Coverage Technical Report

Census of Population, 2021

Release date: October 23, 2024

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 infostats@statcan.gc.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

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.
Cette publication est aussi disponible en français.

Release date: October 23, 2024
Catalogue number 98-303-X, issue 2021001
ISBN 978-0-660-47833-3

Table of contents

Introduction

The 2021 Census required the participation of the entire population of Canada, roughly 37 million persons over a territory of 9 million square kilometres. Although data collection and processing have to meet rigorous quality standards, it is impossible to eliminate all errors. To use census data correctly and appropriately, it is important to understand the conceptual framework and the definitions used to conduct the census, and the data collection and processing methods. Users also need to know the main sources of error and, where possible, the size of the errors and any unusual circumstances that might limit the usefulness or interpretation of census data. With this information, users can assess the risks associated with using census data to draw conclusions or make decisions.

This technical report examines coverage errors in the 2021 Census. There are two types of coverage errors. The first, called **population undercoverage**, refers to excluding persons who should have been enumerated. The second, called **population overcoverage**, refers to either including persons who should not have been enumerated or enumerating persons more than once. In the first instance of overcoverage, the overall error is considered negligible. Undercoverage is generally more common than overcoverage. The net impact of undercoverage and overcoverage on the size of a population of interest is **population net undercoverage**. Net undercoverage is the number of persons excluded who should have been enumerated (undercoverage) less the number of extra enumerations of persons who were enumerated more than once (overcoverage). Coverage errors are among the most significant types of errors, since they affect the accuracy not only of the counts for the various census universes, but also of all the census data that describe the characteristics of these universes.

Census coverage errors are measured using three studies. The first, the Dwelling Classification Survey (DCS), examines coverage errors resulting from errors in classifying dwellings. Census data are adjusted for this type of coverage error. The second, the Census Undercoverage Study (CUS) formerly called the Reverse Record Check (RRC), measures population undercoverage, while the third, the Census Overcoverage Study (COS), measures population overcoverage. Census data are not adjusted for the coverage errors measured by the CUS and the COS. Rather, Statistics Canada uses net undercoverage estimates to produce population estimates. The 2021 Census studies are quite similar to the 2016 studies, with some changes and improvements.

Census data users should be aware that the presence of coverage errors in the 2021 Census means that census products are subject to incomplete enumeration or double-counting. Undercoverage, for example, is higher for young adult males. For 2021 Census coverage error estimates for various demographic and geographic levels and groupings, see Section 1.

Section 2 covers the 2021 Census conceptual framework and provides definitions of the population universe, the dwelling universe and the usual place of residence that the census aims to measure. Section 3 describes coverage errors, their cause, census practices that minimize them and the conceptual framework used to evaluate them. It also introduces census coverage studies. Section 4 and Section 5 describe the data collection and processing methodology used in the 2021 Census.

Section 6 presents the methodology used during the 2021 DCS and the results obtained. Coverage error rate estimates are produced only for the population universe. Section 7 and Section 8 describe the methodology used for and the results of the 2021 CUS and 2021 COS, respectively. Section 9 shows how the results of the CUS and the COS were combined with census data to produce population coverage error estimates and the associated standard errors.

Section 10 covers the results of evaluations performed for the CUS and the COS, and the evaluation of the error of closure. “Error of closure” refers to the difference between population estimates whose base population is the 2016 Census counts, adjusted for net undercoverage, and 2021 Census data, also adjusted for net undercoverage.

Coverage Technical Report

Statistics Canada has conducted census population coverage studies since the first CUS, which dates back to the 1961 Census.[1] Section 11 provides a chronological review of coverage errors from the 1971 Census to the 2021 Census.

Section 12 covers additional topics, including the concept of persons not enumerated, and participation by reserves and settlements in the census.

This report was prepared by Melanie Abeysundera, Karen Bruce, Emily Campling, Monique Daweside, Heather Farr, Vincent Martin, Michel Parenteau, Martin Provost, Hélène St-Jean, Martin St-Pierre and Zhao Zhao of the Statistical Integration Methods Division, and Julien Bérard-Chagnon of the Centre for Demography.

Many individuals made valuable comments on preliminary versions, which improved the content and readability of the final report.

For additional information on census concepts, variables and geography, please see the *Dictionary, Census of Population, 2021*, Statistics Canada Catalogue no. 98-301-X. For additional information about the census process, please see the reference materials, 2021 Census.

1. The first CUS was conducted in 1961, but there was no frame of persons missed in the previous census. The 1966 CUS used the results of the 1961 CUS to build the frame of persons missed by the 1961 Census.

1. Estimates of population coverage errors

1.1 Introduction

The census defines the population to be counted and the rules by which it is to be counted (see Section 3). Coverage errors occur when errors are made relative to these definitions and rules. The main sources of coverage errors include failure to include a dwelling (and, in turn, failure to include its residents) and respondent error (not including all persons who should be included or including persons who should not be included). This section presents estimates of 2021 Census population net undercoverage, undercoverage and overcoverage. Undercoverage and overcoverage may lead to bias in official counts and estimates because the characteristics of the persons not included may differ from those of persons who are included, and the characteristics of those counted more than once (duplicates) may differ from those of persons who were counted only once. Net undercoverage indicates the extent to which the number of enumerations included in census data is higher or lower than complete enumeration.

1.2 Net undercoverage

The 2021 Census population net undercoverage rate was estimated at 3.00%.[2] In other words, the difference between the number of persons who were not included in the census but were members of the census target population and the number of duplicates was estimated to be 3.00% of the census target population. The population undercoverage rate is estimated at 4.98% (1,897,876 persons), while the population overcoverage rate is estimated at 1.98% (755,635 persons). An undercoverage rate of 4.98% indicates that persons who were not included but who were part of the target population represent 4.98% of the census target population. An overcoverage rate of 1.98% indicates that duplicate enumerations represent 1.98% of the census target population.

The estimated undercoverage rate has increased compared with the 2016 Census, while the overcoverage rate has been relatively stable. As a result, the net undercoverage rate, which is the difference between undercoverage and overcoverage, increased in comparison with the 2016 Census.

Since the primary goal of census coverage studies is always to produce the best coverage estimates possible for the most recent census, their methodology for the 2021 Census, as described in Section 7 and Section 8, has been improved. However, these improvements limit the comparability of these estimates with the 2016 net undercoverage estimates and partly explain the difference observed. In addition, since the net undercoverage estimates are measured from samples, the margin of error associated with the 2016 and 2021 estimates may also partly explain the difference.

Table 1.2a
Estimated rates of population coverage error and standard errors for Canada, 2016 and 2021 censuses

	2016 Census		2021 Census	
	Estimated rate	Standard error	Estimated rate	Standard error
Coverage error	percent			
Undercoverage	4.32	0.11	4.98	0.09
Overcoverage	1.96	0.04	1.98	0.02
Net undercoverage	2.36	0.12	3.00	0.10

Sources: Statistics Canada, 2016 and 2021 census coverage studies.

2. Net undercoverage rates in this report could differ slightly from the rates published on September 27, 2023, because incompletely enumerated reserves and settlements are included. All coverage error estimates in this report exclude coverage error for this group.

This section presents estimates of net undercoverage based on the following geographic and demographic variables:

- province or territory of current residence at the time of the census
- age and gender
- marital status and gender
- mother tongue
- census metropolitan area (CMA) of the usual place of residence on Census Day.

Table 1.2b provides an estimate of net undercoverage, standard errors related to the estimate, and the corresponding estimated net undercoverage rate and standard error for various characteristics. Negative net undercoverage estimates indicate that the overcoverage rate was higher than the undercoverage rate. For an explanation of how this can occur, see Section 9.

Table 1.2b
Estimated population net undercoverage and standard errors for various characteristics, 2021 Census

	Population net undercoverage		Population net undercoverage rate	
Characteristics	**Estimated number**	**Standard error**	**Estimated rate (%)**	**Standard error (%)**
Canada	**1,142,241**	**38,241**	**3.00**	**0.10**
Province or territory				
Newfoundland and Labrador	16,235	1,829	3.08	0.34
Prince Edward Island	6,901	995	4.28	0.59
Nova Scotia	27,852	3,452	2.79	0.34
New Brunswick	13,624	2,731	1.73	0.34
Quebec	42,868	16,829	0.50	0.20
Ontario	585,370	28,636	3.95	0.19
Manitoba	39,284	5,145	2.84	0.36
Saskatchewan	34,912	4,464	2.99	0.37
Alberta	153,166	12,527	3.47	0.27
British Columbia	212,846	16,069	4.08	0.30
Yukon	2,467	174	5.78	0.38
Northwest Territories	3,589	228	8.04	0.47
Nunavut	3,128	250	7.82	0.58
Gender and age group				
Total gender	**1,142,241**	**38,241**	**3.00**	**0.10**
0 to 4 years	65,492	6,028	3.45	0.31
5 to 14 years	5,808	13,172	0.14	0.31
15 to 17 years	-2,276	8,846	-0.19	0.73
18 to 19 years	18,027	9,192	2.20	1.10
20 to 24 years	228,413	18,752	9.40	0.70
25 to 34 years	419,936	23,354	7.83	0.40
35 to 44 years	217,802	19,605	4.25	0.37
45 to 54 years	129,102	19,430	2.69	0.39
55 to 64 years	84,286	18,793	1.59	0.35
65 to 74 years	-10,341	13,361	-0.26	0.33
75 years and older	-14,008	12,096	-0.48	0.41

Table 1.2b
Estimated population net undercoverage and standard errors for various characteristics, 2021 Census

Characteristics	Population net undercoverage		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Men+[1]	**758,934**	**34,974**	**4.00**	**0.18**
0 to 4 years	30,137	5,776	3.11	0.58
5 to 14 years	7,324	9,109	0.34	0.42
15 to 17 years	-1,177	6,847	-0.19	1.10
18 to 19 years	9,043	6,155	2.15	1.43
20 to 24 years	118,670	13,585	9.44	0.98
25 to 34 years	272,115	19,190	9.88	0.63
35 to 44 years	158,860	16,321	6.17	0.60
45 to 54 years	85,976	15,141	3.62	0.61
55 to 64 years	76,278	14,916	2.90	0.55
65 to 74 years	8,802	10,430	0.45	0.53
75 years and older	-7,095	7,861	-0.56	0.62
Women+[2]	**383,307**	**29,525**	**2.00**	**0.15**
0 to 4 years	35,355	5,729	3.81	0.59
5 to 14 years	-1,516	9,625	-0.07	0.47
15 to 17 years	-1,099	5,878	-0.19	1.00
18 to 19 years	8,983	6,920	2.26	1.70
20 to 24 years	109,743	14,487	9.35	1.12
25 to 34 years	147,821	15,252	5.67	0.55
35 to 44 years	58,942	10,408	2.31	0.40
45 to 54 years	43,126	12,911	1.78	0.52
55 to 64 years	8,007	11,809	0.30	0.44
65 to 74 years	-19,143	9,139	-0.91	0.44
75 years and older	-6,913	9,066	-0.41	0.54
Marital status and gender for people aged 15 years and older				
Total gender	**1,070,941**	**37,309**	**3.34**	**0.11**
Married (not separated)	104,490	24,944	0.76	0.18
Common law	133,518	18,195	3.31	0.44
Single (never legally married)	654,481	32,073	6.76	0.31
Separated	95,166	12,460	11.38	1.32
Divorced	70,158	15,392	3.52	0.75
Widowed	13,128	12,122	0.78	0.72
Men+[1]	**721,473**	**33,952**	**4.55**	**0.20**
Married (not separated)	76,432	18,026	1.10	0.26
Common law	100,598	14,643	4.90	0.68
Single (never legally married)	421,561	25,951	7.95	0.45
Separated	65,379	10,756	17.01	2.32
Divorced	50,518	11,062	6.23	1.28
Widowed	6,984	7,661	1.88	2.02

Table 1.2b
Estimated population net undercoverage and standard errors for various characteristics, 2021 Census

Characteristics	Population net undercoverage		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Women+[2]	**349,468**	**28,186**	**2.16**	**0.17**
Married (not separated)	28,058	17,248	0.41	0.25
Common law	32,919	10,219	1.66	0.51
Single (never legally married)	232,920	22,326	5.32	0.48
Separated	29,787	6,308	6.59	1.30
Divorced	19,640	10,867	1.66	0.90
Widowed	6,144	9,498	0.47	0.72
Mother tongue				
Total	**1,142,241**	**38,241**	**3.00**	**0.10**
English	523,481	31,784	2.51	0.15
French	23,971	16,633	0.33	0.23
Non-official language	530,185	28,328	6.23	0.31
English and French	10,608	5,006	3.48	1.59
English and non-official language	41,970	9,168	4.34	0.91
French and non-official language	8,609	3,563	6.78	2.62
English, French and non-official language	3,416	1,854	6.05	3.08
Census metropolitan area				
St. John's	6,222	1,651	2.84	0.73
Halifax	9,726	2,694	2.05	0.56
Moncton	885	1,530	0.56	0.96
Saint John	2,078	1,573	1.57	1.17
Fredericton	1,362	1,380	1.24	1.24
Saguenay	3,388	3,444	2.05	2.05
Québec	-4,037	5,843	-0.48	0.70
Sherbrooke	-5,760	2,928	-2.60	1.36
Trois-Rivières	7,930	4,134	4.68	2.33
Drummondville	-949	1,647	-0.94	1.65
Montréal	21,241	13,439	0.49	0.31
Ottawa–Gatineau	40,654	12,023	2.66	0.77
Kingston	4,402	4,764	2.49	2.63
Belleville–Quinte West	1,686	3,373	1.49	2.94
Peterborough	5,257	3,743	3.93	2.69
Oshawa	13,117	6,958	3.06	1.57
Toronto	329,366	26,924	5.04	0.39
Hamilton	30,219	9,093	3.71	1.07
St. Catharines–Niagara	7,349	7,644	1.67	1.70
Kitchener–Cambridge–Waterloo	11,206	6,768	1.91	1.13
Brantford	5,233	3,765	3.50	2.43
Guelph	-1,191	2,771	-0.72	1.70
London	14,807	7,891	2.65	1.38

Table 1.2b
Estimated population net undercoverage and standard errors for various characteristics, 2021 Census

Characteristics	Population net undercoverage		Population net undercoverage rate	
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Windsor	25,248	7,020	5.64	1.48
Barrie	6,320	4,883	2.88	2.16
Greater Sudbury	5,306	4,455	3.02	2.46
Thunder Bay	4,407	4,787	3.45	3.62
Winnipeg	17,674	4,430	2.07	0.51
Regina	3,929	2,620	1.55	1.02
Saskatoon	10,197	3,286	3.11	0.97
Lethbridge	2,127	2,612	1.69	2.04
Calgary	56,836	8,740	3.69	0.55
Red Deer	4,515	3,642	4.29	3.31
Edmonton	35,376	8,761	2.43	0.59
Kelowna	9,470	4,694	4.09	1.94
Kamloops	8,195	3,763	6.70	2.87
Chilliwack	1,076	2,534	0.94	2.19
Abbotsford–Mission	1,237	3,183	0.63	1.61
Vancouver	101,974	14,437	3.72	0.51
Victoria	20,508	6,768	4.91	1.54
Nanaimo	-966	2,172	-0.84	1.91
All census metropolitan areas	817,622	37,380	2.91	0.13
Outside a census metropolitan area	324,618	26,293	3.23	0.25

1. The term “men+” includes men (and boys) and some non-binary people.
2. The term “women+” includes women (and girls) and some non-binary people.
Note: Coverage estimates may not necessarily add up to the totals because of rounding.
Sources: Statistics Canada, 2021 Census and 2021 Census coverage studies.

The standard error provides an indication of the accuracy of sampling-based estimates. An interval covering two standard errors on both sides of the estimate includes the correct value 19 times out of 20. In other words, there are approximately 19 chances out of 20 (95%) that the actual population net undercoverage rate for the 2021 Census was between 2.8% and 3.2% (i.e., 3.0% ± two standard errors) or about 2 chances out of 3 (68%) that the actual rate was between 2.9% and 3.1% (i.e., 3.0% ± one standard error).

Since net undercoverage reflects both undercoverage and overcoverage, readers should also consult the undercoverage and overcoverage estimates in Table 1.3. For example, a low net undercoverage rate may reflect a low undercoverage rate, or a high undercoverage rate combined with a high overcoverage rate.

The highest population net undercoverage rates in the country were observed in the three territories: the Northwest Territories and Nunavut had the highest rates (8.04% and 7.82%, respectively), followed by Yukon (5.78%). Among the provinces, Prince Edward Island had the highest net undercoverage rate, at 4.28%, followed by British Columbia (4.08%) and Ontario (3.95%). Quebec had the lowest population net undercoverage rate, at 0.50%, followed by New Brunswick (1.73%). All other provinces had a net undercoverage rate that was at least one percentage point higher than that of New Brunswick. In 2016, the highest and lowest rates were also observed in the Northwest Territories and Quebec.

Beginning with the 2021 Census, the estimates of census coverage error are reported by gender rather than by sex. Gender refers to an individual's personal and social identity as a man, woman or non-binary person (a person who is not exclusively a man or a woman). Given that the non-binary population is small, data aggregation to a two-category gender variable was used to present estimates in this technical report. Individuals in the category "non-binary persons" were distributed into the other two gender categories and are denoted by the "+" symbol. Throughout the report, "men+" designates men (and boys), as well as some non-binary persons, while "women+" includes women (and girls), as well as some non-binary persons.

Population net undercoverage was generally higher for men+ and highest for young adults. In the general population, net undercoverage was highest in the 20-to-34 age group for both men+ and women+. It stood at 9.40% among persons aged 20 to 24, and 7.83% among those aged 25 to 34. For men+, the highest net undercoverage rates were 9.88% among men aged 25 to 34 and 9.44% among men aged 20 to 24. For women+, these rates reached 9.35% among those aged 20 to 24 and 5.67% among those aged 25 to 34. For women+ aged 65 and older and men+ aged 75 and older, the net undercoverage rate was negative, indicating more excess enumerations than persons who were not enumerated.

The net undercoverage rate of the population aged 15 and older was high for separated persons (11.38%), especially for men+ (17.01%). It was also high for single persons (6.76%), i.e., persons who had never legally married and who were not in a common-law union.

The net undercoverage rate for those whose mother tongue is English only was higher than for those whose mother tongue is French only (2.51%, compared with 0.33%). Excluding persons who declared more than one mother tongue, the net undercoverage rate for allophones, persons whose mother tongue is neither English nor French, was higher (6.23%).

Net population undercoverage was slightly more frequent outside CMAs than within them. Nationally, net undercoverage was 3.23% for those who should have been enumerated outside a CMA. This was slightly higher than the net undercoverage rate for those living in a CMA (2.91%).

1.3 Undercoverage

Undercoverage generally refers to persons who were not included as usual residents in the questionnaire that was completed for their usual residence, or persons for whom no questionnaire was completed for their usual residence. For example, persons who regarded their residence as temporary may not have been included as usual residents elsewhere. Persons without a usual place of residence (e.g., the homeless) were also much more subject to undercoverage.

This section presents undercoverage estimates based on the following geographic and demographic variables:

- province or territory of current residence at the time of the census
- age and gender
- marital status and gender
- mother tongue
- CMA of the usual place of residence on Census Day.

Table 1.3 provides the estimated undercoverage regarding the number of persons not included, the standard error of the estimate, the corresponding estimated undercoverage rate and its standard error.

Table 1.3
Estimated population undercoverage and overcoverage and standard errors for various characteristics, 2021 Census

Characteristics	Population undercoverage				Population overcoverage			
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Canada	**1,897,876**	**37,004**	**4.98**	**0.09**	**755,635**	**9,648**	**1.98**	**0.02**
Province or territory								
Newfoundland and Labrador	26,428	1,775	5.02	0.32	10,194	439	1.94	0.08
Prince Edward Island	10,190	977	6.32	0.57	3,289	191	2.04	0.12
Nova Scotia	47,196	3,373	4.73	0.32	19,344	736	1.94	0.07
New Brunswick	30,165	2,655	3.82	0.32	16,541	641	2.10	0.08
Quebec	225,381	15,756	2.64	0.18	182,513	5,915	2.14	0.07
Ontario	855,470	27,795	5.78	0.18	270,100	6,888	1.82	0.05
Manitoba	62,443	5,089	4.52	0.35	23,160	757	1.68	0.05
Saskatchewan	59,075	4,411	5.06	0.36	24,163	689	2.07	0.06
Alberta	231,249	12,225	5.24	0.26	78,084	2,736	1.77	0.06
British Columbia	338,678	15,827	6.50	0.28	125,832	2,778	2.41	0.05
Yukon	3,294	170	7.71	0.37	827	38	1.94	0.09
Northwest Territories	4,426	227	9.91	0.46	837	15	1.87	0.03
Nunavut	3,879	250	9.70	0.56	751	17	1.88	0.04
Gender and age group								
Total gender	**1,897,876**	**37,004**	**4.98**	**0.09**	**755,635**	**9,648**	**1.98**	**0.02**
0 to 4 years	98,438	5,809	5.19	0.29	32,946	1,610	1.74	0.08
5 to 14 years	140,303	12,630	3.35	0.29	134,496	3,737	3.21	0.09
15 to 17 years	41,091	8,516	3.40	0.68	43,367	2,393	3.58	0.19
18 to 19 years	48,426	8,933	5.91	1.03	30,399	2,168	3.71	0.26
20 to 24 years	313,912	18,381	12.91	0.66	85,499	3,713	3.52	0.15
25 to 34 years	530,315	22,964	9.89	0.39	110,379	4,251	2.06	0.08
35 to 44 years	281,659	19,332	5.49	0.36	63,857	3,261	1.25	0.06
45 to 54 years	192,317	19,130	4.01	0.38	63,215	3,398	1.32	0.07
55 to 64 years	164,211	18,459	3.10	0.34	79,926	3,530	1.51	0.07
65 to 74 years	49,955	13,112	1.23	0.32	60,296	2,567	1.49	0.06
75 years and older	37,248	11,803	1.26	0.40	51,256	2,646	1.74	0.09
Men+[1]	**1,140,736**	**34,194**	**6.01**	**0.17**	**381,802**	**7,344**	**2.01**	**0.04**
0 to 4 years	46,527	5,668	4.80	0.56	16,390	1,113	1.69	0.11
5 to 14 years	76,859	8,684	3.57	0.39	69,535	2,750	3.23	0.12
15 to 17 years	22,004	6,595	3.53	1.02	23,180	1,840	3.72	0.29
18 to 19 years	23,968	5,962	5.70	1.34	14,925	1,530	3.55	0.35
20 to 24 years	163,089	13,315	12.97	0.92	44,420	2,698	3.53	0.21
25 to 34 years	330,497	18,903	12.00	0.60	58,381	3,306	2.12	0.12
35 to 44 years	192,459	16,057	7.48	0.58	33,599	2,925	1.31	0.11

Table 1.3
Estimated population undercoverage and overcoverage and standard errors for various characteristics, 2021 Census

	Population undercoverage				Population overcoverage			
Characteristics	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
45 to 54 years	116,139	15,044	4.89	0.60	30,164	1,710	1.27	0.07
55 to 64 years	116,176	14,720	4.42	0.54	39,897	2,407	1.52	0.09
65 to 74 years	37,494	10,291	1.92	0.52	28,692	1,699	1.47	0.09
75 years and older	15,524	7,677	1.22	0.60	22,619	1,694	1.78	0.13
Women+[2]	**757,140**	**28,656**	**3.95**	**0.14**	**373,833**	**7,113**	**1.95**	**0.04**
0 to 4 years	51,911	5,609	5.60	0.57	16,556	1,164	1.78	0.12
5 to 14 years	63,444	9,289	3.12	0.44	64,961	2,524	3.20	0.12
15 to 17 years	19,088	5,677	3.26	0.94	20,186	1,527	3.44	0.25
18 to 19 years	24,458	6,747	6.14	1.59	15,475	1,539	3.89	0.38
20 to 24 years	150,822	14,260	12.86	1.06	41,079	2,554	3.50	0.21
25 to 34 years	199,818	15,010	7.66	0.53	51,998	2,705	1.99	0.10
35 to 44 years	89,199	10,306	3.49	0.39	30,257	1,448	1.18	0.06
45 to 54 years	76,178	12,572	3.14	0.50	33,052	2,938	1.36	0.12
55 to 64 years	48,036	11,524	1.80	0.42	40,029	2,577	1.50	0.10
65 to 74 years	12,462	8,934	0.60	0.42	31,605	1,928	1.51	0.09
75 years and older	21,724	8,843	1.30	0.52	28,637	1,997	1.71	0.12
Marital status and gender for people aged 15 years and older								
Total gender	**1,659,135**	**36,154**	**5.18**	**0.11**	**588,194**	**9,210**	**1.84**	**0.03**
Married (not separated)	296,693	24,616	2.15	0.17	192,203	4,032	1.39	0.03
Common law	194,311	18,002	4.82	0.42	60,793	2,640	1.51	0.06
Single (never legally married)	917,850	31,336	9.48	0.29	263,369	6,836	2.72	0.07
Separated	106,837	12,428	12.77	1.30	11,671	897	1.40	0.11
Divorced	100,367	15,200	5.04	0.72	30,209	2,425	1.52	0.12
Widowed	43,077	11,956	2.57	0.69	29,949	2,002	1.79	0.12
Men+[1]	**1,017,350**	**33,273**	**6.41**	**0.20**	**295,877**	**6,754**	**1.87**	**0.04**
Married (not separated)	174,151	17,794	2.51	0.25	97,718	2,884	1.41	0.04
Common law	130,893	14,521	6.37	0.66	30,295	1,884	1.47	0.09
Single (never legally married)	565,335	25,393	10.66	0.43	143,774	5,350	2.71	0.10
Separated	70,567	10,737	18.36	2.28	5,188	638	1.35	0.17
Divorced	61,223	11,009	7.55	1.26	10,704	1,075	1.32	0.13
Widowed	15,182	7,576	4.08	1.95	8,198	1,137	2.20	0.30
Women+[2]	**641,785**	**27,411**	**3.96**	**0.16**	**292,317**	**6,562**	**1.81**	**0.04**
Married (not separated)	122,543	17,010	1.78	0.24	94,485	2,855	1.37	0.04
Common law	63,417	10,048	3.20	0.49	30,498	1,860	1.54	0.09
Single (never legally married)	352,515	21,877	8.05	0.46	119,595	4,455	2.73	0.10
Separated	36,270	6,277	8.03	1.28	6,483	631	1.43	0.14
Divorced	39,145	10,647	3.31	0.87	19,505	2,175	1.65	0.18

Table 1.3
Estimated population undercoverage and overcoverage and standard errors for various characteristics, 2021 Census

Characteristics	Population undercoverage				Population overcoverage			
	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Widowed	27,895	9,357	2.14	0.70	21,751	1,627	1.67	0.12
Mother tongue								
Total	**1,897,876**	**37,004**	**4.98**	**0.09**	**755,635**	**9,648**	**1.98**	**0.02**
English	935,492	31,088	4.48	0.14	412,011	6,619	1.97	0.03
French	181,362	15,731	2.48	0.21	157,391	5,402	2.15	0.07
Non-official language	688,896	27,968	8.10	0.30	158,711	4,501	1.87	0.05
English and French	18,386	4,933	6.04	1.52	7,778	856	2.55	0.28
English and non-official language	58,806	9,125	6.08	0.89	16,836	894	1.74	0.09
French and non-official language	10,535	3,550	8.30	2.57	1,925	302	1.52	0.24
English, French and non-official language	4,400	1,843	7.79	3.01	983	201	1.74	0.35
Census metropolitan area								
St. John's	11,044	1,618	5.05	0.70	4,822	329	2.20	0.15
Halifax	18,113	2,639	3.81	0.53	8,387	541	1.76	0.11
Moncton	3,694	1,512	2.33	0.93	2,809	229	1.77	0.14
Saint John	4,595	1,560	3.46	1.13	2,516	198	1.90	0.15
Fredericton	3,813	1,350	3.47	1.19	2,452	284	2.23	0.25
Saguenay	6,799	3,353	4.12	1.95	3,411	790	2.07	0.47
Québec	15,048	5,500	1.80	0.65	19,084	1,970	2.28	0.23
Sherbrooke	1,031	2,718	0.47	1.22	6,791	1,088	3.06	0.48
Trois-Rivières	11,233	4,067	6.63	2.24	3,304	741	1.95	0.43
Drummondville	1,096	1,539	1.09	1.51	2,044	587	2.03	0.57
Montréal	108,911	12,869	2.53	0.29	87,670	3,872	2.03	0.09
Ottawa–Gatineau	67,713	11,882	4.43	0.74	27,060	1,836	1.77	0.12
Kingston	7,590	4,724	4.29	2.56	3,188	611	1.80	0.34
Belleville–Quinte West	3,807	3,324	3.37	2.85	2,120	573	1.88	0.50
Peterborough	7,555	3,718	5.64	2.62	2,298	435	1.72	0.32
Oshawa	20,929	6,892	4.89	1.53	7,812	956	1.82	0.22
Toronto	455,717	26,436	6.98	0.38	126,350	5,102	1.93	0.08
Hamilton	42,294	9,009	5.19	1.05	12,075	1,234	1.48	0.15
St. Catharines–Niagara	15,168	7,573	3.44	1.66	7,819	1,043	1.77	0.23
Kitchener–Cambridge–Waterloo	20,172	6,693	3.44	1.10	8,966	1,006	1.53	0.17
Brantford	7,767	3,667	5.20	2.33	2,534	856	1.70	0.57
Guelph	2,188	2,708	1.33	1.63	3,378	586	2.05	0.35
London	24,481	7,798	4.38	1.34	9,674	1,209	1.73	0.21
Windsor	32,144	6,956	7.18	1.44	6,896	943	1.54	0.21
Barrie	10,079	4,847	4.60	2.11	3,759	597	1.72	0.27
Greater Sudbury	9,323	4,398	5.30	2.37	4,017	711	2.28	0.40

Table 1.3
Estimated population undercoverage and overcoverage and standard errors for various characteristics, 2021 Census

	Population undercoverage				Population overcoverage			
Characteristics	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Thunder Bay	7,022	4,755	5.50	3.52	2,615	553	2.05	0.43
Winnipeg	29,805	4,401	3.50	0.50	12,131	510	1.42	0.06
Regina	9,078	2,593	3.59	0.99	5,150	378	2.03	0.15
Saskatoon	16,775	3,266	5.12	0.95	6,578	357	2.01	0.11
Lethbridge	4,581	2,584	3.64	1.98	2,454	383	1.95	0.30
Calgary	83,937	8,636	5.46	0.53	27,101	1,344	1.76	0.09
Red Deer	7,107	2,650	6.75	2.35	2,592	2,498	2.46	2.31
Edmonton	59,952	8,696	4.12	0.57	24,575	1,064	1.69	0.07
Kelowna	16,229	4,625	7.01	1.86	6,760	802	2.92	0.34
Kamloops	10,217	3,752	8.35	2.81	2,022	293	1.65	0.24
Chilliwack	3,603	2,502	3.14	2.11	2,527	404	2.20	0.35
Abbotsford–Mission	6,273	3,148	3.18	1.55	5,036	472	2.56	0.24
Vancouver	170,741	14,302	6.22	0.49	68,767	1,973	2.51	0.07
Victoria	31,509	6,711	7.54	1.49	11,001	876	2.63	0.21
Nanaimo	2,916	2,053	2.55	1.75	3,881	707	3.39	0.60
All census metropolitan areas	1,372,046	36,333	4.88	0.12	554,424	8,787	1.97	0.03
Outside a census metropolitan area	525,830	25,907	5.24	0.24	201,212	4,492	2.01	0.04

1. The term “men+” includes men (and boys) and some non-binary people.
2. The term “women+” includes women (and girls) and some non-binary people.
Note: Coverage estimates may not necessarily add up to the totals because of rounding.
Sources: Statistics Canada, 2021 Census and 2021 Census coverage studies.

There were some demographic trends in undercoverage. The rate of undercoverage for men+ was one-and-a-half times the rate for women+, or 6.01% compared with 3.95%. The undercoverage rate was highest for young adults aged 18 to 34 (men+ and women+). Among young men+, undercoverage was 12.97% for those aged 20 to 24 and 12.00% for those aged 25 to 34, but it was higher for those aged 35 to 44 than for those aged 18 to 19, contrary to what was typically observed in the past.

In terms of marital status, undercoverage was highest among those aged 15 or older who were separated and not in a common-law union, at 12.77%. It was also high for single persons not in a common-law union, at 9.48%. In both cases, these rates were higher for men+ than for women+.

Table 1.3 shows that the persons most likely to have been missed in the 2021 Census were men+ aged 20 to 34 who were single (who were never married and not in a common-law union), and separated persons. Mother tongue was also significant for undercoverage rates because for persons who declared only one mother tongue, undercoverage rates were lower among those whose first language was French (2.48%), followed by those whose first language was English (4.48%). For persons whose only mother tongue was neither English nor French, the rate was even higher, at 8.10%.

1.4 Overcoverage

Population overcoverage is the number of excess enumerations in the census counts for persons enumerated more than once (usually twice). This error produces bias because these persons should have been enumerated only once. Examples of overcoverage include children whose parents live separately who were included on each parent's census questionnaire, persons who do not live with their family for work reasons who are listed on their family's form and also on the form for the dwelling they live in while working, and students away at school who are listed both by their roommates and by their parents.

This section presents estimates of overcoverage based on the following geographic and demographic variables:

- province or territory of current residence at the time of the census
- age and gender
- marital status and gender
- mother tongue
- CMA of the usual place of residence on Census Day.

Table 1.3 also contains estimates of the number and rate of excess enumerations and the associated standard error.

The standard errors for overcoverage estimates found in Table 1.3 were lower than those for undercoverage. As indicated below, certain demographic trends emerged from overcoverage estimates.

Across the provinces and territories, the estimated overcoverage rate varied less than the estimated undercoverage rate. In fact, the gap between the lowest rate and the highest rate among all provinces and territories was 0.73 percentage points for overcoverage (with the rate varying from 1.68% to 2.41%), whereas this gap was 7.27 percentage points for undercoverage (with the rate varying from 2.64% to 9.91%). Men+ and women+ had similar overcoverage rates, at 2.01% compared with 1.95%. Overcoverage was higher for children and young adults aged 5 to 24, with rates varying from 3.21% to 3.71% for these age groups. As mentioned, these high rates were largely attributable to multiple enumerations of children in shared custody and students.

