

Catalogue no. 13260005
ISSN 2819-4284

Canadian Census Health and Environment Cohort: User Guide, 2021, Version 1b

Release date: June 18, 2025



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, the Agency has developed standards of service which its employees observe in serving its clients. 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

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2025

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

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

Cette publication est aussi disponible en français.

Table of contents

1	Background	4
2	Data	4
2.1	Census of Population, 2021	4
2.2	Canadian Vital Statistics – Death database	5
2.3	Discharge Abstract Database	5
2.4	National Ambulatory Care Reporting System	6
2.5	Ontario Mental Health Reporting System	6
2.6	Annual postal codes for mailing addresses	6
3	Record linkage in the Social Data Linkage Environment	7
3.1	Record linkage method, results, and quality	7
3.2	Supplementary information	8
4	Derivation of the 2021 Canadian Census Health and Environment Cohort	8
5	Data integration	9
5.1	File structures and layouts	9
5.2	Cohort and replicate weights	12
5.3	Limitations of analytical files	12
6	Data access and release guidelines	13
6.1	Custodian contact	13
6.2	Release guidelines	13
	Appendix A – Percentage distributions of select demographic, social and economic characteristics from the 2021 Canadian Census Health and Environment Cohort file and the 2021 Census Research Data Centre file	14
	Appendix A – 2021 Canadian Census Health and Environment Cohort database schema for Research Data Centre data users	15

Canadian Census Health and Environment Cohort: User Guide, 2021, Version 1b

1 Background

The [Canadian Census Health and Environment Cohorts](#) (record number [5422](#)) are population-based [microdata linkages](#) that link the census long-form sample data to the administrative data such as the Canadian Vital Statistics – Death database, the Discharge Abstract Database, the National Ambulatory Care Reporting System, the Ontario Mental Health Reporting System, and the annual postal codes for mailing addresses. These data can be used to examine health outcomes by demographic, social, and economic characteristics including, age, sex at birth and gender, education, geographic areas, housing, immigration and ethnocultural diversity, and income. The microdata linkages ([019-2019](#)) were approved in accordance with the [Statistics Canada Directive on Microdata Linkage](#).

The 2021 Canadian Census Health and Environment Cohort is the latest in a series of population-based microdata linkages. The cohort consists of 8,660,377 persons selected from the 2021 Census of Population long-form sample with mortality, hospitalization, ambulatory care, and mental health services follow-ups. The 2021 Canadian Census Health and Environment Cohort also includes annual postal codes for mailing addresses from as early as 1981 for the cohort members. The linked data can be used to examine health outcomes by demographic, social, and economic characteristics, including, but not limited to, age, sex at birth and gender, education, geographic areas, housing, immigration and ethnocultural diversity, and income. The follow-up periods for version 1b of the 2021 Canadian Census Health and Environment Cohort were as follows:

- Canadian Vital Statistics – Death database (May 11, 2021, to December 31, 2023)
- Discharge Abstract Database (April 1, 2000, to March 31, 2024)
- National Ambulatory Care Reporting System (April 1, 2002, to March 31, 2024)
- Ontario Mental Health Reporting System (April 1, 2006, to March 31, 2024)
- Annual postal codes for mailing addresses (1981 to 2022).

This user guide describes the 2021 Canadian Census Health and Environment Cohort, version 1b.

2 Data

2.1 Census of Population, 2021

The [2021 Census](#) Program enumerated Canadian households using two main types of questionnaires: the short-form questionnaire and the long-form questionnaire. In 2021, a sample of 25% of Canadian households received a long-form questionnaire, which included the questions from the short-form questionnaire. The other households received the short-form questionnaire. For the 2021 Census, the reference date was set to May 11. The information provided in the census questionnaires should reflect each person's situation on May 11, 2021, unless the questions specify otherwise. This reference date ensures that the information collected in the questionnaire provides an accurate snapshot of Canada's society at this point in history.

In addition to the short-form questions, the long-form questionnaire included a series of questions to paint a full portrait of the Canadian population and households, according to their demographic, social and economic characteristics. The subjects included were:

- Commuting to work
- Education, training and learning
- Families, households and housing
- Immigration and ethnocultural diversity
- Income, pensions, spending and wealth
- Indigenous Peoples
- Industries
- Labour
- Languages
- Occupations
- Population and demography
- Population estimates and projections
- Religion
- Society and community.

Beginning in 2021, the census asked questions about both the sex at birth and gender of persons. While data on sex at birth are needed to measure certain indicators, as of the 2021 Census, gender (and not sex) is the standard variable used in concepts and classifications. For more details on the new gender concept, see [Age, Sex at Birth and Gender Reference Guide, Census of Population, 2021](#).

[Table 9.1](#) of the Guide to the Census of Population, 2021, Statistics Canada – Catalogue no. 98-304-X shows the response rates obtained through data processing and data quality assessment. Nationally, the total response rate was 95.7% for the long-form questionnaire – occupied private dwellings (weighted). In 2021, a total of 63 census subdivisions defined as reserves and settlements were incompletely enumerated. For these reserves and settlements, dwelling enumeration was either not permitted or could not be completed because of the various reasons described in [Appendix 1.5](#) of the Guide to the Census of Population, 2021, Statistics Canada – Catalogue no. 98-304-X. For these areas, comparisons between the 2016 and 2021 Census may be less precise due to this missing data.

