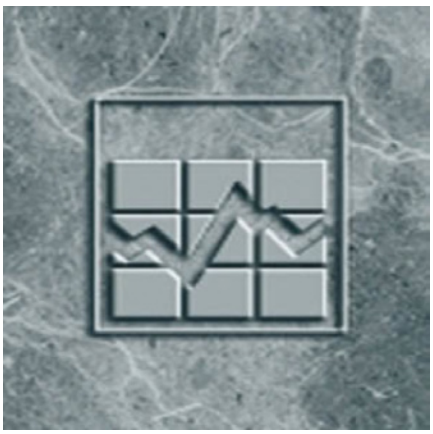
Catalogue no. 89-648-X
ISSN 1927-0100
ISBN 978-0-660-45389-7

Longitudinal and International Study of Adults Research Paper Series

# Attrition in the Longitudinal and International Study of Adults, Wave 1 (2012) to Wave 4 (2018)



by Emmanuel Benhin, Brandon Rouleau, Maaz Shahid, and Gaëlle Simard-Duplain

Release date: November 14, 2022

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*.

## Table of contents

# Attrition in the Longitudinal and International Study of Adults, Wave 1 (2012) to Wave 4 (2018)

by **Emmanuel Benhin; Brandon Rouleau; Maaz Shahid;** and **Gaëlle Simard-Duplain**

## 1. Introduction

The Longitudinal and International Study of Adults (LISA) survey aims to provide data to better understand what is happening in the lives of Canadians. These data can be used to develop effective policies and provide needed services. A growing number of countries conduct similar studies to illuminate the diverse links between work, health, education, and family and to obtain a better sense of individual and family well-being.

The first wave of LISA was conducted from November 2011 to May 2012. Data were collected on topics such as education, training, skills, work histories, current education, work activities, and health. Respondents were re-contacted every two years, from 2014 to 2020, to collect updated data on existing topics and to explore new ones.

One major methodological and analytical challenge to the use of longitudinal data from surveys such as LISA is attrition. Attrition is the loss of respondents over the life of a longitudinal survey. This type of non-sampling error can lead to biased estimates if the respondents who remain in the survey differ systematically from those who drop out. In this case, the attrition is considered non-random. Even when attrition is random, the loss of respondents can result in reduced precision or reliability of the survey estimates. If left untreated, attrition can limit the value of research findings based on longitudinal data. However, the impact of attrition can be mitigated using a non-response weight adjustment and calibration strategy. Where this approach fails to mitigate attrition for some variables, analytical methods that incorporate the effect of attrition in a statistical model may also be helpful.

This report explores the size and nature of the attrition challenges faced by the LISA survey, as well as the use of a non-response weight adjustment and calibration strategy to mitigate the effects of attrition on the LISA estimates. The study focuses on data from waves 1 (2012) to 4 (2018) and uses practical examples based on selected demographic variables, to illustrate how attrition can be assessed and treated.

Section 2 of this technical report provides background on the source and nature of sampling and non-sampling errors, as well as the types of attrition and attrition mechanisms. Section 3 examines the size of the LISA sample and changes in the number of respondents over the four survey waves from 2012 to 2018. Section 4 explores potential measures of attrition for LISA, including sample attrition rates and odds ratios from logistic regression models. The sample attrition rates and odds ratios for two selected demographic variables are discussed. In section 5, the impacts of the LISA non-response weight adjustment and calibration strategy on attrition are evaluated based on changes in sample composition and attrition ratios pre and post-mitigation. Finally, concluding remarks are presented in section 6.

## 2. Background

### 2.1. Objectives

The aim of the Longitudinal and International Study of Adults survey is to collect data that will help improve understanding of what is happening in the lives of Canadians; provide Canadians with information to support their present and future decisions; help policy makers and governments determine what services may be suitable for Canadians and to illuminate the diverse links between work, health, education and family.