For marital status, the overcoverage rate was higher for persons who had never been married and who were not in a common-law union (2.72%). This phenomenon was observed equally among women+ and men+, and it was consistent with the results by age.

In summary, Table 1.3 presents the profile of persons most likely to be counted more than once; they were equally likely to be men+ or women+, and more likely to be a child or a young adult. For persons aged 15 and older, single persons were more likely to be counted more than once.

2. Census universe

2.1 Introduction

While the 2021 Census collects information on the population, dwellings, households and families, the 2021 Census Coverage Error Measurement Program estimates the coverage error of the population universe only. However, the definitions of dwelling concepts and the rules for determining the list of persons who should be enumerated in each dwelling affect coverage of the census target population. As a result, this section describes the concepts of population and dwelling. In addition, since coverage error can be caused by misinterpreting the concept of usual place of residence as defined in census questionnaires, this section also provides the information in the census questionnaires and the 2021 Census definition of usual place of residence.

2.2 Population universe

The 2021 Census target population includes the following groups:

- Canadian citizens and landed immigrants (permanent residents) with a usual place of residence in Canada
- Canadian citizens and landed immigrants (permanent residents) who are abroad, either on a military base or attached to a diplomatic mission
- Canadian citizens and landed immigrants (permanent residents) at sea or in port aboard merchant vessels under Canadian registry or Canadian government vessels
- non-permanent residents with a usual place of residence in Canada
    - who are claiming refugee status (asylum seekers)
    - who hold a study permit (covering Census Day)
    - who hold a work permit (covering Census Day)

  and family members living with them.

The 2021 Census population universe does not include foreign residents, but since 1991, it has included non-permanent residents.

The definition of target population specifies which persons should be included in the census, but not where these persons should be enumerated. The Canadian census uses the modified *de jure* method of enumeration, under which persons are to be enumerated at their usual place of residence, even if they are temporarily away on Census Day. Persons away from their usual place of residence and residing elsewhere in Canada must be enumerated at their usual place of residence and are considered present, but temporarily at the other location. Persons who have no usual place of residence are to be enumerated wherever they happen to be on Census Day. Some countries use the *de facto* method, under which all persons are to be enumerated wherever they are on Census Day, regardless of their usual place of residence.

2.3 Dwelling universe

A dwelling is defined as a set of living quarters. Two types of dwellings are identified in the census: collective dwellings and private dwellings. Census coverage studies include these two types of dwellings, without distinction.

Private dwelling refers to a separate set of living quarters with a private entrance either from outside the building or from a common hall, lobby, vestibule or stairway inside the building. The entrance to the dwelling must be one that can be used without passing through the living quarters of some other person or group of persons.

The dwelling must meet the two conditions necessary for year-round occupancy:

1. a source of heat or power (as evidenced by chimneys, power lines, oil or gas pipes or meters, generators, woodpiles, electric lights, heating pumps, or solar panels)
2. an enclosed space that provides shelter from the elements as evidenced by complete and enclosed walls and a roof, and by doors and windows that provide protection from wind, rain and snow.

Dwellings that do not meet the conditions necessary for year-round occupancy are marginal dwellings. Private dwellings are classified into regular private dwellings and occupied marginal dwellings. Regular private dwellings are further classified into three major groups: occupied dwellings (occupied by usual residents), dwellings occupied solely by foreign residents or by temporarily present persons, and unoccupied dwellings. Marginal dwellings are classified as occupied by usual residents or occupied solely by foreign residents or by temporarily present persons. Marginal dwellings that were unoccupied on May 11, 2021, are not counted in the housing stock.

A collective dwelling refers to a dwelling of a commercial, institutional or communal nature in which a person or group of persons reside or could reside. It must provide care or services or have certain common facilities, such as a kitchen or bathroom, which are shared by the occupants. Examples include lodging or rooming houses, hotels, motels, tourist establishments, nursing homes, residences for senior citizens, hospitals, staff residences, military bases, work camps, correctional facilities and group homes.

Collective dwellings are classified as either occupied dwellings or unoccupied dwellings. Occupied dwellings are either occupied by usual residents or occupied solely by foreign residents or by temporarily present persons. In the case of unoccupied collective dwellings, data on the dwelling, such as the types of services offered, were collected but are not included in census products.

In summary, the dwelling universe includes the following:

- private dwellings occupied by usual residents
- private dwellings occupied solely by foreign residents or by temporarily present persons
- unoccupied private dwellings
- marginal dwellings occupied on Census Day
- collective dwellings occupied by usual residents
- collective dwellings occupied solely by foreign residents or by temporarily present persons.

The dwelling universe does not include the following:

- marginal dwellings that were unoccupied on Census Day
- collective dwellings that were unoccupied on Census Day
- dwellings outside Canada.

2.4 Usual place of residence

Under the *de jure* enumeration method used in the Canadian census, the population is enumerated on a "usual place of residence" basis, that is, at the location where a person lives most of the time. Most individuals have only one residence, and it is easy to enumerate them at their usual place of residence. Enumeration involves listing all the persons with a certain dwelling as their usual place of residence on Census Day by following the step-by-step instructions at the beginning of the census questionnaire: "**Including yourself**", how many persons usually live at this address on May 11, 2021? **Include:** all persons who have their main residence at this address, even if they are temporarily away. **See the instructions on page 3** (joint custody, students, landed immigrants, secondary residence, etc.)." The instructions on page 3 of the 2021 Census questionnaire are presented in Appendix A.

In some cases, it is difficult to determine a person's usual place of residence. That is why special rules were developed for determining the usual place of residence in some cases:

1. Persons with more than one residence

   This category includes all persons who have more than one dwelling in Canada that could be considered their usual place of residence. In this situation, the usual place of residence is the place where a person spends the majority of the year. If the person spends the same amount of time at both residences or is not sure which one to choose, they should choose the residence where they stayed overnight between May 10 and 11, 2021. There are two exceptions to this rule:

   a. Children who live somewhere else while attending school or working at a summer job but return to live with their parents for part of the year should consider the residence they share with their parents to be their usual place of residence, even if they spend most of the year elsewhere.

   b. Spouses or common-law partners who live away from their families while working or studying but return to their families periodically should consider the residence they share with their spouse to be their usual place of residence, even if they spend most of the year elsewhere.

2. Persons in an institution, such as a hospital, home for the aged, prison or correctional institution

   Persons who have been in one or more institutions for a continuous period of six months or longer at the time of the census are to be considered usual residents of the institution.

3. Persons with no usual place of residence

   Persons who do not have a usual place of residence should be enumerated in the dwelling where they stayed overnight between May 10 and 11, 2021.

4. Persons residing outside Canada

   Canadian citizens and landed immigrants (permanent residents) residing outside Canada on Census Day include:

   - persons aboard Canadian vessels or merchant vessels
   - Canadian federal, provincial and territorial government employees and their family members
   - members of the Canadian Armed Forces and their family members.

   These persons should indicate in the census questionnaire the address they use for election purposes or their last permanent address in Canada if they are not already included in the residence of their families. This information is used to assign them a geographic location in Canada for dissemination purposes.

3. Population coverage error

3.1 Sources

Although census data collection and processing have to meet high quality standards, it is very difficult to eliminate all potential errors. There are two kinds of population coverage error. Population undercoverage refers to the exclusion of persons who should have been enumerated, and population overcoverage refers to the inclusion of persons who were enumerated more than once (generally twice). Overcoverage also includes persons who were enumerated but should not have been. However, this type of error is considered negligible; consequently, it is not measured.

Undercoverage can occur in the first stage of the census if the list of dwellings used for the dwelling universe is incomplete. This risk is higher, for example, if a dwelling is under construction. Conversely, overcoverage can occur if a dwelling is listed twice.

Coverage error can also occur during the field data collection stage. Respondent error is responsible for coverage error when the person completing the census form omits someone whose usual place of residence, according to census rules, is the dwelling concerned; this is undercoverage. The person may also include someone whose usual place of residence is not the dwelling concerned; there is overcoverage if this person has already been enumerated at their usual place of residence or somewhere else. In most cases, it is easy to determine a person's usual place of residence. However, as stated in the previous section, the process is sometimes more complex, and special rules have been developed for determining an individual's usual place of residence. The rules are spelled out in the census questionnaire, but the list is long, and there can be comprehension difficulties. Coverage error may result when the rules are not consulted or are incorrectly applied. The idea of using Census Day as the reference date for determining usual residence may also be misunderstood, and this can lead to coverage error.

Coverage errors may also be committed during the processing stage at any point when records for persons or households are added to or removed from the census database. Records can be deleted by mistake. Questionnaires may be linked to the wrong record or returned too late to be included.

Even though efforts are made to enumerate the homeless population, the risk of undercoverage is high. Some other living arrangements are also susceptible to coverage error. For example, young adults newly away from home may be either undercovered, because neither their roommates nor their parents include them in the census questionnaire, or overcovered, because they are included in both census questionnaires. Persons who maintain a second residence because of their employment can also cause coverage error.

Users should also be aware of the extent to which reserves and settlements participated in the 2021 Census. In some cases, enumeration was not permitted by the community or was interrupted before it could be completed. These geographic areas (63 in all in 2021, an increase from 14 in 2016) are considered incompletely enumerated reserves and settlements. There are no 2021 data for incompletely enumerated reserves and settlements, and those areas are not included in the totals. Similar problems have occurred in previous censuses. For example, 22 reserves and settlements were incompletely enumerated in the 2006 Census, and 31 in the 2011 Census.

The population estimates for the 63 incompletely enumerated reserves and settlements are based on a model. However, since no reliable source is available to verify the assumptions in the model, the estimates must be used with caution. For more information, see Section 12.2.

3.2 Control

Potential sources of coverage error were recognized during the planning stage of the 2021 Census, and the following measures were taken to minimize the associated risks:

- Collection unit (CU) boundaries were carefully defined and mapped to ensure that no geographic areas were left out or included twice.
- List/leave areas: The enumerator's manual contained instructions on how to enumerate a CU so as to minimize the risk of missing dwellings. The total number of dwellings from the 2016 Census was provided to field operations supervisors to help them identify significant changes. In addition, when the listing operation resulted in a substantial difference in the number of dwellings relative to the 2016 Census, the listing was checked. Lastly, specific quality control procedures were applied to the CU to evaluate and correct any changes made in the listing.
- Mail-out areas: Mail-out was based on a list of addresses from Statistics Canada's Address Register. This list was updated regularly, and listing activities were carried out (in the field and remotely) mainly in areas where frame errors were more likely. These listing activities were carried out continuously, but more intensively in the two years preceding the census. The work of enumerators was closely monitored.
- In 2021, the mail-out with drop-off (MODO) methodology was introduced. MODO areas are those where all dwellings have addresses, the majority of which are mailable. In these mixed areas, those dwellings with a valid mailing address were mailed the regular mail-out material (just like the mail-out areas), while those that did not have a valid mailing address (that corresponds to the civic address) received an invitation letter dropped at their door by a census employee. The MODO areas were introduced to maximize the number of census mail-out dwellings. As for mail-out areas, MODO was based on a list of addresses from Statistics Canada's Address Register, and the list was updated regularly.
- Special procedures were defined for the enumeration of the population residing on reserves.
- Advertisements informed Canadians about the census and indicated what to do if they did not receive a questionnaire.
- The Census Help Line (CHL) was available to answer any questions about the census, including questions about coverage.
- When calls were received at the CHL regarding a dwelling that may have been missed by the census, a process was in place to examine the surrounding area for other potential missed dwellings.
- There was a "Whom to include" section in the questionnaire so respondents could determine which persons should be included. Also, slightly more than 84% of the responses to the 2021 Census were obtained through the Internet, and the electronic questionnaire included additional verification questions when respondents reported a dwelling as unoccupied or non-existent, or if they had a problem determining whether a person should be included or not.
- In the questionnaire, respondents were asked to indicate whether any persons had not been listed because they were not sure they should be included. The electronic questionnaire provided guidance so respondents could make the right decision. In the other cases, a telephone follow-up was subsequently carried out with the respondent to determine whether the persons in question should be listed in the questionnaire or not.
- Telephone follow-up was carried out after questionnaires were reviewed for coverage inconsistencies or to verify household status, including questionnaires containing only foreign residents or persons temporarily present.

These procedures, along with appropriate staff training, supervisory checks and quality controls during the collection and processing stages, helped to reduce the number of coverage errors.

3.3 Definitions

Algebraic definitions of coverage errors are presented in this section. Let T denote the total or the "actual" number of persons targeted by the Census of Population. Let C denote the published census count of persons in the target population. The error associated with using C instead of T is as follows:

$$N = T - C$$

This error, denoted as N, is the **net population coverage error**.

Let U denote **population undercoverage**, the number of persons not included in C who should have been.

The census count C is composed of two elements:

$$C = E + I$$

where:

E is the number of persons enumerated. This is the number of persons who were listed on a census questionnaire.

I is the number of persons imputed. This is an estimate of the number of persons missed because their dwelling was classified as occupied but non-response or was misclassified as unoccupied (therefore, no follow-up was done). For more information on whole household imputation (WHI), see Section 3.6 of the *Sampling and Weighting Technical Report, Census of Population, 2021*, Statistics Canada Catalogue no. 98-306-X.

Undercoverage compared with the published census count C is therefore what remains of the persons who should have been listed on a census questionnaire and who were not taken into account by the WHI. In other words, it does not include the estimate of the number of persons who were not enumerated either because no completed census questionnaire was returned for the dwelling (non-response dwelling) or because the dwelling was misclassified as unoccupied (classification error) and did not receive a questionnaire.

The concept of undercoverage before the WHI also exists. This is what is referred to as Census of Population collection undercoverage. For more information, see Section 12.1.

Let O denote **population overcoverage**, the number of excess enumerations included in C that should not have been.

O has two components. One is the excess enumerations of persons enumerated more than once. Coverage studies focus on these excess enumerations. The second is persons who were enumerated but who were not in the census target population. For example, foreign residents visiting Canada who are listed on a census questionnaire as usual residents of a dwelling should not be included in C. Fictitious persons are another example. According to previous studies, the number of persons who are enumerated but are not in the census target population is generally very small and can be ignored. Consequently, census coverage does not measure this component of coverage error.

Since U refers to persons who were not enumerated but should be included in C and since O denotes enumerations that should not be included in C, the difference between T and C is U less O. That is:

$$N = U - O$$

The actual number of persons in the census target population is therefore:

$$T = C + N = C + U - O$$

In practice, for reasons of cost and timeliness of the data produced, an estimate of T is given by $\hat{T}$, based on sample studies, where:

$$\hat{T} = C + \hat{N} = C + \hat{U} - \hat{O}$$

$\hat{U}$ is an estimate of the number of persons not included in C who should have been, and $\hat{O}$ is an estimate of the number of persons included in C who should not have been. We can assume that overcoverage from persons included in C who are not in the census target population is zero, since it is negligible. Consequently, $\hat{O}$ is simply an estimate of the number of duplicate enumerations. The purpose of census coverage studies is to determine the values of $\hat{U}$ and $\hat{O}$.

In summary, the actual population T is composed of the census count C and the net undercoverage N. This is referred to as net undercoverage because U is generally larger than O in the context of the current census in Canada. However, the opposite is possible, whereby N would be negative. C consists of E plus the number of persons added in WHI, and this imputation I targets persons living in non-response dwellings or in occupied dwellings misclassified as unoccupied.

Census population coverage errors can generally be expressed as rates relative to the actual population. The undercoverage rate R_U is U as a percentage of T. The overcoverage rate R_O is O as a percentage of T. The net undercoverage rate R_N is the difference between U and O as a percentage of the census target population. These three rates can be estimated by $\hat{R}_U$, $\hat{R}_O$ and $\hat{R}_N$, as follows:

$$\hat{R}_U = 100 * \frac{\hat{U}}{\hat{T}} = 100 * \frac{\hat{U}}{C + \hat{N}}$$

$$\hat{R}_O = 100 * \frac{\hat{O}}{\hat{T}} = 100 * \frac{\hat{O}}{C + \hat{N}}$$

$$\hat{R}_N = 100 * \frac{\hat{N}}{\hat{T}} = 100 * \left(\frac{\hat{U} - \hat{O}}{C + \hat{N}} \right)$$

A positive net undercoverage rate indicates that the undercoverage rate is higher than the overcoverage rate. That is, the number of persons not included in the published census count C is higher than the number of excess enumerations. That is generally the case for all Canadian censuses. For some domains of interest, however, negative net undercoverage is sometimes observed.

3.4 Evaluation

Two postcensal studies were carried out to estimate the 2021 Census population coverage error. The Census Undercoverage Study (CUS) provided estimates for population undercoverage, while the Census Overcoverage Study (COS) estimated population overcoverage. As previously mentioned, the Dwelling Classification Survey (DCS) does not contribute to census coverage error estimates since census counts are already adjusted to take DCS results into account.

The CUS and COS were conducted subsequent to field collection and census processing operations. Preliminary estimates of 2021 Census population coverage error were released on April 28, 2023. Following an in-depth validation exercise with the Centre for Demography and the provincial and territorial statistical focal points, final estimates were released on September 27, 2023. The data were released at the same time as the new official demographic estimates reflecting the update of the base population to the 2021 Census. Census population counts adjusted for net population undercoverage constituted the updated estimates of the base population.

A brief description of the methodology used in the two census coverage studies is presented below:

Census Undercoverage Study

In the CUS, a random sample of individuals representing the 2021 Census target population was selected from frames independent of the census. These frames are described in Section 7.1. The 2021 Census database was then searched to determine whether these persons had indeed been enumerated.

Where necessary, interviews were conducted, mostly via computer-assisted telephone interviewing from the regional offices (ROs), to collect information for use in additional searches of the 2021 Census database. An interview was completed for 60.2% of the 12,787 cases sent to the ROs. The sampling weight was adjusted for non-response. Specifically, the total sampling weight of non-respondents was divided among groups of respondents most like the non-respondents in their response probability.

The estimate of population undercoverage is based on the number of persons in the CUS sample who were classified as “missed.” These persons were part of the target population for the 2021 Census, but no evidence of enumeration could be found in the 2021 Census Response Database. Nationally, 6,212 persons in the CUS sample were classified as missed in the provinces and 1,241 in the territories.

Census Overcoverage Study

Overcoverage was measured by creating a list of potentially duplicated records on the 2021 Census database using probabilistic and deterministic linkage methods to match the final 2021 Census database to itself, and then matching the final 2021 Census database and a list of persons who should have been enumerated according to administrative data sources. Probabilistic linkage estimates the probability that linked pairs are true matches based on the agreement patterns obtained by comparing the linkage variables between records. Deterministic linkage classifies pairs of records as a match when the linkage variables mostly agree and a non-match otherwise. No weight is given to the strength of the match.

A random sample of potential duplicates was selected from the list created from the linkages, and the pairs were manually verified by comparing demographic characteristics and names to identify true cases of overcoverage. Population overcoverage was estimated from this sample.

4. Census data collection

Data collection in the 2021 Census ensured that each of the 16.3 million dwellings in Canada was enumerated and that, for each occupied dwelling, the corresponding household completed a census questionnaire.

In 2021, Canadian households had the option of responding online or contacting the Census Help Line to request a paper questionnaire if they preferred. Households that did not use either of these two response modes received a follow-up by telephone or in person. The four collection methods used for the Canadian census were mail-out, list/leave, mail-out with drop-off (MODO) and canvasser enumeration. To make census collection as efficient as possible, Canada is divided into small geographic units known as collection units (CUs). In the 2021 Census, there were approximately 49,000 CUs in Canada.

- For mail-out CUs, the postal system was used to deliver the census material. Mail-out CUs are typically in urban areas. While mail-out CUs now include about 86% of private Canadian dwellings, they cover only a tiny fraction of the country's land area.
- List/leave CUs are typically in rural areas. In those areas, census employees prepared a list of dwellings and dropped off an invitation letter. About 7% of Canadian dwellings are in list/leave CUs, which cover a large portion of the country's land area.
- In 2021, the MODO methodology was introduced for areas where all dwellings have addresses, the majority of which are mailable. In these mixed areas, those dwellings with a valid mailing address were mailed the regular mail-out material, while those that did not have a valid mailing address (that corresponds to the civic address) received an invitation letter dropped at their door by a census employee. This allowed the mail-out methodology to be extended to about 90% of dwellings.
- Enumeration by canvasser interview has typically occurred in First Nations communities, Métis Settlements, Inuit regions and other remote areas. However, for the first time, in 2021, all First Nations communities, Métis Settlements, Inuit regions and other remote areas were offered the opportunity to self-respond, provided it was operationally feasible (i.e., Internet was accessible in the community). Households in areas where it was not operationally feasible to offer self-response completed their census questionnaire with a census employee (in person or over the phone). In 2021, dwellings in remote, northern and Indigenous communities represented about 1% of dwellings in Canada.

Since the 2011 Census, a wave methodology has been used for census collection. This approach involves contacting non-respondent households at key times to remind them to participate in the census and persuade them to complete the questionnaire. In each wave, households are provided with the information they need to respond. This methodology varies with the collection method used to distribute the census materials for a given region.

Because every Canadian household is required by law to answer the census questions, the wave methodology was designed to encourage them to respond early online, yet offers an alternative for households that do not wish to complete their questionnaire online. The wave methodology has many advantages: it increases the number of online responses, a mode that facilitates the flow of questions and real-time response edits; and it reduces non-response, the need for costly follow-up, and the number of questionnaires to register and the amount of data to capture. In the end, it increases the quality of the census.

The COVID-19 pandemic emerged in Canada in early 2020 and affected all steps of the 2021 Census process, from data collection to dissemination.

No questions were added to the 2021 Census questionnaire to collect information on COVID-19. Nevertheless, Statistics Canada is aware that COVID-19 may have had an impact on answers to some census questions, including those on employment, education, commuting and expenditures. When providing answers to census questions, respondents were instructed to choose the responses that best reflected their situation or the situation of members of the household for the date or time period in question. Additional instructional text was also provided in "help" features in the online questionnaires.

In light of the pandemic, Statistics Canada hired an additional 15 Indigenous Liaison Advisors for an Indigenous Engagement Task Force to bolster engagement efforts with Indigenous and northern communities for the 2021 Census.

The collection strategy for the 2021 Census of Population had to adapt and was turned into a fully contactless process to ensure respondents and census employees were safe. For example, early enumeration was cancelled for First Nations communities, Métis Settlements, Inuit regions and other remote areas, and completion of the census questionnaire online was emphasized. Also, more resources were deployed for telephone follow-up to reduce the number of in-person visits, and an additional reminder letter was sent to non-respondents in July.

Collection procedures for the 2021 Census were redesigned to ensure the safety of respondents and census employees by limiting the contact needed to participate in the census. No census employee from Statistics Canada was permitted to visit or enter institutional collective dwellings, especially the dwellings that house residents who are vulnerable to COVID-19, such as residences for senior citizens and hospitals. For most collective dwellings, the 2021 Census: Collective Dwellings electronic questionnaire was used to collect information on the facility and its usual residents. In April 2021, invitation letters or emails containing a secure access code were sent to the administrators of the collective dwellings for online response.

The pandemic presented some challenges for conducting the 2021 Census of Population, but despite these, the collection response rate for the country as a whole was a resounding success, at 98.0%, thanks to Canadians who completed the census in the midst of the third wave of the pandemic. Additionally, unique challenges were encountered in northern or remote regions of the country, such as travel restrictions, border closures, shorter and shifted collection periods, unavailability of local staff, and wildfires. Ensuring the health and safety of Canadians and Statistics Canada employees by adapting collection operations to ensure high-quality, trusted census data was a high priority for Statistics Canada. For more information on census data collection, see Chapter 1 of the *Sampling and Weighting Technical Report, Census of Population, 2021*, Statistics Canada Catalogue no. 98-306-X.

5. Census data processing

The processing of all the completed questionnaires, regardless of type, includes several steps, from receiving the questionnaires to creating an accurate and complete census database. These steps include questionnaire registration, questionnaire imaging and data capture, editing, error correction, failed edit follow-up, coding, dwelling classification and non-response adjustments, linkage of administrative data, and imputation and weighting (for the long-form questionnaire sample).

Automated processes implemented for the 2021 Census were monitored to ensure that all Canadian residences were enumerated only once. The Master Control System (MCS) was built to control and monitor the process flow from data collection to processing. The MCS held a master list of all the dwellings in Canada, where each dwelling was identified with a unique identifier. This system was updated on a daily basis with information about each dwelling's status in the census process flow (delivered, received, processed, etc.). Reports were generated daily and made available online to managers to ensure that census operations were efficient and effective.

For more information on census data processing, see Chapter 3 of the *Sampling and Weighting Technical Report, Census of Population, 2021*, Statistics Canada Catalogue no. 98-306-X.

6. Dwelling Classification Survey

6.1 Introduction

The Dwelling Classification Survey (DCS) measures dwelling classification errors in the census. The DCS also estimates the number of usual residents in occupied non-response dwellings. Using these results, the occupancy status of non-response dwellings is adjusted, and people are imputed into the occupied dwellings during the whole household imputation (WHI) procedure, based on DCS control total files. In this chapter, the term “dwelling” will always refer to private dwellings.[3]

One of the potential sources of error in a census is the misclassification of a dwelling. When a household does not return a questionnaire, the enumerator has to determine whether the dwelling is occupied. Two types of dwelling classification errors can occur in this situation. **First, an occupied dwelling may be misclassified as unoccupied**. If these dwellings were ultimately considered to be unoccupied, this would result in census population undercoverage because the dwelling occupants would not be counted. **Second, an unoccupied dwelling may be misclassified as occupied**. If these dwellings were ultimately considered to be occupied, this would result in census population overcoverage because people would be counted when, in fact, no one lives at that dwelling. Estimates from the DCS are used to adjust census data for these two coverage errors.

Additionally, a third type of dwelling classification error measured by the DCS is the error incurred when marginal dwellings or dwellings under construction are misclassified as dwellings. This will result in dwelling overcoverage and could result in population overcoverage if the dwelling is classified as occupied. Census data are not adjusted for these dwellings, so census estimates of the housing stock include some degree of overcoverage. This is discussed further in Section 6.3.1.2 and Section 6.3.2.2.

6.2 Methodology

6.2.1 Stratification and sample selection

The DCS target population included all private dwellings classified as either unoccupied or non-response dwellings, excluding dwellings in First Nations communities, Métis Settlements, Inuit regions and other remote areas, as well as private dwellings attached to a collective dwelling.[4] Those areas were excluded because of cost and operational considerations.

The DCS sample size was set at 1,903 collection units (CUs). The sampling frame consisted of all non-remote and non-reserve CUs (i.e., CUs with a collection methodology of mail-out, mail-out with drop-off [MODO], list/leave and seasonal).[5] Consequently, Nunavut had no in-scope CUs, and the DCS was not conducted there. The sample design was as follows. All in-scope CUs in Yukon (42 CUs) and the Northwest Territories (21 CUs) formed two strata, one for each territory. All their in-scope CUs were selected for the DCS sample with certainty. All the CUs in Prince Edward Island formed a third stratum, from which a simple random sample of 51 CUs was selected.

The remaining CUs in each province were grouped into urban and rural strata. A CU was initially considered urban if it had been part of a census metropolitan area (CMA) or a census agglomeration (CA) that had 40,000 or more dwellings. Further, all the CUs within a crew leader district (CLD) were considered urban if more than 50% of the CUs in the CLD were urban. All the remaining CUs formed the rural strata. Urban CUs were stratified by CMA and CA. A simple random sample of at least five CUs was selected within each stratum. From past census data, it was determined that five CUs was an appropriate workload for an interviewer. There were 1,037 urban CUs in the sample. To control field costs, CUs in close proximity to each other were chosen for the rural sample. This

3. Refer to Section 2.3 for the definition of a private dwelling.
4. Refer to Chapter 7 of the *Guide to the Census of Population, 2021*, Statistics Canada Catalogue no. 98-304-X.
5. The seasonal collection methodology is the same as the list/leave collection methodology. However, these areas contain dwellings that are mostly used seasonally and therefore tend to have low occupancy rates. This is important when deciding how many dwellings to sample in these areas.

was done via a two-stage stratified random sampling design. In the first stage, CLDs were selected within each province using simple random sampling. In the second stage, five CUs were randomly chosen from each of the selected CLDs. There were 866 rural CUs in the sample.

Dwelling subsampling within a sampled CU occurred when the number of unoccupied and non-response dwellings exceeded a maximum dwelling parameter, which was 35 in mail-out and list/leave CUs, 40 in MODO CUs, and 135 in seasonal CUs. Subsampling of in-scope dwellings occurred in 584 CUs. Otherwise, all unoccupied dwellings and non-response dwellings[6] in the sampled CUs formed the DCS dwelling sample. A total of 34,459 unoccupied dwellings and 11,346 non-response dwellings were sampled in 2021. Table 6.2.1 shows the distribution of the sample by province and territory.

Table 6.2.1
Sample size for Canada, provinces and territories

Provinces and territories	Number of collection units	Number of unoccupied dwellings	Number of non-response dwellings
Canada	**1,903**	**34,459**	**11,346**
Newfoundland and Labrador	159	4,404	466
Prince Edward Island	51	1,311	175
Nova Scotia	125	2,831	501
New Brunswick	109	1,651	710
Quebec	316	4,987	2,145
Ontario	394	6,700	2,281
Manitoba	164	2,186	957
Saskatchewan	135	2,670	770
Alberta	190	3,118	1,354
British Columbia	197	3,646	1,434
Yukon	42	757	245
Northwest Territories	21	198	308
Nunavut	0	0	0

Source: Statistics Canada, 2021 Dwelling Classification Survey.

6.2.2 Field interviews

A DCS questionnaire was used to verify the true occupancy status on Census Day of sampled dwellings in the sampled CUs. Occupancy status was verified from mid-June to mid-August 2021. DCS operations ideally started once a predetermined response rate threshold for the Census of Population was obtained in the sampled CU, but sometimes they had to start earlier because of operational constraints.

To determine occupancy status and to collect other information, enumerators were instructed to contact current occupants, neighbours, landlords or any other person with knowledge about the dwelling. Up to three contact attempts were made for each dwelling. If the dwelling was found to have been occupied on Census Day, the number of occupants on Census Day was also obtained, along with occupants' sex and age, when possible.

6.2.3 Processing and estimation

All completed questionnaires were sent to head office in Ottawa for processing.

6. Dwelling status is based on the status at the time of the DCS sample selection. This status could change by the end of the census collection period (for example, if a dwelling responds).

The first processing step was data capture. Once this was completed, the questionnaires were subjected to an extensive set of consistency edits. The questionnaires that failed the edits were examined manually to resolve inconsistencies.

At this point in the processing, the unoccupied dwellings and the non-response dwellings in the sample were separated, and the dwellings' classifications were confirmed against the census database. The questionnaires completed for each sampled CU were matched to the final census lists of unoccupied dwellings and non-response dwellings (i.e., no census questionnaire was received for the dwelling). If a match could not be found, meaning that the status of the dwelling had changed after the DCS sample was selected, the sampled dwelling was discarded and no further processing was required. Dwellings listed as unoccupied or as non-response on the census lists in the sampled CUs for which no DCS questionnaire was received were considered as total non-response to the DCS and proceeded to the next processing step.

Total non-response to the DCS was addressed by a weighting adjustment, and item imputation was used for item non-response. The procedure was the same for unoccupied dwellings and non-response dwellings. When there was no information for a dwelling, the design weights of the respondents were adjusted to account for the design weight of the non-respondents. The adjustment was made separately by geographic post-stratum. Refer to Hong (2023) for more details about the post-strata used. Design weights were then adjusted so that the sum of the adjusted weights for each geographic post-stratum equalled the total number of unoccupied or non-response dwellings. Item non-response for occupancy status and the number of usual residents was addressed by imputation. Occupancy status was imputed first and then used in the imputation of the other variables.

Finally, the household size, collected on the DCS questionnaire when a dwelling was found occupied, was used to produce an estimate, by household size, of the occupancy rate of private dwellings classified by the census as unoccupied or as non-response.

6.2.4 Census whole household imputation

The WHI procedure strives to represent non-response private dwellings in the census and, as such, imputes for total non-response to the census. The strategy varies based on whether the DCS was conducted in the area. The WHI procedure results in all private dwellings being classified as either occupied or unoccupied in the census database (i.e., there are no longer any non-response dwellings).

In geographic areas where the DCS was conducted (i.e., CUs with mail-out, MODO, list/leave and seasonal collection types, excluding private dwellings attached to a collective dwelling), the DCS estimates were used as input to the WHI algorithm to impute occupied private dwellings based on post-strata and household size distributions. First, within a DCS geographic post-stratum, all the dwellings without a response were identified (this was done separately for the unoccupied and non-response dwelling universes). Second, for the non-response dwelling universe only, any non-response dwelling for which field collection had obtained the number of usual residents was deemed to be occupied and was assigned the recorded household size. Finally, dwellings without a response were randomly selected and imputed as occupied. The selection was made so that the final number of non-response and unoccupied dwellings converted to occupied dwellings in the post-strata equalled the DCS estimate of occupied dwellings in the non-response and unoccupied dwelling universes. In addition, new for the 2021 cycle, control total files were used such that the DCS estimate of the occupancy rate in each post-stratum was applied in each census subdivision. Once a private unoccupied or non-response dwelling was imputed as occupied, a procedure, subject to the constraints of DCS estimates by post-stratum and household dwelling size, was used to impute the household dwelling size. Then, the geographically nearest dwelling of the predetermined household size was chosen as a donor. The answers to the census short-form questionnaire from this donor household were then assigned to the unoccupied or non-response dwelling that was imputed as occupied.

In geographic areas where the DCS was not conducted (i.e., CUs with remote and reserve collection types, as well as private dwellings attached to a collective dwelling), the vast majority of private dwellings classified as non-response were imputed as occupied, with the household size coming either from field collection or from a donor household. The geographically nearest dwelling of the predetermined household size was chosen as a donor. If the dwelling did not have a household size already assigned, the household size was chosen to be the donor's household size. The information from this donor household was then assigned to the unoccupied or non-response dwelling that was imputed as occupied. For dwellings in remote and reserve collection areas, the

complete record was assigned. For private dwellings attached to a collective dwelling, only the census short-form questionnaire answers were assigned.