To determine the number of people who were missed or counted more than once, Statistics Canada conducts postcensal studies of the coverage of the census population, using representative samples of the population. Results of these studies are usually available two years after Census day. For the 2021 Census, the [net undercoverage rate](#) (percentage of people missed less those enumerated more than once) was 3.1%. This rate is slightly higher than that for the [2016 Census](#), which was 2.4%. [The coverage technical report](#), Census of Population, 2021, Statistics Canada – Catalogue no. 98-304-X, examines coverage errors in the 2021 Census.

For the 2021 Census, the dissemination strategy of quality indicators has been completely revamped, with the aim of providing more detailed information about data quality. For more information on this, researchers should consult the [2021 Census Data Quality Guidelines](#).

The 2021 Canadian Census Health and Environment Cohort (CanCHEC) is a subset of the 2021 Census long-form sample that consisted only of private households, including those living in private dwellings attached to collective dwellings in Canada, but excluding those that are living in incompletely enumerated reserves and settlements. This means the households living in collective dwellings, outside Canada, and in the 63 incompletely enumerated reserves and settlements were considered “out of scope” for the 2021 CanCHEC.

2.2 Canadian Vital Statistics – Death database

The [Canadian Vital Statistics – Death database \(CVSD\)](#) is an administrative survey that collects demographic and medical (cause of death) information monthly and annually from all provincial and territorial vital statistics registries on all deaths in Canada. At the time of the production of version 1b of the 2021 CanCHEC, the 2021 to 2023 death data were considered preliminary due to improvements in methodology and timeliness which shortened the duration of data collection. The 2021 to 2022 death data for Yukon were not available for record linkage purposes at the time of linkage with the Derived Record Depository (DRD).

The cause of death variable in the database is classified according to the World Health Organization “[International Statistical Classification of Diseases and Related Health Problems](#)” (ICD).¹ The cause of death data for 2021 to 2023 are coded to the ICD, tenth revision (ICD-10).

2.3 Discharge Abstract Database

The [Discharge Abstract Database \(DAD\)](#)² captures administrative, clinical and demographic information on hospital discharges (including in-hospital deaths, sign-outs and transfers) from all provinces and territories, except Quebec. Over time, the DAD has been used to capture data on acute care, including day surgery, chronic care, and rehabilitative care. The DAD data for fiscal years 2000–2001 to 2023–2024 were included in the linkage (fiscal years including the period of April 1, 2000, to March 31, 2024). The DAD is an event-based data file meaning that there may be more than one record per person.

In the DAD, jurisdiction-specific instructions for collection of data elements evolve, and the collection of each data element can vary by jurisdiction and over time. Data elements may be mandatory, mandatory if applicable,

1. World Health Organization. (n.d.). [International statistical classification of diseases and related health problems \(ICD\)](#)
2. Canadian Institute for Health Information. [Discharge Abstract Database \(DAD\) metadata](#)

optional or not applicable. Users are recommended to review the listings of the DAD data elements under the heading “Data elements” at the [DAD metadata website](#) for more information on data elements and coverage. More information about methodologies for the DAD, please refer to the [Comprehensive Ambulatory Classification System \(CACS\)](#), the [Case Mix Decision-Support Guide: CMG+](#), and the [Acute and Ambulatory Care Data Content Standard](#).

2.4 National Ambulatory Care Reporting System

The [National Ambulatory Care Reporting System \(NACRS\)](#)³ contains data for hospital-based and community-based ambulatory care including day surgery, outpatient and community-based clinics, and emergency departments. Client visit data are collected at time of service in the participating facilities from several jurisdictions. The NACRS data for fiscal years 2002–2003 to 2023–2024 were included in the linkage, including the period of April 1, 2002, to March 31, 2024. The NACRS is an event-based data file meaning that there may be more than one record per person.

The grouping methodology used in the NACRS is the [Comprehensive Ambulatory Classification System \(CACS\)](#). It is a national grouping methodology for ambulatory care patient data submitted to either the DAD or the NACRS. Specific groupings include emergency visits, ambulatory interventions, rehabilitation and clinic visits, as well as grouping exception cases such as telephone visits and direct diagnostic imaging. Patients are grouped according to a number of data elements including diagnosis or intervention for the DAD as well as emergency visit indicator, mode of visit, visit disposition, ambulatory care type or program area for the NACRS.

The Canadian Institute for Health Information does not mandate data submission to the NACRS, and jurisdiction-specific instructions for collection of data elements evolve over time. Similar to the DAD, collection of each data element may be mandatory, mandatory if applicable, optional or not applicable. For details on the provincial data coverage, and the collection of data elements, please refer to the [Jurisdictional coverage information](#) and the “Data coverage” section of the [NACRS metadata website](#). More information about methodologies for NACRS, please refer to the [Comprehensive Ambulatory Classification System \(CACS\)](#) and the [Acute and Ambulatory Care Data Content Standard](#).