## 2.2. Target population

The Longitudinal and International Study of Adults[1] survey covers the population aged 15 or over that was living in the ten provinces in 2012 (wave 1), plus their future descendants.[2] Excluded from this population were those who, at the time, were living on reserves and in other Aboriginal settlements in the provinces; official representatives of foreign countries living in Canada and their families; members of religious and other communal colonies; members of the Canadian Armed Forces stationed outside of Canada; persons living full-time in institutions; and persons living in other collective dwellings. Altogether, these exclusions represented approximately 2% of the population. For more detailed information about the LISA population, the reader may consult the LISA 2018 User Guide by Statistics Canada (2020) or Gossen and Lévesque (2013).

## 2.3. Sample design

The LISA sample was selected based on a stratified multistage, multi-phase sampling design with probability proportional to size. The sampling frame for the LISA population was first stratified by province and within each province was further stratified into rural and urban areas.

For the rural areas, smaller geographical areas or clusters were first created within each rural area and then a sample of clusters was selected with probability proportional to size of the cluster. A sample of dwellings was selected with probability proportional to the size of dwellings within each cluster.

For the urban areas, there was no geographical clustering of dwellings before a sample of these was selected with probability proportional to their size.

From among the selected dwellings for urban and rural areas, a further subsample of dwellings was selected. This subsample of dwellings made up the LISA person-level sample, which is the primary focus of this attrition study.

For more detailed information about the LISA sample design, the reader may consult the LISA 2018 User Guide by Statistics Canada (2020) or Gossen and Lévesque (2013).

## 2.4. Data collection periods

Data collection for LISA wave 1 was conducted from November 2011 to May 2012. The wave 2 collection period was from January to June 2014, wave 3 from January to May 2016 and wave 4 from January to June 2018. The wave 5 collection period was scheduled for January to June 2020. However, due to the COVID-19 pandemic, collection was suspended in March, resumed in July and completed in September 2020. Data collection for LISA was done primarily by computer-assisted personal interview (CAPI). Computer-assisted telephone interviews (CATI) were used under exceptional circumstances, such as the 2020 (wave 5) collection from July to September 2020.

## 2.5. Sampling and non-sampling errors

Since LISA is a sample survey, its estimates are subject to both sampling and non-sampling errors.

Sampling errors occur because inferences about the survey population are based on data from a sample of that population rather than the entire population. The sample design, the variability of the characteristic being measured, the population and sample sizes, and the survey response rate will impact the size of the sampling error and the precision of the survey estimates. Data precision can be improved by incorporating auxiliary data into the estimation process (known as calibration). For LISA, the auxiliary data, the estimator and the method used to estimate sampling variance were all chosen to minimise sampling errors and increase the precision of the LISA estimates.[3]

1. Being a longitudinal survey that collects information on education, employment and overall well-being, the purpose of LISA is to improve understanding of how Canadian households change over time.
2. Members aged 0-14 at the time of the 2012 interview and any future children of permanent sample members will be interviewed in future cycles of the survey after they turn 15 years of age. In addition, in future cycles, any household members 15 years of age or older living with the permanent sample members will also be interviewed (Statistics Canada, 2014).
3. See Benhin, Frolova & Quesnel (2019) for additional details on the weighting strategy in LISA.

Non-sampling errors can occur at any point in the collection and processing of survey data. They can take the form of coverage errors, nonresponse errors, response errors, capture errors, coding errors and other types of processing errors. Over a large number of observations, randomly occurring non-sampling errors will have a minimal impact on estimates derived from the survey. On the other hand, systematically occurring errors will contribute to bias in the survey estimates and can have a more significant impact on their reliability.[4]

## 2.6. Attrition

Attrition is an example of a non-sampling error that results from the loss of respondents over the life of a longitudinal survey. Attrition, as with non-response in general in surveys, may be non-random or random. It is non-random if the respondents who remain in the survey differ significantly from those who drop out. This kind of attrition may lead to biased estimates of population parameters. Attrition is random when the respondents who remain in the survey are similar to those who drop out. Even though random attrition may result in unbiased estimates, the loss of respondents can reduce the precision of the survey estimates. Whether attrition is random or non-random, if left untreated, it can impact the usefulness of the survey results.