Several adaptations were implemented in the 2021 Census collection plan to mitigate the impact of the COVID-19 pandemic and other potential risks. A statistical contingency plan was developed based on the secure, responsible and appropriate use of administrative data to support the 2021 Census in the event of disruptions to census collection. Linked administrative data from federal and provincial data sources were used in specific areas to improve the imputation of non-response households at the imputation stage after the 2021 Census collection ended.[7] Households constructed from administrative data were added to the donor pool in those specific areas.

More information on WHI can be found in the report by Hong (2023).

6.3 Estimates

Census data were adjusted using DCS estimates, via the WHI procedure, for non-response dwellings and for occupied dwellings that were misclassified as unoccupied. The estimates are provided in sections 6.3.1.1 and 6.3.2.1. Census data were not adjusted for marginal dwellings or dwellings under construction that were misclassified as dwellings. Sections 6.3.1.2 and 6.3.2.2 present the estimates of the number of marginal dwellings and dwellings under construction that were wrongly classified as dwellings and were therefore erroneously included in the housing stock.

6.3.1 Unoccupied dwellings

6.3.1.1 Occupied dwellings misclassified as unoccupied

Table 6.3.1.1.1 provides the estimated number of occupied dwellings that were misclassified as unoccupied and the corresponding error rate for unoccupied dwellings by urban and rural areas[8] and by province and territory. For comparison, Table 6.3.1.1.2 provides the estimates for the same variables for the 2016 Census.

Table 6.3.1.1.1 shows that 17.3% of all dwellings classified as unoccupied in 2021 were truly occupied. This is a slight increase from 15.0% in 2016 and is a statistically significant change. Dwelling misclassification was more prevalent in urban areas (23.8%) than in rural areas (10.9%). Both urban and rural areas show an increase from 2016. The misclassification rate increased or remained nearly unchanged for all provinces and territories except for the Northwest Territories, where it decreased.

Among the provinces and territories, New Brunswick had the highest misclassification rate (20.2%), followed by Ontario and British Columbia (18.9%), and Quebec (17.7%). The rates for the other provinces and territories ranged from 8.1% for Prince Edward Island to 16.6% for Alberta.

Table 6.3.1.1.3 provides the estimated number of occupied dwellings that were misclassified as unoccupied, the estimated number of people living in those dwellings and the standard error. These estimates were used to decide how many dwellings to impute as occupied in WHI. Table 6.3.1.1.4 provides the estimates for the same variables for the 2016 Census. Because of errors in the initial dwelling classification, approximately 218,042 households were not enumerated in the 2021 Census. This is the number of households added to the census during WHI.

7. Refer to Appendix 1.7 of the *Guide to the Census of Population, 2021*, Statistics Canada Catalogue no. 98-304-X.
8. The terms urban and rural are defined by the geographic post-strata used for estimation. Refer to Hong (2023) for more information.

Table 6.3.1.1.1
Number of occupied dwellings misclassified as unoccupied dwellings for various geographic areas, 2021 Census

Geographic Area	Number of unoccupied dwellings	Occupied dwellings misclassified as unoccupied			
		Estimated number	Standard error	Estimated rate[1] (%)	Standard error (%)
Canada	**1,259,149**	**218,042**	**10,097**	**17.3**	**0.8**
Urban	626,525	148,973	5,918	23.8	0.9
Rural	632,624	69,069	5,565	10.9	0.9
Atlantic provinces	**132,141**	**16,504**	**1,359**	**12.5**	**1.0**
Newfoundland and Labrador	44,142	3,818	586	8.6	1.3
Prince Edward Island	10,345	836	127	8.1	1.2
Nova Scotia	47,584	5,778	743	12.1	1.6
New Brunswick	30,070	6,073	764	20.2	2.5
Quebec	**285,920**	**50,477**	**5,283**	**17.7**	**1.8**
Ontario	**433,952**	**82,077**	**7,862**	**18.9**	**1.8**
Prairies	**237,232**	**36,947**	**2,875**	**15.6**	**1.2**
Manitoba	49,464	6,346	1,445	12.8	2.9
Saskatchewan	60,916	9,575	1,373	15.7	2.3
Alberta	126,852	21,026	2,005	16.6	1.6
British Columbia	**168,515**	**31,810**	**2,774**	**18.9**	**1.6**
Territories[2]	**1,389**	**226**	**0**	**16.3**	**0.0**
Yukon	1,016	166	0	16.4	0.0
Northwest Territories	373	59	0	15.9	0.0

1. The rate is the estimated number of occupied dwellings misclassified as unoccupied dwellings as a percent of all dwellings classified as unoccupied.
2. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2021 Dwelling Classification Survey.

Table 6.3.1.1.2
Number of occupied dwellings misclassified as unoccupied dwellings for various geographic areas, 2016 Census

Geographic Area	Number of unoccupied dwellings	Occupied dwellings misclassified as unoccupied			
		Estimated number	Standard error	Estimated rate[1] (%)	Standard error (%)
Canada	**1,187,392**	**178,219**	**5,520**	**15.0**	**0.5**
Urban	680,629	145,308	4,819	21.3	0.7
Rural	506,763	32,911	2,692	6.5	0.5
Atlantic provinces	**144,073**	**12,577**	**816**	**8.7**	**0.6**
Newfoundland and Labrador	43,770	2,830	310	6.5	0.7
Prince Edward Island	10,909	646	117	5.9	1.1
Nova Scotia	51,940	5,351	633	10.3	1.2
New Brunswick	37,454	3,750	395	10.0	1.1
Quebec	**289,593**	**41,544**	**2,240**	**14.3**	**0.8**
Ontario	**378,298**	**62,536**	**4,309**	**16.5**	**1.1**
Prairies	**204,641**	**30,779**	**1,637**	**15.0**	**0.8**
Manitoba	36,147	4,680	460	12.9	1.3
Saskatchewan	54,414	6,839	1,022	12.6	1.9
Alberta	114,080	19,260	1,193	16.9	1.0
British Columbia	**169,340**	**30,526**	**1,881**	**18.0**	**1.1**
Territories[2]	**1,447**	**256**	**0**	**17.7**	**0.0**
Yukon	874	143	0	16.3	0.0
Northwest Territories	573	113	0	19.8	0.0

1. The rate is the estimated number of occupied dwellings misclassified as unoccupied dwellings as a percent of all dwellings classified as unoccupied.
2. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2016 Dwelling Classification Survey.

Table 6.3.1.1.3
Number of occupied dwellings misclassified as unoccupied and number of persons living in them for various geographic areas, 2021 Census

	Occupied dwellings misclassified as unoccupied dwellings		Persons living in occupied dwellings misclassified as unoccupied dwellings	
Geographic Area	Estimated number	Standard error	Estimated number	Standard error
Canada	**218,042**	**10,097**	**406,008**	**19,153**
Urban	148,973	5,918	286,311	12,305
Rural	69,069	5,565	119,698	9,311
Atlantic provinces	**16,504**	**1,359**	**29,463**	**2,321**
Newfoundland and Labrador	3,818	586	6,750	1,052
Prince Edward Island	836	127	1,533	241
Nova Scotia	5,778	743	10,350	1,317
New Brunswick	6,073	764	10,830	1,195
Quebec	**50,477**	**5,283**	**81,818**	**8,503**
Ontario	**82,077**	**7,862**	**157,184**	**15,138**
Prairies	**36,947**	**2,875**	**70,823**	**5,650**
Manitoba	6,346	1,445	12,132	2,710
Saskatchewan	9,575	1,373	18,597	2,946
Alberta	21,026	2,005	40,093	3,966
British Columbia	**31,810**	**2,774**	**66,263**	**6,532**
Territories[1]	**226**	**0**	**457**	**0**
Yukon	166	0	330	0
Northwest Territories	59	0	128	0

1. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2021 Dwelling Classification Survey.

Table 6.3.1.1.4
Number of occupied dwellings misclassified as unoccupied and number of persons living in them for various geographic areas, 2016 Census

	Occupied dwellings misclassified as unoccupied dwellings		Persons living in occupied dwellings misclassified as unoccupied dwellings	
Geographic Area	Estimated number	Standard error	Estimated number	Standard error
Canada	**178,219**	**5,520**	**338,246**	**11,110**
Urban	145,308	4,819	277,996	9,818
Rural	32,911	2,692	60,250	5,199
Atlantic provinces	**12,577**	**816**	**23,220**	**1,577**
Newfoundland and Labrador	2,830	310	5,042	623
Prince Edward Island	646	117	1,257	248
Nova Scotia	5,351	633	9,638	1,155
New Brunswick	3,750	395	7,283	838
Quebec	**41,544**	**2,240**	**73,306**	**4,200**
Ontario	**62,536**	**4,309**	**120,951**	**8,429**
Prairies	**30,779**	**1,637**	**58,289**	**3,684**
Manitoba	4,680	460	8,502	906
Saskatchewan	6,839	1,022	12,395	2,134
Alberta	19,260	1,193	37,392	2,863
British Columbia	**30,526**	**1,881**	**62,003**	**4,323**
Territories[1]	**256**	**0**	**477**	**0**
Yukon	143	0	251	0
Northwest Territories	113	0	226	0

1. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2016 Dwelling Classification Survey.

6.3.1.2 Housing stock overcoverage in dwellings misclassified as unoccupied

Table 6.3.1.2 shows the estimated number of dwellings classified as unoccupied that were not in the housing stock and the corresponding error rate for unoccupied dwellings for various geographic areas. No adjustments were made to the census database to account for dwellings not in the housing stock that were misclassified as unoccupied.

The enumeration of unoccupied dwellings that fall outside the housing universe results in dwelling overcoverage. Dwellings are outside the housing universe if they are used for commercial purposes, are not habitable year round or are double counted in the census. Double counting can occur when the dwelling appears to have two addresses associated with it or when two questionnaires are mistakenly returned for a dwelling that no longer contains a separate apartment.

The DCS estimates of the number of dwellings not in the housing stock that were misclassified as unoccupied dwellings were not used to adjust the census database because of the degree of subjectivity associated with classifying a dwelling as suitable for year-round occupancy. A dwelling must have a source of heat or power and provide complete shelter from the elements to be considered suitable for year-round occupancy. It is sometimes difficult to tell whether a dwelling is habitable, such as when a dwelling is a cottage, is under construction and almost complete, or has deteriorated. Note that the conditions used to identify dwellings that fall outside the housing universe were updated since the 2016 cycle to better reflect the definition provided in this paragraph. With the updated conditions, dwellings outside the housing stock would have accounted for 27.5% of all dwellings classified as unoccupied in 2016.

This percentage was similar in 2021, where dwellings outside the housing stock accounted for 26.3% of all dwellings classified as unoccupied. Among the provinces and territories, the incidence of dwellings outside the housing stock being classified as unoccupied ranged from 15.1% in Newfoundland and Labrador to 38.0% in New Brunswick. The problem was more prevalent in rural areas (27.7%) than in urban areas (24.8%).

Dwellings actually outside the housing stock represented 2.1% of all private dwellings in the 2021 Census. This was the same as the 2016 error rate (2.1%). Among the provinces and territories, the error ranged from 0.8% in the Northwest Territories to 4.1% in Prince Edward Island.

Table 6.3.1.2
Dwellings not in the housing stock misclassified as unoccupied dwellings for various geographic areas, 2021 Census

Geographic Area	Number of unoccupied dwellings	Dwellings not in housing stock misclassified as unoccupied dwellings			
		Estimated number	Standard error	Estimated rate[1] (%)	Standard error (%)
Canada	**1,259,149**	**330,777**	**15,092**	**26.3**	**1.2**
Urban	626,525	155,519	8,518	24.8	1.4
Rural	632,624	175,258	11,166	27.7	1.8
Atlantic provinces	**132,141**	**32,440**	**2,419**	**24.5**	**1.8**
Newfoundland and Labrador	44,142	6,686	1,030	15.1	2.3
Prince Edward Island	10,345	3,084	496	29.8	4.8
Nova Scotia	47,584	11,253	1,181	23.6	2.5
New Brunswick	30,070	11,417	1,036	38.0	3.4
Quebec	**285,920**	**68,354**	**6,645**	**23.9**	**2.3**
Ontario	**433,952**	**129,372**	**10,506**	**29.8**	**2.4**
Prairies	**237,232**	**60,729**	**5,986**	**25.6**	**2.5**
Manitoba	49,464	14,586	3,371	29.5	6.8
Saskatchewan	60,916	17,947	2,362	29.5	3.9
Alberta	126,852	28,195	4,169	22.2	3.3
British Columbia	**168,515**	**39,570**	**3,517**	**23.5**	**2.1**
Territories[2]	**1,389**	**312**	**0**	**22.4**	**0.0**
Yukon	1,016	247	0	24.3	0.0
Northwest Territories	373	65	0	17.4	0.0

1. The rate is the estimated number of dwellings not in the housing stock misclassified as unoccupied dwellings as a percent of all dwellings classified as unoccupied.
2. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2021 Dwelling Classification Survey.

6.3.2 Non-response dwellings

6.3.2.1 Occupancy rate among non-response dwellings

Table 6.3.2.1.1 provides the estimated number and rate of occupied non-response dwellings in the census by urban and rural area[9] and by province and territory. Table 6.3.2.1.2 provides this information for the 2016 Census. Table 6.3.2.1.1 shows that 61.5% of all dwellings classified as non-response were occupied. This is a slight decrease from the 63.1% occupancy rate in 2016, though the difference is not statistically significant. The

9. The terms urban and rural are defined by the subprovincial post-strata used for estimation. Refer to Hong (2023) for more information.

2021 Census was relatively consistent in classifying non-response dwellings in urban (61.2%) and rural (62.2%) areas. At the provincial and territorial level in 2021, Yukon had the highest occupancy rate for non-response dwellings, at 69.5%, while Newfoundland and Labrador had the lowest rate, at 41.6%.

Table 6.3.2.1.3 shows the number of occupied non-response dwellings in the 2021 Census and provides the number of people added in those dwellings through the DCS. Table 6.3.2.1.4 shows the data from the 2016 Census. In 2021, 421,202 people were added to the census in 210,370 dwellings. The comparable 2016 numbers were 357,666 people in 179,823 dwellings.

Table 6.3.2.1.1
Occupied non-response dwellings for various geographic areas, 2021 Census

Geographic Area	Number of non-response dwellings	Occupied non-response dwellings			
		Estimated number	Standard error	Estimated rate[1] (%)	Standard error (%)
Canada	**342,162**	**210,370**	**4,198**	**61.5**	**1.2**
Urban	242,381	148,282	3,699	61.2	1.5
Rural	99,781	62,088	1,801	62.2	1.8
Atlantic provinces	**22,470**	**13,288**	**804**	**59.1**	**3.6**
Newfoundland and Labrador	5,751	2,392	470	41.6	8.2
Prince Edward Island	1,044	671	44	64.3	4.2
Nova Scotia	8,787	6,094	266	69.4	3.0
New Brunswick	6,888	4,131	265	60.0	3.8
Quebec	**79,340**	**51,664**	**1,709**	**65.1**	**2.2**
Ontario	**109,090**	**64,555**	**2,826**	**59.2**	**2.6**
Prairies	**81,640**	**47,810**	**1,914**	**58.6**	**2.3**
Manitoba	12,948	8,705	367	67.2	2.8
Saskatchewan	13,558	7,632	611	56.3	4.5
Alberta	55,134	31,472	1,761	57.1	3.2
British Columbia	**48,986**	**32,621**	**1,030**	**66.6**	**2.1**
Territories[2]	**636**	**434**	**0**	**68.2**	**0.0**
Yukon	360	250	0	69.5	0.0
Northwest Territories	276	183	0	66.5	0.0

1. The rate is the estimated number of occupied non-response dwellings as a percent of all non-response dwellings.
2. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2021 Dwelling Classification Survey.

Table 6.3.2.1.2
Occupied non-response dwellings for various geographic areas, 2016 Census

Geographic Area	Number of non-response dwellings	Occupied non-response dwellings			
		Estimated number	Standard error	Estimated rate[1] (%)	Standard error (%)
Canada	**284,966**	**179,823**	**2,525**	**63.1**	**0.9**
Urban	227,692	144,678	2,219	63.5	1.0
Rural	57,274	35,145	1,205	61.4	2.1
Atlantic provinces	**20,914**	**11,371**	**487**	**54.4**	**2.3**
Newfoundland and Labrador	4,779	2,637	313	55.2	6.5
Prince Edward Island	1,302	802	64	61.6	4.9
Nova Scotia	7,839	4,199	272	53.6	3.5
New Brunswick	6,994	3,733	248	53.4	3.6
Quebec	**58,039**	**39,376**	**1,110**	**67.8**	**1.9**
Ontario	**91,159**	**58,195**	**1,536**	**63.8**	**1.7**
Prairies	**64,103**	**38,020**	**1,264**	**59.3**	**2.0**
Manitoba	10,811	6,375	586	59.0	5.4
Saskatchewan	11,143	6,453	333	57.9	3.0
Alberta	42,149	25,191	1,070	59.8	2.5
British Columbia	**50,016**	**32,418**	**973**	**64.8**	**1.9**
Territories[2]	**735**	**444**	**0**	**60.4**	**0.0**
Yukon	405	227	0	56.1	0.0
Northwest Territories	330	217	0	65.7	0.0

1. The rate is the estimated number of occupied non-response dwellings as a percent of all non-response dwellings.
2. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2016 Dwelling Classification Survey.

Table 6.3.2.1.3
Persons living in occupied non-response dwellings for various geographic areas, 2021 Census

Geographic Area	Occupied non-response dwellings		Persons living in occupied non-response dwellings	
	Estimated number	Standard error	Estimated number	Standard error
Canada	**210,370**	**4,198**	**421,202**	**10,025**
Urban	148,282	3,699	302,456	8,927
Rural	62,088	1,801	118,745	4,406
Atlantic provinces	**13,288**	**804**	**24,111**	**1,507**
Newfoundland and Labrador	2,392	470	4,039	766
Prince Edward Island	671	44	1,322	109
Nova Scotia	6,094	266	10,977	626
New Brunswick	4,131	265	7,774	559
Quebec	**51,664**	**1,709**	**88,176**	**3,728**
Ontario	**64,555**	**2,826**	**135,062**	**6,867**
Prairies	**47,810**	**1,914**	**106,973**	**5,000**
Manitoba	8,705	367	21,176	1,179
Saskatchewan	7,632	611	18,042	1,649
Alberta	31,472	1,761	67,755	4,433
British Columbia	**32,621**	**1,030**	**66,025**	**3,556**
Territories[1]	**434**	**0**	**855**	**0**
Yukon	250	0	498	0
Northwest Territories	183	0	357	0

1. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2021 Dwelling Classification Survey.

Table 6.3.2.1.4
Persons living in occupied non-response dwellings for various geographic areas, 2016 Census

Geographic Area	Occupied non-response dwellings		Persons living in occupied non-response dwellings	
	Estimated number	Standard error	Estimated number	Standard error
Canada	**179,823**	**2,525**	**357,666**	**7,800**
Urban	144,678	2,219	286,286	7,204
Rural	35,145	1,205	71,380	2,990
Atlantic provinces	**11,371**	**487**	**22,686**	**1,163**
Newfoundland and Labrador	2,637	313	5,054	686
Prince Edward Island	802	64	1,459	130
Nova Scotia	4,199	272	8,651	783
New Brunswick	3,733	248	7,523	503
Quebec	**39,376**	**1,110**	**68,744**	**2,761**
Ontario	**58,195**	**1,536**	**125,071**	**5,541**
Prairies	**38,020**	**1,264**	**78,093**	**3,226**
Manitoba	6,375	586	13,312	1,239
Saskatchewan	6,453	333	12,795	996
Alberta	25,191	1,070	51,986	2,807
British Columbia	**32,418**	**973**	**62,089**	**3,278**
Territories[1]	**444**	**0**	**983**	**0**
Yukon	227	0	452	0
Northwest Territories	217	0	531	0

1. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2016 Dwelling Classification Survey.

6.3.2.2 Housing stock overcoverage in dwellings misclassified as non-response

Table 6.3.2.2 shows the 2021 Census dwelling classification error from dwellings misclassified as non-response when they should not have been included in the housing stock. Section 6.3.1.2 provides the definition of dwellings outside the housing universe and comments on the difficulty in determining whether a dwelling should be included in the housing stock. Note that the conditions used to identify dwellings that fall outside the housing universe were updated since the 2016 cycle to better reflect the definition provided in this paragraph. With the updated conditions, non-response dwellings that were outside the housing stock accounted for 7.0% of non-response dwellings in 2016.

At the national level, dwellings outside the housing stock accounted for 7.2% of all non-response dwellings in 2021. The error rate was higher in urban areas (7.6%) than in rural areas (6.2%). For provinces and territories, the incidence of dwellings outside the housing stock being classified as non-response ranged from 4.5% in Newfoundland and Labrador to 15.4% in Saskatchewan. At the national level, non-response dwellings outside the housing stock accounted for 0.2% of all private dwellings. This was similar to the 2016 error rate of 0.1%. Among the provinces and territories, the error ranged from 0.1% to 0.4% when looking at the error by province or territory.

Table 6.3.2.2
Dwellings not in the housing stock misclassified as non-response dwellings for various geographic areas, 2021 Census

Geographic Area	Number of non-response dwellings	Dwellings not in housing stock misclassified as non-response dwellings			
		Estimated number	Standard error	Estimated rate[1] (%)	Standard error (%)
Canada	**342,162**	**24,485**	**1,894**	**7.2**	**0.6**
Urban	242,381	18,327	1,614	7.6	0.7
Rural	99,781	6,159	875	6.2	0.9
Atlantic provinces	**22,470**	**1,561**	**199**	**6.9**	**0.9**
Newfoundland and Labrador	5,751	258	53	4.5	0.9
Prince Edward Island	1,044	111	32	10.7	3.1
Nova Scotia	8,787	473	113	5.4	1.3
New Brunswick	6,888	719	139	10.4	2.0
Quebec	**79,340**	**6,222**	**865**	**7.8**	**1.1**
Ontario	**109,090**	**7,051**	**1,000**	**6.5**	**0.9**
Prairies	**81,640**	**5,929**	**1,086**	**7.3**	**1.3**
Manitoba	12,948	757	140	5.8	1.1
Saskatchewan	13,558	2,083	725	15.4	5.3
Alberta	55,134	3,090	679	5.6	1.2
British Columbia	**48,986**	**3,689**	**769**	**7.5**	**1.6**
Territories[2]	**636**	**34**	**0**	**5.3**	**0.0**
Yukon	360	20	0	5.6	0.0
Northwest Territories	276	13	0	4.9	0.0

1. The rate is the estimated number of dwellings not in the housing stock misclassified as non-response dwellings as a percent of all non-response dwellings.
2. Nunavut is not included, because there are no in-scope dwellings in this territory.
Source: Statistics Canada, 2021 Dwelling Classification Survey.

7. Census Undercoverage Study

The primary objective of the Census Undercoverage Study (CUS) is to estimate the number of persons in the 2021 Census target population who were not enumerated at the national, provincial and territorial levels. A sample of individuals was drawn from six sampling frames independent of the 2021 Census. The data for the selected persons (SPs) were linked with tax data and other administrative sources to obtain recent information about their usual residence, contact addresses, household members, and related groups of persons.

A set of complex automated linkages and manual searches was done to find the SP in the 2021 Census Response Database (RDB). The census coverage studies (CCS), including the CUS, were carried out based on the version of the RDB that was available in mid-October 2021 (i.e., before the end of census processing). This version, which predates the final 2021 RDB, was called the CCS-RDB. There are a few minor differences between the CCS-RDB and the later versions of the census databases. The CCS-RDB, a database of persons, comprises all the records of enumerated persons, except three record groups: census records imputed through whole household imputation (WHI); all census records that were added late (after processing for the CUS began), but this did not happen in 2021 compared with the two preceding cycles; and, census records called “incomplete enumerations.” Section 7.4.6 provides more information on incomplete enumerations.

When a search produces no matches, multimode collection is done to determine whether the SP was a member of the target population and to get additional information (including addresses) to help find the SP in the CCS-RDB. At the end of the search, each SP is classified as out-of-scope (deceased, emigrated, temporarily outside Canada), enumerated or missed. A small number of non-response cases, consisting mostly of persons who could not be traced through collection, must be processed and are used to adjust respondent weights based on a non-response adjustment model.

7.1 Sampling

The sampling frame for the CUS target population, which includes all persons who should have been enumerated in the 2021 Census, is constructed from six frames independent of the 2021 Census. The first five frames were used to select a sample to estimate undercoverage in the 10 provinces, while estimates for the three territories were calculated using samples from the last frame only.

At the provincial level, sampling began with the persons who were in the 2016 Census target population. This includes all persons enumerated in the 2016 Census and all persons missed by the 2016 Census, represented by the portion of the sample of SPs in the 2016 CUS who were classified as missed. To account for persons added to the target population since the last census, intercensal (i.e., between the 2016 and 2021 censuses) births and immigrants were added, as were non-permanent residents as of Census Day in 2021. The data sources for these frames are as follows:

- Census frame: Persons who were enumerated in the 2016 Census and appear in the 2016 CCS-RDB.
- Missed frame: There is no comprehensive list of missed persons. However, there is a representative sample of these persons: the 2016 CUS sample of SPs classified as missed. They are all included in the 2016 sample with their 2016 weights.
- Birth frame: Vital statistics data on intercensal births. Since the final vital statistics file on births is only available late, the CUS sample of births is drawn from a mix of preliminary, final and raw vital statistics data files.
- Immigrant frame: Administrative data from Immigration, Refugees and Citizenship Canada (IRCC) on immigrants who arrived in Canada during the intercensal period.
- Non-permanent resident frame: Administrative data from IRCC on persons claiming refugee status on Census Day and persons with a valid work or study permit on Census Day.

For each territory, the main survey frame consisted of health insurance files for persons eligible for health care on Census Day. Although this frame has excellent coverage, it is incomplete, so the sampling weight must be adjusted. Each frame for a given territory is independent of the other territory frames and is used to estimate the undercoverage only for that given territory. In addition, the territory frames are not used to estimate undercoverage in the provinces. In the 2021 CUS, non-permanent residents in the territories who had work or study permits and were not already included in health insurance files were added to the territory frames.

None of the first five frames for the provinces covered persons who had emigrated or who were outside Canada during the 2016 Census and did not complete a 2016 Census questionnaire and who returned during the intercensal period ("returning Canadians within a province"). According to the 2021 Census long-form questionnaire, the number of persons in this group was estimated at 252,089. In addition, the number of persons returning from a territory to a province was estimated at 13,426. Added to this number were 120 persons from reserves and settlements that were incompletely enumerated in 2016 and enumerated in 2021, and 8,489 persons from reserves or settlements who had returned in 2016 and were enumerated in 2021, but who were excluded from the 2016 Census frame. Also, persons born after the 2016 Census outside Canada or in the territories who have Canadian citizenship and who returned to one of Canada's 10 provinces by Census Day in 2021 were not covered by the first five CUS census frames. According to the 2021 Census long-form questionnaire, the number of persons in this group was estimated at 16,925. Coverage error estimates do not include these populations, estimated at a total of 291,049 persons.

One problem with using multiple sampling frames is the possibility that the same person could be included in more than one frame. For example, a person in the immigrant frame may have been in Canada on a work permit in May 2016 and therefore may have been enumerable in the 2016 Census. That person would then be in both the immigrant frame and the census frame if they were enumerated, or in the missed frame if they were not enumerated. Consequently, it is important to identify all cases of frame overlap. Otherwise, estimates may be too high because some people are included twice in the frames. Whenever possible, this overlap is identified when the sampling frames are constructed, but some overlap is also identified later using information provided by respondents.

The sample design varied by frame depending on the type of list used. A one-stage stratified design was used for the 2016 Census frame. The stratification methodology was significantly changed during the 2021 CUS. Prior to stratification, several deterministic linkages were done. First, there was a linkage of the frame with the tax data, and over 96% of the persons were linked. Then there was a linkage with the vital statistics death files. There was also a linkage with IRCC files to find non-permanent residents in the frame. Finally, there was a linkage with the 2021 RDB using the monster match program, which is also used for the processing of the CUS sample. This process provides suggestions for potential enumeration and an indicator of the strength of this suggestion. Some suggestions are strong enough to consider the enumerated person without having to check the suggestion. These cases are called self-enumerations. Following these linkages, the frame was stratified. Two take-all strata were created: the deceased stratum and the self-enumerated stratum. Next, six take-some strata were created taking into account the probability of enumeration of persons (strength of the suggestion in the 2021 RDB), the tax situation and the likelihood of being out of scope of the census. However, enumerated persons on reserves and settlements in the 2016 Census were placed in separate strata using the same criteria, but by grouping some strata together as the population is smaller and more homogeneous.

Second, the take-some strata were stratified by province. For those residing in the six smallest provinces in 2016, the stratification province was the province of residence in 2016 (in the 2016 RDB). For persons in the four largest provinces in 2016, the derivation of the stratification province varied by stratum. In the strata with high probability of enumeration in the 2021 RDB, the province of potential enumeration in the 2021 RDB was used. Otherwise, where the person was linked to the tax data, the most recent province of residence based on these data was used. As a last resort, the province listed in the 2016 RDB was used.

The missed frame is a sample-based frame because there is no list of all persons missed in the 2016 Census. The sample for this frame consists of all cases classified as "missed" in the 2016 CUS. Although the sample was not stratified as such, implicit stratification was inevitable because the 2016 missed cases were from different frames and strata.

To construct the birth frame, copies of intercensal birth registrations were obtained from vital statistics through the National Routing System, which provides faster access to these data. The frame contains all births between May 10, 2016, and May 10, 2021, inclusively. The frame was then stratified by the mother's province of residence or province of birth if this data was not available.

The immigrant frame was constructed with records from IRCC. The immigrant frame contains all persons who immigrated to Canada between May 10, 2016, and May 10, 2021, inclusively. Those who were non-permanent residents on Census Day in 2016 were removed from the 2016 immigrant frame because they were already covered by the 2016 Census frame or by the 2016 missed frame. The immigrant frame was stratified by province. The province was derived based on information available in an address file provided by IRCC and in the IRCC immigration file. The most likely province of residence on Census Day in 2021 was selected. Then, immigrants from all provinces were separated into two strata by their immigration date. The first stratum consisted of immigrants who arrived between May 10, 2016, and April 30, 2020, and the second consisted of immigrants who arrived between May 1, 2020, and May 10, 2021, because newer immigrants are usually more likely to be missed in the census.

The non-permanent resident frame (persons who hold a work or study permit and refugee claimants) was constructed with IRCC records. Non-permanent residents as of Census Day in 2016 and intercensal immigrants were removed from the 2021 non-permanent resident frame. The frame was stratified by province, according to the most likely province of residence on Census Day in 2021. To this end, a deterministic linkage of the frame with the tax data was done. The IRCC address file and the various IRCC non-permanent resident files were also used. At the end of the process, a number of non-permanent residents had no associated provinces of residence (residents with an open permit), so they were placed in a national stratum.

In the provinces, the total size of the 2021 sample was determined to achieve two main objectives. First, the 2021 CUS collection budget was to remain the same as the 2016 CUS collection budget (but adjusted for unit cost increases between 2016 and 2021). Only a portion of the persons in the sample required collection, and proportions varied by frame and stratum. Second, the CUS sought to obtain standard errors in the rate of similar undercoverage among provinces of comparable size. The aim was to produce smaller standard errors for the larger provinces than for the small provinces as this would help to obtain a small standard error at the national level. Where possible, standard errors were not to be higher than those obtained in 2016.

Starting in 2020, by constantly updating the parameters used to calculate the standard error of undercoverage and the number of persons requiring collection, sample size simulations by frame and stratum were done to calculate the appropriate standard errors at all levels (national, provincial, age and gender). The frames and results of the 2016 CUS were used to make these simulations. Since some survey frames were ready before others, sample sizes were determined for these frames before establishing sizes for other frames and strata. Among other things, the sample size of the stratum for the 2016 missed frame was already set because everyone who was classified as "missed" in the 2016 CUS was selected. Then, the size of the first stratum of the immigrant frame was determined in the summer of 2020, and so on for the other strata and frames (births and non-permanent residents). The sample allocation was completed in November 2021 with the stratification of the 2016 Census frame as described above.

In several strata, a total size was determined for all ten provinces, and then a power-allocation scheme was used to allocate the total sample among the provinces. Minimum sample sizes were also set in the smallest provinces.

In addition, for some strata of the sampling frame, sub-stratification by sex and age group was performed to ensure that there were sufficient numbers of persons missed from these domains. Similarly, the allocation of the sample to the reserve strata of the census frame was carried out to obtain clarification on the undercoverage in the reserves at least as good as in the 2016 CUS. The final total allocated sample was 32,534 SPs across the frames in the provinces. Table 7.1.1 shows the final sample allocation by stratum for all provinces. According to this sample allocation, the target standard errors for the undercoverage rate ranged from 0.16% to 0.42% at the provincial level, and was 0.09% for the provinces as a whole. It should be noted that the resulting allocation does not guarantee that this level of precision will necessarily be achieved, because assumptions have been made about several parameters that are included in the calculation of the standard error of the undercoverage (strata and frame sizes, missed rate, CUS collection response rates, etc.). In addition, the effects of the COVID-19 pandemic may have affected the accuracy of these assumptions, including the number of immigrants and non-permanent residents, interprovincial migration and missed rates in the 2021 Census.

Table 7.1.1
Sample allocation, sampling frames and strata for all provinces

Sampling frames	Strata within each province	Number of people
Take-all total	**…**	**26,944,027**
2016 Census	Deceased	1,239,662
	Auto-enumerated in a province	25,704,365
Take-some total	**…**	**32,534**
2016 Census	Off reserves TS_1: Strong suggestions of enumeration	5,559
	Off reserves TS_2: Strong suggestions of incomplete enumeration	369
	Off reserves TS_3: High probability of being out of scope	510
	Off reserves TS_4: Medium suggestions of enumeration	757
	Off reserves TS_5: High probability of being missed	5,041
	Off reserves TS_6: Others	1,712
	Reserves TS_7: Strong or medium suggestions of enumeration	270
	Reserves TS_8: High probability of being missed	505
	Reserves TS_9: Others	200
	Reserves TS_10: Newfoundland and Labrador and Prince Edward Island	60
2016 missed	No further stratification	4,821
Births	No further stratification	5,978
Immigrants	Between May 10, 2016, and April 30, 2020	2,593
	Between May 1, 2020, and May 10, 2021	588
Non-permanent residents	No further stratification	3,571

... not applicable

TS = take-some

Source: Statistics Canada, 2021 Census Undercoverage Study.

