

Canadian Internet Use Survey (CIUS)

Methodology and imputation strategy

Presentation for the ITU EGH

September 2022



Statistics Canada

A better Canada driven by our agency's leadership in expanding knowledge and insights.



Statistics
Canada

Statistique
Canada

Canada



In this presentation:
A simplified
introduction to the
methodology of the
CIUS

1. Survey background
.....
2. Sampling methodology
.....
3. Imputation strategy
.....



Survey background





Survey objectives

The CIUS is designed to measure **the adoption and use of digital technologies** by Canadians, with the goal of publishing estimates at various geographic levels (e.g., provincial, rural/urban).

Target population



Persons **15 years of age and older** living in the ten provinces of Canada

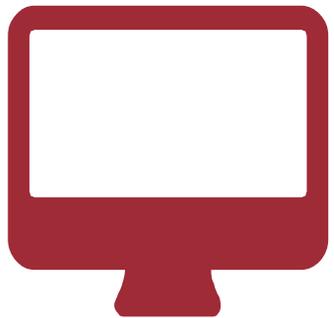
Survey gathers information about:



- Digital technologies
- E-commerce
- Social connections online
- The use of government online services
- Security and privacy online

History and purpose

The CIUS is a voluntary social survey conducted by **Statistics Canada** on behalf of **Innovation, Science and Economic Development Canada (Department of Industry)**



History:

- First launched in 2005
- Underwent major redesigns in 2010 and 2018
- Most recent data published for 2020
- Collection for 2022 in-progress

Purpose:

- Collected data are used to inform evidence-based policymaking, research and program development related to the digital economy.
- The CIUS also provides internationally comparable statistics on the use of digital technologies.



Sampling methodology





Sampling

The **CIUS** is a sample survey with a **cross-sectional design**

Two-stage sampling design:

1. Household
2. Individual(s) within the household

Sample based on a stratified design employing **probability sampling**:

- Stratification done at the province/census metropolitan area (CMA) and census agglomeration (CA) level
- Weights used to account for the uneven probability of selection

Complex frame:

In order to identify households, telephone numbers (landline and cellular) have historically been used.

The 2020 sample included an initial sample size of 45,000 units.

The survey frame was changed in 2022 to use the building register instead of telephone numbers.

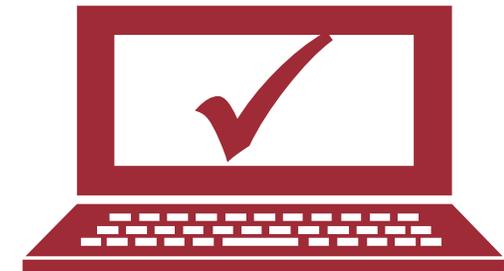
Data collection



For the 2020 iteration, data collection took place from **November 2020** to **March 2021**



Data are collected via electronic questionnaire first, followed by telephone interviewing to resolve issues with responses



In 2020, the collection **response rate** was **42%**



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Sampling errors and data accuracy

Survey estimates can be affected by two types of errors: **sampling** and **non-sampling**.

Sampling error

Measured by the coefficient of variation (CV)

Quality thresholds used on the CIUS:

Acceptable estimate	➡	16.5% and below
Marginal estimate	➡	16.6% to 33.3%
Unacceptable estimate	➡	Above 33.3%

Non-sampling error

Non-response

➡ Managed by redistributing the survey weights of non-responding units to responding units.

Coverage error (or imperfect coverage)

➡ Survey weights are used to allow the sample to better represent all persons in the sampling population. However, they only help to reduce the impact of coverage error, not eliminate it.



Weighting

Weights for the CIUS are generated in several steps:

1. Household weights are derived from the probability of selection from the stratified frame;
2. Household weights are transformed into person weights by accounting for the number of in-scope persons in each household; and
3. The person weights are calibrated to province/age/sex projections based on the Canadian Census of Population.

The result is that each respondent represents a multitude of people in the target population.

For example: in a random 2% sample of the population, each person represents 50 persons in the population





Imputation strategy





Data processing and imputation

The CIUS microdata are cleaned using **editing** and **imputation**.

Editing

.....

- Edits are a series of logical rules used to identify and treat:
 - Errors in questionnaire flow; and
 - Erroneous or missing information in questions which should have been answered
 - For the latter, a non-response code is assigned to the question for the majority of questions

Imputation

.....

- The process that replaces invalid or missing data with valid values to preserve the structure of the data
- The objective is to establish internally consistent data records
- Imputation produces datasets that are easier for data users to use



Three types of non-response

Complete non-response

- When the respondent does not provide the minimum set of answers
- Records are dropped in the weighting process

Item non-response

- When the respondent does not provide an answer to a particular question
- Imputed or coded as “not stated”

Partial non-response

- When the respondent provides the minimum set of answers, but does not complete the survey
- Handled as complete non-response or multiple item non-response

For the CIUS, donor imputation is used to fill in missing data for some cases of item and partial non-response





Imputation for item non-response

For the CIUS, imputation use the **nearest-neighbour method** which searches for “donor” records to replace missing records



The recipient records are imputed by a donor chosen from a group of records with **similar demographic characteristics**

Donors are selected using a score function. A value is added for each match between the donor and recipient.



Limited imputation strategy

For variables on the CIUS, imputation was limited to addressing **item non-response** for:

The gender of
the respondent

The value of the
amount spent on
physical goods
over the Internet

The value of the
amount spent
over the Internet
on other services

Age, location and
income



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Imputation of gender

Due to the limited number of gender-diverse respondents in the sample, data could not be published for gender-diverse individuals without risking disclosure.

Where the gender of the respondent was **diverse or missing**, imputation was used to assign a gender to such respondents in order to match the proportion of male and female individuals in the population.



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



Imputation of Internet spending

Donor records were chosen from a group of records with **similar demographics**. The relative imputation rate serves as a **data quality indicator**:

- In 2020, for every \$100 spent in online orders for physical goods, about \$8.20 was imputed.
- Similarly, with the value of online orders for other services, for every \$100 spent, \$11.70 was imputed.
- For nearly all imputed cases, the range of the amount spent by the respondent was known.





Imputation of age, location and income

Respondents are asked to report/confirm their age and location during collection and reported values are verified by linking to other data sources Statistics Canada has access to. Erroneous or missing responses are replaced with the values from alternative sources.

Since 2020, income has not been collected on the survey. Instead, it is imputed by either directly **linking to tax records** or using **nearest-neighbour imputation** when no linkage can be made.



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada



In conclusion

The CIUS is continually changing and improving due to **complexities** regarding the survey frame and collection.

Survey design challenges

- Non-response (around 58%)
- Uneven coverage (non-sampling errors)
- Weighting and imputation are not a perfect solution for handling errors

Other challenges

- Response mechanism correlated with certain key indicators (e.g., measuring Internet use primarily via the Internet)
- Voluntary survey
- Subject to memory bias from respondents





Thank you

Howard Bilodeau

Economist

Investment, Science and Technology Division

Statistics Canada

howard.bilodeau@statcan.gc.ca



Statistics
Canada

Statistique
Canada

Delivering insight through data for a better Canada

Canada