Attrition is a concern common to all longitudinal surveys. In their study of the factors associated with non-response in longitudinal surveys, Watson and Wooden (2009) reference that, over an eight-year period, both the German Socio-economic Panel (GSOEP) and the British Household Panel Survey (BHPS) reported losses of nearly 34 percent of their original samples. They provide evidence that the issue of attrition has increased over time, with newer panel surveys experiencing a loss of respondents at a higher rate than those in the past. In the Household, Income and Labour Dynamics in Australia (HILDA) survey, which began in 2001, roughly 25% of the original sample were not re-interviewed after a four-year period. Given that the original cohorts of the GSOEP, BHPS and the HILDA started in 1984, 1991 and 2001, respectively, this study provides an additional opportunity to explore attrition for a longitudinal survey that began at least 10 years after the aforementioned surveys.

As expected, at each wave of LISA, there has been a loss of respondents. If the losses were random then the remaining sample members in the subsequent waves would be as representative of the Canadian population in 2012 as the LISA sample in 2012 (wave 1). This means that estimates based on subsequent waves would be as unbiased as estimates from the sample in wave 1. If the loss of respondents is not random, then the remaining respondents in the survey may not be as representative of the Canadian population in 2012 as the original 2012 sample. Estimates based on their responses may be biased if sample attrition is untreated.

For LISA, the effect of attrition has been mitigated primarily using a nonresponse weight adjustment and calibration strategy. For specific analytical variable(s) where the weight adjustment strategy does not mitigate the effect of attrition, there are other analytical methods that can be explored. However, these are not covered in this report.[5]

### 2.6.1. Attrition mechanisms

Attrition mechanisms describe the patterns by which the loss of respondents or respondents' data over the life of a survey are assumed to have occurred. These patterns may or may not be related to the survey, the analytical variables of interest or other variables in the survey. In the longitudinal survey context, the loss of respondents over time may fall into three attrition mechanisms or patterns. For graphical representations of each pattern, see Thoemmes and Mohan (2015).

The first attrition mechanism is missing completely at random (MCAR) or unsystematic where the loss of respondents is independent of the observed and unobserved (missing) data. In other words, the probability that data are missing for a given variable is unrelated to any other surveyed variable, including the variable with missing values itself.

An attrition pattern is missing at random (MAR) if the loss of respondents depends on the observed data but not on the unobserved (missing) data. In this case, the probability that data are missing for a given variable is related to some other surveyed variable but not to the variable with missing values itself. For example, young males are more likely to have higher nonresponse. The probability of missing data for a respondent is related to their age group and sex, which are assumed to be observed for all individuals in the survey.

4. See the LISA 2018 Data Quality Report (Statistics Canada, 2020) for information on non-sampling errors in LISA.
5. See Woolridge (2010) for details on other methods used to address attrition.

The third attrition pattern is not missing at random (NMAR) or missing not at random (MNAR) whereby the loss of respondents depends on unobserved data. The missing data for a variable are related to the values of that variable itself even after controlling for other variables. For example, when data are missing for the variable low employment and only the people with low employment (e.g., young males) have missing observations for the employment variable.

For LISA, missing at random (MAR) is the assumed attrition mechanism for the purpose of implementing a nonresponse weight adjustment and calibration strategy to mitigate the effect of attrition. The limitation of this strategy is that for specific individual analytical variable(s) whose pattern of missing respondents is not MAR, valid statistical analyses and inferences cannot be made. Other methods outside the scope of this report must be used to treat attrition in these cases.