2.5 Ontario Mental Health Reporting System

The [Ontario Mental Health Reporting System \(OMHRS\)](#)⁴ contains patient record data for all individuals receiving adult mental health services in Ontario, in addition to some individuals receiving services in youth inpatient beds and selected facilities in other provinces starting in fiscal year 2006–2007. The OMHRS data are collected using the Resident Assessment Instrument – Mental Health (RAI-MH©) version 2.0, and includes information relating to patient admissions and assessment, care planning, and outcome measurement. For information about methodologies for the OMHRS, such as case mix methodologies, see the [System for Classification of In-Patient Psychiatry \(SCIPP\)](#) that was developed for the RAI-MH. The OMHRS is an event-based data file meaning that there may be more than one record per person.

For this record linkage, the OMHRS records covering the fiscal years from 2006–2007 to 2023–2024 were included. Data are currently submitted to Canadian Institute for Health Information from participating hospitals in Ontario, as well as from three facilities in Newfoundland and Labrador and one facility in Manitoba.

Researchers will find the listings of the OMHRS data elements and data quality under the “Data elements” and “Data quality” sections of the [OMHRS metadata website](#).

2.6 Annual postal codes for mailing addresses

In Canada, income tax returns are submitted annually to the Canada Revenue Agency (CRA). The T1 Personal Master File (T1PMF), also known as the T1 General and Schedules, is a collection of the income tax returns shared by the CRA with Statistics Canada, and it provides income and demographic (e.g., date of death) information on tax filers in Canada. Every resident of Canada who earns taxable income is required to complete an income

3. Canadian Institute for Health Information. [National Ambulatory Care Reporting System \(NACRS\) metadata](#)

4. Canadian Institute for Health Information. [Ontario Mental Health Reporting System \(OMHRS\) metadata](#)

tax return, known as a T1 form, at the end of the year in which the income was received. Therefore, the T1PMF includes almost all persons who filed an individual T1 tax return for the year of reference (i.e., some late filers may not be included) or those who received Canada Child Benefits and their non-filing spouses.

The T1PMF (1981 to 2022) is the principal data source for the annual postal codes for mailing addresses. Mailing address postal codes reported on these tax files were extracted to estimate a person's place of residence for that reference year. Note that for some tax filers, the mailing addresses used for filing T1 tax records may not be associated with their place of residence (e.g., P.O. box, accountants' or lawyers' offices, parents' addresses for young adults, children's addresses for elderly parents).⁵

3 Record linkage in the Social Data Linkage Environment

The [Social Data Linkage Environment \(SDLE\)](#) has created a series of linked population files for social analysis using the [Derived Record Depository \(DRD\)](#), a dynamic relational database containing only basic personal identifiers. Survey and administrative data are linked to the DRD using SAS and [G-Link \(Record Linkage – Generalized System\)](#), a SAS-based generalized record linkage software that supports deterministic and probabilistic linkage developed at Statistics Canada. See the Statistics Canada website for additional information about the SDLE.

Section 3.1 of this user guide highlights relevant information extracted from the SDLE methodology reports. Section 3.2 provides supplementary information about the data that is not available in these reports.

3.1 Record linkage method, results, and quality

Briefly, 96.1%⁶ of the person records on the 2021 Census Response Database (RDB) file were probabilistically linked to the person records on the DRD. This corresponds to the linkage rate for the census short-form data. The linkage rate for the census long-form sample data was higher than that of the short-form data by 0.6 percentage point at 96.7%. The probabilistic record linkage relied on information such as last name, first name, sex at birth, date of birth, telephone number, and geography (as detailed as 6-character postal code). The overall estimated percentage of false matches was 0.39% based on the census short-form data.

The death records on the Canadian Vital Statistics – Death database (CVSD) for 2021 and 2023 were linked to the person records on the DRD through a probabilistic record linkage method using information such as last name, first name, sex at birth, date of birth, date of death, and geography (as detailed as 6-character postal code). About 99.1% of death records were linked to the person records on the DRD. The overall estimated percentage of false matches is expected to be less than 0.3%.

For event-based data such as the Discharge Abstract Database (DAD), National Ambulatory Care Reporting System (NACRS), and the Ontario Mental Health Reporting System (OMHRS), a two-phase deterministic record linkage methodology was employed to match records between each of the data files with the person records on the DRD using the sex at birth, date of birth, and postal code, and/or health insurance/card number and issuing province or territory. Each event (hospital discharge, ambulatory visit, and/or mental health assessment) was deterministically linked to the DRD to account for time-varying linkage variables such as patient's postal code of usual residence and health card information. The cumulative linkage rates for the DAD, the NACRS and the OMHRS are as follows:

- 95.2% of the DAD records from 2000–2001 to 2023–2024 were linked to the DRD
- 95.6% of the NACRS records from 2002–2003 to 2023–2024 were linked to the DRD
- 94.6% of the OMHRS records from 2006–2007 to 2023–2024 were linked to the DRD.

The percentages of false matches are not estimated for these deterministic record linkages.

