
Agriculture–Population Linkage: Data quality report, 2021



Release date: August 25, 2023

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

© His Majesty the King in Right of Canada as represented by the Minister of Industry, 2023

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

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

Cette publication est aussi disponible en français.

Release date: August 25, 2023

Catalogue number 32260006, issue 2021001

Table of Contents

About Agriculture–Population Linkage	1
New for 2021	1
Sources of error	2
Automated matching process	2
Response and imputation rates	2
Sampling and weighting	4
Quality and data suppression	4

About Agriculture–Population Linkage

An important benefit of conducting the Census of Agriculture at the same time as the Census of Population is that information from these sources can be linked by means of an automated matching process to create the Agriculture–Population Linkage database. This database contains selected variables from the Census of Agriculture and selected variables (education, occupation, etc.) included on the Census of Population long-form questionnaire. The Agriculture–Population Linkage database permits the cross-tabulation of socioeconomic characteristics of farm operators and their families (for example, the age, education and income of operators) with the agricultural characteristics of farm operations (for example, farm type, operating arrangement, farm area, total gross farm receipts, total farm business operating expenses and total farm capital).

The 2021 Agriculture–Population Linkage database follows the Agriculture–Population Linkage database initially created for the 1971 Census, and also available for the 1981 and subsequent censuses. The 2021 database targets farm operators who were identified on the 2021 Census of Agriculture and their families, except those residing in Canada’s three territories or in [collective dwellings](#).

The Agriculture–Population Linkage database is an amalgamation of information from two data sources. Until 2006, the population information came from the Census of Population long-form questionnaire, which was a mandatory questionnaire distributed to about 20% of Canadian households. In 2011, the population data source was the voluntary National Household Survey, which was distributed to approximately 33% of Canadian private households (excluding households occupying collective dwellings). Since 2016, the household data source is once again the mandatory Census of Population long-form questionnaire, distributed to approximately 25% of Canadian private households. Thus, residents of collective dwellings are not represented in the Agriculture–Population Linkage database from 2016 onwards.

Because the Agriculture–Population Linkage database is an amalgamation of information from two data sources, users are encouraged to refer to the reference material from the [Census of Population](#) and the [Census of Agriculture](#) for further information on the data collection, processing and dissemination methods used.

New for 2021

The creation of the 2021 Agriculture–Population Linkage database follows the same methodology as in 2016, except for changes in the following variables.

1. A gender variable will replace the sex at birth variable that had been used in 2016 and earlier. Given that the non-binary population is small, data aggregation to a two-category gender variable is necessary to protect the confidentiality of responses provided. In these cases, individuals in the “non-binary persons” category are distributed into the other two gender categories and are denoted by the “+” symbol. In the Agriculture–Population Linkage database, the gender variable is aggregated to the binary form, “women+” and “men+.” The “women+” category includes women (and girls) and may include some non-binary people as well. The “men+” category includes men (and boys) and may include some non-binary people as well.
2. The marital status variable has a new separate category for “living common law.”
3. The farm operating arrangement variable includes only one category for partnerships. Previous censuses included two categories: partnerships with a written agreement and partnerships without a written agreement.

Sources of error

In a sample survey such as the Census of Population long-form questionnaire, there can be two types of errors—sampling errors and non-sampling errors. In a census such as the Census of Agriculture, only non-sampling errors exist.

Sampling error arises from estimating a population characteristic by measuring only a portion of the population rather than the entire population. The error can be controlled by the sample size, sample design and method of estimation.

Non-sampling errors are errors that are unrelated to sampling. They can include errors in the frame from which the sample is drawn; inadequate collection tools; survey non-response; and errors in data capture, editing, coding and other processing steps. During the planning stages, steps were taken to reduce non-sampling error through questionnaire testing, interviewer training, and quality control of data capture and coding, as well as many other approaches.

Automated matching process

The fundamentals of the agriculture–population automated matching process are simple. A farm operator completes a Census of Agriculture questionnaire as well as a Census of Population questionnaire. The operator may also be selected to complete a Census of Population long-form questionnaire, distributed to approximately one-quarter of all households. During early Census of Agriculture data processing, farm operators are linked to the Census of Population databases using information that is common to both questionnaires, such as name, sex, date of birth and address to establish a one-to-one match. This match allows for linking Census of Agriculture data to data from the Census of Population long form, which is required by the Agriculture–Population Linkage database creation process. The 1991 to 2021 censuses of agriculture allowed respondents to report up to three operators per farm, and all farm operators were included in the matching process. With this additional information, the relationship between family members living in the same household and operating the same farm can be analyzed. As well, operators in different households operating the same farm can be included in the analysis.

In some cases, the linkage between the Census of Agriculture operator and the Census of Population long-form questionnaire household was unsuccessful because of non-response to the Census of Population long-form questionnaire, incomplete operator information on the Census of Agriculture, special farms that are included with certainty in the linkage database but not sampled by the Census of Population long-form questionnaire, or other factors. In these cases, either the long-form questionnaire data were imputed on the Agriculture–Population Linkage database with the information from a similar responding household from the Census of Population or, in rare cases, the weights were adjusted to account for the unlinked records.

Response and imputation rates

The following tables present the weighted response rate for the entire Census of Population long-form questionnaire and the weighted imputation rate for the Agriculture–Population Linkage database population in 2021.