### 2.6.2. Measurement of attrition in LISA

In this report, attrition is measured using a cumulative attrition rate for a person-level sample. As defined here and in other longitudinal studies,[6] cumulative attrition (at wave $t = 2,3,4,...$) is the loss of respondents from the survey since wave 1. Respondents to the initial phase (wave 1) of the survey represent the longitudinal sample. The sample attrition rate (SAR) or unweighted attrition rate (UAR) at wave $t$ is the proportion of the initial longitudinal sample lost to the survey at wave $t$. This is usually expressed as a percentage based on the following formula:

(1)

$$\mathrm{UAR}_t = \left(1 - \frac{\sum_{i \in s_1} R_{ti}}{\sum_{i \in s_1} R_{1i}}\right) \times 100, \qquad t = 2,3,4,\ldots$$

where

$$R_{ti} = \begin{cases} 1, & \text{if person } i \text{ responded to the survey in all waves up to wave } t = 2,3,4,... \\ 0, & \text{otherwise} \end{cases}$$

$$R_{1i} = \begin{cases} 1, & \text{if person } i \text{ responded to the survey in wave 1} \\ 0, & \text{otherwise} \end{cases}$$

$s_1$ is the longitudinal sample at wave 1.

Cumulative attrition for the population could also be estimated and measured by a weighted attrition rate. A weighted attrition rate (WAR) at wave $t$ is the proportion of the initial longitudinal population members lost to the study at wave $t$. This may be expressed as a percentage based on the following formula:

(2)

$$\mathrm{WAR}_t = \left(1 - \frac{\sum_{i \in s_1} w_{1i} R_{ti}}{\sum_{i \in s_1} w_{1i} R_{1i}}\right) \times 100, \qquad t = 2,3,4,\ldots$$

6. See National Centre for Longitudinal Data (2018) or Statistics Canada (2010)

where

$$R_{ti} = \begin{cases} 1, & \text{if person } i \text{ responded to the survey in all waves up to wave } t = 2,3,4,... \\ 0, & \text{otherwise} \end{cases}$$

$$R_{1i} = \begin{cases} 1, & \text{if person } i \text{ responded to the survey in wave 1} \\ 0, & \text{otherwise} \end{cases}$$

$s_1$ is the longitudinal sample at wave 1 and $w_{1i}$ is the survey weight at wave 1 for person $i$.

The analyses of attrition in this report are based on measures of sample attrition.

For any given analytical variable with two or more categories, attrition rates may be computed for each of these categories. For example, aboriginal status has two categories: aboriginal and non-aboriginal. If the attrition rates for these two categories remain essentially the same over time, this provides evidence that attrition does not lead to bias that would limit the use of the variable (e.g., aboriginal status) for statistical analysis. On the other hand, if the attrition rates differ significantly among the categories, then attrition may have an effect on the use of the variable for statistical analysis.

Another well-known statistical tool used to measure the impact of attrition on any given analytical variable of interest is logistic regression analysis. This may be done by fitting a two-category dependent variable representing the occurrence or non-occurrence of attrition to an analytical (independent) variable of interest, say aboriginal status. Refer to Appendix A for details on the use of logistic regression methods for attrition analysis.

## 3. Data

The data used to analyse the effects of attrition in this study are from the Longitudinal and International Study of Adults survey from 2012 (wave 1) to 2018 (wave 4). The longitudinal sample consisted of 23,926 respondents to the initial survey in 2012 (wave 1). The sample represents all persons 15 years of age and older living in the ten provinces of Canada in 2012. The variables selected for the attrition analysis in this study include age group and aboriginal status.

### 3.1. LISA weighting and calibration

Survey design weights and corresponding bootstrap weights were assigned to each person in the initial responding sample of the LISA survey for 2012 for use in producing survey estimates.[7] Unfortunately, LISA estimates for subsequent survey waves were impacted by nonresponse and attrition. If the initial survey design weights were used in estimation for other waves, sample attrition would have resulted in invalid and unreliable statistical estimates and inferences. For these reasons, a nonresponse weight adjustment and calibration strategy was put in place and the initial LISA survey design weights were adjusted for each new wave of the survey. To adjust for nonresponse, it was first assumed that LISA attrition was missing at random (MAR) and that the departure of respondents from the survey was related to variables such as their age, sex, marital status, province or region of residence.[8] These variables and others were used in the nonresponse weight adjustment and calibration strategy to produce the new LISA sampling weights for each wave. In addition, the weights were calibrated to ensure that population counts generated from the LISA sample at each wave would be consistent with the actual population count observed at wave 1 (2012).[9] This nonresponse adjustment and calibration strategy helped to minimize the effect of nonresponse and attrition on estimates for variables that were directly included in the adjustment process, as well as for other strongly correlated variables. For specific analytical variable(s) that were not correlated with those included in the adjustment process, the impact of nonresponse and attrition may not have been reduced. In this situation, analytical methods may be needed to directly account for the effect of attrition on each specific analytical variable. However, discussions of this second corrective measure were out of scope for this report.[10]