Recall that the mailing addresses used for filing T1 tax records are the principal data source for the postal code history. The person records from the T1 Person Master File (T1PMF) were linked to the person records on the DRD

5. [Comparison of Place of Residence between the T1 Family File and the Census: Evaluation using record linkage](#)

6. Duplicate person records are excluded from the calculation.

using a probabilistic method based on Social Insurance Number, last name, first name, sex, date of birth, date of death, telephone number, and geography (as detailed as 6-character postal code). About 99.7% of 46 million persons on the T1PMF (1981 to 2022) were either linked to the person records on the DRD or added to the DRD as new persons. The estimated percentages of false matches over the tax filing years were consistently no worse than 0.21%.

3.2 Supplementary information

In this section, additional information about the 2021 Census long-form sample data required to create the 2021 Canadian Census Health and Environment Cohort (CanCHEC) is provided. These data were extracted from the following files:

- 2021 Census Response Database (RDB) file
- 2021 Census Research Data Centre (RDC) file,

where the latter is the disseminated version of the census long-form data. About 1.17% of the person records on the 2021 Census RDC file did not exist on the version of 2021 Census RDB file that had undergone the linkages described in the methodology reports. Over 90% of these were persons who had completed the long-form questionnaire (form [2A-R](#)) that was sent to private households in First Nations communities, Métis Settlements, Inuit regions and other remote areas to enumerate 100% of the population. As these person records were never linked to the person records on the DRD, these were considered “out of scope” for the 2021 CanCHEC.

4 Derivation of the 2021 Canadian Census Health and Environment Cohort

The 2021 Canadian Census Health and Environment Cohort (CanCHEC) was created based on the 2021 Census Research Data Centre (RDC) file and the linkage keys between the 2021 Census Response Database (RDB) file and the Canadian Vital Statistics – Death database (CVSD) files (2021 to 2023) produced using the Derived Record Depository (DRD). This final cohort, 8,660,377 persons, consisted of 95.45% of all person records on the 2021 Census RDC file. This is a fixed cohort, which means the cohort membership is defined at the creation of the cohort and no new members are added during the follow-up period. The steps for deriving the 2021 CanCHEC are outlined in Table 1.

Table 1
Derivation of the 2021 Canadian Census Health and Environment Cohort

Description	Size ¹
Base cohort	Number of person records
Person records on the 2021 Census Research Data Centre (RDC) file	9,072,940
Exclusion criteria (in the order of deletions from the base cohort)	Percentage of base cohort size
1. Person records not on the version of 2021 Census Response Data Base (RDB) file that was linked to the Derived Record Depository (DRD)	-1.17
2. Person records not linked to the DRD	-3.06
3. Presumed duplicate person records according to the internal record linkage of the 2021 Census RDB file	-0.30
4. Person records linked to a death preceding the Census day (May 11, 2021)	-0.01
5. Person records presumed to be falsely linked to the DRD among those linked to the deaths that occurred from the Census day onwards	-0.01
Final cohort	Number of person records
Person records on the 2021 Canadian Census Health and Environment Cohort (CanCHEC)	8,660,377

1. Enumerated by census identifier for person table (PP_ID).

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)).

5 Data integration

The 2021 Canadian Census Health and Environment Cohort (CanCHEC) files are designed to be merged with each other, as well as with the 2021 Census Research Data Centre (RDC) file and the health outcome files, to create files for research and analytical purposes.

Due to the large number of records on these files – for example, the RDC file for the 2021 Census contains more than 9 million records and 907 variables - it is recommended that researchers only retain the variables on the 2021 Census RDC file, the historical postal code file, and the health outcome files that are required to answer their specific research questions.

Note that some source files may have the same variable concepts across years but use different variable names, formats, or codes (e.g., gender, age, date of birth). Note that files might contain the same variable names with different formats. Consult the codebooks and rename and/or drop unnecessary variables prior to merging.

The variable **UniqID** holds unique identification numbers for each of the cohort members. Note that these identification numbers are unique within this cohort.

For RDC data users, a diagram of merging the 2021 CanCHEC files in the RDC is shown in Figure 1, Appendix B. If you have any questions or concerns about merging files, please contact your RDC Analyst.

5.1 File structures and layouts

Table 2 lists the latest 2021 CanCHEC files released in the RDCs.

Table 2

List of the latest 2021 Canadian Census Health and Environment Cohort (CanCHEC) files released in the Research Data Centres