The sampling methodology for the territories was similar to that of the census frame for the provinces. The persons included in the sampling frame for each of the territories were linked to the tax data and then to the 2021 RDB, using the monster matching process, which is also used for the processing of the CUS sample (see Section 7.2.1). Following these steps, the frame was stratified, taking into account the strength of the linkage with the 2021 RDB, the location of the enumeration and recent fiscal activity. A take-all self-enumeration stratum in the territory was formed, and six take-some strata were formed (see Table 7.1.2). For the first and sixth strata, a sub-stratification by sex and three age groups (0 to 17 years, 18 to 29 years and 30 years of age and older) was performed.

For sample allocation to the territories, the first step was to determine the total sample to be allocated to each territory in order to achieve similar and adequate precision of the undercoverage. In 2021, the target standard error for the undercoverage rate was approximately 0.40% in Yukon and the Northwest Territories (an improvement from 2016) and 0.60% in Nunavut (similar to 2016). Using the results of the 2016 CUS, assumptions of missed rates, undercoverage rates, and others were calculated for each stratum. For the first take-some stratum, the sample size was set manually in each territory as this stratum had very little effect on the accuracy of the undercoverage rate but more impact on the accuracy of the enumeration rate. This is important for the calculation of a calibration factor at the time of weighting. In addition, the workload of the employees who had to check the sample of this stratum had to be taken into account. Similarly, a sample was manually set for the fourth stratum as it represented persons who are almost certainly out of scope, but who are subject to some research work by CUS's employees. Then, iteratively, an optimal distribution of the total sample was made among the other take-some strata, including the six substrata of the last stratum. An approximate total size was initially set, then the accuracy of the optimal distribution was calculated, and this was repeated by increasing or decreasing the total size until the desired precision for the undercoverage rate in each territory was obtained. The final total allocated sample was 4,285 SPs across the frames in the territories.

Table 7.1.2 shows the allocation by stratum for all territories.

Table 7.1.2
Sample allocation, strata by territory

Strata	Yukon	Northwest Territories	Nunavut	Total
Take-all: Auto-enumerated within its territory	**27,881**	**26,696**	**16,981**	**71,558**
Take-some total	**1,156**	**1,331**	**1,798**	**4,285**
TS_1: Strong suggestions of enumeration	530	440	468	1,438
TS_2: Medium suggestions of enumeration	57	196	356	609
TS_3: Strong suggestions of incomplete enumeration	30	30	44	104
TS_4: Strong suggestions of enumeration outside its territory	53	78	70	201
TS_5: High probability of being out of scope	97	83	96	276
TS_6: High probability of being missed (substratification)				
Females, 0 to 17 years	30	59	158	247
Females, 18 to 29 years	48	44	69	161
Females, 30 years and older	109	117	157	383
Males, 0 to 17 years	33	61	132	226
Males, 18 to 29 years	54	61	65	180
Males, 30 years and older	115	162	183	460

TS = take-some
Source: Statistics Canada, 2021 Census Undercoverage Study.

Table 7.1.3 shows the sample allocation for Canada, the provinces and the territories.

Table 7.1.3
Sample size for Canada, provinces and territories

Provinces and territories	Take-all strata (number of people)	Take-some strata (number of people)
Canada	**27,015,585**	**36,819**
All provinces	**26,944,027**	**32,534**
Newfoundland and Labrador	393,554	1,551
Prince Edward Island	106,063	1,437
Nova Scotia	696,275	1,943
New Brunswick	579,964	1,680
Quebec	6,668,208	4,298
Ontario	10,415,555	7,126
Manitoba	947,750	2,579
Saskatchewan	794,538	2,540
Alberta	2,940,437	4,215
British Columbia	3,401,683	5,015
NPR-CA	0	150
All territories	**71,558**	**4,285**
Yukon	27,881	1,156
Northwest Territories	26,696	1,331
Nunavut	16,981	1,798

NPR-CA = non-permanent residents without a known province
Source: Statistics Canada, 2021 Census Undercoverage Study.

A systematic sampling method within the strata was used to select samples. Here is the list of sorting variables used to obtain an efficient sample (implicit stratification), classified by sampling frame:

- 2016 Census frame: sex, age, Code M,[10] 2016 geography, tax situation, reason for potentially being out of scope and likely province in 2021 (if stratified in the six smallest provinces);
- Birth frame: age on Census Day, sex, age group of mother and postal code;
- Immigrant frame: age group, sex and country of birth;
- Non-permanent resident frame: type of permit, age group, sex and country of birth;
- Territories frame: sex, age, code M, tax situation and municipality of residence.

No sampling was required for the 2016 missed frame, as all persons missed in the 2016 CUS were selected from the 2021 CUS sample.

Following the selection of provincial and territorial samples, these samples must be prepared by checking the quality of information for the different variables of interest (i.e., geographic and demographic variables); for example, the accuracy of names and the validity of birth dates were checked. Addresses were standardized to facilitate subsequent processing activities. To update the geographic information, especially for the census sample and the missed persons whose information was from 2016, these were linked with the Canada Revenue Agency (CRA) records, including personal income tax records for 2015 to 2021 and Canada Child Tax Benefit records for 2016 to 2022. CRA files and vital statistics data were also used to check whether any selected persons had died. This preparation stage was important because it helped to determine the persons enumerated in the census frames, and to contact persons not found and interview them.

7.2 Processing and classification

7.2.1 Processing

The objective of processing is to provide information for the classification of SPs for the purposes of non-response adjustment and estimation. Specifically, processing is carried out to:

- determine whether the SPs are enumerated in the Census Response Database
- determine whether the SPs are in the census target population
- provide further information for non-response adjustment.

The processing results were recorded in a classification assigned to each SP for estimation and tabulation purposes (see Section 7.4 and Section 9).

Most of the processing work involved automated and computer-assisted searching of the census coverage studies version of the 2021 Census Response Database (CCS-RDB) to determine whether the SP was enumerated.

Various elements of information were used for searching, including surnames, given names and birth dates. Telephone numbers and addresses associated with the SP or members of their household were also used. Questionnaires in which the SP could have been listed were identified from a variety of sources, including the following:

- matches with the CCS-RDB using the birth date and sex of the SP and members of the household, or the SP's name, postal code or telephone number;
- selection addresses from the sampling frame;
- address updates from tax records;
- information from the computer-assisted telephone interview (CATI) (see Section 7.3).

10. Code M is a code associated with each enumeration suggestion from monster matching, based on the strength of the link produced.

The first step after sample preparation was to search the CCS-RDB for each SP by processing all SPs with the addresses available from the sampling frame and tax data. There were two outcomes. When the SP was found, they were usually classified as “enumerated,” and no further processing was required, except for SPs who were later identified through vital statistics information as being deceased before the census. When the SP was not found, the case was sent for collection. While collection was taking place, the CCS-RDB search continued. When CATI data were available, researchers could determine whether each SP was part of the census target population. If so, the CATI data could enable further searching.

Searching for the SP was done both automatically and manually by coding staff guided by subject matter experts. To ensure coding uniformity, coding staff were provided with a highly detailed procedure manual that spelled out the specific steps for coding the search results. Automated searches were conducted first. For addresses obtained from a match with the CCS-RDB, there was a corresponding census questionnaire. A measure of similarity between the census questionnaire and the data available for the survey was calculated. When this measure was above a specified threshold, it was automatically concluded that the SP was enumerated at that address. In these cases, neither this address nor the SP’s other addresses needed to be processed by the coding staff. Computer programs also determined when one address was a duplicate of another. These duplicate addresses also did not need to be processed.

For other cases, a manual linkage was conducted using DocLink’s Interactive Verification Application (DIVA), an application developed specifically for this operation. The coding staff used a number of tools for this process, such as Geographical Reference Files, electronic telephone directories and the Street Attributes File. There were often suggested census questionnaires or census collection units that matched the address that was used as the first step for searching. Staff could also search the CCS-RDB using flexible parameters further in the process (searching by name, date of birth, etc.). The results of the manual search were then automatically edited via DIVA built-in edits to minimize errors. A file containing the search results was then produced. The data from this file were used to classify SPs.

7.2.2 Classification

Processing provides the information required to determine whether SPs were:

- included in the “census target population” or “out of scope” (not included)
- “classified” or “not classified”
- “listed” or “not listed”
- “identifiable” or “non-identifiable”
- “enumerated”
- “missed.”

Some SPs fit into more than one category, which will be explained in greater detail in this section.

7.2.2.1 “Target population” or “out-of-scope” classification

The “census target population” includes the group of persons mentioned in Section 2.2. An SP is considered “out of scope” if they are not in the census target population. Each SP classified as “out of scope” is assigned one of the following statuses: deceased, emigrated or represented in another frame. For a person to be classified as deceased, they must appear as deceased in at least two administrative sources (vital statistics death files, income tax files, death files), or in the CUS collection interview. Permanent or temporary emigrants were also determined through a collection interview based on certain criteria and the response on their place of residence on Census Day, the amount of time spent outside Canada, their intention to return to live in Canada and the reason they were outside Canada on Census Day. Other SPs were also classified as “listed emigrants,” regardless of whether they were respondents during collection. These are non-permanent residents (from the 2016 Census and missed frames) who no longer had a work or study permit in 2021 or immigrant status since 2016.

SPs classified as “represented in another frame” includes cases selected in a province but classified in one of the three territories. Cases selected in a territory but classified in a province or another territory are also classified as “represented in another frame.”

SPs classified in the census target population were either “enumerated,” “missed” or “not classified” (see Section 7.2.2.2). An SP was considered “enumerated” if they were in the CCS-RDB. SPs in the census target population were classified as “missed” if they were not enumerated or “not classified.”

7.2.2.2 Classification for non-response and non-response adjustment

Whether an SP was classified as “listed” or “not classified” depended on the usefulness of the addresses provided and the CATI information. In many cases, collection provided information and one or more addresses that could not be found from other sources. In other cases, all the addresses and all the information obtained through collection could be found from other sources.

An SP was “listed” if they were classified without using CATI data; even if data were collected, the addresses and information collected through the interview were not required.

A person was considered “not classified” if it was possible to determine whether they were in the target population but not whether they were missed. This occurred when the place of residence on Census Day, as defined in Section 2.4, was known but not identified in the CCS-RDB. Persons whose place of residence on Census Day was not specific enough (e.g., only the name of a large city) and persons without a fixed address were included in this category.

SPs for whom one or more of the characteristics in the list above could not be determined were considered non-respondents. There are three types of non-respondents:

- An SP was “not identified” when it could not be determined whether they were listed. In other words, since the information about the SP was incomplete, it was impossible to link the SP with the CCS-RDB or to collect their information through an interview.
- An SP was “not traced” when it could not be determined whether they were included in the census target population.
- A “not classified” SP was deemed to be partial non-response. It was known that the person was in the target population but not whether they were missed or enumerated.

7.2.2.3 Distribution of the sample by classification

Table 7.2 shows the distribution of the sample by classification and sampling frame. This table excludes persons in the take-all strata as these persons were classified (enumerated or deceased) prior to sample selection. Classification is determined from specific combinations of the characteristics of the list presented above. Initially, a total sample of 36,819 SPs was selected in the provinces and territories. Of that number, 22,083 SPs were classified as “enumerated,” 7,453 as “missed,” and 5,171 as non-respondents, of which 169 were classified as “not classified.” The other 2,112 SPs were classified as “out of scope,” specifically 583 “deceased,” 938 “emigrants” (permanent or temporary), 405 persons outside the universe of the territories or provinces, and 186 persons, for other reasons. A non-response adjustment was made during estimation (see Section 7.4). It is important to note that for the purposes of classification and, therefore, estimation, the definition of a non-respondent differs from the usual definition of a non-respondent that data collection is attempted but not completed. This is because classification is based on data from several sources, including collection. To prevent any confusion, Section 7.3 on collection refers to “completed collection” rather than “response.”

7.2.2.4 Implications of the classification

"Traced" SPs are SPs for whom it was possible to determine whether they were included in the census target population. For purposes of estimation and tabulation, traced SPs who were also classified were the respondents. Since names, including those of household members, and addresses were available in the CCS-RDB, and since the tools for consulting the database were sufficiently powerful, it was possible to verify whether an SP was enumerated at an address even if the address provided was vague.

The usefulness of knowing whether an SP was enumerated is self-evident. SPs who were in the census target population but who were not enumerated and were therefore classified as "missed" formed the basis for the undercoverage estimate. We also wanted to classify SPs according to the above-mentioned characteristics so that the most appropriate respondents could be chosen to represent non-respondents.

Lastly, except for SPs who were not classified, the Census Day address (usual place of residence) of each SP in the census target population was determined. This is the address where, according to census instructions, the SP should have been enumerated. If the SP was enumerated, the enumeration address was considered to be the Census Day address, despite other information provided that may suggest that the census instructions were not well understood.

For more information on processing and classification, see Parenteau (2023).

Table 7.2
Classification of selected people, sampling frames for Canada

	Provincial strata									
	2016 Census[1]		2016 missed		Births		Immigrants		Non-permanent residents	
Classification	number	percent	number	percent	number	percent	number	percent	number	percent
Total	**14,983**	**100.0**	**4,821**	**100.0**	**5,978**	**100.0**	**3,181**	**100.0**	**3,571**	**100.0**
Enumerated	**7,354**	**49.1**	**3,122**	**64.8**	**5,210**	**87.2**	**2,610**	**82.0**	**2,015**	**56.4**
Listed	7,201	48.1	3,112	64.6	5,206	87.1	2,604	81.9	1,997	55.9
Not listed	153	1.0	10	0.2	4	0.1	6	0.2	18	0.5
Missed	**4,156**	**27.7**	**710**	**14.7**	**432**	**7.2**	**284**	**8.9**	**630**	**17.6**
Listed	821	5.5	86	1.8	68	1.1	22	0.7	49	1.4
Not listed	3,335	22.3	624	12.9	364	6.1	262	8.2	581	16.3
Out of scope	**882**	**5.9**	**433**	**9.0**	**102**	**1.7**	**100**	**3.1**	**188**	**5.3**
Listed	505	3.4	327	6.8	79	1.3	10	0.3	104	2.9
Not listed	377	2.5	106	2.2	23	0.4	90	2.8	84	2.4
Non-response	**2,591**	**17.3**	**556**	**11.5**	**234**	**3.9**	**187**	**5.9**	**738**	**20.7**
Traced not classified	87	0.6	17	0.4	17	0.3	2	0.1	10	0.3
Identified not traced	2,492	16.6	539	11.2	217	3.6	185	5.8	728	20.4
Not identified	12	0.1	0	0.0	0	0.0	0	0.0	0	0.0

Table 7.2
Classification of selected people, sampling frames for Canada

	Territorial strata			
	Territorial frames[1]		Total	
Classification	number	percent	number	percent
Total	**4,285**	**100.0**	**36,819**	**100.0**
Enumerated	**1,772**	**41.4**	**22,083**	**60.0**
Listed	1,760	41.1	21,880	59.4
Not listed	12	0.3	203	0.6
Missed	**1,241**	**29.0**	**7,453**	**20.2**
Listed	238	5.6	1,284	3.5
Not listed	1,003	23.4	6,169	16.8
Out of scope	**407**	**9.5**	**2,112**	**5.7**
Listed	293	6.8	1,318	3.6
Not listed	114	2.7	794	2.2
Non-response	**865**	**20.2**	**5,171**	**14.0**
Traced not classified	36	0.8	169	0.5
Identified not traced	829	19.3	4,990	13.6
Not identified	0	0.0	12	0.0

1. Excluding the take-all strata.
Source: Statistics Canada, 2021 Census Undercoverage Study.

7.3 Collection

7.3.1 Overview

Head office staff in Ottawa worked closely with staff in the Statistics Canada regional offices (ROs) to collect data during the survey phase of the Census Undercoverage Study (CUS). The suggestions and recommendations made by the ROs as a result of conducting the 2016 CUS were incorporated into the design and operations of the 2021 survey.

The main purpose of the CUS is to find (trace) the correct selected persons (SPs) and collect demographic and address information so they can be classified as enumerated, missed or out of scope for the census. The classification results are used to estimate the number of persons who were missed, or undercovered, in the census. To help find and classify the SPs, the Census Day address and household composition were collected, as well as any other address where the SP may have been enumerated. Other information, such as the SP's mother tongue, was also collected for the coverage study tables.

The CUS take-some sample size was 36,819 (Section 7.1 describes the sample design). Pre-collection processing attempted to find these cases on the CCS-RDB, in vital statistics and in other administrative files. The cases that were matched or found in those files, and that could thus be classified as either enumerated or deceased before Census Day, were not sent to collection. All other cases that were not classified were sent to collection. The total number of cases sent to collection (the collection sample size) was 13,096. During the collection period, the processing team continued to try to match some of the cases, and those that could be classified were removed from collection (see Table 7.3.2 for these counts).

By design, collection was by proxy for SPs who were younger than 18 years. Proxy respondents were also used when the SP was not available during the collection period or was difficult to reach. Overall, 34% of the completed cases were by proxy, and a higher percentage of proxy cases were completed by interviewers than by self-response.

For deceased SPs, it was important to determine whether they had died before, on or after Census Day, since different questionnaire flows were used, depending on the date of death. In some cases—for example, by matching tax records and vital statistics—SPs were determined to be deceased before Census Day, prior to collection. These cases were not sent for collection. However, when in doubt, cases were sent for collection with a note indicating that the SP may be deceased.

It was imperative that the correct SP (or a proxy for the correct SP) be interviewed. If data were collected about the wrong person, the matching and resulting classification would be incorrect. The computer-assisted telephone interview (CATI) system was designed to instruct interviewers to verify that the person they were interviewing was the correct SP at the beginning of the interview. If an interview was completed with someone other than the SP (e.g., someone with a similar name and date of birth), the case was sent back to collection to be completed with the correct person.

The CUS is a mandatory multi-mode survey. The main data collection mode is CATI, and the secondary mode is self-enumeration. For 2021, the CUS used web-based electronic questionnaires for both modes as it transitioned to the Integrated Collection and Operation System, which is a standardized collection application developed at Statistics Canada. Previously, the CUS self-response mode used paper questionnaires. The transition to an electronic questionnaire was a big improvement, as it decreased respondent burden and reduced operating time and costs associated with mailing out paper questionnaires and manually entering the returned data.

The third collection mode was personal visits by field interviewers. The plan for the 2021 CUS was to continue to use field interviews in a limited scope, as in previous cycles (in the 2016 CUS, only 0.5% of cases were completed by field interviewers), but instead of the paper questionnaires that were used in the past, field interviewers would have used a laptop and the same application as telephone interviewers. However, all in-person interviews were cancelled at the collection planning stage because of the COVID-19 pandemic.

7.3.2 Operations

Data collection for the CUS began in all ROs on March 28, 2022. The last day of active collection was November 4, 2022. Table 7.3.2 shows the distribution of cases loaded into CATI from head office over time. The majority of cases were sent at the start of collection on March 28 and consisted of adult cases from all frames except Nunavut. The adjusted total represents the number of cases sent to collection, excluding the cases removed from collection.

Table 7.3.2
Total cases in collection

Description	Count
Cases started March 28, 2022: Adults in all frames except Nunavut	9,922
Cases started April 27, 2022: Minors in all frames (including most of the birth frame) except Nunavut	1,822
Cases started June 6, 2022: Nunavut frame and remaining birth frame cases	1,352
Total cases sent	**13,096**
Cases dropped by head office: Collection no longer required (classified in processing as either enumerated or out of scope)	309
Adjusted total	**12,787**

Source: Statistics Canada, 2021 Census Undercoverage Survey.

Introductory letters explaining the CUS and advising the SP (or proxy) that they had been selected for the survey were sent for all cases that started collection in March and April and that had a valid mailing address. A phone number was provided if they had any questions or if they wanted to call the RO to complete the survey. Cases without a contact phone number (requiring tracing) were also provided with a secure access code and a link to the self-response questionnaire. Introductory letters were not sent for the cases starting in June; instead, they received the reminder letters sent in July. These reminder letters were sent for all cases not yet completed near the midway point of collection. A second reminder letter was sent one month later. All reminder letters contained secure access codes and links to the self-response questionnaire. New for the 2021 CUS, near the end of collection, email reminders were sent for all incomplete cases that had a valid email address.

Near the end of collection, in an effort to boost response rates, the Toronto and Western ROs began a process similar to the field interview visits done in the past. If there was an address for an SP close to where an interviewer was visiting for another survey, they would visit the address to try to find the SP. If they located the SP or confirmed that the address was the SP's residence, they requested a phone number and time for the RO to call back to complete the interview. If they were speaking to the SP, they could also provide a secure access code to complete the questionnaire online. If the SP was not there, the interviewer tried to collect any contact information that could be useful for tracing.

Data quality analysis was performed to verify the completeness and accuracy of each case. Cases with missing or ambiguous data in key fields, or where the data collected were for someone other than the SP, were reactivated and sent back to collection for follow-up. There were 41 reactivated cases in the 2021 CUS. Cases that passed the data quality analysis were compiled into batches for processing, as described in Section 7.2.1.

Quality management of the collection operation involved a two-day virtual training session for regional data collection managers, who in turn trained their interviewers. Weekly meetings between head office and ROs were held during collection to discuss progress and address any issues that arose. A ticket-based communication tool was used to centralize and facilitate communication between head office and ROs. It tracked all questions and issues and ensured that each one was resolved in a timely manner. RO managers allocated resources to the survey while balancing the needs of other surveys taking place in their region. Sustained efforts to interview persons who initially refused to participate in the survey improved response rates.

Detailed management reports were created at head office on a daily and weekly basis to document survey collection progress. The reports presented the number of cases collected and response rates by province of selection and sampling frame.

7.3.3 Tracing

As part of the sample preparation, cases were linked to tax and other administrative data to provide updated contact information for the SP and their household members. In some cases, initial CATI data were outdated or incomplete, and tracing was required. Tracing is the process of searching for contact information for either an SP or a suitable proxy, and it is a major part of the CUS.

Tracing leads were loaded into the CATI application as alternate contacts prior to collection, and additional leads were sent to the ROs as they were found in processing during the collection period. More tracing source files were sent to collection for the 2021 CUS (29 files, compared with 13 in 2016), and an improvement in processing meant that only new phone numbers and addresses were sent to the ROs, with no duplication of previous sources.

The CUS had agreements with and received tracing information from 11 provinces and territories, 9 of which used deemed employees. Head office sent files containing names of SPs, which were matched with health care files and sent back with updated contact information. Having a deemed employee meant that both the name and date of birth of the SP could be supplied, making it easier to match the files.

At the start of data collection, only 2.1% of the cases had insufficient contact information and needed to be traced. Because of the quality and quantity of tracing sources provided by head office, 90.6% of the completed cases used phone numbers that were provided by head office. Another 8.6% of the completed cases were contacted with a new phone number that was found by the RO tracing efforts, and a final 0.8% were completed when respondents called in to the RO.

7.3.4 Collection statistics

Many statistics were monitored throughout the data collection period, and they were analyzed after collection was completed.

Table 7.3.4.1 shows the provincial and territorial completion rates by collection method. Of the 7,702 completed cases, 87.6% were completed by CATI and 12.4% by online self-response.

Table 7.3.4.1
Completion counts and rates by collection method for Canada, provinces and territories of selection

Provinces and territories	Cases sent	Interviewer		Self-response		Total	
		Cases completed	Completion rate (%)	Cases completed	Completion rate (%)	Cases completed	Completion rate (%)
Canada	**12,787**	**6,745**	**52.7**	**957**	**7.5**	**7,702**	**60.2**
Newfoundland and Labrador	503	291	57.9	33	6.6	324	64.4
Prince Edward Island	487	278	57.1	52	10.7	330	67.8
Nova Scotia	620	374	60.3	37	6.0	411	66.3
New Brunswick	522	286	54.8	34	6.5	320	61.3
Quebec	1,315	769	58.5	93	7.1	862	65.6
Ontario	2,406	1,214	50.5	235	9.8	1,449	60.2
Manitoba	852	451	52.9	47	5.5	498	58.5
Saskatchewan	832	425	51.1	51	6.1	476	57.2
Alberta	1,375	703	51.1	100	7.3	803	58.4
British Columbia	1,746	828	47.4	152	8.7	980	56.1
Yukon	460	239	52.0	29	6.3	268	58.3
Northwest Territories	632	345	54.6	33	5.2	378	59.8
Nunavut	950	529	55.7	56	5.9	585	61.6
NPR-CA	87	13	14.9	5	5.7	18	20.7

NPR-CA = non-permanent residents without a known province
Source: Statistics Canada, 2021 Census Undercoverage Survey.

Table 7.3.4.2 shows the completion rates by sampling frame and collection method. As expected historically, the non-permanent resident frame had the lowest completion rate, 49.4%, as SPs in this frame tend to be more mobile and have less contact information, making tracing more difficult.

Table 7.3.4.2
Completion counts and rates by sampling frame and collection method for Canada

Sampling frames	Cases sent	Interviewer		Self-response		Total	
		Cases completed	Completion rate (%)	Cases completed	Completion rate (%)	Cases completed	Completion rate (%)
Total	**12,787**	**6,745**	**52.7**	**957**	**7.5**	**7,702**	**60.2**
2016 Census	6,773	3,720	54.9	482	7.1	4,202	62.0
2016 missed	1,310	691	52.7	84	6.4	775	59.2
Births	671	377	56.2	40	6.0	417	62.1
Immigrants	553	280	50.6	87	15.7	367	66.4
Non-permanent residents	1,438	564	39.2	146	10.2	710	49.4
Yukon	460	239	52.0	29	6.3	268	58.3
Northwest Territories	632	345	54.6	33	5.2	378	59.8
Nunavut	950	529	55.7	56	5.9	585	61.6

Source: Statistics Canada, 2021 Census Undercoverage Survey.

Table 7.3.4.3 shows the completion rates by sex and age group. The lowest completion rates were for both sexes aged 20 to 44 years, and the best rate was for females aged 45 years and older.

Table 7.3.4.3
Completion counts and rates by collection method, sex and age group for Canada

		Interviewer		Self-response		Total	
Sex and age groups	Cases sent	Cases completed	Completion rate (%)	Cases completed	Completion rate (%)	Cases completed	Completion rate (%)
Both sexes	**12,783**	**6,745**	**52.8**	**957**	**7.5**	**7,702**	**60.3**
0 to 19 years	1,930	1,062	55.0	140	7.3	1,202	62.3
20 to 29 years	2,420	1,198	49.5	169	7.0	1,367	56.5
30 to 44 years	4,697	2,303	49.0	389	8.3	2,692	57.3
45 years and older	3,736	2,182	58.4	259	6.9	2,441	65.3
Males	**6,952**	**3,609**	**51.9**	**496**	**7.1**	**4,105**	**59.0**
0 to 19 years	963	530	55.0	74	7.7	604	62.7
20 to 29 years	1,273	623	48.9	99	7.8	722	56.7
30 to 44 years	2,678	1,305	48.7	199	7.4	1,504	56.2
45 years and older	2,038	1,151	56.5	124	6.1	1,275	62.6
Females	**5,831**	**3,136**	**53.8**	**461**	**7.9**	**3,597**	**61.7**
0 to 19 years	967	532	55.0	66	6.8	598	61.8
20 to 29 years	1,147	575	50.1	70	6.1	645	56.2
30 to 44 years	2,019	998	49.4	190	9.4	1,188	58.8
45 years and older	1,698	1,031	60.7	135	8.0	1,166	68.7

Note: This table excludes four cases for which the sex was unknown.
Source: Statistics Canada, 2021 Census Undercoverage Survey.

7.4 Estimation

The CUS estimate was divided into two parts. First, the SPs were weighted, and then the census undercoverage was calculated. Weighting involves determining the initial sampling weights of SPs, and all adjustments made to these initial weights, to create the SPs' final weights. Weighting involves several steps that are described in Sections 7.4.1 to 7.4.4. The methodology for calculating census undercoverage is described in Section 7.4.6.

7.4.1 Calculating the initial weights

For SPs of all sampling frames except the 2016 missed frame, initial weights were based on the inverse of the probability of being selected in the sample. However, the initial weight of an SP from the 2016 missed frame corresponds to the final weight assigned to it during the 2016 CUS when the SP was classified as "missed."

7.4.2 Initial weight adjustments

The weights of SPs from the 2016 Census frame who were enumerated more than once in 2016 were adjusted downward to account for the fact that these individuals had more than one chance of being selected.

Then, the initial influential weights in the 2016 missed frame were adjusted. The objective was to reduce the effect of high and influential weights on estimates and standard errors through the trimming of their initial weights. Some of the 4,821 people in the 2016 missed frame had a very high initial weight. The method used was to truncate weights to a multiplier of the median of weights in each trimming group formed. The trimming groups were formed by the province of selection and five age groups. The weight of a person with a weight above the threshold was reduced to that value. The truncated weights were redistributed evenly to other persons in the trimming group.

7.4.3 Non-response adjustment

To reduce statistical bias, the initial weights of respondents had to be adjusted to account for non-response. The weight of persons who could not be classified (non-respondents) was redistributed among persons who were classified (respondents). There are three types of non-response. First, there are the unidentified persons (only 12 SPs). The initial weights of these persons were transferred to identified persons in each sampling stratum.

The second type of non-response involves untraced persons (4,990 SPs). The adjustment involved forming response homogeneity groups (RHGs) among unlisted persons (listed persons being the persons classified without the help of CUS collection) and transferring the weight of untraced persons to unlisted traced persons within the RHGs.

The first step in the creation of the RHGs was to group unlisted persons (12,337 SPs) into main groups based on their estimated propensity to be in the target population. The groups were formed based on an analysis of the correlation between several tax indicators, particularly those for 2020 and 2021, and the final classification for unlisted traced persons. Up to seven main groups were created based on the sampling frame. These main groups were also strongly correlated with the likelihood to respond. The second step in creating RHGs was to group unlisted persons based on their likelihood to respond in each domain, with a domain being defined by crossing a sampling frame with a main group. In each domain, the likelihood to respond was analyzed using a national logistic regression model (and regional, when the data allowed it) and an analysis of multi-level, cross-frequency tables. For the models, several auxiliary variables available for both traced and untraced persons were used: variables available in the sampling frames (e.g., age, sex, relationship to other household members, country of origin, and type of non-permanent resident), variables available in the tax data for related persons (e.g., whether they were in certain files, frequency of address changes since 2016, and type of address), variables related to contact information (e.g., number and sources of telephone numbers, address availability and link of last known address with the 2021 Census), and a few other variables. Thus, the auxiliary variables that were significantly correlated with the likelihood to respond were determined and used to form the RHGs. In most domains, the RHGs were formed within the province or territory of selection. Therefore, the adjustment consisted of transferring the weight of untraced persons to unlisted traced persons within each RHG.

The third non-response adjustment was the adjustment for unclassified persons (169 SPs). An unclassified person is a person who had their primary residence in a given province or territory on Census Day (thus in the census target population), but for whom it was not certain whether they were missed or enumerated. Using the same principle as with untraced persons, homogeneous groups of classified persons were formed within each sampling frame and province of classification. The adjustment consisted of transferring the weight of unclassified persons to unlisted classified persons within each homogeneous group.

7.4.4 Final adjustments to the weights for classified persons

7.4.4.1 Adjustment for influential weights

At this stage, some SPs have a weight that is high and considered influential in their province of classification. To reduce the effect of high and influential weights on provincial estimates and their standard errors, an adjustment to influential weights was made in the five frames for the provinces. The method used was to trim weights by a multiplier of the median of weights in each trimming group formed. There are two types of influential weights at this stage.

First, there are SPs whose province of classification is different from the province of selection. Therefore, the weight is very high compared to other SPs in this province of classification. Consider, for example, an SP selected in Ontario with a large weight, who is classified in Prince Edward Island. In this situation, the weight is truncated according to the threshold established by trimming group. A factor between four and six times the median for each group was used as a pruning threshold. The trimming groups were formed according to the province of classification and five age groups. The truncated weights of an SP were redistributed evenly to the other SPs in the same province of selection, the same sampling frame, the same classification (enumerated, missed or out-of-scope person), the same status (listed or unlisted) and by age group. Therefore, the influential weight of a missed SP in a given province of classification was allocated to other missed persons, but in the province of selection of the SP. For this first type of influential weight, there were 49 SPs whose weight was truncated (i.e., 33 enumerated persons and 16 missed persons).

The second type of influential weight relates to the SPs from the 2016 missed frame only, who still had a high and influential weight within their province of classification even though it was identical to the province of selection (which is, in fact, the province of classification in 2016). For this type of influential weight, the threshold was set at four times the median weight in the trimming group. The truncated weights of SPs were redistributed evenly to the other SPs in the same province of classification and the same classification, thus having no effect on the estimate of provincial undercoverage. For this first type of influential weight, there were 95 SPs whose weight was truncated (i.e., 10 enumerated persons, 55 missed persons and 30 out-of-scope persons).

7.4.4.2 Weight calibration for the birth frame

For the birth frame sample, enumerated persons were calibrated to take into account cases where a provincial sample would contain too many or too few enumerated persons. An automated deterministic linkage applied to the 2021 CCS-RDB helped to determine the control totals per province for the enumerated persons calibration group. Then, for the other persons in the frame, a linkage to the tax data determined their province of residence on Census Day (otherwise, the province of selection was used) to determine the control totals per province for the non-enumerated persons calibration group. In addition, control totals by year of age (0-4 years) were calculated. The calibration was carried out using a raking mechanism for the margins using the 20 control totals described above as the first margin, and 5 calibration groups by age as the second margin. To this end, Statistics Canada's Generalized Estimation System (G-EST) was used.

7.4.4.3 Weight calibration for the immigrant frame

For the immigrant frame sample, a calibration of the number of persons in certain calibration groups was carried out to take into account cases where a provincial sample would contain too many or too few enumerated persons or persons in other groups. An automated deterministic linkage applied to the 2021 CCS-RDB helped to determine the control totals per province for the enumerated persons calibration group. Then, for the other persons in the frame, a linkage to the tax data determined their tax status (active or non-active) and their province of residence on Census Day (otherwise, the province of selection was used) to determine control totals by province for the other non-enumerated persons calibration groups. In the four largest provinces, three control totals were determined: for enumerated persons, for persons with recent fiscal activities, and for other persons. However, in the other six provinces, only two control totals were determined: for enumerated persons and for other persons. Thus, 24 control totals were formed. A simple poststratification method was then used to calibrate the immigrant frame.