7. For additional details on the survey design weights in LISA, see Appendix B.
8. See Benhin, Frolova & Quesnel (2019) for additional details on how non-response is addressed in LISA.
9. See Bocci (2014) and Benhin (2018) for more details on the LISA weighting process.
10. See Wooldridge (2010) for further information on methods used to correct for attrition.

## 3.2. Overview of sample changes

The LISA sample consists of a household-level sample and a person-level sample. The LISA household-level sample over the four waves is presented in Table 3.2.1. It shows that the number of responding households decreased across survey waves from 2012 (wave 1) to 2018 (wave 4). Response rates also declined from 74% in wave 2 to 60% in wave 4. For the LISA survey, a responding household had at least one household member who completed the survey.

**Table 3.2.1**
**LISA household-level response outcomes, 2012 to 2018, Canada**

| | Wave 1 (2012) | Wave 2 (2014) | Wave 3 (2016) | Wave 4 (2018) |
|---|---|---|---|---|
| **Household response outcome (number of households)** | | number | | |
| Responses | 11,458 | 9,680 | 9,414 | 9,132 |
| Nonresponses | ... | 3,407 | 4,871 | 6,091 |
| **Total households** | **11,458** | **13,087** | **14,285** | **15,223** |

... not applicable

**Note:** Members of the wave 1(2012) responding households form the basis for the LISA longitudinal study.
**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

The LISA person-level sample from 2012 (wave 1) to 2018 (wave 4) is shown in Table 3.2.2.[11] In wave 1, there were 32,133 persons from the LISA responding households. Over time, the person-level sample increased with the addition of both new children (Permanent Sample Member 2 (PSM2)) born or adopted by members of the original sample (Permanent Sample Member 1 (PSM1)) and temporary sample members (TSM). The latter are individuals, who at the time of the LISA data collection, were living with an original sample member or their children.

**Table 3.2.2**
**LISA person-level sample counts, 2012 to 2018, Canada**

| | Wave 1 (2012) | Wave 2 (2014) | Wave 3 (2016) | Wave 4 (2018) |
|---|---|---|---|---|
| **LISA person-level sample (number of persons)** | | number | | |
| Wave 2 (2014) | | | | |
| PSM1(original sample member) | 32,133 | 32,133 | 32,133 | 32,133 |
| PSM2 (new child of the original sample member) | ... | 440 | 440 | 440 |
| TSM (temporary sample member) | ... | 1,300 | 1,300 | 1,300 |
| Wave 3 (2016) | | | | |
| PSM2 (new child of the original sample member) | ... | ... | 343 | 343 |
| TSM (temporary sample member) | ... | ... | 1,266 | 1,266 |
| Wave 4 (2018) | | | | |
| PSM2 (new child of the original sample member) | ... | ... | ... | 329 |
| TSM (temporary sample member) | ... | ... | ... | 1,143 |
| **Total household members** | **32,133** | **33,873** | **35,482** | **36,954** |

... not applicable

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

For attrition analysis purposes in this report, the LISA person-level sample consists of 23,926 individuals aged 15 years or older in 2012 who responded to the survey in wave 1. The number of respondents and attrition rates from 2012 to 2018 are reported in Table 3.2.3.[12] Respondents in the table responded to all waves of the LISA survey from 2012 to 2018.

11. Counts in the table include original sample members at wave 1 (2012), new children born or adopted by the original sample members after 2012 and temporary sample members living with original sample members at the time of the data collections.
12. Counts in the table include respondents to the survey exclusively.