Filename	Number of records	Description
canchech_c21keyfile_f3_v1b	8,660,377	Linkage key file between the 2021 Census Research Data Centre (RDC) file and the Canadian Vital Statistics – Death database (CVSD) files (2021 to 2023), consisting of persons included in the 2021 Canadian Census Health and Environment Cohort (CanCHEC)
canchech_c21ltfufile_f3_v1b	8,660,377	Lost to follow-up file according to the linkage key file between the 2021 Census RDC file and the Derived Record Depository (DRD), extracted on January 23, 2025
canchech_c21weights_f3_v1b	8,660,377	File containing cohort and replicate weights for the 2021 CanCHEC
canchech_c21mobility_f3_v1	8,624,545	File containing annual postal codes for mailing addresses (1981 to 2022) for the 2021 CanCHEC
canchech_c21dad00_f3_v1–canchech_c21dad23_f3_v1	12,644,011	Discharge Abstract Database (DAD) analytical files by fiscal year for the 2021 CanCHEC
canchech_c21dadinstcode00_f3_v1–canchech_c21dadinstcode23_f3_v1	1,2644,011	Institution numbers assigned to the DAD records in the analytical files (see Section 6.2)
canchech_dadinstcode0023_f3_v1	1,472	List of participating facilities in the DAD analytical files (see Section 6.2)
canchech_c21nacrs02_f3_v1–canchech_c21nacrs23_f3_v1	58,123,412	National Ambulatory Care Reporting System (NACRS) analytical files by fiscal year for the 2021 CanCHEC
canchech_c21nacrsinstcode00_f3_v1–canchech_c21nacrsinstcode23_f3_v1	58,123,412	Facility numbers assigned to NACRS records in the analytical files (see Section 6.2)
canchech_nacrsinstcode0223_f3_v1	618	List of participating facilities in the NACRS analytical files (see Section 6.2)
canchech_c21omhrs0623_f3_v1	277,529	Ontario Mental Health Reporting System (OMHRS) analytical file for the 2021 CanCHEC
canchech_omhrsinstcode0623_f3_v1	109	List of participating facilities in the OMHRS analytical file (see Section 6.2)

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)); Canada Revenue Agency, T1 Personal Master File; and Canadian Institute for Health Information, Discharge Abstract Database, 2000–2001 to 2023–2024 ([metadata](#)), National Ambulatory Care Reporting System, 2002–2003 to 2023–2024 ([metadata](#)), and Ontario Mental Health Reporting System, 2006–2007 to 2023–2024 ([metadata](#)).

canchec_c21keyfile_f3_v1b is the linkage key file between the 2021 Census RDC file and the Canadian Vital Statistics – Death database (CVSD) files from 2021 to 2023. The file contains 8,660,377 records, which correspond to the cohort size. Moreover, the file contains record identifiers for 150,065 deaths. Note that **UniqID** and **PP_ID** have a one-to-one relationship. Moreover, **EVENT_YEAR**, **PLACEOFDEATH_PROVINCE**, and **REGISTRATION_NUMBER** together uniquely identify a CVSD record. Table 3 describes the variables in the file.

Table 3

List of variables and their descriptions for the linkage key file between the 2021 Census Research Data Centre file and the Canadian Vital Statistics – Death database files (2021 to 2023), consisting of persons included in the 2021 Canadian Census Health and Environment Cohort (CanCHEC)

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
PP_ID	Num	8	Census identifier for Person table
PRCDDA	Num	8	Province, Census Division, Dissemination Area
HhNum	Num	8	Key for Households
PpNum	Num	8	Key for Persons
EVENT_YEAR	Char	4	Year when death occurred (YYYY)
PLACEOFDEATH_PROVINCE	Char	3	Province/territory where death occurred
REGISTRATION_NUMBER	Char	6	Death registration number

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)).

canchec_c21lftfile_f3_v1b v1b is the file that contains the full cohort (8,660,377 records) with a flag that identifies persons who are no longer linked to the Derived Record Depository (DRD) due to data cleaning and periodic updates to the depository over time. These cohort members are therefore considered “lost to follow-up.” The percentage of persons lost to follow-up as of version 1b of the 2021 CanCHEC is less than 0.01%⁷. Table 4 describes the variables in the file

Table 4

List of variables and their descriptions for the lost to follow-up file for the 2021 Canadian Census Health and Environment Cohort (CanCHEC), according to the linkage key file between the 2021 Census Research Data Centre file and the Derived Record Depository extracted on January 23, 2025

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
LTFU	Num	8	Indicator to flag persons lost to follow-up, where 1 = lost to follow-up and 0 = not lost to follow-up

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)).

canchec_c21weights_f3_v1b is the file containing the cohort and replicate weights for the 2021 CanCHEC. The file also contains 8,660,377 records. Table 5 describes the variables in the file. For more information on these weights, please see Section 5.2.

7. More than half of the records lost to follow-up were women+ (58.0%), while 45.4% of these records were in Ontario on Census day, and almost all aged 15 years or older on that day.

Table 5**List of variables and their descriptions for the file containing cohort and replicate weights for the 2021 Canadian Census Health and Environment Cohort (CanCHEC)**

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
CanCHECW2	Num	8	Cohort weight
CanCHEC_REPWT1–CanCHEC_REPWT100	Num	8	100 replicate weights

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)).

canchech_c21mobility_f3_v1 contains annual postal codes for mailing addresses from 1981 to 2022 for the 2021 CanCHEC. The file contains 8,624,545 records. Table 6 describes the variables in the file.

Table 6**List of variables and their descriptions for the file containing annual postal codes for mailing addresses (1981 to 2022) for the 2021 Canadian Census Health and Environment Cohort (CanCHEC)**

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
PC1981–PC2022	Char	6	Postal code by validity year according to the Derived Record Depository

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)); and Canada Revenue Agency, T1 Personal Master File.

canchech_c21dad00_f3_v1–canchech_c21dad23_f3_v1 contain the Discharge Abstract Database (DAD) records by fiscal year for the 2021 CanCHEC. The files contain 12,644,011 records, though the yearly number of records varies by fiscal year. Table 7 describes the variables in the files.

