses were compared with the residential addresses used in CHS's survey frame (see Table 5).

Table 5 shows that about 6% of respondents have a different mailing ARUID than their residential ARUID, compared with 54% of respondents that have the same ARUID.

³ A geocoded address is represented by a unique record in Statistics Canada's Address Register identified by ARUID.

Inferring dwelling moves

Housing mobility is not about whether or not a household moves but rather under what circumstances and how often the household is moving. Inferring a dwelling move requires knowledge of the geolocation of an individual for the two years that are compared. A move cannot be identified if geolocation for either year is missing.

For illustrative purposes, Table 6 presents an interpretation of changes in the address marker⁴ for CHS respondents between the years 2012 and 2017, and the inferred mobility status.

Based on data limitations and mitigation strategies, information on CHS non-movers by design is more complete and more accurate than for movers. For example, for one in ten individuals, their 1 year mobility status cannot be successfully inferred, and for one in five their 5 year mobility status cannot be inferred. Both non-mover proportions are comparable to 2016 Census figures (see Table 7).

Conclusion

This new data will allow researchers across the country to better understand housing transitions, respondents' historical income trends and the incidence of chronic movers, and potentially help inform policies concerning the trajectories of households receiving support.

⁴ The address marker variable describes the reverse chronological order of different dwellings the person has lived in between 2008 and 2018. The dwelling respondent was reached at his dwelling number 1, the current residence. Code A corresponds to current residence (dwelling number 1), code B corresponds to prior residence (dwelling number 2), up to dwelling number 10 (code J) if the respondent has linked to 10 different addresses. There are three other reserve codes, 6 (Aged 15 and under for this period), 9 (dwelling not found for this period), Z (prior residence/dwelling number unknown).

Project Managers

George Ngoundjou Nkwinkem
Senior Specialist, Housing Research
Canada Mortgage and Housing Corporation

Mike Edwards
Senior Analyst, Housing Research
Canada Mortgage and Housing Corporation

Consultant

Statistics Canada

Appendix Tables

Table 1: Proportion of CHS respondents matched to tax records by adjusted age range

CHS respondents	All, adjusted age ⁵ ranges	0 to 14 years	15 to 64 years	65 years and over
All persons (weighted)	36,464,900	6,237,600	24,428,700	5,798,700
Percentage				
Linked to the tax data	84.1%	72.9%	85.7%	89.3%
T1FF matched (2017)	66.8%	0.7%	78.7%	87.8%
T4 matched (2017)	44.3%	0.5%	63.0%	13.0%

Source: Statistics Canada, 2020

Table 2: Proportion of CHS respondents aged 16 and over matched to 2008 and 2017 tax records by selected characteristics

Selected characteristics for persons age 16 and over (self-reported)	CHS roster Weighted number of persons	Percentage matched to tax records			
		2008		2017	
		T1FF	T4	T1FF	T4
All, gender	30,227,400	64.0	49.5	80.4	53.4
Male	14,902,700	62.6	50.7	79.1	55.3
Female	15,307,000	65.4	48.2	81.7	51.5
Other	17,600	46.7	52.5	78.0	61.8
All, region	30,227,400	64.0	49.5	80.4	53.4
Atlantic provinces	1,976,800	71.0	54.6	83.3	55.5
Québec	6,927,400	71.2	53.5	87.0	56.3
Ontario	11,858,700	61.3	46.5	78.1	50.8
Prairie provinces	5,343,200	60.6	50.4	77.8	55.0
British Columbia	4,016,000	61.2	47.5	78.3	52.5
Northern territories	105,500	54.4	51.8	68.8	60.4
All, tenure⁶	30,227,400	64.0	49.5	80.4	53.4
Owner	22,141,300	68.2	52.9	81.9	54.4
Renter not in SAH	7,115,100	51.8	41.7	76.0	53.6
Renter in SAH	970,900	57.8	26.7	79.3	30.0

Source: Statistics Canada, 2020

⁵ Adjusted age describes the age of the person in a tax reference year (e.g. CHS respondent who is 16 has an adjusted age of 15 in 2017).

⁶ Including renters living in Social and Affordable Housing (SAH).

Table 3: Proportion of CHS respondents with assigned geolocation by adjusted age range

CHS respondents	All, adjusted age ⁷ ranges	0 to 14 years	15 to 64 years	65 years and over
All persons (weighted)	36,464,900	6,237,600	24,428,700	5,798,700
Percentage				
Linked to the tax data	84.1%	72.9%	85.7%	89.3%
Dwelling sequence assigned (2018)	77.6%	-	93.3%	94.6%
DBUID available (2018)	61.9%	-	73.3%	80.4%
Dwelling sequence assigned (2008)	55.6%	-	61.3%	91.4%
DBUID available (2008)	47.5%	-	52.8%	76.6%

Source: Statistics Canada, 2020

Table 4: Proportion of CHS respondents aged 16 and over with assigned DBUID in 2018 by selected characteristics

Selected characteristics for persons age 16 and over (self-reported)	CHS roster Weighted number of persons	Address assigned in 2018	DBUID available in 2018
		Percentage of weighted persons	
All, region	30,227,400	93.6%	74.6%
Atlantic provinces	1,976,800	94.6%	73.9%
Québec	6,927,400	94.9%	85.3%
Ontario	11,858,700	93.1%	75.4%
Prairie provinces	5,343,200	93.5%	65.4%
British Columbia	4,016,000	92.0%	67.5%
Northern territories ⁸	105,500	95.5%	42.1%
All, tenure⁹	30,227,400	93.6%	74.6%
Owner	22,141,300	94.0%	75.6%
Renter not in SAH	7,115,100	92.3%	71.6%
Renter in SAH	970,900	93.3%	75.6%

Source: Statistics Canada, 2020

⁷ Adjusted age describes the age of the person in a tax reference year (e.g. CHS respondent who is 16 has an adjusted age of 15 in 2017).