Table 1

Weighted response rate for the Census of Population long-form questionnaire, Canada and provinces^{1,2}

Census of Population long-form questionnaire weighted response rate	
Geography	percent
Canada³	95.7
Newfoundland and Labrador	95.6
Prince Edward Island	96.8
Nova Scotia	96.1
New Brunswick	95.7
Quebec	96.3
Ontario	96.2
Manitoba	94.4
Saskatchewan	93.5
Alberta	94.4
British Columbia	95.1

1. In 2021, total non-response rate (TNR) replaces global non-response rate (GNR) in measuring non-response. For more details, refer to the [Total non-response rate](#) section in chapter 9 of the *Guide to the Census of Population, 2021*.

2. For more information on the calculation of the Census of Population weighted response rates, consult the [2021 Census of Population response rate](#) section in chapter 9 of the *Guide to the Census of Population, 2021*.

3. The Census of Population long-form response rate for Canada includes households from Canada's three territories.

There is non-response bias when the characteristics of interest of non-respondents of a survey are different from those of the respondents. Therefore, the higher a survey's non-response is, the greater the risk of non-response bias. The quality of the estimates can be affected if such a bias is present.

Table 2

Weighted imputation rate for the Agriculture–Population Linkage database, Canada and provinces

Agriculture–Population Linkage database weighted imputation rate	
Geography	percent
Canada¹	0.5
Newfoundland and Labrador	5.0
Prince Edward Island	2.1
Nova Scotia	0.9
New Brunswick	1.5
Quebec	0.5
Ontario	0.4
Manitoba	0.4
Saskatchewan	0.7
Alberta	0.4
British Columbia	0.7

1. The Agriculture–Population Linkage database weighted imputation rate for Canada excludes [collective households](#) and Canada's three territories.

The Agriculture–Population Linkage database weighted imputation rate represents the weighted percentage of households in the database for which the Census of Population long-form questionnaire data were imputed. The Agriculture–Population Linkage database weighted imputation rates are very low, reducing the risk of bias caused by imputation of data.

Sampling and weighting

The Agriculture–Population Linkage database contains agricultural data (farm operations and farm operators) and population data (person, household, census family and economic family). Because only a sample of Canadian households was selected to receive the Census of Population long-form questionnaire, weights were assigned to the records on the Agriculture–Population Linkage database to represent the entire farming population. Weights have been calculated at the farm level, person level, household level, census family level and economic family level.

The weights were calculated independently within each province. For most records¹ an initial weight was obtained for each household from the Census of Population based on the number of households in a subprovincial area and the number that responded to the Census of Population long-form questionnaire. Then, characteristics referred to as “constraints” were identified. These were agricultural and population characteristics of primary importance to data users and were fully enumerated on either the Census of Population or Census of Agriculture.

Five basic weights were created (one each at the farm level, person level, household level, census family level and economic family level). Constraints associated with the level of the basic weight were used in calculating the basic weight. For example, only household-level constraints were used to calculate the household-level basic weight. For each province, a method known as generalized regression ensured that the Agriculture–Population Linkage database estimates of these constraints would match the known population counts at the provincial level. From these five basic weights, a total of six weights were produced. Once again, the weight to be used depended on the type of analysis.

For any given geographic area, the weighted population, household, family or farm totals or subtotals may differ from similar estimates presented in previous Census of Agriculture data releases. This is because the Census of Agriculture collected data from all farming operations, whereas the estimates from the Agriculture–Population Linkage database come from a sample. The discrepancies for any variables highly correlated with at least one of the variables used to define a constraint should, in general, be quite small. For other variables, discrepancies will depend on the relationship with the variable used to define a constraint and could be large if no relationship exists. Estimates from the Agriculture–Population Linkage database may also differ from those from the Census of Agriculture because collective dwellings (and the farms associated with these dwellings) are not included in the Agriculture–Population Linkage database.

Quality and data suppression

Results from the Agriculture–Population Linkage database may be suppressed for two reasons: (1) to protect confidentiality of individual respondent data and (2) to limit the dissemination of data of poor quality (which will subsequently be referred to as data quality).

Confidentiality of respondent data is controlled through random rounding, which transforms all estimates of counts to random rounded counts at a base 5 level. Employing this technique, all figures in each table, including totals, are randomly rounded either up or down to a multiple of five. While providing protection against disclosure, this procedure does not add significant error to the data. The random rounding algorithm uses a random seed value to initiate the rounding pattern for tables. In these routines, the method used to seed the pattern can result in the same count in the same table being rounded up in one execution and rounded down in the next.

The data quality of the estimates from the Agriculture–Population Linkage database will be described through the use of the coefficient of variation (CV). The CV of an estimate is the ratio of the estimate of standard error and the estimate itself, expressed as a percentage. The lower the CV, the more accurate the estimate.

1. A small number of records on the Agriculture–Population Linkage database were automatically assigned a weight of 1 and were not weighted as described here. These are households associated with agricultural operations with special characteristics.

The CV is an interesting measure of variability, since it does not depend on the estimate's unit of measure. This makes it possible to compare the accuracy of estimates that have different units of measure. Please note that the CV accounts only for sampling error and does not account for non-sampling errors, such as the error caused by imputation.

To ease interpretation, in the publication tables, the CV value is replaced with a letter indicator. More specifically, estimates with a CV between 25.0% and 49.9% are accompanied by a superscript E to indicate that the user should use caution when interpreting these results because of a moderate level of variability associated with them. Estimates that have a CV of 50.0% and over are not published because of data quality concerns and are replaced in the tables by a letter F.

Table 3
Coefficient of variation ranges and associated letter codes used for estimates from the Agriculture–Population Linkage database

Letter	Coefficient of variation range	Data quality interpretation
[Blank]	0% to 24.9%	Acceptable or better
^E (superscript)	25.0% to 49.9%	Use with caution
F	50.0% and over	Too unreliable to be published