Table 7**List of variables and their descriptions for the Discharge Abstract Database (DAD) analytical files (masked) for the 2021 Canadian Census Health and Environment Cohort (CanCHEC)**

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
DAD_ID	Char	15	DAD record identifier (randomized)
FISCAL_YEAR and other DAD variables	Num/Char	Varies	Fiscal year of the hospital discharge record, and other DAD variables

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)); and Canadian Institute for Health Information, Discharge Abstract Database, 2000–2001 to 2023–2024 ([metadata](#)).

canchech_c21nacrs02_f3_v1–canchech_c21nacrs23_f3_v1 contain the National Ambulatory Care Reporting System (NACRS) records by fiscal year for the 2021 CanCHEC. The files contain 58,123,412 records, though the number of records varies by fiscal year. Table 8 describes the variables in the files.

Table 8**List of variables and their descriptions for the National Ambulatory Care Reporting System (NACRS) analytical files (masked) for the 2021 Canadian Census Health and Environment Cohort (CanCHEC)**

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
NACRS_ID	Char	15	NACRS record identifier (randomized)
FISCAL_YEAR and other NACRS variables	Num/Char	Varies	Fiscal year of the ambulatory or emergency care record, and other NACRS variables

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)); and Canadian Institute for Health Information, National Ambulatory Care Reporting System, 2002–2003 to 2023–2024 ([metadata](#)).

canchec_c21omhrs0623_f3_v1 contains the Ontario Mental Health Reporting System (OMHRS) records from 2006–2007 to 2023–2024 by assessment reference date for the 2021 CanCHEC. The file contains 277,529 records. Table 9 describes the variables in the file.

Table 9

List of variables and their descriptions for the Ontario Mental Health Reporting System (OMHRS) analytical file (masked) for the 2021 Canadian Census Health and Environment Cohort (CanCHEC)

Variable	Type	Length	Description
UniqID	Num	8	CanCHEC unique identifier
OMHRS_ID	Char	15	OMHRS record identifier (randomized)
FACILITY_PROVINCE and other OMHRS variables	Num/Char	Varies	Reporting facility's province/territory and other OMHRS variables

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)); and Canadian Institute for Health Information, Ontario Mental Health Reporting System, 2006–2007 to 2023–2024 ([metadata](#)).

5.2 Cohort and replicate weights

A weight file (**canchec_c21weights_f3_v1b**) was created for the cohort to make it more representative of the target population and to reduce bias due to missed links. Briefly, the main weight for the long-form sample survey (**CompW2**) was adjusted by model parameters on the probability of linking to a person record on the Derived Record Depository (DRD). This adjusted weight was then calibrated to census counts to produce the cohort weight (**CanCHECW2**).

Recall that the [replicate estimator chosen for the 2021 long-form sample survey was derived from Fay's balanced half-sample method](#). This method determined the creation of replicates, the calculation of replicate weights (100) and the multiplication factor used to estimate variance. The set of 100 replicate weights for the long-form sample (**MAIN_REPWT1–MAIN_REPWT100**) was adjusted using the same procedure as above to produce the final set of 100 replicate weights for the cohort (**CanCHEC_REPWT1–CanCHEC_REPWT100**).

To use these weights, a survey-aware analysis procedure must be used. The variance estimation method should be Balanced Repeated Replication (BRR), and one must specify the Fay adjustment of 2, or epsilon value of 0.2928932 depending on the software:

- In SAS 9.3 or later SURVEY procedure, specify the epsilon value by **VARMETHOD=BRR (FAY=0.2928932)**.
- In SUDAAN, use **REPWGT CanCHEC_REPWT1-CanCHEC_REPWT100 / ADJFAY=2**.
- Using the R package [survey: Analysis of Complex Survey Samples](#), the arguments for the type of replication weights and the shrinkage factor for the weights in Fay's variance estimator in a BRR design are specified as **type="Fay"** and **rho=0.2928932**, respectively.

The cohort and replicate weights for the 2021 CanCHEC are available in **canchec_c21weights_f3_v1b** and this file can be merged with other files using the **UniqID** variable.

Appendix A shows the percentage distributions of select demographic, social and economic characteristics from 2021 CanCHEC file (unweighted and weighted) and 2021 Census RDC file (weighted).

5.3 Limitations of analytical files

There is potential inclusion bias associated with data linkage. Although the cohort weights were designed to help mitigate this bias, unknown bias might exist if those missing from the cohort differed systematically from those who were included.

The 2021 CanCHEC should be considered a healthier population than the general population due to the exclusion of the institutional population and non-institutional collectives at baseline.

During the record linkage, the 2021 to 2023 data from the CVSD were considered preliminary due to improvements in methodology and timeliness, which shortened the duration of data collection. Data for Yukon between 2021 and 2022 in the CVSD were not available at the time of the record linkage. Data will be revised with subsequent releases. Though the overall linkage rate between the CVSD and the DRD is 99.1% for 2021 to 2023, the linkage

rate is known to be lower for decedents aged 7 years and under, particularly those under the age of 1 year, due to undercoverage of the DRD.

The postal code file may not necessarily reflect where a person lives, owing to the fact that for some tax filers, the mailing addresses used for filing T1 tax records may not be associated with their place of residence (e.g., P.O. box, accountants' or lawyers' offices, parents' addresses for young adults, children's addresses for elderly parents).