⁸ This refers to either mailing ARUID or resident ARUID. Resident ARUID is not known for respondents in the Northwest Territories since the data were collected on behalf of CHS by the NWT Community Survey.

⁹ Including renters living in Social and Affordable Housing (SAH).

Table 5: Comparing 2018 tax derived ARUID to CHS survey frame ARUID

	Unweighted frequency	Weighted frequency	Column % weighted
All persons	146.2	36,464,900	100.0
Person aged under 16	26.1	6,237,600	17.1
Person aged 16 and over with missing ARUID	41.1	8,505,700	23.3
Mailing ARUID is the same as resident ARUID	71.3	19,634,400	53.8
Mailing ARUID is not the same as resident ARUID	7.7	2,087,300	5.7
CHS respondent moved within the last year	1.6	365,700	1.0
CHS respondent did not move within the last year	6.1	1,721,600	4.7

Source: Statistics Canada, 2020

Table 6: Mobility status, 1 year and 5 year, derived using historical address

Mobility Status	2012 address marker	2016 address marker	2017 address marker
1 year mover, 5 year mover	B	B	A
1 year mover, 5 year non-mover	A	B	A
1 year unknown, 5 year mover	B	9	A
1 year non-mover, 5 year unknown	6	A	A
1 year unknown, 5 year unknown	B	B	9
1 year non-mover¹⁰, 5 year non-mover	6	6	6

Source: Statistics Canada, 2020

Table 7: Inferring dwelling moves using historical addresses

Mobility Status	Census (2016)	Historical Address (2017)
1 year mover	13.0	6.3
1 year non-mover	87.0	85.3
1 year unknown	-	8.4
5 year mover	38.2	22.3
5 year non-mover	62.8	61.1
5 year unknown	-	16.4

Source: Statistics Canada, 2020

¹⁰ For illustrative purposes, persons 15 and under are grouped with non-movers.

CMHC helps Canadians meet their housing needs

Canada Mortgage and Housing Corporation (CMHC) has been helping Canadians meet their housing needs for more than 70 years. As Canada's authority on housing, we contribute to the stability of the housing market and financial system, provide support for Canadians in housing need, and offer unbiased housing research and advice to Canadian governments, consumers and the housing industry. Prudent risk management, strong corporate governance and transparency are cornerstones of our operations.

For more information, visit our website cmhc.ca or follow us on [Twitter](#), [LinkedIn](#), [Facebook](#), [Instagram](#) and [YouTube](#).

You can also reach us by phone at 1-800-668-2642 or by fax at **1-800-245-9274**.

Outside Canada call **613-748-2003** or fax to **613-748-2016**.

Canada Mortgage and Housing Corporation supports the Government of Canada policy on access to information for people with disabilities. If you wish to obtain this publication in alternative formats, call **1-800-668-2642**.

©2021, Canada Mortgage and Housing Corporation. All rights reserved. CMHC grants reasonable rights of use of this publication's content solely for personal, corporate or public policy research, and educational purposes. This permission consists of the right to use the content for general reference purposes in written analyses and in the reporting of results, conclusions, and forecasts including the citation of limited amounts of supporting data extracted from this publication. Reasonable and limited rights of use are also permitted in commercial publications subject to the above criteria, and CMHC's right to request that such use be discontinued for any reason.

Any use of the publication's content must include the source of the information, including statistical data, acknowledged as follows:

Source: CMHC (or "Adapted from CMHC," if appropriate), name of product, year and date of publication issue.

Other than as outlined above, the content of the publication cannot be reproduced or transmitted to any person or, if acquired by an organization, to users outside the organization. Placing the publication, in whole or part, on a website accessible to the public or on any website accessible to persons not directly employed by the organization is not permitted. To use the content of this CMHC publication for any purpose other than the general reference purposes set out above or to request permission to reproduce large portions of, or the entire content of, this CMHC publication, please send a Copyright request to the Housing Knowledge Centre at Housing_Knowledge_Centre@cmhc.ca. Please provide the following information: Publication's name, year and date of issue.

Without limiting the generality of the foregoing, no portion of the content may be translated from English or French into any other language without the prior written permission of Canada Mortgage and Housing Corporation.

The information, analyses and opinions contained in this publication are based on various sources believed to be reliable, but their accuracy cannot be guaranteed. The information, analyses and opinions shall not be taken as representations for which Canada Mortgage and Housing Corporation or any of its employees shall incur responsibility.