**Table 3.2.3**
**LISA respondents (all waves) and attrition rates, 2012 to 2018, Canada**

| Survey year | Number of respondents | Wave-to-wave attrition | Cumulative attrition | Wave-to-wave attrition rates | Cumulative attrition rates |
|---|---|---|---|---|---|
| | number | | | percent | |
| Wave 1 (2012) | 23,926 | ... | ... | ... | ... |
| Wave 2 (2014) | 16,895 | 7,031 | 7,031 | 29.4 | 29.4 |
| Wave 3 (2016) | 12,991 | 3,904 | 10,935 | 23.1 | 45.7 |
| Wave 4 (2018) | 10,640 | 2,351 | 13,286 | 18.1 | 55.5 |

... not applicable

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

For some national and international longitudinal surveys, the attrition rate is defined in an alternative way as the proportion of respondents remaining in the survey at a given wave relative to the number of respondents at wave 1.[13] This definition includes respondents who may have left the survey in one wave but rejoined in another. Table 3.2.4 presents the response rates and proportion of respondents remaining in the LISA from 2012 (wave 1) to 2018 (wave 4) corresponding to the above definition. Using this definition, the attrition rate at wave 4 (2018) for LISA is 43.4% (Table 3.2.4) compared with 55.5% (Table 3.2.3) using the cumulative definition and requiring that respondents had responded in all previous waves. The latter is the definition (see section 2.6.2 in the report) primarily employed for the analysis of attrition in this report.

**Table 3.2.4**
**Proportion of respondents[1] remaining in the LISA from 2012 (wave 1) to 2018 (wave 4), Canada**

| Survey year | Total sample | Respondents | Response rate | Attrition rate |
|---|---|---|---|---|
| | number | | percent | |
| Wave 1 (2012) | 23,926 | 23,926 | ... | ... |
| Wave 2 (2014) | 23,926 | 16,895 | 70.6 | 29.4 |
| Wave 3 (2016) | 23,926 | 14,852 | 62.1 | 37.9 |
| Wave 4 (2018) | 23,926 | 13,531 | 56.6 | 43.4 |

... not applicable
1. Respondents may have left and rejoined the survey.

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

## 4. Attrition for selected variables

For the purposes of attrition analyses, respondents to the LISA 2012 (wave 1) aged 15 and older form the basis of the longitudinal sample. This sample of respondents consists of 23,926 persons. These sample members represent the Canadian population of all persons 15 years of age and older living in the 10 provinces in 2012.

As discussed in Section 2, two tools can be used to quantify the effect of attrition on longitudinal estimates: the unweighted attrition rates and the logistic regression model for the analysis of attrition for any analytical variable of interest. Both are employed to measure the effect of attrition on LISA for the age group and aboriginal status[14] variables in sections 4.1 and 4.2 respectively.

### 4.1. Attrition by age group

Table 4.1.1 presents the cumulative unweighted attrition rates (UAR) by age group for LISA from 2012 to 2018 calculated using equation (1) in section 2.6.2 of this report. The cumulative attrition rates[15] at the Canada level increased from 29% in 2014 (wave 2) to 56% in 2018 (wave 4). Those in age group 15 to 24 had the highest attrition rates, increasing from 40% in 2014 (wave 2) to 72% in 2018 (wave 4). Attrition rates were negatively correlated with age group. This means that the LISA survey was much more likely to lose respondents who were 15 to 24 years of age in 2012 (wave 1) than respondents who were 65 years of age and older.

13. See Statistics Canada (2010), Taylor (2018) or National Centre for Longitudinal Data (2018). However, depending on the specific needs of the longitudinal survey, attrition rates can be defined differently.
14. Excluded from the survey's coverage are those who at the time of LISA 2012 (Wave 1) were living on reserves and other Aboriginal settlements in the provinces.
15. Cumulative attrition represents all respondents lost to the survey since wave 1. Non-responding individuals who returned to the survey are considered lost to the survey for the cumulative rates calculations.