Data included in the DAD, the NACRS, and the OMHRS do not cover data from all provinces and territories, and coverage may vary by fiscal year and may not be comparable across jurisdictions. Specifically, the DAD does not contain hospital discharges and day surgery data from Quebec.

The DAD, the NACRS and the OMHRS analytical data files contain data elements pertaining to death information (e.g., date of death) and this information may be inconsistent within or across these data files when events are grouped at the patient-level. Furthermore, death information captured in the DAD, the NACRS, and the OMHRS data files may differ from the linked death information captured through the CVSD. Death information captured through the CVSD should be considered as the standard for analyses of the linked data involving mortality.

Finally, emigration status of cohort members is not known.

6 Data access and release guidelines

6.1 Custodian contact

The Centre for Health Data Integration and Direct Measures is the custodian division for the Canadian Census Health and Environment Cohorts (CanCHECs). Questions about the CanCHEC files can be sent to the Client Services via hd-ds@statcan.gc.ca.

Please refer to the [Research Data Centres \(RDCs\) Program](#) for access via the RDCs.

6.2 Release guidelines

For information on the vetting process and vetting guidelines, please see the document Disclosure Control Guidelines for the Canadian Census Health and Environment Cohorts. The document was last updated in August 2024.

The CanCHEC data should not be used to produce census tabulations beyond a descriptive "Table 1" that would be traditionally found in a journal article. All other census tabulations must be obtained via the full census data.

The linked data should not be used to produce official mortality statistics. Official counts and rates of mortality are available on the Statistics Canada website or can be generated by requesting use of the Canadian Vital Statistics – Death database, which is accessible through the RDCs or by requesting a custom tabulation from Statistics Canada (hd-ds@statcan.gc.ca). Furthermore, the linked data should not be used to produce official hospitalization statistics (acute care, ambulatory care, mental health services). Official counts and rates are available from the Canadian Institute for Health Information.

NOTE: All tables and analyses produced from the 2021 to 2023 death data must be accompanied by the following message: "During the record linkage, the 2021 to 2023 death data were considered preliminary due to improvements in methodology and timeliness which shortened the duration of data collection. Data for Yukon were not available before 2023."

Furthermore, the linked data cannot be used to produce statistics related to the institutions reported in the Discharge Abstract Database (DAD), the National Ambulatory Care Reporting System (NACRS), and the Ontario Mental Health Reporting System (OMHRS) and any outputs at the institution level will be restricted as per the vetting guidelines. Institution information such as postal code captured in the DAD, the NACRS and the OMHRS can be used to derive information such as distances to the acute care facilities, but this cannot be used as an outcome of interest.

Appendix A – Percentage distributions of select demographic, social and economic characteristics from the 2021 Canadian Census Health and Environment Cohort file and the 2021 Census Research Data Centre file

Table 10 shows the percentage distributions of select demographic, social and economic characteristics from the 2021 Canadian Census Health and Environment Cohort file and the 2021 Census Research Data Centre file.

Table 10
Percentage distributions of select demographic, social and economic characteristics from the 2021 Canadian Census Health and Environment Cohort (CanCHEC) file and the 2021 Census Research Data Centre file

Population characteristic	2021 CanCHEC		2021 Census
	Unweighted	Weighted	Weighted
Size	number (rounded)		
	8,660,375	36,328,475	36,328,475
	percentage of size		
Gender¹			
Woman+ ²	50.7	50.6	50.6
Man+ ³	49.3	49.4	49.4
Age by age groups			
0 to 14 years	16.8	16.5	16.5
15 to 24 years	11.4	11.5	11.5
25 to 34 years	13.3	13.5	13.5
35 to 44 years	13.4	13.4	13.4
45 to 54 years	12.8	12.8	12.8
55 to 64 years	14.2	14.2	14.2
65 to 74 years	11.0	11.0	11.0
75 to 84 years	5.4	5.4	5.4
85 years and over	1.7	1.7	1.7
Marital status (de facto)			
Married or living common law	48.6	48.4	48.2
Not living common law (never married, separated, divorced, or windowed)	51.4	51.6	51.8
Mobility status: Place of residence 1 year ago (2020)			
Same address (dwelling)	87.5	87.0	87.0
Different address (dwelling)	10.8	11.3	11.3
Outside Canada	0.7	0.8	0.8
Not applicable (age exclusion)	1.0	0.9	0.9
Education: Highest certificate, diploma or degree			
No certificate, diploma or degree	13.7	13.4	13.5
High (secondary) school diploma or equivalency certificate	22.0	22.2	22.3
Apprenticeship or trades certificate or diploma	7.3	7.3	7.3
College, CEGEP or other non-university certificate or diploma	15.7	15.8	15.7
University certificate or diploma below bachelor level	2.4	2.5	2.5
Bachelor's degree or higher	22.1	22.4	22.3
Not applicable (< 15 years)	16.8	16.5	16.5
Labour force status during the week of Sunday, May 2 to Saturday, May 8, 2021			
Employed	47.5	47.7	47.7
Unemployed	5.4	5.5	5.5
Not in labour force	30.3	30.3	30.3
Not applicable (< 15 years)	16.8	16.5	16.5
Low-income status based on low-income measure, after tax			
Not in low income	88.9	89.0	88.9
In low income	11.1	11.0	11.1
Indigenous identity			
First Nations (North American Indian)	4.4	2.9	2.9
Métis	1.7	1.7	1.7
Inuk (Inuit)	0.5	0.2	0.2
Multiple Indigenous responses	0.1	0.1	0.1
Indigenous responses not included elsewhere	0.1	0.1	0.1
Non-Indigenous identity	93.3	95.0	95.0