7.4.4.4 Post-stratification adjustment for the territories

After the initial weight adjustment, the estimated number of enumerated persons in the territories was observed to be traditionally lower than the comparable census count. This was due to undercoverage of the census target population in health insurance files. To address this undercoverage, the weights of the SPs selected in each territory were adjusted so that the estimated number of enumerated persons equalled the comparable census count for that territory. The adjustments were made for six calibration groups (by age and gender) in each territory.

7.4.4.5 Adjustment for overlap of frames or strata

For a small number of SPs in the five provincial frames, the weight is not the final weight, as another adjustment must be made to take into account the overlap between the sampling frames or, in some cases, the overlap between the census frame strata (i.e., overcoverage in 2016), but which was noted only after the CUS collection in 2021. As for the few SPs who overlap frames, it is mostly SPs from the immigrant frame or the non-permanent resident frame who were finally taken into account in the 2016 Census frame (i.e., enumerated in 2016). This information was not known when these sampling frames were prepared. Therefore, an adjustment factor was calculated taking into account the probability of selection in both sampling frames.

7.4.5 Weighted distribution by classification

Table 7.4.5 shows the weighted distribution of SPs by classification and sampling frame. For a reminder of the definitions, see Section 7.2. Only SPs found in the CCS-RDB were classified as "enumerated." Persons who were in the target population but not in the CCS-RDB were classified as "missed." The remaining SPs were classified as "out of scope" (e.g., deceased or emigrated).

Table 7.4.5
Weighted classification of selected people, sample frames for Canada

	Provincial strata									
	2016 Census		2016 missed		Births		Immigrants		Non-permanent residents	
Classification	number	%	number	%	number	%	number	%	number	%
Total	**32,933,387**	**100.0**	**2,830,944**	**100.0**	**1,855,111**	**100.0**	**1,072,833**	**100.0**	**1,140,539**	**100.0**
Enumerated	**29,127,257**	**88.4**	**1,784,797**	**63.0**	**1,646,438**	**88.8**	**874,651**	**81.5**	**639,236**	**56.0**
Listed	29,023,031	88.1	1,773,009	62.6	1,643,876	88.6	871,537	81.2	626,405	54.9
Not listed	104,226	0.3	11,788	0.4	2,562	0.1	3,114	0.3	12,831	1.1
Missed	**2,083,885**	**6.3**	**662,494**	**23.4**	**164,767**	**8.9**	**130,942**	**12.2**	**387,586**	**34.0**
Listed	243,914	0.7	41,300	1.5	16,954	0.9	7,987	0.7	14,693	1.3
Not listed	1,839,971	5.6	621,194	21.9	147,813	8.0	122,955	11.5	372,893	32.7
Out of scope	**1,722,245**	**5.2**	**383,653**	**13.6**	**43,906**	**2.4**	**67,240**	**6.3**	**113,717**	**10.0**
Listed	1,402,710	4.3	206,632	7.3	25,675	1.4	3,964	0.4	32,613	2.9
Not listed	319,535	1.0	177,021	6.3	18,231	1.0	63,276	5.9	81,104	7.1

	Territorial strata			
	Territorial frames		Total	
Classification	number	%	number	%
Total	**137,867**	**100.0**	**39,970,681**	**100.0**
Enumerated	**94,583**	**68.6**	**34,166,962**	**85.5**
Listed	94,272	68.4	34,032,130	85.1
Not listed	311	0.2	134,832	0.3
Missed	**32,760**	**23.8**	**3,462,434**	**8.7**
Listed	5,567	4.0	330,415	0.8
Not listed	27,193	19.7	3,132,019	7.8
Out of scope	**10,524**	**7.6**	**2,341,285**	**5.9**
Listed	7,768	5.6	1,679,362	4.2
Not listed	2,756	2.0	661,923	1.7

Source: Statistics Canada, 2021 Census Undercoverage Study.

7.4.6 Calculating census undercoverage

Note the following definitions:

C = published census count of the number of persons in the target population

$\hat{U}$ = undercoverage estimate

= estimate of the number of persons not included in C who should have been

$\hat{M}$ = estimate of the number of persons in the CUS target population who were not enumerated

= sum of the final weight of persons considered to be missed

X = the number of persons included in C who could not be identified with certainty as enumerated in the CUS.

Census population undercoverage was estimated by the number (weighted) of missed persons less the number of persons counted in the census (term C) but excluded from the CCS-RDB:

$$\hat{U} = \hat{M} - X$$

X has three components: imputations, incomplete enumerations and late enumerations.

The SP's address on Census Day refers to a dwelling for which an enumeration was imputed. This was the case in particular for non-response dwellings for which another household's data were used in WHI.

Some enumerations in the census database were deemed too incomplete to be used by the CUS to determine whether an SP was enumerated. Incomplete enumerations in this context usually involve missing or invalid date of birth or name data (e.g., "?", "Mr.", "Unknown" or "Person 1"). An SP enumerated in this manner was classified as "missed." This was referred to as a "CUS incomplete enumeration." This category of enumeration also includes certain types of collective dwellings for which only the number of usual residents was collected in the census (no names or dates of birth). Data of people living in these collective dwellings was imputed from the RDB.

At the national level, X made up slightly less than half of $\hat{M}$. The value of X increased from 2016 because of an increase in the number of persons imputed as part of the WHI and the increase in imputations in certain types of collective dwellings (incomplete enumerations).

Table 7.4.6 shows the national numbers for the various components of the population undercoverage estimate, namely the numbers for the three components of the term X.

Table 7.4.6
Components of the population undercoverage estimate for Canada

Components	Number of people
Estimate of M	**3,462,434**
Total X	**1,564,558**
X for imputed people	931,346
X for late enumerations	0
X for CUS incomplete enumerations	633,212
Estimate of U	**1,897,876**

CUS = Census Undercoverage Study
M = number of people in the Census Undercoverage Study (CUS) target population who were not enumerated
X = number of people included in the published census count but who could not be identified with certainty as enumerated in the CUS
U = undercoverage
Source: Statistics Canada, 2021 Census Undercoverage Study.

Lastly, the variance of the undercoverage estimates was calculated as follows:

$$v(\hat{U}) = v(\hat{M} - X) = v(\hat{M})$$

$v(\hat{M})$ = estimated variance of $\hat{M}$ as determined by the CUS design.

The variance was calculated using the classic bootstrap resampling method. To that end, weights of 500 bootstrap replicates were produced.

8. Census Overcoverage Study

8.1 Overview

Overcoverage error occurs when in-scope individuals are enumerated more than once or when individuals who should not have been enumerated are included in the target population of a survey or a census. The purpose of the Census Overcoverage Study (COS) is to estimate the number of persons enumerated more than once in the Canadian Census of Population.[11]

The 2021 COS consisted of two types of linkages, namely deterministic and probabilistic. The deterministic linkage (DL) identified definite pairs of duplicate persons, meaning those persons were enumerated more than once and hence represent overcoverage. The methodology was based on a modification of the Automated Match Study (AMS), which had been used in previous census cycles to evaluate the COS. The probabilistic linkage (PL) identified possible pairs of duplicate persons and was based on the methods used in past cycles of the COS. The COS used data from the 2021 Census Response Database and administrative data from the Canadian Statistical Demographic Database provided by the Census Research section of the Statistical Integration Methods Division. The COS sampling frame was created in multiple steps and includes definite and possible pairs of duplicate persons identified with both the DL and PL, along with an extension of the sampling frame based on households. A sample of possible pairs of duplicate persons was drawn from the COS frame and sent for manual verification to determine whether the sampled pairs were indeed duplicate persons. With the result of the manual verification of the sampled pairs, and definite pairs of duplicate persons identified by the DL, an estimate of overcoverage was then obtained.

8.2 Linkage steps

8.2.1 Data used for the linkages

Two sources of data were used for the linkages.

Firstly, the Census Coverage Studies version of the Census Response Database (CCS-RDB, referred to as the RDB in this chapter) was a version of the Census Response Database that did not include late or incomplete enumerations, or persons added through the whole household imputation process. The RDB contained a little over 35 million records and included responses from individuals living in both private and collective dwellings. It contained names (including given names and surnames), demographic information (including date of birth and sex) and geographic information (including province or territory and postal code, and census geographic variables such as collection unit (CU), census subdivision (CSD) and census metropolitan area (CMA)).

Secondly, administrative (ADM) data were used, based on the Canadian Statistical Demographic Database provided by the Census Research section of the Statistical Integration Methods Division. They comprised records from multiple ADM data sources and aimed to represent persons in scope for the census. The ADM data consisted of around 53 million records. They included names (given names and surnames), demographic information (including date of birth and sex) and geographic information (including province or territory and postal code).

The following matching variables were used in the linkages (when applicable):

- names: given name(s) and surname(s) variables
- demographic data: date of birth and sex variables
- geographic data: province or territory and postal code, and census geographic variables.

11. Prior to 2016, studies conducted to determine the impact of including out-of-scope individuals have shown this error to be negligible, so it is not treated by the COS. However, a pilot study is under consideration after the 2021 production cycle to re-evaluate this assumption.

8.2.2 Deterministic linkage

The purpose of the DL was to identify high-quality pairs of duplicate persons, consisting of two records from the RDB, which were classified as definite pairs of overcoverage. The deterministic matching programs traditionally used for the AMS were modified to include, as part of the linkage criteria, a comparison of names and also considered matches between a household living in a private dwelling and a household living in a collective dwelling.

The DL was based on the following series of operations:

- Deterministic matching programs were used to identify household pairs that were "similar." Similarity was described in terms of their relative geographic proximity (households within the same CU, households in different CUs within the same CSD, etc.) and the number of persons matched between them. Persons were matched based on the variables of name, sex and date of birth. Two persons were said to be an exact match if they had the same sex; day, month and year of birth; and names also match. Two persons were said to be a near match if their names matched and three of the four other components (sex and day, month and year of birth) agreed or just the day and month of birth were reversed. Household pairs consisted of one or both households living in a private dwelling.
- An initial list of possible pairs of duplicate persons was created from household pairs.
- A verification sample was taken from the initial list of possible pairs of duplicate persons for manual verification purposes to confirm their high quality before classifying them as definite pairs of duplicate persons (i.e., overcoverage).
- A final list of pairs of duplicate persons was determined, and they were classified as definite pairs of duplicate persons resulting from the DL.

There were 460,572 definite pairs of duplicate persons resulting from the DL.

8.2.3 Probabilistic linkage

The purpose of the PL was to identify possible pairs of duplicate persons. The PL consisted of an internal probabilistic record linkage of the entire RDB to itself, referred to as the RDB–RDB linkage, and an external probabilistic record linkage of the RDB to ADM data, referred to as the RDB–ADM linkage. The RDB–RDB linkage resulted in pairs of RDB records, whereas the RDB–ADM linkage resulted in pairs where one record was from the RDB and the other record was from ADM data, and pairs of RDB records were later derived.

PL is conducted with G-Link, a probabilistic record linkage system designed at Statistics Canada that uses the Fellegi–Sunter method to solve large file linkage problems when there are no direct identifiers common to both sources (Fellegi and Sunter, 1969). As in past cycles, G-Link was used in 2021, and the following series of operations were done separately for the RDB–RDB and RDB–ADM linkages.

The first task in a probabilistic linkage is to build a set of potential pairs (also known as a **linked set**), which is used to estimate the characteristics of the set of **true matched** pairs. To do this, a set of selection criteria was applied, which reduced the Cartesian product of all the possible matches to a more manageable comparison space. Improvements were made in the 2021 selection criteria to overcome challenges that arose with the 2016 selection criteria. In addition, rather than use identical selection criteria for both the internal and external linkages, criteria were developed, tested and optimized separately for these two linkages. Many of the RDB pairs derived from the RDB–ADM linked set had corresponding RDB–ADM pairs that were captured by different criteria, suggesting that a direct comparison would not be able to capture them. The selection criteria of the internal RDB–RDB linkage returned a linked set of 86,429,651 RDB–RDB pairs. The selection criteria of the external RDB–ADM linkage returned 70,274,756 pairs. Of these, 41,474,581 involved multiple RDB records linked to the same ADM record; hence, the RDB–ADM linked set contained these 41,474,581 pairs.

Once a linked set of pairs was obtained, the records of the pairs were compared by applying linkage rules in G-Link, which calculated the weights of the results of the linkage rules. Quality linkage rules that address all sets of characteristics for which two records agree were necessary to ensure the completeness of the COS sampling frame resulting from the PL. If some sets of characteristics are not addressed by the linkage rules, then pairs with such characteristics are likely to be assigned a lower linkage weight and to be rejected when thresholds are applied. Many improvements were made in the 2021 linkage rules to ensure that estimated linkage weights were well correlated with the likelihood of a pair being a true match. In 2021, more linkage variables were added to the rules, and the outcomes for existing rules that had been used in 2016 were modified, such as the rules on names. Outcomes based on census-specific geographic variables—such as the unique identifier of a dwelling (known as the FRAME_ID), CU and CSD—were added in 2021 and were applicable only to the RDB–RDB linkage.

A linkage weight threshold for each province and territory was then established separately for the RDB–RDB and RDB–ADM linkages. The objective in choosing linkage weight thresholds was to optimally partition the pairs from the linked set into two classes: potential matched pairs and unmatched pairs. As in previous cycles, provincial and territorial thresholds were chosen outside G-Link because the built-in tools for finding thresholds did not work well with the many user-defined linkage rules used by the COS PL. The thresholds were selected in two steps. First, a set of preliminary thresholds was selected. In general, choosing a lower threshold is somewhat subjective. The COS employed the guidelines for profile reviews developed by the Social Data Linkage Environment (SDLE) experts at Statistics Canada to assist in choosing a preliminary lower threshold. To avoid missing potential overcoverage, a fairly low threshold was initially selected. Then a sample of pairs from above and below the preliminary threshold was selected, and the threshold was adjusted as required. The final threshold was chosen to minimize a target missed match rate of 0.01. Because a set of definite pairs of duplicate persons was obtained through the DL step, no upper threshold was selected. All the pairs from the RDB–RDB and RDB–ADM linked sets whose weight was greater than the threshold were selected and considered to be potential pairs.

8.3 Creation of the Census Overcoverage Study sampling frame

The COS sampling frame was created in multiple steps and included linked pairs identified with the DL and the PL, along with an extension of the sampling frame based on households. Then, sampling units are created.

As previously described, the DL was used to identify a set of RDB–RDB pairs that were classified as definite pairs of duplicate persons. The PL was used to identify a set of potential pairs. The internal linkage of the RDB to itself identified potential RDB–RDB pairs that were classified as possible pairs of duplicate persons. RDB–RDB pairs needed to be derived from the RDB–ADM linked set to include them in the COS sampling frame. Potential pairs identified through the RDB–ADM linkage were converted into RDB–RDB pairs. Where two RDB records were linked to the same ADM record, those two RDB records became an RDB–RDB pair. One-to-one RDB–ADM pairs were not of interest, as the goal was to measure overcoverage (i.e., duplicate persons) on the RDB. The final set of RDB–RDB pairs derived from the RDB–ADM linkage contained 4,301,512 pairs, which were classified as possible pairs of duplicate persons.

As in previous cycles, the frame was then enriched with additional pairs not already identified by the PL but created from the households of pairs linked by the internal and external linkage steps. The purpose of this step was to identify additional possible pairs of duplicate persons in the households of captured pairs that may not have been caught with the PL, because the PL was based on comparisons of individuals rather than households. Potential pairs from this step were known as extension pairs and classified as possible pairs of duplicate persons. To construct the set of extension pairs, a household pair was first produced for each RDB–RDB pair classified as a possible pair of duplicate persons resulting from the PL by adding the other household members to it. Second, sex and date of birth were used as variables to identify new RDB–RDB pairs by comparing the persons present in the household pair. Comparison rules were applied to identify pairs that might represent overcoverage cases. The extension pairs included pairs from two private households, or pairs where an individual from a private household was linked to an individual from a collective dwelling. Pairs where both records were from collective dwellings were excluded.

The final linked set comprised pairs from the DL, extension pairs, pairs from the PL of the RDB–ADM and pairs from the PL of the RDB–RDB.

Table 8.3.1
Breakdown of possible pairs of duplicate people by linkage type

Linkage type	Frequency	Percent
DL	460,572	3.62
Extension	471,688	3.71
RDB-ADM	4,301,351	33.80
RDB-RDB	7,491,998	58.87

DL = deterministic linkage
RDB-ADM = probabilistic linkage of Census Response Database to administrative data
RDB-RDB = probabilistic linkage of Census Response Database to itself
Source: Statistics Canada, 2021 Census Overcoverage Study.

When pairs in the PL set were also found by the DL, the linkage type was set to DL. Then, the possible pairs of duplicate persons obtained from the DL, the PL and the extension were combined and deduplicated.

Since 2011, the COS has used interconnected record groups to estimate overcoverage in the census rather than record pairs. This is because overcoverage estimated by record pairs would be positively biased in the presence of triple or higher-order enumerations. Thus, mutually exclusive groups of connected RDB records were formed, where most of the groups of records on the frame resulted in one or two pairs (involving two or three records). For cases where the groups of records contained more than 10 links, a graph theory method was applied to reduce the group into small subgroups called “neighbourhoods” (Dasylva et al., 2015) to facilitate manual verification.

Lastly, the COS sampling frame consisted of three types of sampling units: pairs, groups and neighbourhoods. Sampling units were categorized by three process types: (1) DL-only, composed of pairs and groups or neighbourhoods of RDB records resulting from the DL; (2) PL-only, composed of pairs and groups or neighbourhoods of RDB records resulting from the RDB–RDB linkage, RDB–ADM linkage and extension pairs; and (3) PL–DL, composed of groups or neighbourhoods of RDB records from both the PL and DL (including extension pairs).

Table 8.3.2
Distribution of deterministic linkage-only, probabilistic linkage-only and probabilistic linkage-deterministic linkage pairs, groups and neighbourhoods in the 2021 Census Overcoverage Study sampling frame

	Process type			
Sampling unit types	DL-only	PL-only	PL-DL	Total
Group	4,822	1,635,296	86,930	1,727,048
Neighbourhood	64	161,641	6,493	168,198
Pair	345,243	5,931,084	0	6,276,327
Total	**350,129**	**7,728,021**	**93,423**	**8,171,573**

DL = deterministic linkage
PL = probabilistic linkage
PL-DL = probabilistic linkage-deterministic linkage (some of the pairs in the group were identified by the probabilistic linkage only, while others were identified by the deterministic linkage)
Source: Statistics Canada, 2021 Census Overcoverage Study.

8.4 Sample design

The first level of stratification was by linkage process type, resulting in three strata:

- Stratum 1 consisted of DL pairs and groups or neighbourhoods made up of DL pairs only. This was treated as a take-all stratum, and sampling units in this stratum were classified as definite pairs of duplicate persons.
- Stratum 2 consisted of PL pairs and groups or neighbourhoods that contained only PL pairs. A probabilistic sample was drawn from this stratum, and the pairs were sent for manual verification.
- Stratum 3 consisted of groups or neighbourhoods that had a combination of PL and DL pairs. This stratum was further divided into two substrata. The first substratum was composed of groups and neighbourhoods that contained at least one DL pair that was sampled as part of the DL verification sample used to confirm the quality of these pairs. This substratum was treated as take-all. The second substratum was composed of groups and neighbourhoods that did not contain any DL pairs that were part of the DL verification sample. It had a probabilistic sample of PL–DL groups or neighbourhoods drawn from it. PL pairs in groups with DL pairs belonging to the first substratum were sent for manual verification, along with the PL and DL pairs selected from the second substratum.

The targeted sample size was approximately 55,000 pairs from the PL-only stratum and around 4,500 pairs from the PL–DL stratum. In this section, intraprovincial means all RDB records in a sampling unit are from the same province or territory, and interprovincial means RDB records in a sampling unit are from more than one province or territory. Tables in this section present counts of pairs whether the sampling unit is a pair, group or neighbourhood. Groups and neighbourhoods are broken down into their constituent pairs to derive the count of pairs. For simplicity, sampled pairs were sent for manual verification rather than groups of records.

For the PL-only stratum, the sampling unit type substrata were further stratified into 14 strata: 13 provincial strata containing sampling units (pairs or interconnected record groups or neighbourhoods) where all of the records belong to the same province or territory, and an interprovincial stratum where the sampling units have records from different provinces or territories. As in 2016, the interprovincial units may be groups that also contain some intraprovincial pairs. This was unavoidable when using interconnected record groups. To better control the sample size, the group and neighbourhood sampling units were further stratified by the number of pairs in the group. Finally, the sampling units were sorted by the estimated overcoverage propensity in the case of groups or neighbourhoods and by their conditional match probabilities in the case of pairs, and a systematic sample was then drawn.[12]

For the first PL–DL substratum, the DL pairs that were part of the verification sample had already been verified and so were not sent for manual verification. This was advantageous and allowed for a larger sample in the PL–DL substratum with at least one DL pair in the verification sample. The PL–DL groups for which none of the DL pairs were part of the verification sample were further stratified into 14 strata: 13 intraprovincial strata and an interprovincial stratum. As with the PL-only stratum, these 14 substrata were further stratified by the number of links to better control the sample size. The sampling units were then sorted by the estimated overcoverage propensity, and a systematic sample was drawn.

8.4.1 Deterministic linkage-only stratum

As mentioned above, the DL-only pairs and groups or neighbourhoods were considered definite matches and were not sent for manual verification. As shown in Table 8.3.2, there were fewer interconnected record groups among the DL pairs than among the PL pairs. In Table 8.4.1.1, which shows the breakdown of DL-only pairs by province or territory and interprovincial pairs, there were also fewer interprovincial DL-only pairs than interprovincial PL-only pairs (1.39% from Table 8.4.1.1 versus 18.34% from Table 8.4.2.2). This was what would be expected for pairs that were true matches.

12. For further details on how the overcoverage propensity was computed please request the Statistics Canada Internal document "*2021 Census Overcoverage Study (COS), Methodology Report". (2023)* by contacting infostats@statcan.gc.ca.

Table 8.4.1.1
Frequency of deterministic linkage-only pairs by province or territory and interprovincial strata

Provinces and territories	Frequency	Percent
Newfoundland and Labrador	5,133	1.42
Prince Edward Island	1,678	0.47
Nova Scotia	9,145	2.54
New Brunswick	8,230	2.28
Quebec	77,599	21.54
Ontario	122,913	34.12
Manitoba	11,908	3.31
Saskatchewan	13,159	3.65
Alberta	36,290	10.07
British Columbia	67,876	18.84
Yukon	463	0.13
Northwest Territories	494	0.14
Nunavut	391	0.11
Interprovincial	5,001	1.39

Source: Statistics Canada, 2021 Census Overcoverage Study.

8.4.2 Probabilistic linkage-only stratum

Table 8.4.2.1 shows the number of pairs for each sampling unit type and an estimate of the number of sampling units needed to obtain approximately that many pairs in the sample. The allocation to pairs and groups or neighbourhoods was proportional to size.

Table 8.4.2.1
Frequency of pairs, sampling units and sample sizes by sampling unit type

Sampling unit types	Number of pairs	Number of sampling units	Sample size (in terms of pairs)	Percent of sample (in terms of pairs)	Sample size (in terms of sampling units)
Group or neighbourhood	6,411,761	1,796,937	28,110	52	9,599
Pair	5,931,084	5,931,084	25,920	48	25,920
Total	**12,342,845**	**7,728,021**	**54,030**	**100**	**35,519**

Source: Statistics Canada, 2021 Census Overcoverage Study.

Probabilistic linkage-only pairs

The PL-only pairs were first stratified by intraprovincial and interprovincial pairs. Table 8.4.2.2 below gives the breakdown of intra- and interprovincial pairs among the PL-only pairs. Sample allocation to the intra- and interprovincial substrata was proportional to size.

Table 8.4.2.2
Frequency of intraprovincial and interprovincial pairs among probabilistic linkage-only pairs and sample sizes

Types of pairs	Frequency of pairs	Percent	Number of sampled pairs
Intraprovincial	4,843,438	81.66	22,004
Interprovincial	1,087,646	18.34	4,753

Source: Statistics Canada, 2021 Census Overcoverage Study.

Within the intraprovincial pair stratum, a power allocation was used to allocate the PL-only pairs across provinces, with the measure of size taken to be the number of pairs in each province and q = ½. The pairs were then sorted by their conditional match probabilities, and a systematic sample was drawn. Note that the three territories were take-all. Table 8.4.2.3 shows the allocation of PL-only intraprovincial pairs by province or territory.

Table 8.4.2.3
Probabilistic linkage-only intraprovincial pairs sample allocation by province or territory

Provinces and territories	Frequency
Newfoundland and Labrador	568
Prince Edward Island	300
Nova Scotia	760
New Brunswick	748
Quebec	6,915
Ontario	5,600
Manitoba	730
Saskatchewan	701
Alberta	1,740
British Columbia	2,655
Yukon	383
Northwest Territories	448
Nunavut	456
Total sample size	**22,004**

Source: Statistics Canada, 2021 Census Overcoverage Study.

The PL-only interprovincial pairs were further stratified by unique province combination and allocated proportional to size. There were 78 unique province combinations among the interprovincial pairs. Within the provincial combination substrata, pairs were sorted by their conditional match probabilities, and systematic sampling was used to draw the sample.

Probabilistic linkage-only groups and neighbourhoods

For the groups and neighbourhoods, the pairs were first stratified by intraprovincial and interprovincial groups. A group was considered interprovincial if it contained at least one interprovincial pair. Table 8.4.2.4 shows the breakdown of intra- and interprovincial groups in the PL-only stratum. The sample was allocated proportional to size between the intra- and interprovincial strata.

Table 8.4.2.4
Frequency of intraprovincial and interprovincial groups or neighbourhoods and sample sizes

Group types	Frequency of pairs	Percent	Number of sampled pairs	Number of sampled groups
Intraprovincial	4,070,750	63.49	17,970	6,656
Interprovincial	2,341,011	36.51	10,140	3,022

Source: Statistics Canada, 2021 Census Overcoverage Study.

Within the intraprovincial stratum, groups were allocated to provinces using a power allocation. Table 8.4.2.5 shows the allocation of PL-only intraprovincial sampling units by province or territory. As there were so few sampling units in the territories, these substrata were take-all. To better control the final sample size, the provincial strata were further stratified by group size in terms of the number of pairs in the group. The sample within each provincial stratum was allocated among group sizes proportional to the size. A minimum of one sampling unit was sampled within each stratum.

Table 8.4.2.5
Allocation of probabilistic linkage-only intraprovincial sampling units by province or territory

Group levels (provinces and territories)	Number of sampled pairs	Number of sampled groups
Newfoundland and Labrador	270	112
Prince Edward Island	100	43
Nova Scotia	439	187
New Brunswick	450	184
Quebec	7,594	2,514
Ontario	4,949	1,899
Manitoba	385	162
Saskatchewan	342	143
Alberta	1,119	467
British Columbia	2,123	866
Yukon	54	24
Northwest Territories	68	27
Nunavut	64	28
Total sample size	**17,957**	**6,656**

Note: The three territories are take-all strata.
Source: Statistics Canada, 2021 Census Overcoverage Study.

The interprovincial group and neighbourhood stratum was divided into two substrata: those with a majority province or territory (i.e., a province or territory to which most pairs in the group belong) and those without a majority province or territory (i.e., the pairs within the group are split evenly among the provinces or territories involved). The breakdown of pairs by majority and no majority groups and neighbourhoods is given in Table 8.4.2.6.

Table 8.4.2.6
Frequency of pairs within probabilistic linkage-only interprovincial groups or neighbourhoods by group with a majority province or territory or group without a majority province or territory, and sample sizes

Group types	Frequency of pairs	Percent	Number of sampled pairs	Number of sampled groups
With a majority province or territory	515,949	90	9,099	2,702
Without a majority province or territory	57,923	10	1,041	422

Source: Statistics Canada, 2021 Census Overcoverage Study.

Interprovincial groups and neighbourhoods with a majority province or territory were further stratified by dominant province or territory in the group and allocated using a power allocation. The sampling units within the provincial substrata were then stratified by the number of pairs in the groups. Allocation to group size was proportional to size. The sampling units were then sorted by expected overcoverage in the group and the proportion of intraprovincial pairs in the group, and a systematic sample was drawn. Because there were only 102 groups with a majority territory, these strata were take-all. A minimum of at least four sampling units were drawn from the other strata. Table 8.4.2.7 shows the allocation of PL-only interprovincial sampling units with a majority province or territory by majority province or territory.

Table 8.4.2.7
Allocation of probabilistic linkage-only interprovincial sampling units with a majority province or territory by majority province or territory

Group levels (provinces and territories)	Number of sampled pairs	Number of sampled groups
Newfoundland and Labrador	282	91
Prince Edward Island	127	37
Nova Scotia	481	156
New Brunswick	433	144
Quebec	2,000	540
Ontario	2,593	685
Manitoba	346	108
Saskatchewan	264	99
Alberta	898	292
British Columbia	1,399	448
Yukon	129	49
Northwest Territories	102	39
Nunavut	36	14
Total sample size	**9,090**	**2,702**

Note: The three territories are take-all strata.
Source: Statistics Canada, 2021 Census Overcoverage Study.

Groups with no majority province or territory were stratified by group size, and the sample was allocated proportional to size. Table 8.4.2.8 shows the allocation of PL-only interprovincial groups with no dominant province or territory by group size.

Table 8.4.2.8
Allocation of probabilistic linkage-only interprovincial groups with no dominant province or territory by number of pairs in groups

Number of pairs	Number of sampled pairs	Number of sampled groups
2	548	274
3	372	124
4	36	9
5	55	11
6	6	1
7	7	1
8	8	1
9	9	1
Total sample size	**1,041**	**422**

Source: Statistics Canada, 2021 Census Overcoverage Study.

8.4.3 Probabilistic linkage–deterministic linkage stratum

The breakdown of PL pairs and DL pairs in the PL–DL groups and neighbourhoods is shown in Table 8.4.3.1.

Table 8.4.3.1
Probabilistic linkage pairs and deterministic linkage pairs in the probabilistic linkage-deterministic linkage groups and neighbourhoods

Linked by	Frequency	Percent
PL	250,270	70.64
DL	104,020	29.36

PL = probabilistic linkage
DL = deterministic linkage
Note: The term "probabilistic linkage-deterministic linkage" means some of the pairs in the group were identified by the probabilistic linkage only, while others were identified by the deterministic linkage.
Source: Statistics Canada, 2021 Census Overcoverage Study.

As previously mentioned, a sample of DL pairs was drawn during the DL step and sent for manual verification to evaluate the quality of DL pairs and ensure that all DL pairs could be classified as definite pairs of duplicate persons. This sample was referred to as the DL verification sample. To use the DL verification sample, groups to which these sampled pairs belonged were treated as take-all strata, and the corresponding PL pairs, and any corresponding DL pairs not part of the DL verification sample, were sent for manual verification.

There were 1,010 sampled DL pairs among the pairs in the PL–DL interconnected record groups. These pairs belonged to 929 groups. The breakdown of PL and DL pairs among these 929 groups is shown in Table 8.4.3.2.

Table 8.4.3.2
Probabilistic linkage pairs and deterministic linkage pairs among the 929 probabilistic linkage-deterministic linkage groups that contained deterministic linkage pairs that were part of the deterministic linkage verification sample

Linked by	Frequency	Percent
PL	2,553	71.31
DL	1,027	28.69

PL = probabilistic linkage
DL = deterministic linkage
Note: The term "probabilistic linkage-deterministic linkage" means some of the pairs in the group were identified by the probabilistic linkage only, while the deterministic linkage identified others.
Source: Statistics Canada, 2021 Census Overcoverage Study.

There were 17 DL pairs and 2,553 PL pairs sent for manual verification. The 1,010 DL pairs that were part of the DL verification sample had already been verified. Hence, they were not sent for manual verification.

An additional sample of 533 groups (1,930 pairs) was selected from the PL–DL stratum. The PL–DL stratum was stratified by group-level province or territory and group size, and the sample was selected so that the full PL–DL sample was approximately proportional to size. The pair-level provincial breakdown of the full PL–DL sample is given in Table 8.4.3.3.

Table 8.4.3.3
Breakdown of pairs in the probabilistic linkage-deterministic linkage sample by pair-level province or territory

Provinces and territories	Frequency	Percent
Newfoundland and Labrador	62	1.13
Prince Edward Island	32	0.58

Table 8.4.3.3
Breakdown of pairs in the probabilistic linkage-deterministic linkage sample by pair-level province or territory

Provinces and territories	Frequency	Percent
Nova Scotia	87	1.58
New Brunswick	105	1.91
Quebec	1,583	28.76
Ontario	1,683	30.57
Manitoba	91	1.65
Saskatchewan	83	1.51
Alberta	311	5.65
British Columbia	714	12.97
Yukon	15	0.27
Northwest Territories	15	0.27
Nunavut	9	0.16
Interprovincial	715	12.99
Total	**5,505**	**100.00**

Note: The term "probabilistic linkage-deterministic linkage" means some of the pairs in the group were identified by the probabilistic linkage only, while others were identified by the deterministic linkage.
Source: Statistics Canada, 2021 Census Overcoverage Study.

The provincial strata were further stratified by number of links, and a systematic sample was drawn.

8.4.4 Final sample sizes (by pairs)

Table 8.4.4.1 below shows the final sample sizes for the PL-only and PL–DL strata that were sent for manual verification. The DL-only stratum consisted of 360,280 pairs that were classified as definite pairs of duplicate persons.

Table 8.4.4.1
Final sample sizes for probabilistic linkage-only and probabilistic linkage-deterministic linkage strata sent for manual verification

Strata	Number of pairs by stratum	Number of sampled pairs sent for manual verification (after de-duplication for overlapping neighbourhoods)
PL-DL (without 1,010 DL pairs that were part of the DL verification sample)	92,494	4,495
PL-only interprovincial pairs	1,087,646	4,753
PL-only intraprovincial pairs	4,844,635	22,004
PL-only intraterritorial groups (take-all)	4,070,750	186
PL-only intraprovincial groups	484	18,153
PL-only interterritorial groups with a majority territory (take-all)	266	266
PL-only interprovincial groups with a majority province	2,184,665	8,815
PL-only interprovincial groups with no majority province or territory	156,079	1,040
Total size	**12,437,019**	**59,712**

PL = probabilistic linkage
DL = deterministic linkage
PL-DL = probabilistic linkage-deterministic linkage (some of the pairs in the group were identified by the probabilistic linkage only, while others were identified by the deterministic linkage)
Source: Statistics Canada, 2021 Census Overcoverage Study.