**Table 4.1.1**
**Respondents and cumulative sample attrition rates by age group, 2012 to 2018, Canada**

| | Wave 1 (2012) | | Wave 2 (2014) | | Wave 3 (2016) | | Wave 4 (2018) | |
|---|---|---|---|---|---|---|---|---|
| **Age group** | number | percent | number | percent | number | percent | number | percent |
| 15 to 24 | 4,556 | 0.0 | 2,722 | 40.0 | 1,792 | 61.0 | 1,280 | 72.0 |
| 25 to 34 | 3,155 | 0.0 | 1,978 | 37.0 | 1,479 | 53.0 | 1,223 | 61.0 |
| 35 to 44 | 3,755 | 0.0 | 2,662 | 29.0 | 2,032 | 46.0 | 1,678 | 55.0 |
| 45 to 54 | 5,360 | 0.0 | 3,973 | 26.0 | 3,151 | 41.0 | 2,599 | 52.0 |
| 55 to 64 | 4,168 | 0.0 | 3,288 | 21.0 | 2,691 | 35.0 | 2,364 | 43.0 |
| 65 + | 2,932 | 0.0 | 2,272 | 23.0 | 1,846 | 37.0 | 1,496 | 49.0 |
| **Totals** | **23,926** | **0.0** | **16,895** | **29.0** | **12,991** | **46.0** | **10,640** | **56.0** |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

To test whether attrition for different age groups is statistically significant, a logistic regression method was employed. This involved fitting a two-category outcome (dependent) variable representing the occurrence (assigned value 1) or non-occurrence (assigned value 0) of attrition to an independent (explanatory) variable representing the age group. Using the LISA data for 2012 (wave 1) and 2018 (wave 4), the wave 1 responding-person weights (RPW) and the corresponding bootstrap weights, a logistic regression model was implemented using a SAS procedure SURVEYLOGISTIC (see Appendix A for details on the logistic regression model). The results from this model are presented in Table 4.1.2. The results show that compared to the age group 65 years of age and older, the LISA survey is more likely to lose respondents for all age groups except those aged 55 to 64. It is worth noting that respondents who were 15 to 24 years of age in 2012 (wave 1) were about three times more likely to leave the survey than respondents who were 65 years of age and older.[16] This means that attrition left untreated or not accounted for in any analysis involving the variable age group, could lead to invalid and unreliable statistical inferences.

**Table 4.1.2**
**Odds of attrition by age group, 2018 (wave 4), Canada**

| Age group | Odds ratio |
|---|---|
| 15 to 24 | 3.06* |
| 25 to 34 | 1.74* |
| 35 to 44 | 1.38* |
| 45 to 54 | 1.25* |
| 55 to 64 | 0.87* |
| 65 + (reference group) | |

* statistically significant as confidence interval does not contain '1'.
**Note:** All estimates use wave 1 longitudinal weights.
**Source:** Statistics Canada, Longitudinal and International Study of Adults (2018).

## 4.2. Attrition by aboriginal status

Table 4.2.1 presents the cumulative unweighted sample attrition rates by aboriginal status for LISA from 2012 to 2018. The attrition rates for aboriginals increased from 39% in 2014 (wave 2) to 65% in 2018 (wave 4) compared with 29% to 55% for non-aboriginals. To test the significance of this finding, a logistic regression analysis was used to measure the odds of attrition for aboriginal and non-aboriginal LISA respondents. The results in Table 4.2.2 show that attrition for aboriginals is significantly different from that of non-aboriginals. In fact, the odds of an aboriginal leaving the survey was about 1.5 times that of non-aboriginals.

It can also be observed that for the aboriginal status variable, there were a small number of respondents for which data were missing. Researchers are advised to determine the best approach to include the effect of this item non-response in their analysis of LISA data. One approach is the use of imputation to replace the missing data of the aboriginal status with data from a respondent with similar characteristics.

16. The odds ratio is the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure. So in the context of the age group table, the odds ratio compares the odds of losing respondents who are 15 to 24 years of age with the odds of losing respondents who are 65 years of age or older.