1. Given that the non-binary population is small, data aggregation to a two-category gender variable is sometimes necessary to protect the confidentiality of responses. In these cases, individuals in the category "non-binary persons" are distributed into the other two gender categories and are denoted by the "+" symbol.

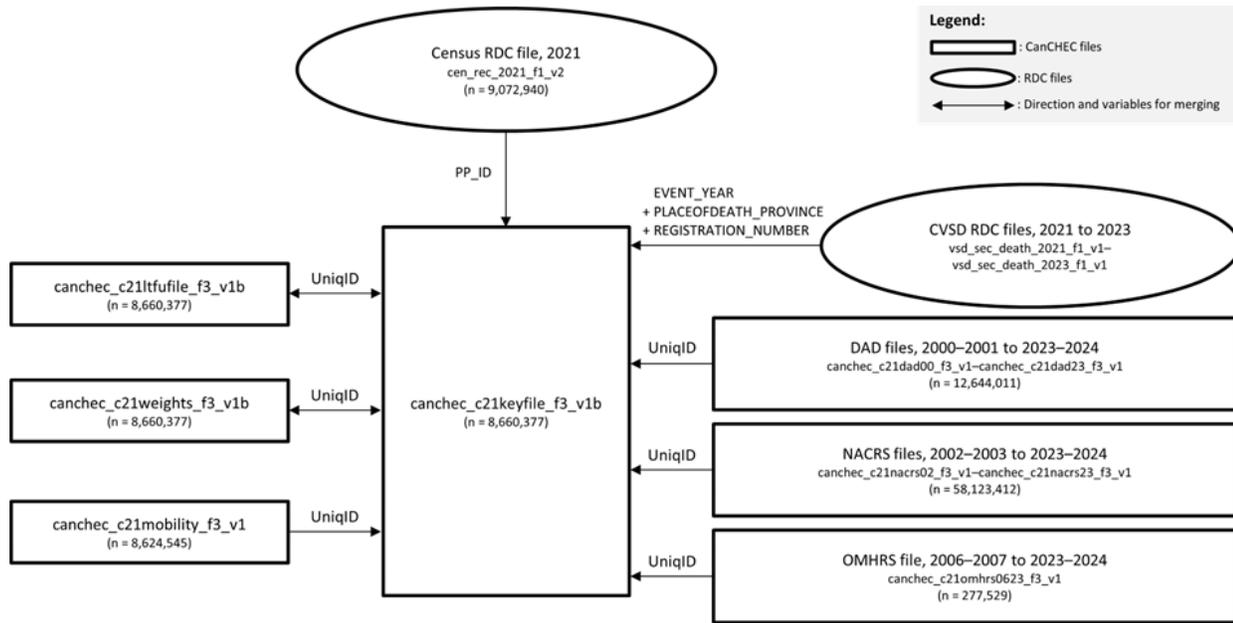
2. This category includes women (and/or girls), as well as some non-binary persons.

3. This category includes men (and/or boys), as well as some non-binary persons.

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number [5422](#)), Census of Population, 2021 (record number [3901](#)), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number [3233](#)).

Appendix B – 2021 Canadian Census Health and Environment Cohort database schema for Research Data Centre data users

Figure 1
2021 Canadian Census Health and Environment Cohort database schema for Research Data Centre data users



Notes: The objects are not drawn to scale.

Acronyms and abbreviations: CanCHEC: Canadian Census Health and Environment Cohort; CVSD: Canadian Vital Statistics – Death database; CEN: Census of Population; DAD: Discharge Abstract Database; NACRS: National Ambulatory Care Reporting System; OMHRS: Ontario Mental Death Reporting System; RDC: Research Data Centres.

Version 1b of the 2021 CanCHEC was prepared using cen_rec_2021_f1_v2 and vsd_sec_death_2021_f1_v1–vsd_sec_death_2023_f1_v1; however, other versions of the 2021 Census of Population and Canadian Vital Statistics - Death database can be used with version 1b of the 2021 CanCHEC linkage keys.

Sources: Statistics Canada, Canadian Census Health and Environment Cohort, 2021 (record number 5422), Census of Population, 2021 (record number 3901), and Canadian Vital Statistics – Death database, 2021 to 2023 (record number 3233); Canada Revenue Agency, T1 Personal Master File; and Canadian Institute for Health Information, Discharge Abstract Database, 2000–2001 to 2023–2024 (metadata), National Ambulatory Care Reporting System, 2002–2003 to 2023–2024 (metadata), and Ontario Mental Health Reporting System, 2006–2007 to 2023–2024 (metadata).