8.5 Manual verification operation

The manual verification operation was a clerical operation and had several objectives:

- independently verify sampled pairs to determine whether they are overcoverage
- review the household members associated with the sampled pairs to potentially identify additional cases of overcoverage not on the COS frame
- code the potential cause of the overcoverage (i.e., overcoverage scenario).

Manual verification was done pair by pair. When a group or neighbourhood was sampled, all of the pairs that it contained were examined manually. However, coders were not provided the grouping information for the pairs in groups and neighbourhoods. Each pair was verified on its own. The pairs were examined only once, even if they belonged to more than one sampled neighbourhood.

The manual verification process consisted of a comprehensive examination of all available information on the RDB. As in 2016, it consisted of the following steps:

1. comparing the sampled RDB persons based on the names, sex, birth date and relationships, as well as some additional information added in 2021
2. comparing the RDB household members based on the same criteria
3. weighing the evidence for or against overcoverage between two RDB persons and between two RDB households
4. determining the overcoverage scenario if there was overcoverage (Table 8.5.1 provides a list of overcoverage scenario codes and their description).

Table 8.5.1
Overcoverage scenario codes

Codes	Description
1.1	Two different FRAME_IDs for the same household; same or similar address
1.2	Two different FRAME_IDs for the same household; different address
2.1	Child of parents in separate households
2.2	Child (age 0 to 17) with other relative(s)
2.3	Child (age 0 to 17) with other unrelated adult(s)
3.1	Student or young adult (age 18 to 24) newly away from home
3.2	Young adult (age 18 to 24) entering or leaving married or common law relationship
3.3	Young adult (age 18 to 24) with other relative(s)
3.4	Young adult (age 18 to 24) with other unrelated adult(s)
4.1	Adult (age 25 or older) newly away from home
4.2	Adult (age 25 or older) entering or leaving married or common law relationship
4.3	Adult (age 25 or older) with other relative(s)
4.4	Adult (age 25 or older) with other unrelated adult(s)
5.1	One household not a private dwelling
6.1	Intrahousehold overcoverage (same Frame_ID)
7.1	Other

FRAME_ID = unique household identifier
Source: Statistics Canada, 2021 Census Overcoverage Study.

The sample was divided into batches of 500 household pairs (household A, household B). Each batch was assigned to a clerk (verifier), who examined and decided whether the selected person of household A was duplicated (overcovered) with the selected person of household B for each household pair in the batch. A selected pair of records was the sampled pair of interest. Furthermore, the verifier identified additional pairs of duplicate persons (if any) from each household pair and within each household.

When verifiers were uncertain of how to code a case, they were instructed to refer it to their supervisor, who in turn consulted with the Data Quality (DQ) team (a team of subject matter experts in the Coverage Measurement Section of the Statistical Integration Methods Division) or referred the case to the DQ team. In 2021, some complex sampled pairs were sent directly to the DQ team to verify. Complex sampled cases included

- intra-household cases, that is, when the pair is from within a single household (for example, the same person is listed twice)
- one-person households (when the pair comes from two different households, each of which has a household size of 1).

Experience from past cycles showed that these complex sampled cases required the expertise of the DQ team to code them properly. The DQ team was also able to consult additional sources of information to help make an accurate decision, such as consulting the current and/or past census cycle's questionnaire data and using linkages conducted by the SDLE team at Statistics Canada. All sampled cases had to be coded with certainty, as no non-response was permitted.

Confidence in the coded results was required for the manual operation since the results directly contributed to the estimate of overcoverage. Thus, a 100% verification was implemented. This means two different verifiers coded the same batch. Once a batch had been coded by two different verifiers, their results were compared. All coded fields were compared. If any of the coding did not match, then the case was sent to the DQ team to make an informed decision. The 100% verification strategy ensured high-quality coded results, and continuous feedback was also provided to the clerks throughout the manual verification operation.

8.6 Weighting and estimation

8.6.1 Weighting

The initial weight of a sampling unit was simply the inverse of its selection probability. The sampling units that were groups and neighbourhoods varied in terms of the number of pairs they contained. These units were stratified by the number of pairs during sampling to better control the final sample size. However, for the interprovincial groups and neighbourhoods, the weighted provincial or territorial counts may have differed from what was on the frame. Therefore, a calibration step was added to ensure correct representation of the number of pairs in each province and territory. The sampling weights of the interprovincial groups and neighbourhoods were calibrated so that the estimated number of intraprovincial and interprovincial pairs in each province and territory matched the corresponding frame counts. Statistics Canada's Generalized Estimation System (G-EST) was used to perform the calibration. Table 8.6.1.1 shows the calibration factors for each province and territory.

Table 8.6.1.1
Average calibration factor (ratio of frame total to weighted estimate) by stratum and type of pair for intraprovincial and interprovincial groups and neighbourhoods

Provinces and territories	Intraprovincial	Interprovincial
Newfoundland and Labrador	0.76	0.74
Prince Edward Island	0.69	1.30
Nova Scotia	1.01	0.96
New Brunswick	0.99	0.88
Quebec	0.98	1.08
Ontario	0.99	0.99

Table 8.6.1.1
Average calibration factor (ratio of frame total to weighted estimate) by stratum and type of pair for intraprovincial and interprovincial groups and neighbourhoods

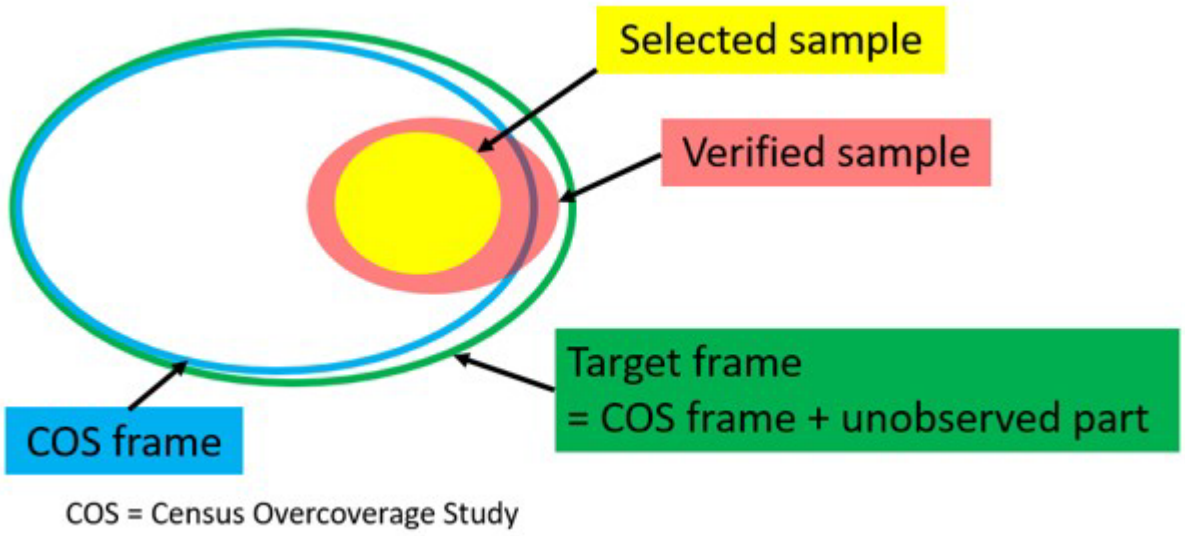
Provinces and territories	Intraprovincial	Interprovincial
Manitoba	1.36	1.07
Saskatchewan	1.13	0.90
Alberta	1.08	1.08
British Columbia	1.07	0.97
Yukon	1.47	1.21
Northwest Territories	1.40	0.42
Nunavut	1.22	2.80

Source: Statistics Canada, 2021 Census Overcoverage Study.

During the manual verification operation, verifiers identified cases of overcoverage in the households of sampled pairs that were not covered by the COS frame, and these pairs of duplicate persons were referred to as additional pairs of overcoverage found during manual verification. This occurred when the differences between the two records were too great for the pair to have been captured by the linkage processes. For example, if there were multiple typos, errors or too many differences in the fields used during the linkage process, the overcoverage pair was not on the COS frame.

This situation is illustrated in Figure 1 below. The oval with a blue outline represents the COS frame, while the oval with a green outline represents the target frame, which includes a small number of pairs that could not be captured with the linkage processes (i.e., the unobserved part of the target frame). The solid yellow oval represents the selected sample, which includes sampled person pairs, while the solid red oval represents the verified sample, which includes sampled person pairs and their household members. There are no weights directly associated with those pairs in the solid red oval that fall outside the COS frame (i.e., a small portion of the solid red oval falls in the unobserved portion of the target frame). The Generalized Weight Share Method (GWSM) (Lavallée, P. 2007) was used to assign weights from the weights of sampled pairs, through which these were indirectly sampled. Hence, all the additional pairs of overcoverage found during manual verification had a weight derived for them, and they were added to the sample for the purpose of estimation. This replaced the adjustment based on the AMS, which took into account overcoverage measured by the AMS outside the COS frame. This had been used since the 2006 COS.

Figure 1
Illustration of selected sample, verified sample, Census Overcoverage Study frame and target frame



Source: Statistics Canada, 2021 Census Overcoverage Study.

There were some limitations associated with the way additional pairs of overcoverage were identified. Duplicated single-person households or duplicated persons whose other household members have nothing in common within the unobserved part of the target frame would not be captured by manually verifying all household members of a sampled pair. Thus, it is acknowledged that the 2021 COS may still not represent the entire target frame of duplicate persons in the census. This would have also been the case when using the AMS to adjust the COS in previous cycles. However, the unobserved portion of the target frame is expected to be extremely small.

8.6.2 Estimation

The results from the manual verification operation were processed to create overcoverage groups that were used for estimation. Overcoverage groups consisted of all RDB records that were linked together by verified overcoverage. The COS estimates were based on the sum of the overcoverage estimate counted in each overcoverage group. For an overcoverage group that was a pair, the overcoverage count was simply 1. If the overcoverage group was contained within a small group of records (i.e., a group not broken into neighbourhoods), then:

Overcoverage = number of records in overcoverage group – 1.

For overcoverage groups broken down into neighbourhoods, overcoverage was counted in the following two steps:

1. Calculate overcoverage in each neighbourhood whose anchor (i.e., the RDB record acting as the centre of the neighbourhood) was involved in verified overcoverage for that overcoverage group as follows:

$$\text{Overcoverage in the neighbourhood} = \frac{(\text{number of records belonging to the overcoverage group-1})}{\text{number of records belonging to the overcoverage group}}$$

2. Add up the neighbourhood overcoverage to obtain the total overcoverage in the overcoverage group.

Domain overcoverage was obtained by prorating the total pair, group or neighbourhood overcoverage by the proportion of RDB records in the given domain among those that belonged to the overcoverage group.

For interprovincial groups and neighbourhoods, the overcoverage calculated for a unit was multiplied by the calibrated weight to obtain the weighted estimate. Additional pairs of overcoverage found during manual verification were multiplied by their derived sampling weight from the use of the GWSM, to obtain the weighted estimate. Otherwise, the overcoverage calculated for a unit was multiplied by its initial sampling weight to obtain the weighted estimate. The variance of the estimate was calculated using G-EST.

8.7 Results

The 2021 COS estimated that 755,635 persons were enumerated more than once in the 2021 Census of Population. The results were examined by each of the components that led to the construction of the sampling frame and its contribution to the overall estimation of census overcoverage. Potential reasons why persons were counted more than once in the census were also examined.

8.7.1 Overcoverage by component

Each case of overcoverage (definite or manually verified) was characterized by the COS components that identified the pairs in its sampling unit. They are of four types:

- DL-only: all the pairs in the overcoverage group were identified by the DL
- PL-only: all the pairs in the overcoverage group were identified only by the PL
- PL–DL: some of the pairs in the group were identified by the PL only, while others were identified by the DL
- overcoverage manual verification (OCMV): all the pairs in the overcoverage group were additional pairs of duplicate persons found during manual verification that were not on the COS sampling frame and for which an indirect sampling weight was derived using the GWSM.

It is important to remember that pairs identified by both the PL and DL steps were classified as DL, so the "DL-only" category includes all the groups that are made up only of pairs that were identified by the DL, even though some of those same pairs could also have been identified by the PL.

Table 8.7.1.1 presents the number of overcoverage cases estimated by each of the COS components, as well as the percentage of the total estimated overcoverage that it represented, for Canada, as well as by province or territory.

Table 8.7.1.1
Contribution of each 2021 Census Overcoverage Study component to the total estimated overcoverage for each province and territory

Provinces and territories	DL-only		PL-only		PL-DL		OCMV		Total	
	Estimated number	% of total	Estimated number	% of total	Estimated number	% of total	Estimated number	% of total	Estimated number	Standard error
Canada	**352,059**	**46.6**	**318,459**	**42.1**	**81,172**	**10.7**	**3,946**	**0.5**	**755,635**	**9,648**
Newfoundland and Labrador	5,148	50.5	4,664	45.8	382	3.7	0	0.0	10,194	439
Prince Edward Island	1,678	51.0	1,284	39.0	311	9.5	16	0.5	3,289	191
Nova Scotia	9,200	47.6	8,412	43.5	1,639	8.5	94	0.5	19,344	736
New Brunswick	8,079	48.8	6,890	41.7	1,440	8.7	132	0.8	16,541	641
Quebec	76,760	42.1	80,126	43.9	25,242	13.8	385	0.2	182,513	5,915
Ontario	120,765	44.7	118,619	43.9	29,382	10.9	1,334	0.5	270,100	6,888
Manitoba	11,930	51.5	10,231	44.2	970	4.2	29	0.1	23,160	757
Saskatchewan	13,210	54.7	9,501	39.3	1,194	4.9	258	1.1	24,163	689
Alberta	36,902	47.3	35,085	44.9	5,527	7.1	570	0.7	78,084	2,736
British Columbia	66,976	53.2	42,762	34.0	15,017	11.9	1,078	0.9	125,832	2,778
Yukon	479	57.9	315	38.1	21	2.5	12	1.4	827	38
Northwest Territories	508	60.6	293	35.0	25	3.0	12	1.4	837	15
Nunavut	423	56.3	277	36.9	23	3.1	28	3.7	751	17

DL-only = all the pairs in the overcoverage group were identified by the deterministic linkage
PL-only = all the pairs in the overcoverage group were identified only by the probabilistic linkage
PL-DL = some of the pairs in the group were identified by the probabilistic linkage only, while others were identified by the deterministic linkage
OCMV = all the pairs in the overcoverage group were additional pairs of duplicate persons found during manual verification that were not on the Census Overcoverage Study sampling frame and for which an indirect sampling weight was derived using the generalized weight share method
Note: Coverage estimates may not necessarily add up to the totals because of rounding.
Source: Statistics Canada, 2021 Census Overcoverage Study.

At the national level, the DL-only and PL-only components represented 46.6% and 42.1%, respectively, of the total estimate of overcoverage, while the PL–DL component represented 10.7%, and the OCMV component accounted for 0.5%.

The DL-only contribution to the total provincial or territorial estimate was higher for the Northwest Territories (60.6%) and Yukon (57.9%) and lower for Ontario (44.7%) and Quebec (42.1%). The PL-only contribution to the total provincial or territorial estimate was higher for Newfoundland and Labrador (45.8%) and Alberta (44.9%) and lower for the Northwest Territories (35.0%) and British Columbia (34.0%). As for the PL–DL component, its contribution was higher in Quebec (13.8%) and British Columbia (11.9%) and lower for the three territories (ranging from 2.5% to 3.1%). Lastly, for the OCMV component, its contribution was higher for the three territories (ranging from 1.4% to 3.7%) and lower for Manitoba (0.1%) and Newfoundland and Labrador (0.0%), where no additional pairs of duplicate persons were identified during the manual verification operation that were not already on the COS sampling frame.

8.7.2 Overcoverage by scenario

Table 8.7.2.1 shows the estimated overcoverage by potential reason why the overcoverage occurred, called the overcoverage scenario, at the national and provincial and territorial levels for 2021. It is important to mention that these results are not comparable to the 2016 overcoverage results by scenario for two reasons:

- The overcoverage scenario was coded during the manual verification operation. Since the DL-only pairs were considered as definite pairs of duplicate persons without manual verification, an overcoverage scenario is not available for those pairs.
- The codes used for the scenarios were modified for the 2021 cycle to improve the consistency of the coding and the usefulness of the results.

Excluding the DL-only cases, almost 25% of all overcoverage at the national level is between two identical households. This proportion is a little lower for Newfoundland and Labrador and higher for British Columbia.

When only overcoverage within non-identical households is considered and the DL-only cases are excluded again, the most frequent overcoverage scenario is a child enumerated by both parents in separate households, as was the case in 2016 and previous cycles. This is true for every province and territory, except for Nova Scotia and Nunavut. In Nova Scotia, the most frequent scenario was a student or young adult (age 18 to 24) newly away from home, while in Nunavut, it was a child (age 0 to 17) with other relative(s).

Table 8.7.2.1
Distribution of 2021 Census overcoverage by scenario for each province and territory

Provinces and territories	Overcoverage scenario				
	1.1	2.1	2.2	2.3	3.1
	Identical households	Child of parents in separate households	Child (age 0 to 17) with other relative(s)	Child (age 0 to 17) with other unrelated adult(s)	Student or young adult (age 18 to 24) newly away from home
	percent				
Canada	**12.5**	**11.3**	**0.8**	**0.3**	**5.8**
Newfoundland and Labrador	9.5	11.5	1.2	0.0	7.7
Prince Edward Island	11.2	11.5	0.3	0.4	9.0
Nova Scotia	10.8	9.9	2.0	0.0	15.5
New Brunswick	11.0	10.5	1.7	0.5	6.4
Quebec	12.1	16.6	0.7	0.2	6.0
Ontario	13.0	10.6	0.5	0.1	5.4
Manitoba	10.5	8.8	2.4	0.5	5.1
Saskatchewan	9.3	11.1	1.5	0.8	4.0
Alberta	11.5	9.3	0.7	0.8	6.6
British Columbia	14.6	7.0	0.7	0.5	4.7
Yukon	9.2	13.1	0.4	0.2	4.0
Northwest Territories	10.8	8.2	2.1	1.0	1.4
Nunavut	13.6	4.5	8.8	1.2	1.6

Table 8.7.2.1
Distribution of 2021 Census overcoverage by scenario for each province and territory

Provinces and territories	Overcoverage scenario					
	3.2	3.3	3.4	4.1	4.2	4.3
	Young adult (age 18 to 24) entering or leaving married or common law relationship	Young adult (age 18 to 24) with other relative(s)	Young adult (age 18 to 24) with other unrelated adult(s)	Adult (age 25 or older) newly away from home	Adult (age 25 or older) entering or leaving married or common law relationship	Adult (age 25 or older) with other relative(s)
	percent					
Canada	**1.2**	**0.5**	**0.7**	**3.6**	**3.5**	**3.9**
Newfoundland and Labrador	2.5	0.3	0.0	1.6	4.9	3.6
Prince Edward Island	1.1	0.8	0.3	3.7	1.8	1.2
Nova Scotia	1.9	0.0	0.6	1.9	3.0	2.1
New Brunswick	3.4	0.4	0.4	2.8	4.5	2.7
Quebec	1.7	0.6	0.5	4.1	4.7	3.9
Ontario	0.6	0.2	0.9	4.1	3.0	4.9
Manitoba	1.7	1.1	0.9	3.2	3.5	2.4
Saskatchewan	0.6	1.5	1.1	2.4	1.5	3.7
Alberta	2.1	0.5	0.5	3.3	3.8	3.9
British Columbia	0.5	0.5	0.9	3.0	2.8	2.5
Yukon	0.8	0.8	0.3	0.8	3.1	1.9
Northwest Territories	0.7	0.6	0.5	0.9	2.7	2.9
Nunavut	1.1	2.7	0.7	1.2	0.8	4.8

Table 8.7.2.1
Distribution of 2021 Census overcoverage by scenario for each province and territory

Provinces and territories	Overcoverage scenario				
	4.4	5.1	6.1	7.1	8.1
	Adult (age 25 or older) with other unrelated adult(s)	One household not a private dwelling	Intrahousehold overcoverage (same FRAME_ID)	Other	Deterministic linkage
	percent				
Canada	**1.6**	**2.4**	**0.6**	**3.2**	**48.2**
Newfoundland and Labrador	1.0	3.0	0.3	3.3	49.8
Prince Edward Island	1.3	2.5	0.0	0.8	54.0
Nova Scotia	1.2	1.0	0.4	1.8	47.9
New Brunswick	0.4	2.2	0.0	3.1	50.0
Quebec	0.9	2.3	0.7	2.9	41.9
Ontario	1.6	2.3	0.7	3.0	49.2
Manitoba	1.8	4.1	0.3	2.5	51.2
Saskatchewan	1.7	3.1	0.2	2.8	54.6
Alberta	2.5	2.7	0.5	4.4	47.0
British Columbia	2.1	2.3	0.4	4.0	53.6
Yukon	1.4	1.4	0.1	3.0	59.5
Northwest Territories	2.7	1.8	0.5	1.7	61.4
Nunavut	1.2	3.2	0.4	2.3	51.7

FRAME_ID = unique household identifier
Note: Overcoverage by scenario is estimated at the pair level rather than the group level hence there is a small difference in the percentages when compared to Table 8.7.1.1.
Source: Statistics Canada, 2021 Census Overcoverage Study.

9. Estimation

Estimation for the Dwelling Classification Survey, the Census Undercoverage Study (CUS) and the Census Overcoverage Study (COS) is covered in Sections 6.2, 7.4 and 8.6, respectively. This section describes how the results of census coverage studies are combined to produce estimates of population undercoverage (*U*), population overcoverage (*O*) and population net undercoverage (*N*) in different domains. The impact of sampling errors on the quality of the estimates is also measured by an estimated standard error for each estimate. CUS results and census data are used to produce undercoverage estimates, while COS results estimate overcoverage. Net undercoverage is the difference between undercoverage and overcoverage. This section expands on how these estimates and the associated standard errors are calculated.

The following definitions are used:

C = published census count of the number of persons in the target population

$\hat{U}$ = undercoverage estimate

= estimated number of persons not included in C who should have been included

$\hat{O}$ = overcoverage estimate

= estimated number of enumerations included in C that should not have been included

$\hat{N}$ = net undercoverage estimate

= estimated number of enumerations not included in C that should have been included, less the number of enumerations included in C who should not have been included

= $\hat{U} - \hat{O}$

$\hat{T}$ = estimated number of persons in the census target population based on census enumerations and the estimate of population net undercoverage

= $C + \hat{N}$

$\hat{R}_U$ = estimated undercoverage rate

= $100 * \frac{\hat{U}}{\hat{T}} = 100 * \frac{\hat{U}}{C + \hat{N}}$

$\hat{R}_O$ = estimated overcoverage rate

= $100 * \frac{\hat{O}}{\hat{T}} = 100 * \frac{\hat{O}}{C + \hat{N}}$

$\hat{R}_N$ = estimated net undercoverage rate

= $100 * \frac{\hat{N}}{\hat{T}} = 100 * \frac{\hat{U} - \hat{O}}{C + \hat{N}}$

$\hat{U}$ is calculated using CUS results and census data, and $\hat{O}$ is produced from the COS, as shown below:

Table 9.1
Components of the population coverage error estimates for Canada

Components	Number of people
Estimate of U	1,897,876
Estimate of O	755,635
Estimate of N	1,142,241
C	36,991,981
C + estimate of N	38,134,222

U = undercoverage
O = overcoverage
N = net undercoverage
C = published census count
Sources: Statistics Canada, 2021 Census coverage studies and 2021 Census.

The estimated standard errors are defined as follows:

By definition, we have $v(\hat{U}) = v(\hat{M})$ (refer to Section 7.4.6).

$v(\hat{M})$ = estimated variance of $\hat{M}$ based on the CUS design

$v(\hat{O})$ = estimated variance of $\hat{O}$ based on the COS design

Therefore:

$$se\ (\hat{U}) = \sqrt{v(\hat{M})}$$

$$se\left(\hat{R}_U\right) = \sqrt{\left(\frac{\hat{U}^2 + \hat{T}^2 - 2\hat{U}\hat{T}}{\hat{T}^4}\right)v\left(\hat{M}\right) + \frac{\hat{U}^2}{\hat{T}^4}v\left(\hat{O}\right)}$$

$$se\left(\hat{O}\right) = \sqrt{v(\hat{O})}$$

$$se\left(\hat{R}_O\right) = \sqrt{\left(\frac{\hat{O}^2}{\hat{T}^4}\right)v(\hat{M}) + \left(\frac{\hat{U}^2 + \hat{T}^2 - 2\hat{O}\hat{T}}{\hat{T}^4}\right)v(\hat{O})}$$

$$se\left(\hat{N}\right) = \sqrt{v(\hat{M}) + v(\hat{O})}$$

$$se\left(\hat{R}_N\right) = \sqrt{\left(\frac{(\hat{U} - \hat{O})^2 + \hat{T}^2 - 2(\hat{U} - \hat{O})\hat{T}^2}{\hat{T}^4}\right)\left[v(\hat{M}) + v(\hat{O})\right]}$$

10. Evaluation of coverage studies

10.1 Census Undercoverage Study

10.1.1 Introduction

The results of the largest coverage study, the Census Undercoverage Study (CUS), can be assessed by comparing its estimates with data on the same characteristics from other sources, such as the 2021 Census database and administrative data used by the Demographic Estimates Program (DEP). The purpose of making comparisons with CUS estimates is to evaluate the CUS estimates and to quantify conceptual and measurement differences.

Despite some conceptual differences between the CUS and the 2021 Census, the CUS estimates of persons enumerated in the 2021 Census can be compared with the census counts. To make the two numbers comparable, certain adjustments were first made to the census counts.

Estimates of the components of intercensal demographic growth can be compared with estimates from other sources. The CUS estimates of the number of persons who died between the 2016 Census and the 2021 Census can be compared with the counts from vital statistics files. Estimates of net interprovincial migration calculated by the DEP based on Canada Revenue Agency data can be compared with CUS estimates. Lastly, CUS estimates of the components of demographic growth can be compared with similar estimates from administrative data.

10.1.2 Comparisons with census counts

Since the CUS's single-stage stratified sampling design produces unbiased estimates, differences between CUS estimates and census counts are mainly attributable to sampling error in the CUS estimates; conceptual differences between the two sources; or systematic biases that have impacts on the two sources, resulting in an underestimate or overestimate of the characteristic being studied.

Enumerated persons

Provincial and national comparisons are presented in Table 10.1.2.1, along with the standard error of the CUS estimate and the *t*-value used to test the hypothesis that there is no difference between the CUS estimate and the comparable census count. The adjustments below were made to the published census counts to account for conceptual differences between the two sources:

- Adjustments based on whole household imputation were excluded because, while they were included in the census counts, they were not part of the CUS estimate of enumerated persons.
- The 2021 Census overcoverage estimate was subtracted because the census database contained overcovered persons, whereas the CUS estimate was based on the number of unique persons enumerated (and not on the number of enumerations).
- The estimate of the number of persons living outside Canada five years earlier (excluding intercensal immigrants and non-permanent residents [NPRs]) from the 2021 Census long-form questionnaire was also subtracted because the CUS estimates did not include the majority of these persons. For the same reason, the estimated number of children aged 0 to 4 years who were born outside Canada but had Canadian citizenship was also subtracted.
- Similarly, for the provinces, the number of persons living in a territory five years earlier was subtracted because they were not covered by the CUS provincial sampling frames.
- The number of persons from reserves (who participated in the 2021 Census, but not in the 2016 Census) was also subtracted because the CUS estimates did not include the majority of these persons.

Table 10.1.2.1
Comparison of the estimated number of enumerated people from the Census Undercoverage Study with comparable census counts for Canada, the provinces and territories

	Enumerated people				
	CUS				
Provinces and territories	Estimated number	Standard error	Comparable census count	Difference	*t*-value[1]
Canada	**34,166,964**	**40,467**	**34,380,739**	**-213,775**	**-5.28**
Newfoundland and Labrador	478,988	2,432	480,370	-1,382	-0.57
Prince Edward Island	142,075	1,033	144,380	-2,305	-2.23
Nova Scotia	899,614	4,153	907,675	-8,061	-1.94
New Brunswick	724,776	3,361	725,234	-458	-0.14
Quebec	7,912,911	18,532	7,976,614	-63,703	-3.44
Ontario	13,243,488	29,802	13,306,050	-62,562	-2.10
Manitoba	1,217,061	5,621	1,231,102	-14,041	-2.50
Saskatchewan	1,033,774	5,457	1,031,302	2,472	0.45
Alberta	3,894,128	13,778	3,932,868	-38,740	-2.81
British Columbia	4,525,565	15,461	4,550,560	-24,995	-1.62
Yukon	34,815	0	34,815	0	…
Northwest Territories	34,360	0	34,360	0	…
Nunavut	25,409	0	25,409	0	…

... not applicable
CUS = Census Undercoverage Study
1. A *t*-value greater than 1.96 or less than -1.96 indicates that the difference is significant at the 95% level.
Sources: Statistics Canada, 2021 Census coverage studies and 2021 Census.

Nationally, the CUS estimate of the number of persons enumerated in the 2021 Census was lower than the comparable census count (-0.62%). For the 1996 to 2016 censuses, the national difference between the CUS estimate and the comparable census count was between -0.09% and 0.12%. It is the first time since that comparison has been performed that the difference is statistically significant at the national level. At the provincial level, the CUS estimate is lower than the comparable census count for every province except Saskatchewan, and that difference is statistically significant for five provinces: Prince Edward Island, Quebec, Ontario, Manitoba and Alberta.

In previous cycles, significant differences were also observed. The differences were investigated to make sure that there was no bias in the CUS classification (including, for example, province of residence on Census Day). Other factors may also play an important role in the observed differences. Apart from sampling error, biases in the adjustments (e.g., returning Canadians) applied to the published census counts to obtain conceptually comparable figures may be responsible for the differences. CUS non-response bias may also have played a role since the non-response adjustment was designed to obtain the best result for estimating missed persons rather than enumerated persons. Regular checks and quality controls were performed for all steps in the CUS.

In view of the more significant differences than typically observed, a thorough investigation was conducted. In past censuses, there have always been persons enumerated in the census who were not covered by the CUS frames, either because of limitations with those frames or because they were not part of the census target population. On top of the returning Canadians already mentioned above, the first category also includes dependants (children and spouses) of NPRs. The second category includes persons deceased before Census Day, postcensal births, foreign visitors, immigrants who arrived in Canada after Census Day, and NPRs and their dependants who arrived after Census Day or did not have a valid permit. In past cycles, the number of such persons was deemed small enough to not have a significant impact on the comparison between the CUS estimate of enumerated persons and the comparable census count. With the large increase in the number of NPRs in recent years, it seems that this is no longer the case. A large portion of the difference between the two numbers can be explained by the enumeration of dependants of NPRs or out-of-scope census records.

An analysis of the differences by age group showed that the negative differences were concentrated among children (ages 0 to 14, with the differences for the 0-to-4 and 5-to-9 age groups being statistically significant) and 25- to 29-year-olds. Differences are minimal, or at least non-significant, for the other groups. This is consistent with the fact that a significant portion of the enumerated persons not covered by the CUS frames would be children or spouses of NPRs with permits.

None of the investigative work raised any concern with the CUS classification, weighting or estimation steps.

10.1.3 Comparison with demographic estimates

Deceased persons

Table 10.1.3.1a provides a comparison of the estimated number of persons who died during the intercensal period (May 10, 2016, to May 10, 2021) by CUS province of classification with counts from vital statistics files. The CUS estimate excludes persons who died outside Canada when the country of death is known. At the national level, the CUS estimate exceeded the vital statistics count by 6,492 persons (0.5%), and this difference was not statistically significant. At the provincial level, the greatest percentage differences were noted in Prince Edward Island (648, or 9.7%) and Quebec (12,731, or 3.7%), but only the first one was statistically significant (*t*-value of 2.28). In the other provinces, the relative differences were between -1.7% and 2.1%. They were not statistically significant, and most differences were smaller than what was observed in 2016.

Table 10.1.3.1a
Comparison of the estimated number of deceased people from the Census Undercoverage Study with the vital statistics counts for the provinces

	People deceased May 10, 2016, to May 10, 2021				
	CUS				
Provinces	Estimated number	Standard error	Vital statistics count	Difference	*t*-value[1]
Total	**1,442,017**	**13,967**	**1,435,525**	**6,492**	**0.46**
Newfoundland and Labrador	25,999	626	26,255	-256	-0.41
Prince Edward Island	7,317	285	6,669	648	2.28
Nova Scotia	47,877	933	48,349	-472	-0.51
New Brunswick	38,084	918	37,748	336	0.37
Quebec	355,031	7,366	342,300	12,731	1.73
Ontario	531,134	9,193	540,387	-9,253	-1.01
Manitoba	55,461	1,656	56,058	-597	-0.36
Saskatchewan	48,750	1,419	48,835	-85	-0.06
Alberta	132,746	3,842	133,471	-725	-0.19
British Columbia	199,619	5,001	195,453	4,166	0.83

CUS = Census Undercoverage Study
1. A *t*-value greater than 1.96 or less than -1.96 indicates that the difference is significant at the 95% level.
Note: Coverage estimates may not necessarily add up to the totals because of rounding.
Sources: Statistics Canada, 2021 Census Undercoverage Study, Vital Statistics Program and Demographic Estimates Program.

Certain reasons may explain a significant difference. Firstly, the CUS estimate may include deaths that occur abroad, which are not included in vital statistics. In the CUS, if the country of death is known and is abroad, then the death is not included in the comparison of deceased persons in Table 10.1.3.1a. However, if the person is not found in the vital statistics files and the country of death is unknown, then the death would be filed by default in the person's most recent province of residence in Canada. There were only 12 selected persons in this situation, and none were in Prince Edward Island; therefore, it is not the main reason for the significant difference in that province. Another reason that may explain the difference is not being able to find the person in the vital statistics

because of differences in personal information. The last reason is underreporting in vital statistics. There are 34 selected persons for whom the death was confirmed to be in Canada but who were not found in the vital statistics, and among them 10 were in Prince Edward Island. Table 10.1.3.1b provides a comparison of the CUS estimate of the number of persons who died during the intercensal period (May 10, 2016, to May 10, 2021) by province of residence indicated in the vital statistics files (therefore, only for persons found in these files) with vital statistics counts. The Prince Edward Island difference that had been significant no longer is, with a *t*-value of 0.45. However, the difference becomes significant in Ontario (-19,386, with a *t*-value of -2.47). Even if these last results do not seem indicative of issues related to the CUS estimates of the number of deceased persons, a more detailed investigation was conducted to confirm that no classification or other error was involved in the operations or estimates. No such errors or problems were detected.

Table 10.1.3.1b
Comparison of Census Undercoverage Study estimated number of deceased people linked to vital statistics and vital statistics count for the provinces

	People deceased May 10, 2016, to May 10, 2021				
	CUS				
Provinces	Estimated number	Standard error	Vital statistics count	Difference	*t*-value[1]
Total	**1,419,908**	**12,750**	**1,435,525**	**-15,617**	**-1.22**
Newfoundland and Labrador	25,966	626	26,255	-289	-0.46
Prince Edward Island	6,773	232	6,669	104	0.45
Nova Scotia	47,877	933	48,349	-472	-0.51
New Brunswick	37,123	887	37,748	-625	-0.70
Quebec	347,817	7,081	342,300	5,517	0.78
Ontario	521,001	7,859	540,387	-19,386	-2.47
Manitoba	54,605	1,628	56,058	-1,453	-0.89
Saskatchewan	48,750	1,419	48,835	-85	-0.06
Alberta	131,732	3,823	133,471	-1,739	-0.45
British Columbia	198,263	4,869	195,453	2,810	0.58

CUS = Census Undercoverage Study
1. A *t*-value greater than 1.96 or less than -1.96 indicates that the difference is significant at the 95% level.
Note: Coverage estimates may not necessarily add up to the totals because of rounding.
Sources: Statistics Canada, 2021 Census Undercoverage Study, Vital Statistics Program and Demographic Estimates Program.

Interprovincial migration

Table 10.1.3.2 compares CUS estimates of net interprovincial migration for the intercensal period with corresponding figures calculated by the DEP based on Canada Revenue Agency files. In general, data on interprovincial migrants were not comparable because the CUS only took into account migration flows that occurred between the sampling frame reference date (e.g., May 10, 2016, for the census frame) and Census Day in 2021, whereas the DEP estimates took annual migration into account. For this reason, only net interprovincial migration estimates are presented.

Although the estimates differ, the CUS and demographic estimates of net interprovincial migration go in the same direction (positive or negative net migration) for every province. The difference between the CUS and demographic estimates is smaller than it was in 2016 for 6 of the 10 provinces, and roughly the same for another one, but the standard errors of the CUS estimates are also much smaller than they were in 2016. As a result, two provinces show a statistically significant difference between the CUS and demographic estimates of net migration—Ontario (with *t*-value of 3.27) and Manitoba (-2.27). A significant difference was observed for one province in 2016.

Table 10.1.3.2
Comparison of Census Undercoverage Study estimates and demographic estimates (Canada Revenue Agency definition) of net interprovincial migration for the provinces

	Net interprovincial migration					
	CUS[1]					
Provinces	Sample size	Estimated number	Standard error	Demographic estimate	Difference	*t*-value[2]
Newfoundland and Labrador	27,245	-6,477	2,405	-8,555	2,078	0.86
Prince Edward Island	13,650	1,616	1,093	2,993	-1,377	-1.26
Nova Scotia	69,283	14,315	4,119	21,720	-7,406	-1.80
New Brunswick	50,933	6,938	2,918	7,491	-553	-0.19
Quebec	111,790	-39,080	9,831	-26,852	-12,227	-1.24
Ontario	291,826	73,340	15,334	23,157	50,183	3.27
Manitoba	57,416	-41,156	4,598	-30,715	-10,441	-2.27
Saskatchewan	65,324	-42,617	5,066	-42,019	-599	-0.12
Alberta	232,028	-58,339	13,084	-35,046	-23,293	-1.78
British Columbia	211,119	91,460	13,209	87,826	3,634	0.28

CUS = Census Undercoverage Study
1. The Census Undercoverage Study (CUS) excludes people living in a province on May 11, 2021, who had lived in one of the three territories five years before, on May 10, 2016.
2. A *t*-value greater than 1.96 or less than -1.96 indicates that the difference is significant at the 95% level.
Sources: Statistics Canada, 2021 Census Undercoverage Study and Demographic Estimates Program.

10.2 Census Overcoverage Study

The validation of the results of the two coverage studies was guided by Statistics Canada's Directive for the Validation of Statistical Outputs. Among various validation steps, it was possible to evaluate the results from the 2021 Census Overcoverage Study (COS) by assessing how each of the components that led to the construction of its sampling frame contributed to the overall estimation of census overcoverage. It was also possible to look at the potential reasons why persons were counted more than once in the census. Refer to Section 8.7 for more information.

During the validation process, another step was to compare the COS frame and estimates with other available sources, to analyze what might be missing from the COS frame and identify systemic issues, if any. One such source is the Social Data Linkage Environment (SDLE). The SDLE team at Statistics Canada is in charge of all census linkages used to derive information from administrative data sources (to gather income information, for example). As a first step to its linkages, the SDLE team does an internal probabilistic linkage of the entire Census Response Database (RDB) to itself to identify potential duplicate persons. There are some important differences between the objective of the SDLE's internal record linkage of the RDB and the objective of the COS. The objective of the SDLE is to identify with certainty duplicate records, and, therefore, it uses a conservative approach in duplicate identification. The objective of the COS is to construct a frame of all possible duplicate pairs from the internal linkage of the RDB.

The SDLE list of duplicated persons was compared with the COS frame, and all SDLE duplicates were on the 2021 COS frame. Also, a comparison of COS estimates and SDLE potential duplicate persons was done to investigate trends between 2016 and 2021. Although the results of the COS are an estimate of overcoverage and contain sampling error, whereas the results of the SDLE are counts and are expected to represent a subset of the COS estimate because of their nature, comparing the percentage change between 2016 and 2021 for both sources was still a useful exercise. The two sources showed consistent results, with an increase in overcoverage from 2016 to 2021, and a similar pattern was observed for each province and territory.

Another evaluation was done by examining the correlation between the overcoverage status of a pair of potentially duplicate RDB records and the linkage weight that was derived for this pair, for pairs that had been identified by the probabilistic linkage of the RDB to itself (refer to Section 8.2.3 for more information on this probabilistic linkage step). In a probabilistic linkage, a linkage weight is calculated for each pair of linked records, based on the strength of that link. As described in Section 8, in the COS, a sample of the pairs identified as potential duplicates by the probabilistic linkage of the RDB to itself was subject to a manual verification process, where coders had to determine whether each pair was a true duplicate (verified overcoverage) or not. The distribution of linkage weights for the overcoverage cases and the non-overcoverage cases was compared. If the linkage performed as expected, there should be a difference between the weights of the two groups. As expected, the linkage weights were much larger on average for the overcoverage cases, and this was true for every province and territory.

Also, the additional cases of overcoverage that were identified during manual verification operations but were not on the final COS frame were evaluated to understand why they were not captured. There does not appear to be a systemic reason why the additional pairs of overcoverage were not on the COS frame. In general, these pairs were too different, meaning they had multiple typos, errors or too many differences in the fields used during the linkage processes, resulting in them not being on the COS frame. As is done for every cycle, the initial selection criteria and linkage rules will be reviewed, revised and tested before the next cycle.

The 2021 COS validation activities did not raise any concern about the methodology of the study, and the evaluation of the COS estimates showed results that were consistent with past results and with what was to be expected.

10.3 Population estimates

10.3.1 Error of closure

Statistics Canada's DEP determines provincial and territorial population counts on Census Day by summing census population counts, estimates of census net undercoverage (CNU), and the population estimate for incompletely enumerated reserves and settlements. The DEP then extends these adjusted counts to July 1, 2021, and they become the base for postcensal population estimates.

When determining these adjusted counts, the DEP evaluates the quality of the postcensal estimates that it produced in the five-year period preceding the census. The evaluation focuses on the difference between the postcensal estimates for Census Day and the adjusted population counts for this census. This difference is referred to as the error of closure. The detailed examination of this error is the main quality measure of the postcensal estimates.

Table 10.3.1 shows the errors of closure for 2006, 2011, 2016 and 2021 by province and territory, and for Canada. Note that a positive error of closure means that the postcensal population estimate is higher than the adjusted census count. At the national level, the error of closure for 2021 was -41,269 persons, for an error rate of -0.11%. The national population estimates therefore underestimated Canada's population. The error rate in 2021 was lower than from 2006 to 2016.[13] Four provinces and one territory had errors of closure greater than 1% or less than -1% in 2021: Newfoundland and Labrador (-1.24%), Prince Edward Island (1.59%), Nova Scotia (-1.00%), Saskatchewan (1.06%), and the Northwest Territories (2.57%). By comparison, in 2016, five provinces and one territory had similar errors of closure. In 2021, eight provinces and one territory had smaller errors of closure (in absolute value terms) than in 2016.

13. Errors of closure for 2006 to 2016 are calculated using postcensal estimates from the 2006, 2011 and 2016 censuses, updated in 2023 following a revision of the components.

Table 10.3.1
Error of closure for Canada, provinces and territories, 2006, 2011, 2016 and 2021

Provinces and territories	2006		2011		2016		2021	
	number	rate (%)	number	rate (%)	number	rate (%)	number	rate (%)
Canada	**39,409**	**0.12**	**158,558**	**0.46**	**120,044**	**0.33**	**-41,269**	**-0.11**
Newfoundland and Labrador	-1,821	-0.36	-11,121	-2.12	1,097	0.21	-6,540	-1.24
Prince Edward Island	-31	-0.02	2,096	1.46	2,906	1.99	2,564	1.59
Nova Scotia	-3,997	-0.43	5,075	0.54	7,395	0.79	-9,944	-1.00
New Brunswick	2,673	0.36	1,432	0.19	-5,992	-0.79	-317	-0.04
Quebec	19,776	0.26	-23,207	-0.29	89,035	1.08	33,890	0.40
Ontario	24,532	0.19	121,217	0.92	68,329	0.49	-43,978	-0.30
Manitoba	-5,977	-0.51	21,464	1.74	5,358	0.41	-3,084	-0.22
Saskatchewan	-3,691	-0.37	-7,779	-0.73	12,492	1.10	12,402	1.06
Alberta	-50,869	-1.49	-3,345	-0.09	43,891	1.05	2,013	0.05
British Columbia	61,120	1.44	52,325	1.16	-104,201	-2.15	-29,372	-0.56
Yukon	-1,027	-3.19	103	0.29	-391	-1.02	150	0.35
Northwest Territories	-857	-1.99	758	1.74	-47	-0.11	1,146	2.57
Nunavut	-422	-1.37	-460	-1.35	172	0.47	-199	-0.50

Source: Statistics Canada, Centre for Demography.

10.3.2 Accuracy of postcensal estimates

For the purposes of producing the DEP estimates, the census coverage studies are used to adjust census counts for CNU. However, since these studies are based in part on sample surveys, the CNU results contain some statistical variability attributable to sampling. To determine whether the errors of closure discussed above are statistically significant, the standard error of the adjusted census counts must be taken into account. Moreover, since the 2016 adjusted census counts were used as the base population for the 2016 to 2021 postcensal estimates, a standard error that combines the statistical variability of the adjusted census counts for 2016 and 2021 was calculated for Canada and for each province and territory.

Table 10.3.2 shows the 2021 error of closure for Canada and the provinces and territories, the combined standard error of the 2016 and 2021 adjusted census counts, and the *t*-value.[14] The error of closure is statistically significant at a 95% confidence level for Newfoundland and Labrador, Prince Edward Island, Nova Scotia, Saskatchewan, and the Northwest Territories. For these jurisdictions, the variability attributable to sampling of the 2016 and 2021 adjusted census counts therefore does not explain the majority of the error of closure.

14. If the *t*-value is greater than 1.96 or less than -1.96, the DEP estimate is statistically different from the adjusted census count at a 95% confidence level.

Table 10.3.2
Impact of the adjusted censuses' statistical variability on the accuracy of population estimates for Canada, provinces and territories, 2016 to 2021

Provinces and territories	Error of closure	Combined standard error of the 2016 and 2021 adjusted censuses	*t*-value[1]
	number		
Canada	**-41,269**	**49,096**	**-0.84**
Newfoundland and Labrador	-6,540	2,456	-2.66
Prince Edward Island	2,564	1,238	2.07
Nova Scotia	-9,944	4,372	-2.27
New Brunswick	-317	3,542	-0.09
Quebec	33,890	23,487	1.44
Ontario	-43,978	38,965	-1.13
Manitoba	-3,084	6,420	-0.48
Saskatchewan	12,402	5,646	2.20
Alberta	2,013	16,334	0.12
British Columbia	-29,372	20,052	-1.46
Yukon	150	262	0.57
Northwest Territories	1,146	320	3.58
Nunavut	-199	331	-0.60

1. A *t*-value either greater than 1.96 or less than -1.96 indicates that the difference is signficant at the 95% level.
Source: Statistics Canada, Demography Division.

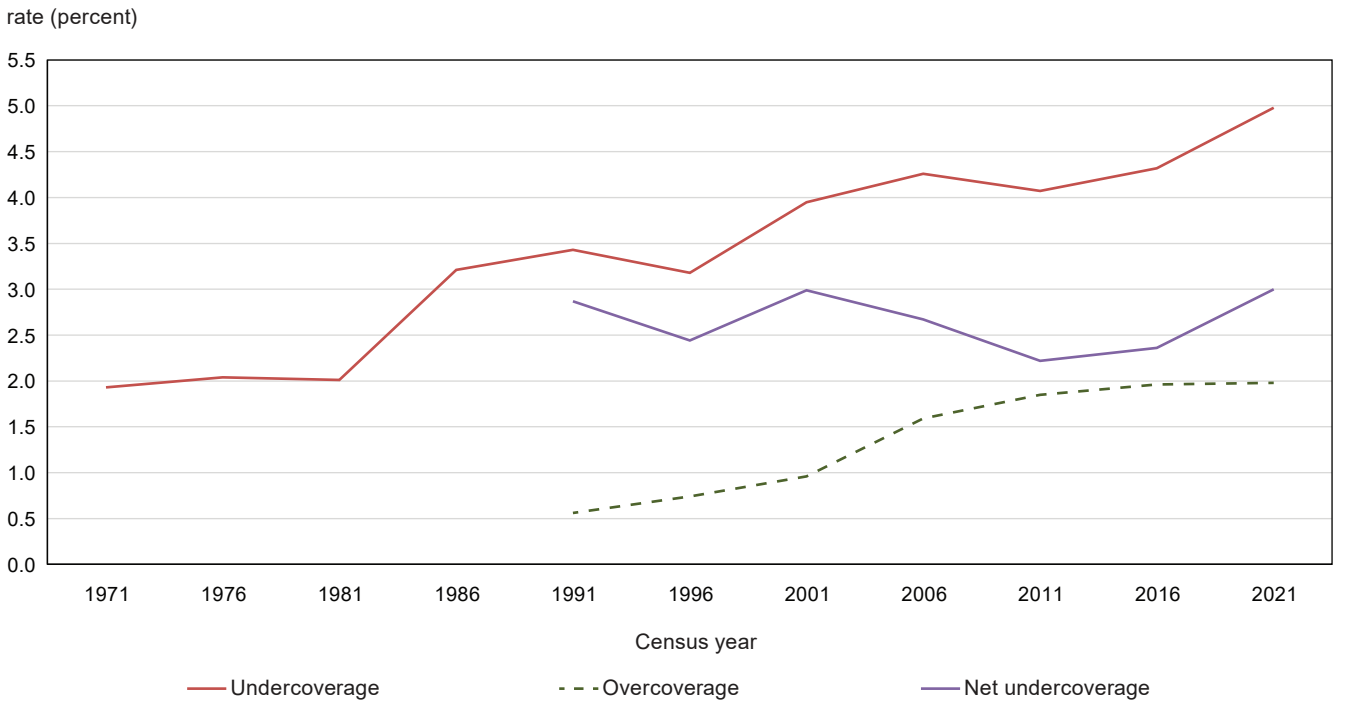
The components of demographic growth estimated by the DEP were compared with those from other sources, notably the CUS, to determine the components that could be more closely linked to the error of closure. This analysis focused on the five jurisdictions for which the error was statistically significant. Interprovincial migration, particularly that of recent immigrants, could explain part of the error of closure calculated for Prince Edward Island, Saskatchewan and the Northwest Territories. The impacts of the CUS sampling and several components of demographic growth could help to explain the error calculated for Newfoundland and Labrador, as well as for Nova Scotia. However, it is difficult to identify a primary factor for these two provinces. Lastly, emigration and the number of non-permanent residents generally remain demographic phenomena that are particularly difficult to measure.

11. Historical estimates of population coverage error

11.1 Estimates

This section presents historical estimates of population coverage error. Chart 11.1a shows the estimated population undercoverage rate $\hat{R}_U$ for the 1971 Census to the 2021 Census, as well as the estimated population overcoverage rate $\hat{R}_O$ and the estimated population net undercoverage rate $\hat{R}_N$ for the 1991 Census to the 2021 Census. Overcoverage and net undercoverage begin in 1991 because the overcoverage rate was first estimated for the 1991 Census, after an experimental study was conducted for the 1986 Census.

Chart 11.1a
Estimated rates of population coverage error for Canada, 1971 to 2021 censuses



Sources: Statistics Canada, 1971 to 2021 census coverage studies.

Population coverage error is a major data quality concern; undercoverage has increased by a factor of 2.5 since 1981, and overcoverage is two-and-a-half times higher than it was in 1996, although it has been fairly stable in the last three censuses. Changes in net undercoverage from census to census reflect changes in undercoverage and overcoverage, which in turn reflect changes in the demographic situation, in the living arrangements of Canadians, in census methodology and in the methodology of the coverage studies. The last issue is discussed in Section 11.2.

As shown in Chart 11.1a, the undercoverage rate rose sharply in 2021, one of the largest census-to-census increases, while the overcoverage rate changed very little. The undercoverage rate changed little in the 1971 to 1981 censuses, before increasing 1.2 percentage points in the 1986 Census. It held relatively steady from 1986 to 1996, then increased again in 2001, by 0.78 percentage points. It remained fairly stable again in the censuses from 2001 to 2016, with alternating small increases and declines, then rose again by 0.66 percentage points in 2021. As mentioned, some of these fluctuations can be explained by the improvements made to the coverage study methodology described in Section 11.2.

The overcoverage rate has increased in each census since it was first measured. The largest recorded increase in the overcoverage rate was from 0.96% in 2001 to 1.59% in 2006. In 2021, the rate peaked at 1.98%, although the small increase of 0.02 percentage points from 2016 was not statistically significant.

In 2021, net undercoverage increased significantly, the second census in a row in which it rose, following declines in the previous two censuses. It is the highest net undercoverage rate observed since it started being measured, although roughly equivalent to the rate observed in 2001. The increase from 2016 was statistically significant.

An examination of undercoverage since the 1981 Census shows that the increase observed in the 1986 Census led to the creation of the Address Register (AR) for the 1991 Census. The AR provided a separate list of urban dwellings that should have been enumerated. For the 1996 Census, the use of enumerators instead of self-enumeration in some central parts of large cities reduced undercoverage. In addition, moving Census Day from early June to mid-May helped to control undercoverage because individuals were more likely to be at home and less likely to be moving. In 2006, mailing out the questionnaires in urban areas reduced the number of employees required for collection. The introduction of online questionnaires also reduced data capture problems. In 2011, the adoption of a wave methodology made it possible to target census follow-up activities more effectively, and a sharp increase in online responses further reduced data capture problems. The elimination of the long-form questionnaire probably resulted in a slight decrease in the census non-response rate. In 2016, the reintroduction of the mandatory long-form questionnaire, the upward trend in online responses and the movement of public support for the census reduced the census non-response rate and improved the overall quality of the census. In 2021, a large increase in the number of non-permanent residents, a population with a typically much higher undercoverage rate, contributed to the rise in undercoverage. The pandemic could also have played a role in that result.

Estimates of undercoverage are presented in Table 11.1a and Table 11.1b. Note that 1971 is not included in Table 11.1b because estimates were produced for different age groups for persons older than 24.

Table 11.1a
Estimated population undercoverage rates and standard errors for Canada, provinces and territories, 1971 Census to 2021 Census

Provinces and territories	1971		1976		1981	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Canada	**1.93**	**0.09**	**2.04**	**0.10**	**2.01**	**0.09**
Newfoundland and Labrador	2.25	0.72	1.10	0.39	1.74	0.45
Prince Edward Island	1.23	1.13	0.38	0.25	1.17	0.54
Nova Scotia	1.33	0.45	0.86	0.34	1.05	0.34
New Brunswick	1.65	0.56	2.16	0.37	1.81	0.30
Quebec	2.10	0.19	2.95	0.25	1.91	0.21
Ontario	1.68	0.12	1.52	0.17	1.94	0.14
Manitoba	1.13	0.38	1.07	0.33	0.98	0.35
Saskatchewan	1.00	0.37	1.33	0.34	0.99	0.37
Alberta	2.55	0.44	1.49	0.26	2.54	0.36
British Columbia	2.89	0.39	3.13	0.31	3.16	0.33
Yukon	..	..	..	..	..	..
Northwest Territories	..	..	..	..	..	..
Nunavut	…	…	…	…	…	…

Table 11.1a
Estimated population undercoverage rates and standard errors for Canada, provinces and territories, 1971 Census to 2021 Census

Provinces and territories	1986		1991		1996		2001	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent							
Canada	**3.21**	**0.13**	**3.43**	**0.12**	**3.18**	**0.09**	**3.95**	**0.13**
Newfoundland and Labrador	1.92	0.33	2.47	0.30	2.45	0.29	2.43	0.32
Prince Edward Island	2.14	0.80	1.67	0.23	1.76	0.28	1.89	0.53
Nova Scotia	2.15	0.34	2.25	0.36	2.70	0.27	3.44	0.41
New Brunswick	2.71	0.33	3.71	0.42	2.49	0.28	3.57	0.42
Quebec	2.91	0.31	3.18	0.20	2.46	0.18	2.93	0.26
Ontario	3.43	0.19	4.23	0.28	3.40	0.18	4.56	0.25
Manitoba	2.94	0.40	2.31	0.36	2.55	0.29	3.49	0.43
Saskatchewan	2.38	0.37	2.15	0.32	3.30	0.32	3.18	0.37
Alberta	3.00	0.32	2.51	0.27	2.99	0.24	3.18	0.33
British Columbia	4.48	0.36	3.42	0.24	4.58	0.24	5.30	0.34
Yukon	..	..	4.12	0.58	3.92	0.51	5.59	1.16
Northwest Territories	..	..	5.73	0.57	4.28	0.67	9.10	0.80
Nunavut	…	…	…	…	6.54	0.63	5.07	1.39

Provinces and territories	2006		2011		2016		2021	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent							
Canada	**4.26**	**0.17**	**4.07**	**0.16**	**4.32**	**0.11**	**4.98**	**0.09**
Newfoundland and Labrador	2.62	0.54	3.70	0.53	3.94	0.36	5.02	0.32
Prince Edward Island	3.04	0.52	3.90	0.62	4.01	0.57	6.32	0.57
Nova Scotia	4.02	0.54	4.04	0.54	3.70	0.30	4.73	0.32
New Brunswick	3.56	0.43	2.64	0.43	4.24	0.34	3.82	0.32
Quebec	2.46	0.32	2.99	0.29	2.57	0.21	2.64	0.18
Ontario	5.18	0.34	4.47	0.32	4.63	0.22	5.78	0.18
Manitoba	4.32	0.57	3.11	0.48	3.95	0.35	4.52	0.35
Saskatchewan	3.81	0.50	4.43	0.57	4.99	0.37	5.06	0.36
Alberta	4.74	0.49	5.11	0.45	4.51	0.30	5.24	0.26
British Columbia	4.83	0.41	4.31	0.41	6.31	0.31	6.50	0.28
Yukon	7.23	0.64	6.30	0.81	8.42	0.45	7.71	0.37
Northwest Territories	5.74	0.57	5.99	0.69	7.83	0.53	9.91	0.46
Nunavut	5.55	0.60	7.39	1.65	4.25	0.59	9.70	0.56

.. not available for a specific reference period
... not applicable
Notes: Excludes incompletely enumerated reserves. The 1991 Census and subsequent censuses include non-permanent residents and territories. Figures include revisions to the original 1986 publication. Figures exclude estimates of people missed in dwellings misclassified as unoccupied in 1971 and 1976.
Sources: Statistics Canada, 1971 to 2021 census coverage studies.

Table 11.1b
Estimated population undercoverage rates and standard errors, sex and age group for Canada, 1976 Census to 2021 Census

Sex[1] and age groups	1976		1981		1986	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Both sexes	**2.04**	**0.10**	**2.01**	**0.09**	**3.21**	**0.13**
0 to 4 years	2.31	0.28	1.21	0.22	2.14	0.49
5 to 14 years	1.20	0.16	1.23	0.21	2.08	0.26
15 to 17 years[2]	1.99	0.38	2.96	0.52	3.58	0.60
18 to 19 years	..	..	..	..	..	..
20 to 24 years	5.31	0.38	5.51	0.29	8.66	0.46
25 to 34 years	2.85	0.28	2.31	0.28	4.51	0.35
35 to 44 years	1.54	0.26	2.20	0.26	2.32	0.31
45 to 54 years	1.22	0.33	0.81	0.23	1.58	0.29
55 to 64 years	0.92	0.20	0.91	0.29	2.06	0.31
65 years and older	1.20	0.25	0.71	0.30	1.76	0.31
Males[3]	**2.46**	**0.17**	**2.37**	**0.13**	**3.75**	**0.16**
0 to 4 years	2.53	0.46	1.32	0.33	2.22	0.67
5 to 14 years	1.14	0.21	1.27	0.29	1.98	0.32
15 to 17 years[2]	1.93	0.48	3.12	0.68	4.09	0.74
18 to 19 years	..	..	..	..	..	..
20 to 24 years	5.99	0.52	6.03	0.48	10.36	0.57
25 to 34 years	3.64	0.46	2.70	0.44	5.43	0.45
35 to 44 years	2.33	0.48	3.42	0.40	3.29	0.51
45 to 54 years	1.63	0.41	1.21	0.38	1.95	0.52
55 to 64 years	1.28	0.34	0.91	0.40	1.88	0.47
65 years and older	1.90	0.44	0.69	0.47	1.57	0.50
Females[4]	**1.61**	**0.10**	**1.65**	**0.12**	**2.68**	**0.17**
0 to 4 years	2.07	0.36	1.10	0.33	2.06	0.62
5 to 14 years	1.26	0.27	1.19	0.31	2.20	0.33
15 to 17 years[2]	2.05	0.51	2.80	0.73	3.05	0.76
18 to 19 years	..	..	..	..	..	..
20 to 24 years	4.62	0.48	4.98	0.43	6.89	0.72
25 to 34 years	2.03	0.38	1.92	0.32	3.59	0.45
35 to 44 years	0.72	0.24	0.93	0.31	1.33	0.32
45 to 54 years	0.81	0.38	0.41	0.26	1.20	0.35
55 to 64 years	0.58	0.25	0.92	0.34	2.23	0.50
65 years and older	0.64	0.38	0.71	0.42	1.89	0.44

Table 11.1b
Estimated population undercoverage rates and standard errors, sex and age group for Canada, 1976 Census to 2021 Census

Sex[1] and age groups	1991		1996		2001		2006	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent							
Both sexes	**3.43**	**0.12**	**3.18**	**0.09**	**3.95**	**0.13**	**4.26**	**0.17**
0 to 4 years	3.55	0.49	2.89	0.36	4.42	0.71	4.07	0.65
5 to 14 years	2.49	0.27	1.45	0.14	2.90	0.38	3.10	0.46
15 to 17 years[2]	3.75	0.42	3.48	0.42	4.36	0.53	1.56	0.60
18 to 19 years	..	..	..	..	..	..	8.86	1.58
20 to 24 years	8.18	0.52	8.00	0.34	9.85	0.62	10.50	0.74
25 to 34 years	5.65	0.35	5.81	0.29	8.07	0.36	9.43	0.56
35 to 44 years	2.84	0.29	2.78	0.24	4.04	0.33	5.36	0.50
45 to 54 years	1.61	0.27	1.90	0.21	1.79	0.29	2.64	0.43
55 to 64 years	1.69	0.28	2.23	0.34	1.22	0.37	0.95	0.53
65 years and older	1.51	0.28	1.52	0.26	1.29	0.34	0.21	0.40
Males[3]	**3.95**	**0.16**	**3.89**	**0.14**	**4.90**	**0.19**	**5.51**	**0.26**
0 to 4 years	2.79	0.58	2.56	0.47	3.36	0.89	4.24	0.95
5 to 14 years	2.32	0.34	1.46	0.24	2.38	0.49	3.04	0.64
15 to 17 years[2]	3.55	0.60	3.68	0.43	5.49	0.80	1.88	0.88
18 to 19 years	..	..	..	..	..	..	10.06	2.45
20 to 24 years	8.98	0.81	9.48	0.50	11.68	0.92	12.21	1.12
25 to 34 years	7.28	0.56	7.74	0.42	10.67	0.55	11.42	0.86
35 to 44 years	3.65	0.41	3.94	0.39	5.71	0.51	7.77	0.79
45 to 54 years	2.05	0.45	2.12	0.27	2.50	0.44	4.14	0.69
55 to 64 years	2.04	0.44	2.50	0.54	1.35	0.54	2.13	0.77
65 years and older	1.41	0.50	1.64	0.45	1.50	0.53	-0.05	0.56
Females[4]	**2.93**	**0.17**	**2.49**	**0.12**	**3.02**	**0.18**	**3.04**	**0.23**
0 to 4 years	4.35	0.71	3.24	0.55	5.50	1.14	3.88	0.92
5 to 14 years	2.65	0.39	1.45	0.22	3.44	0.58	3.17	0.66
15 to 17 years[2]	3.96	0.54	3.28	0.55	3.13	0.69	1.23	0.83
18 to 19 years	..	..	..	..	..	..	7.58	1.96
20 to 24 years	7.36	0.71	6.45	0.48	7.91	0.84	8.70	0.98
25 to 34 years	3.98	0.37	3.84	0.40	5.41	0.46	7.43	0.73
35 to 44 years	2.01	0.35	1.62	0.28	2.35	0.43	2.90	0.61
45 to 54 years	1.16	0.34	1.68	0.33	1.09	0.37	1.13	0.51
55 to 64 years	1.35	0.33	1.97	0.40	1.09	0.52	-0.22	0.73
65 years and older	1.58	0.36	1.43	0.32	1.13	0.45	0.40	0.56

Table 11.1b
Estimated population undercoverage rates and standard errors, sex and age group for Canada, 1976 Census to 2021 Census

Sex[1] and age groups	2011		2016		2021	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Both sexes	**4.07**	**0.16**	**4.32**	**0.11**	**4.98**	**0.09**
0 to 4 years	3.36	0.62	3.79	0.51	5.19	0.29
5 to 14 years	2.61	0.42	2.94	0.37	3.35	0.29
15 to 17 years[2]	3.83	0.85	4.35	0.78	3.40	0.68
18 to 19 years	6.28	0.93	7.71	1.25	5.91	1.03
20 to 24 years	9.60	0.69	9.64	0.58	12.91	0.66
25 to 34 years	8.96	0.48	8.60	0.42	9.89	0.39
35 to 44 years	4.66	0.45	5.07	0.41	5.49	0.36
45 to 54 years	2.95	0.42	3.86	0.39	4.01	0.38
55 to 64 years	1.02	0.41	2.21	0.42	3.10	0.34
65 years and older	1.19	0.45	0.77	0.29	1.25	0.24
Males[3]	**5.07**	**0.24**	**5.27**	**0.18**	**6.01**	**0.17**
0 to 4 years	3.14	0.82	4.22	0.68	4.80	0.56
5 to 14 years	3.00	0.62	2.35	0.54	3.57	0.39
15 to 17 years[2]	4.31	1.13	4.55	1.11	3.53	1.02
18 to 19 years	5.42	1.19	7.38	1.82	5.70	1.34
20 to 24 years	9.37	0.88	11.46	0.88	12.97	0.92
25 to 34 years	10.54	0.73	10.44	0.63	12.00	0.60
35 to 44 years	6.34	0.68	6.24	0.63	7.48	0.58
45 to 54 years	4.69	0.66	5.16	0.59	4.89	0.60
55 to 64 years	2.58	0.69	3.11	0.67	4.42	0.54
65 years and older	1.32	0.60	1.23	0.41	1.64	0.39
Females[4]	**3.08**	**0.22**	**3.39**	**0.18**	**3.95**	**0.14**
0 to 4 years	3.59	0.95	3.34	0.87	5.60	0.57
5 to 14 years	2.20	0.57	3.56	0.56	3.12	0.44
15 to 17 years[2]	3.31	1.28	4.13	1.06	3.26	0.94
18 to 19 years	7.17	1.45	8.07	1.77	6.14	1.59
20 to 24 years	9.83	1.07	7.66	0.77	12.86	1.06
25 to 34 years	7.37	0.63	6.71	0.56	7.66	0.53
35 to 44 years	2.99	0.58	3.93	0.53	3.49	0.39
45 to 54 years	1.21	0.52	2.56	0.55	3.14	0.50
55 to 64 years	-0.52	0.44	1.32	0.50	1.80	0.42
65 years and older	1.08	0.66	0.39	0.42	0.91	0.33

.. not available for a specific reference period
1. For 2021, estimates were produced by gender rather than by sex.
2. Data for all years except 2006 to 2021 are for people aged 15 to 19.
3. For 2021, data are for men+. The term “men+” includes men (and boys) and some non-binary people.
4. For 2021, data are for women+. The term “women+” includes women (and girls) and some non-binary people.
Notes: Excludes incompletely enumerated reserves. The 1991 Census and subsequent censuses include non-permanent residents and the territories. Rates include revisions to the original 1986 publication. Rates exclude estimates of people missed in dwellings misclassified as unoccupied in 1976.
Sources: Statistics Canada, 1976 to 2021 census coverage studies.

The following can be observed from these tables.

Undercoverage is usually higher in the three territories. Among the provinces, undercoverage is generally higher in British Columbia and Ontario. However, in recent censuses, the undercoverage rate in Alberta and Saskatchewan has been close to Ontario's rate, and has even exceeded it a few times. Between 1971 and 2001, British Columbia was the province with the highest undercoverage rate in every census, except in 1991, when Ontario had the highest rate. Ontario had the highest rate in 2006 as well, while Alberta had it in 2011. In 2016 and 2021, British Columbia again had the highest rate, followed in 2021 by Prince Edward Island, a province that never had one of the highest rates in the past. Undercoverage rates for Quebec and the Atlantic provinces tend to be lower than the national rate.

Undercoverage was higher for young adults and higher for men+ than for women+ in 2021, two persistent demographic trends. Prior to the 2021 coverage studies, coverage errors were estimated by sex rather than by gender. As shown in Table 11.1b, undercoverage for males was higher than undercoverage for females for every census since 1971. Table 11.1b also shows that undercoverage for young men aged 20 to 24 was higher than undercoverage for all males. This was also the case for women aged 20 to 24, but the rate for women aged 20 to 24 was lower than the rate for men in the same age group in every census except 2011, when the rates were 9.83% and 9.37%, respectively. In 2021, the undercoverage rate for women+ aged 20 to 24 was very close to the one for men+, and both were higher than any rate observed for males or females in previous censuses. The undercoverage rates for adults aged 25 to 34 were also high. The undercoverage rate was higher for young adults, partly because of their less stable living arrangements. Young adults are more likely than older adults or children to change their living arrangements because they are moving away from home to work or to attend a postsecondary institution, or they are moving in with friends or spouses.

Estimates of overcoverage rates are presented in Table 11.1c and Table 11.1d.

Table 11.1c
Estimated population overcoverage rates and standard errors for Canada, provinces and territories, 1991 Census to 2021 Census

Provinces and territories	1991		1996		2001	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Canada	**0.56**	**0.04**	**0.74**	**0.04**	**0.96**	**0.05**
Newfoundland and Labrador	0.48	0.09	0.77	0.12	0.63	0.10
Prince Edward Island	0.74	0.15	0.91	0.14	0.92	0.18
Nova Scotia	0.36	0.09	0.47	0.07	0.81	0.14
New Brunswick	0.46	0.09	0.60	0.10	0.89	0.19
Quebec	0.51	0.07	0.85	0.08	1.03	0.10
Ontario	0.59	0.07	0.67	0.07	0.88	0.09
Manitoba	0.45	0.11	0.88	0.15	0.80	0.15
Saskatchewan	0.35	0.08	0.55	0.11	1.06	0.20
Alberta	0.51	0.09	0.59	0.10	0.89	0.13
British Columbia	0.68	0.10	0.89	0.09	1.26	0.12
Yukon	0.29	0.07	0.70	0.17	0.86	0.16
Northwest Territories	0.29	0.07	1.32	0.22	1.00	0.11
Nunavut	…	…	0.99	0.22	0.59	0.10

Table 11.1c
Estimated population overcoverage rates and standard errors for Canada, provinces and territories, 1991 Census to 2021 Census

Provinces and territories	2006		2011		2016		2021	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent							
Canada	**1.59**	**0.01**	**1.85**	**0.02**	**1.96**	**0.04**	**1.98**	**0.02**
Newfoundland and Labrador	1.63	0.05	1.76	0.04	2.09	0.07	1.94	0.08
Prince Edward Island	1.66	0.06	1.54	0.04	1.65	0.07	2.04	0.12
Nova Scotia	1.40	0.03	1.72	0.04	1.81	0.07	1.94	0.07
New Brunswick	1.41	0.03	2.12	0.05	2.18	0.08	2.10	0.08
Quebec	1.66	0.02	2.07	0.03	2.15	0.12	2.14	0.07
Ontario	1.49	0.02	1.67	0.04	1.87	0.08	1.82	0.05
Manitoba	1.42	0.04	1.35	0.04	1.51	0.06	1.68	0.05
Saskatchewan	1.53	0.04	1.65	0.04	1.91	0.14	2.07	0.06
Alberta	1.47	0.02	1.70	0.05	1.74	0.06	1.77	0.06
British Columbia	1.96	0.03	2.28	0.05	2.24	0.07	2.41	0.05
Yukon	1.62	0.08	2.45	0.05	2.22	0.08	1.94	0.09
Northwest Territories	1.98	0.08	1.44	0.11	1.25	0.06	1.87	0.03
Nunavut	1.44	0.07	1.17	0.07	1.73	0.07	1.88	0.04

… not applicable
Notes: Excludes incompletely enumerated reserves. Figures include non-permanent residents.
Sources: Statistics Canada, 1991 to 2021 census coverage studies.

Table 11.1d
Estimated population overcoverage rates and standard errors, sex and age groups for Canada, 1996 Census to 2021 Census

Sex[1] and age groups	1996		2001		2006	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Both sexes	**0.74**	**0.04**	**0.96**	**0.05**	**1.59**	**0.01**
0 to 4 years	0.61	0.10	0.96	0.18	1.35	0.07
5 to 14 years	0.96	0.09	1.52	0.15	2.24	0.07
15 to 17 years[2]	1.24	0.15	1.85	0.26	2.33	0.14
18 to 19 years	..	..	..	..	2.65	0.17
20 to 24 years	2.44	0.28	2.66	0.32	2.88	0.11
25 to 34 years	0.66	0.08	0.92	0.09	1.43	0.06
35 to 44 years	0.38	0.06	0.49	0.06	1.05	0.05
45 to 54 years	0.48	0.11	0.39	0.04	1.13	0.05
55 to 64 years	0.52	0.11	0.38	0.05	1.24	0.06
65 years and older	0.36	0.07	0.77	0.21	1.60	0.06

Table 11.1d
Estimated population overcoverage rates and standard errors, sex and age groups for Canada, 1996 Census to 2021 Census

Sex[1] and age groups	1996		2001		2006	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Males[3]	**0.70**	**0.04**	**0.92**	**0.06**	**1.62**	**0.02**
0 to 4 years	0.52	0.09	0.69	0.07	1.35	0.09
5 to 14 years	0.99	0.15	1.59	0.21	2.25	0.10
15 to 17 years[2]	1.12	0.24	1.45	0.31	2.37	0.20
18 to 19 years	..	..	..	..	2.28	0.21
20 to 24 years	2.34	0.34	2.44	0.45	2.75	0.15
25 to 34 years	0.65	0.11	1.03	0.14	1.51	0.08
35 to 44 years	0.38	0.06	0.46	0.06	1.10	0.06
45 to 54 years	0.35	0.07	0.34	0.03	1.16	0.07
55 to 64 years	0.37	0.12	0.33	0.04	1.30	0.09
65 years and older	0.33	0.02	0.74	0.21	1.69	0.10
Females[4]	**0.77**	**0.06**	**1.00**	**0.08**	**1.56**	**0.01**
0 to 4 years	0.69	0.18	1.25	0.36	1.35	0.10
5 to 14 years	0.92	0.14	1.44	0.21	2.23	0.10
15 to 17 years[2]	1.36	0.29	2.27	0.43	2.28	0.19
18 to 19 years	..	..	..	..	3.04	0.28
20 to 24 years	2.55	0.46	2.89	0.46	3.01	0.17
25 to 34 years	0.66	0.11	0.81	0.12	1.35	0.08
35 to 44 years	0.37	0.10	0.53	0.11	0.99	0.06
45 to 54 years	0.61	0.20	0.43	0.07	1.11	0.06
55 to 64 years	0.66	0.19	0.42	0.09	1.18	0.07
65 years and older	0.38	0.11	0.80	0.33	1.53	0.08

Sex[1] and age groups	2011		2016		2021	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Both sexes	**1.85**	**0.02**	**1.96**	**0.04**	**1.98**	**0.02**
0 to 4 years	1.61	0.10	1.76	0.19	1.74	0.08
5 to 14 years	2.79	0.10	3.49	0.20	3.21	0.09
15 to 17 years[2]	2.98	0.23	3.25	0.35	3.58	0.19
18 to 19 years	3.37	0.27	3.27	0.40	3.71	0.26
20 to 24 years	3.11	0.13	3.51	0.26	3.52	0.15
25 to 34 years	1.69	0.08	2.19	0.17	2.06	0.08
35 to 44 years	1.23	0.06	1.12	0.11	1.25	0.06
45 to 54 years	1.36	0.06	1.18	0.10	1.32	0.07
55 to 64 years	1.50	0.07	1.44	0.14	1.51	0.07
65 years and older	1.64	0.08	1.55	0.10	1.59	0.05

Table 11.1d
Estimated population overcoverage rates and standard errors, sex and age groups for Canada, 1996 Census to 2021 Census

Sex[1] and age groups	2011		2016		2021	
	Estimated rate	Standard error	Estimated rate	Standard error	Estimated rate	Standard error
	percent					
Males[3]	**1.86**	**0.04**	**1.95**	**0.07**	**2.01**	**0.04**
0 to 4 years	1.65	0.13	1.56	0.24	1.69	0.11
5 to 14 years	2.77	0.14	3.50	0.28	3.23	0.12
15 to 17 years[2]	3.04	0.27	4.14	0.58	3.72	0.29
18 to 19 years	3.06	0.33	2.56	0.44	3.55	0.35
20 to 24 years	3.03	0.19	3.27	0.37	3.53	0.21
25 to 34 years	1.75	0.11	2.29	0.24	2.12	0.12
35 to 44 years	1.26	0.09	1.21	0.16	1.31	0.11
45 to 54 years	1.32	0.09	1.17	0.13	1.27	0.07
55 to 64 years	1.54	0.11	1.22	0.20	1.52	0.09
65 years and older	1.68	0.14	1.47	0.15	1.59	0.08
Females[4]	**1.83**	**0.04**	**1.98**	**0.07**	**1.95**	**0.04**
0 to 4 years	1.57	0.15	1.96	0.31	1.78	0.12
5 to 14 years	2.81	0.15	3.49	0.29	3.20	0.12
15 to 17 years[2]	2.93	0.38	2.32	0.38	3.44	0.25
18 to 19 years	3.69	0.43	4.02	0.68	3.89	0.38
20 to 24 years	3.19	0.18	3.77	0.38	3.50	0.21
25 to 34 years	1.63	0.11	2.08	0.23	1.99	0.10
35 to 44 years	1.20	0.10	1.04	0.16	1.18	0.06
45 to 54 years	1.39	0.09	1.19	0.14	1.36	0.12
55 to 64 years	1.46	0.10	1.66	0.20	1.50	0.10
65 years and older	1.60	0.10	1.63	0.14	1.60	0.08

.. not available for a specific reference period
1. For 2021, estimates were produced by gender rather than by sex.
2. Data for 1996 and 2001 are for people aged 15 to 19.
3. For 2021, data are for men+. The term “men+” includes men (and boys) and some non-binary people.
4. For 2021, data are for women+. The term “women+” includes women (and girls) and some non-binary people.
Note: Excludes incompletely enumerated reserves.
Sources: Statistics Canada, 1996 to 2021 census coverage studies.

The following can be observed from these tables.

Overcoverage is consistently higher for British Columbia than for the other provinces. British Columbia has been the province with the highest rate of population overcoverage for the past five censuses. Of the three territories, Yukon had the highest overcoverage rate for the third consecutive census, whereas the Northwest Territories had the highest rate from 1996 to 2006. Overcoverage was up in eight provinces and territories in 2021 and down in the five others, but the difference was statistically significant for less than half of those. The national overcoverage rate remained almost identical from 2016 to 2021. Manitoba had the lowest rate of all the provinces in each of the last three censuses, second-lowest only to the Northwest Territories in 2016 among the provinces and territories.

As shown in Table 11.1d, overcoverage was more common for school-aged children and young adults. For school-aged children, this situation was largely because children whose parents do not live together are often enumerated by both parents. Overcoverage for young adults is probably attributable to the same less stable living arrangements that can also lead to undercoverage. Nationally, overcoverage rates were above 3% for young adults (ages 18 to 24) and for youth aged 5 to 17. The 15-to-19 age group saw the largest increase in overcoverage relative to 2016.

11.2 Changes in the design of population coverage studies

Because of differences in the design of the coverage studies over time, the rates in Table 11.1a, Table 11.1b, Table 11.1c and Table 11.1d are not strictly comparable. A list of the methodological changes made since 1976 is provided below. It is worth noting that the fundamentals of the approach to measuring undercoverage in the Census Undercoverage Study (CUS) (known as the Reverse Record Check until the 2021 Census) have not changed much since the 1966 Census. A sample is selected from frames that cover the target population independent of the census. Census records are then checked to determine whether the sampled persons were actually enumerated. More changes were made to the overcoverage measurement. Multiple studies were carried out in 1991, 1996 and 2001. In 1996, the CUS scope was expanded to include overcoverage. In 2006, the CUS was no longer used to estimate overcoverage, and a new study was introduced to measure all overcoverage cases based on probabilistic and exact matches using name, date of birth and sex.

2021 Census coverage studies

a) The stratification for the CUS census and territorial frames has been improved. These whole frames have been linked to the 2021 Census Response Database (RDB) and to tax files, as well as to death records from vital statistics in the case of the census frame. With information from these linkages, units on the frames were then stratified based on their likelihood of being enumerated, missed or out of scope for the census. This made for a more efficient design.

b) Changes were made to the CUS collection operations. The CUS transitioned from the Blaise system to the Integrated Collection and Operation System, a standard web-based application developed at Statistics Canada and in use since the 2016 Census. The paper questionnaire that was mailed out to respondents upon request or during non-response follow-up operations was replaced by a self-response electronic questionnaire. Also, more tracing updates were provided during collection to help find and interview the selected respondents.

c) To improve the classification of persons whose main residence was outside Canada on Census Day 2021, donor imputation was used for item non-response to the out-of-Canada module of the CUS questionnaire.

d) In the CUS, an adjustment for some influential weights and a weight calibration for the birth and immigrant frames were added to the weighting methodology.

e) A deterministic linkage was added to the steps used to create the Census Overcoverage Study (COS) sampling frame of potential census duplicates. This deterministic linkage is based on a modification of the Automated Match Study (AMS) methodology.

f) The Generalized Weight Share Method was used to derive a weight for pairs of overcoverage not on the COS frame that were found while manually verifying household pairs associated with sampled person pairs, so that these pairs could be included in the COS estimates.

g) The estimates of coverage errors by age group and sex have been replaced with estimates by age group and gender, with two gender categories (men+ and women+).

2016 Census coverage studies

a) For the CUS census frame, the stratification was based on the province of residence, which was updated using tax data. Also, a deceased stratum was determined prior to sample selection. These improvements made for a more efficient sample.

b) An “administrative” household was created using tax data for each selected person from the CUS to make processing and tracing more efficient.

c) The CUS non-response adjustment for untraced persons no longer uses the subsample. The adjustment is made within response homogeneity groups composed of persons with similar probabilities of being in the target population and responding to the survey, as estimated with available data for both respondents and non-respondents. Since the non-response adjustment model was finalized after collection, the subsample that was used for the adjustment for untraced persons in previous CUSs was still sent to collection, but was not used. It will not be required for the next CUSs.

d) A calibration was performed to adjust the weights from the census frame when a provincial sample contained too many or not enough enumerated or deceased persons. For the three territories, a calibration to the comparable number of enumerated persons was performed, like in previous CUSs, except it was done separately for six age–sex groups in each territory, rather than using a single overall calibration group for each of them, as had been done in the past.

e) Hierarchical deterministic waves were used to form blocks for the probabilistic linkages for steps 1 and 2 of the COS to improve frame coverage.

f) The linkage rules for steps 1 and 2 were expanded to identify as many potential overcoverage cases as possible.

g) Although Step 3 (extension) of the COS was restricted to private household pairs in 2011, it was expanded to include the household pairs produced for each person pair created in steps 1 and 2 and for which one of the two households was in a collective dwelling.

h) Instead of being created separately for each step, the groups of records that could represent the same person were created by combining steps 1, 2 and 3 of the COS to better account for cases of multiple overcoverage.

Like every CUS since 1996, the 2016 CUS did not estimate the number of persons missed on incompletely enumerated reserves and settlements. For more information on this subject, see Section 12.2.

2011 Census coverage studies

a) The 2011 CUS was very similar to the 2006 CUS. To make it more efficient, some changes were introduced, including improvements in the monster match program, more effective strategies for searching the census RDB and the use of new birth frames.

b) For the first time, the census frame sample weighting considered the overcoverage in the census frame.

c) With automated methods, it was possible to use provincial and territorial parameters instead of national parameters to develop the COS frame.

2006 Census coverage studies

For the CUS and the COS, the name field added to the 2006 Census RDB was used optimally in matching and searching. In addition, the following changes were made:

a) The measurement of overcoverage was restricted to the COS. The CUS methodology was subsequently changed so that not all cases were sent for field collection. Since 2006, the CUS has had a processing step that is carried out prior to collection to determine whether collection is required. The version of the census RDB used for the census coverage studies (CCS-RDB) was searched for the sampled persons using information from the sampling frames and various update sources, such as tax data. If the search located the sampled person in the CCS-RDB, collection was not required. The only exception was a sample of persons who had been located to collect data required for the non-response adjustment.

b) The three coverage studies conducted in 2001 to measure overcoverage were replaced by the COS in 2006. The COS used a different methodology from the one used in previous overcoverage studies. Essentially, it employed a linkage technique based on surnames, given names, sex and date of birth and manual verification to measure overcoverage.

2001 Census coverage studies

a) The institutional component of the Collective Dwelling Study (CDS) was dropped, and overcoverage estimates for this population were produced by the CUS.

b) The Dwelling Classification Study (DCS) replaced the vacancy check (VC), which was used in previous censuses to check the classification of dwellings that the enumerator had determined to be unoccupied. The DCS is an extension of the VC, which estimated the number of persons living in non-response dwellings.

1996 Census coverage studies

a) The 1996 CUS did not estimate the number of persons missed on incompletely enumerated reserves.

b) The Temporary Resident Study was cancelled because of concerns about data quality, and because it was recognized that the CUS would measure this type of undercoverage appropriately.

c) A measure of overcoverage that was more comprehensive than the 1991 measure was produced by incorporating the Private Dwelling Study (PDS) into the CUS so that each sampled person could be identified as having been enumerated more than once. This approach resulted in an increase in the number of addresses to be processed where overcoverage could have occurred. Also, compared with 1991, the AMS was expanded substantially so that overcoverage could be determined not only for an enumeration area, but also for a large region (Atlantic provinces, Quebec, Ontario, Western Canada and the territories).

1991 Census coverage studies

a) Non-permanent residents were included in the target population for the first time.

b) Following experimental studies in 1986, the measurement of population overcoverage began in 1991. The results of three studies—the Private Dwelling Study, the Collective Dwelling Study and the Automated Match Study—were combined to form a comprehensive estimate.

1986 Census coverage studies

The rates shown in Table 11.1a and Table 11.1b for the 1986 Census differ from the results published in the *User's Guide to the Quality of 1986 Census Data: Coverage*, which included revisions made after the 1986 publication, when incompletely enumerated reserves were considered missed. In the original 1986 publication, they were counted as "enumerated" since provincial data included an estimate of persons missed for reserves.

1976 Census coverage studies

Census data did not include an estimate from the VC of persons missed in dwellings incorrectly classified as unoccupied. The 1976 population undercoverage rate would have been 1.78% if it had included the results of the 1976 VC. There was no VC in the 1971 Census.

For more details on the history of coverage studies, see Dolson (2010).

12. Special topics

12.1 Collection undercoverage

Up to this point, this report focused on undercoverage in the census population counts. This section deals with the concept of Census of Population collection undercoverage. It is useful to expand the concept of undercoverage to include persons not enumerated for any reason. Undercoverage is defined as the number of persons not included in the census counts. As discussed in Section 3.3, the census counts C are composed of two elements, $C = E + I$, where E = the number of enumerations and I = the number of persons imputed.

Undercoverage, therefore, represents only part of all persons who were not listed on a census form but should have been. It does not include persons who were not enumerated either because no census form was returned for the dwelling (non-response dwelling) or because the dwelling was erroneously classified as unoccupied (occupied dwelling misclassified as vacant) and was not covered by non-response follow-up.

Also discussed in Section 3.3, an estimate of the actual number of persons in the census target population is given by:

$$\hat{T} = C + \hat{N} = C + \hat{U} - \hat{O}$$

If we combine these two equations, we get:

$$\hat{T} = C + \hat{N} = C + \hat{U} - \hat{O} = E + \left(I + \hat{U}\right) - \hat{O}$$

This formulation of $\hat{T}$ has three components:

E = the number of persons listed on a census form[15] (enumerations)

$\hat{O}$ = an estimate of the number of excess enumerations[16]

$\left(I + \hat{U}\right)$ = an estimate of the number of persons who were not listed on a census form but should have been.

The last component, $\left(I + \hat{U}\right)$, is an estimate of the number of persons missed in the census for any reason. The Census of Population collection undercoverage, (L), refers to persons not enumerated for any reason. The estimate of census collection undercoverage is given by:

$$\hat{L} = \left(I + \hat{U}\right)$$

The corresponding estimate of the Census of Population collection undercoverage rate is:

$$\hat{R}_L = 100 * \frac{\hat{L}}{\hat{T}} = 100 * \left(\frac{I + \hat{U}}{C + \hat{N}}\right)$$

Census of Population net collection undercoverage can be calculated by subtracting overcoverage $\hat{O}$ from $\hat{L}$:

$$\hat{T} = C + \hat{N} = E + \left(I + \hat{U}\right) - \hat{O} = E + \hat{L} - \hat{O}$$

15. It is possible that some of the persons listed on a census form may not appear in the final census database. Therefore, the term “persons listed on a census form” is used in this section to refer to persons in the final census database.

16. Most cases of overcoverage involve duplicate enumerations, where the same person appears twice in the database. However, in a small number of cases, the same person can appear more than twice. The variable $\hat{O}$ denotes the estimate of the number of excess enumerations rather than the number of persons involved in multiple enumerations.

Although net collection undercoverage cannot be applied to census data to adjust for coverage error, $\hat{L}$ and $\hat{R}_L$ provide a broader picture of how well the census was able to enumerate the target population. In fact, they include persons not enumerated, whether or not they were taken into account in the census through imputations. Table 12.1 shows the 2021 Census of Population collection undercoverage estimates $\hat{L}$ and $\hat{R}_L$ and the population undercoverage estimates $\hat{U}$ and $\hat{R}_U$ (also see Table 1.3). It also shows their estimated standard errors. The results are shown by province or territory and for Canada, and by age group and gender. At the national level, the 2021 Census collection enumerated 92.58% of the target population ($100 - \hat{R}_L$), whereas the official published 2021 Census count represented 95.02% of the persons who should have been included in this figure $(100 - \hat{R}_U)$. The difference between these two rates is simply attributable to the inclusion of the imputations in $\hat{R}_L$. The 2016 Census enumerated 93.63% of the target population, when 95.68% of the persons who should have been included in the official 2016 Census counts were included. Therefore, the 2021 Census collection enumerated a slightly lower percentage of the target population than the previous census, and there were more imputations than in 2016.

Table 12.1
Estimated population collection undercoverage, population undercoverage and standard errors for various characteristics, 2021 Census

	Population collection undercoverage				Population undercoverage			
Characteristics	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
Canada	**2,829,222**	**37,004**	**7.42**	**0.09**	**1,897,876**	**37,004**	**4.98**	**0.09**
Provinces and territories								
Newfoundland and Labrador	39,055	1,775	7.41	0.31	26,428	1,775	5.02	0.32
Prince Edward Island	13,177	977	8.17	0.56	10,190	977	6.32	0.57
Nova Scotia	70,323	3,373	7.05	0.31	47,196	3,373	4.73	0.32
New Brunswick	50,361	2,655	6.38	0.31	30,165	2,655	3.82	0.32
Quebec	410,770	15,756	4.81	0.18	225,381	15,756	2.64	0.18
Ontario	1,162,640	27,795	7.85	0.17	855,470	27,795	5.78	0.18
Manitoba	106,952	5,089	7.74	0.34	62,443	5,089	4.52	0.35
Saskatchewan	106,650	4,411	9.14	0.34	59,075	4,411	5.06	0.36
Alberta	357,305	12,225	8.09	0.25	231,249	12,225	5.24	0.26
British Columbia	488,394	15,827	9.37	0.28	338,678	15,827	6.50	0.28
Yukon	4,768	170	11.17	0.35	3,294	170	7.71	0.37
Northwest Territories	7,571	227	16.95	0.42	4,426	227	9.91	0.46
Nunavut	11,254	250	28.14	0.45	3,879	250	9.70	0.56
Gender and age groups								
Total gender, all ages	**2,829,222**	**37,004**	**7.42**	**0.09**	**1,897,876**	**37,004**	**4.98**	**0.09**
0 to 4 years	143,827	5,809	7.58	0.28	98,438	5,809	5.19	0.29
5 to 14 years	231,406	12,630	5.53	0.29	140,303	12,630	3.35	0.29
15 to 17 years	65,929	8,516	5.45	0.67	41,091	8,516	3.40	0.68
18 to 19 years	66,568	8,933	8.13	1.00	48,426	8,933	5.91	1.03
20 to 24 years	375,100	18,381	15.43	0.64	313,912	18,381	12.91	0.66

Table 12.1
Estimated population collection undercoverage, population undercoverage and standard errors for various characteristics, 2021 Census

	Population collection undercoverage				Population undercoverage			
Characteristics	Estimated number	Standard error	Estimated rate (%)	Standard error (%)	Estimated number	Standard error	Estimated rate (%)	Standard error (%)
25 to 34 years	673,702	22,964	12.57	0.37	530,315	22,964	9.89	0.39
35 to 44 years	402,817	19,332	7.85	0.35	281,659	19,332	5.49	0.36
45 to 54 years	301,743	19,130	6.28	0.37	192,317	19,130	4.01	0.38
55 to 64 years	299,714	18,459	5.65	0.33	164,211	18,459	3.10	0.34
65 to 74 years	158,130	13,112	3.91	0.31	49,955	13,112	1.23	0.32
75 years and older	110,285	11,803	3.74	0.39	37,248	11,803	1.26	0.40
Men+,[1] all ages	**1,606,471**	**34,194**	**8.46**	**0.16**	**1,140,736**	**34,194**	**6.01**	**0.17**
0 to 4 years	69,726	5,668	7.20	0.54	46,527	5,668	4.80	0.56
5 to 14 years	123,836	8,684	5.75	0.38	76,859	8,684	3.57	0.39
15 to 17 years	35,012	6,595	5.61	1.00	22,004	6,595	3.53	1.02
18 to 19 years	33,034	5,962	7.85	1.31	23,968	5,962	5.70	1.34
20 to 24 years	194,168	13,315	15.44	0.90	163,089	13,315	12.97	0.92
25 to 34 years	403,628	18,903	14.66	0.59	330,497	18,903	12.00	0.60
35 to 44 years	254,696	16,057	9.90	0.56	192,459	16,057	7.48	0.58
45 to 54 years	171,964	15,044	7.24	0.59	116,139	15,044	4.89	0.60
55 to 64 years	183,775	14,720	6.99	0.52	116,176	14,720	4.42	0.54
65 to 74 years	89,756	10,291	4.59	0.50	37,494	10,291	1.92	0.52
75 years and older	46,876	7,677	3.68	0.58	15,524	7,677	1.22	0.60
Women+,[2] all ages	**1,222,751**	**28,656**	**6.39**	**0.14**	**757,140**	**28,656**	**3.95**	**0.14**
0 to 4 years	74,101	5,609	7.99	0.56	51,911	5,609	5.60	0.57
5 to 14 years	107,570	9,289	5.29	0.43	63,444	9,289	3.12	0.44
15 to 17 years	30,918	5,677	5.28	0.92	19,088	5,677	3.26	0.94
18 to 19 years	33,534	6,747	8.42	1.55	24,458	6,747	6.14	1.59
20 to 24 years	180,931	14,260	15.42	1.03	150,822	14,260	12.86	1.06
25 to 34 years	270,074	15,010	10.36	0.52	199,818	15,010	7.66	0.53
35 to 44 years	148,120	10,306	5.80	0.38	89,199	10,306	3.49	0.39
45 to 54 years	129,779	12,572	5.35	0.49	76,178	12,572	3.14	0.50
55 to 64 years	115,940	11,524	4.34	0.41	48,036	11,524	1.80	0.42
65 to 74 years	68,375	8,934	3.26	0.41	12,462	8,934	0.60	0.42
75 years and older	63,409	8,843	3.78	0.51	21,724	8,843	1.30	0.52

1. The term “men+” includes men (and boys) and some non-binary people.
2. The term “women+” includes women (and girls) and some non-binary people.
Note: Coverage estimates may not necessarily add up to the totals because of rounding.
Source: Statistics Canada, 2021 Census Undercoverage Study.

12.2 Participation of reserves and settlements

12.2.1 Introduction

In 2021, there were 1,026 reserves and settlements, 63 of which were considered incompletely enumerated. For the 63 incompletely enumerated reserves and settlements, dwelling enumeration either was not permitted or was interrupted before it could be completed. There were no 2021 data for the incompletely enumerated reserves and settlements and, as a result, they were not included in any calculations. In the 2016 Census, 14 reserves and settlements were declared incompletely enumerated and, of these, 1 took part in the 2021 Census. In 2011, there were 31 incompletely enumerated reserves and settlements. The large increase in the number of incompletely enumerated reserves and settlements for the 2021 Census relative to the 2016 Census was mainly because of health and safety measures related to the COVID-19 pandemic.

Further information can be found here: Incompletely enumerated reserves and settlements.

The estimates for incompletely enumerated reserves and settlements were based on a model. Since no reliable source is available to verify the assumptions used in the model, these estimates must be used with caution.

12.2.2 Incompletely enumerated reserves and settlements

For 63 incompletely enumerated reserves and settlements, the 2021 Census was not able to produce population counts, and the coverage studies could not directly estimate net population undercoverage. This was because of a lack of 2021 Census data. The counts and net undercoverage were estimated using alternative methods.

To estimate census population counts, a model-based methodology was used for these incompletely enumerated reserves and settlements. For the estimation model, linear regression without intercept was chosen. The model parameter was estimated based on reserves that were completely enumerated in both the 2016 and the 2021 censuses. The model assumed that the 2021 Census count was a linear function of the 2016 Census count for all provinces, with separate estimates for the slope for each province. In particular, incompletely enumerated reserves and settlements were assumed to follow the same growth pattern as completely enumerated reserves and settlements. The model was evaluated for the basic regression assumptions: independence of errors, homogeneity of variances and normality of errors. For each of the 63 incompletely enumerated reserves and settlements, the input variable for the regression model was either the actual census count in 2016 or the estimated census count from the 2016 model. The output of the model was the estimated census count in 2021 for these 63 communities. The resulting estimates should be used with caution as they are based entirely on a model whose assumptions cannot be verified. The validity of these model-based estimates depends on the extent to which the model assumptions capture the true underlying situation.

In the 2016 Census, 14 reserves, with approximately 28,000 persons, were classified as incompletely enumerated. Among the 63 reserves and settlements for which model-based estimates were produced for the 2021 Census, 50 were considered to have been completely enumerated in the 2016 Census, while the other 13 were classified as incompletely enumerated. The total population for the 63 incompletely enumerated reserves and settlements was estimated at 58,480—more than double from 2016.

The model-estimated population counts for the 63 incompletely enumerated reserves and settlements are subject to coverage errors, just like the census population counts for the rest of the country. Net undercoverage for these 63 areas was estimated by calculating the net undercoverage rate for all completely enumerated reserves in the country. That rate was then applied to the estimated census counts of all the incompletely enumerated reserves and settlements. The estimated total population of 58,480 for incompletely enumerated reserves and settlements in 2021 includes this adjustment for coverage errors.

The population estimates for incompletely enumerated reserves and settlements were not included in the estimates of undercoverage, collection undercoverage, overcoverage and net undercoverage presented in this report because they were based on a model, not on census coverage studies. In addition, they do not provide the same level of detail (e.g., estimates by mother tongue or marital status) as the other estimates.

The population estimates for incompletely enumerated reserves and settlements are included in demographic estimates. The estimates, broken down by province and territory, can be found in Table 1 of the “Data quality concepts and methodology” section in the following demographic estimates publication: *Annual Demographic Estimates: Canada, Provinces and Territories (Total Population only), 2023*, Statistics Canada Catalogue no. 91-215-X.

Appendix A—Whom to include in the census questionnaire

The following instructions were provided on page 3 of the 2021 Census short- and long-form questionnaires to help determine who should be included in the questionnaire.

1. Whom to include in Step B

- All persons who have their **main residence** at this address on May 11, 2021, including newborn babies, roommates and persons who are temporarily away
- **Canadian citizens, landed immigrants** (permanent residents), persons who have claimed **refugee status** (asylum seekers), persons from **another country** with a **work** or **study permit** and family members living here with them
- Persons staying at this address temporarily on May 11, 2021, who have **no main residence elsewhere**.

2. Where to include persons with more than one residence

- **Children in joint custody** should be included in the home of the parent where they live most of the time. Children who spend equal time with each parent should be included in the home of the parent with whom they are staying on May 11, 2021.
- **Students** who return to live with their parents during the year should be included at their parents' address, even if they live elsewhere while attending school or working at a summer job.
- **Spouses or common-law partners temporarily away** who stay elsewhere while working or studying should be listed at the main residence of their family, if they return periodically.
- **Persons in an institution** for **less than six months** (for example, in a home for the aged, a hospital or a prison) should be listed at their usual residence.

Bibliography

Abeysundera, M., Zhao, Z. and Farr, H. (2023). *2021 Census Overcoverage (COS) Methodology Report*. Statistics Canada.

Dasylva, A., Abeysundera, M., Haddou, M. and Lachance, M. (2015). *Sampling and Estimation Based on a Frame of Connected Record-Groups Methodology*. Statistics Canada.

Dolson, D. (2010). "Census coverage studies in Canada: A history with emphasis on the 2011 Census." *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Fellegi, I.P. and Sunter, A.B. (1969). *A Theory for Record Linkage*. Journal of the American Statistical Association. Vol. 64, 1183-1210.

Hong, R. (2023). *The 2021 Dwelling Classification Survey*. Statistics Canada.

Lavallée, P. (2007). *Indirect Sampling*. Springer, New York.

Parenteau, M. 2023. "Processing Steps for the 2021 CUS." *Meeting of the Federal-Provincial-Territorial Committee on Demography*, January 24 to 26, 2023, Statistics Canada, Ottawa, Ontario, 38 slides. (French meeting January 31st to February 2nd, 2023).