**Table 4.2.1**
**Respondents and attrition rates by aboriginal status, 2012 to 2018, Canada**

| | Wave 1 (2012) | | Wave 2 (2014) | | Wave 3 (2016) | | Wave 4 (2018) | |
|---|---|---|---|---|---|---|---|---|
| **Aboriginal status** | number | percent | number | percent | number | percent | number | percent |
| Aboriginal | 911 | 0.0 | 553 | 39.0 | 392 | 57.0 | 320 | 65.0 |
| Non-aboriginal (reference) | 22,965 | 0.0 | 16,313 | 29.0 | 12,576 | 45.0 | 10,300 | 55.0 |
| Missing | 50 | 0.0 | 29 | 42.0 | 23 | 54.0 | 20 | 60.0 |
| **Totals** | **23,926** | **0.0** | **16,895** | **29.0** | **12,991** | **46.0** | **10,640** | **56.0** |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

**Table 4.2.2**
**Odds of attrition by aboriginal status, 2018 (wave 4), Canada**

| **Aboriginal status** | **Odds ratio** |
|---|---|
| Aboriginal | 1.53* |
| Non-aboriginal (reference) | |

* statistically significant as confidence interval does not contain '1'.
**Note:** All estimates use wave 1 longitudinal weights.
**Source:** Statistics Canada, Longitudinal and International Study of Adults (2018).

# 5. Managing attrition with a nonresponse weight adjustment and calibration strategy

A nonresponse adjustment and calibration strategy has been used to minimise or eliminate the effect of attrition on LISA estimates. The impact of this strategy can be measured by looking at how the sample composition over the four LISA waves changed before and after the non-response adjustment and calibration. This is discussed in sections 5.1 and 5.2 with a focus on the age group and aboriginal status variables. Attrition ratios can also be used to measure the impact of the nonresponse adjustment strategy, as discussed in sections 5.3 and 5.4. It should be noted that for weighting purposes after 2012 (wave 1), persons who died, were institutionalized or who moved out of Canada were considered respondents and assigned a longitudinal weight. This allowed for the estimation of the total number of individuals in the LISA population who have been lost due to death, institutionalization or departure from Canada over the course of the survey.

## 5.1. Sample composition by age group from 2012 (wave 1) to 2018 (wave 4)

Provided that sample attrition is random, the sample composition for a longitudinal survey should not change over time. This means that for pre-defined categories of any analytical variable, the number or proportion of respondents in each category does not change from one wave to the next. If the sample composition changes systematically for certain dimensions (e.g., aboriginal status) more than others across waves, attrition may be non-random and estimates based on the sample may be biased and result in invalid statistical inferences.

To evaluate whether the remedial nonresponse weight adjustment and calibration for LISA attrition worked, the sample composition was calculated with and without weights for the age group and aboriginal status variables. The unweighted sample composition represents the case where no mitigation strategy was implemented and the weighted sample composition reflects the implementation of a nonresponse weight adjustment and calibration mitigation strategy. The unweighted sample compositions by age group over the four waves from 2012 (wave 1) to 2018 (wave 4) are shown in Table 5.1.1. The results show that 19.0% of the respondents in 2012 (wave 1) were individuals age 15 to 24 but this proportion decreased consistently across all waves to 11.6% for 2018 (wave 4). On the other hand, 12.3% of the respondents in 2012 were individuals aged 65 years and older and the share of this group in the population increased steadily to 17.3% in 2018. Since the respondents' composition in the sample by age group has either decreased or increased systematically over time, the sample attrition is non-random. Unmitigated, sample attrition for age group would lead to biased estimates and invalid statistical inferences.

**Table 5.1.1**
**Unweighted respondents' composition in the sample by age group, 2012 to 2018, Canada**

| | Wave 1 (2012) | | Wave 2 (2014) | | Wave 3 (2016) | | Wave 4 (2018) | |
|---|---|---|---|---|---|---|---|---|
| **Age group** | number | percent | number | percent | number | percent | number | percent |
| 15 to 24 | 4,556 | 19.0 | 2,760 | 16.0 | 1,845 | 13.6 | 1,322 | 11.6 |
| 25 to 34 | 3,155 | 13.2 | 1,997 | 11.6 | 1,503 | 11.1 | 1,243 | 10.9 |
| 35 to 44 | 3,755 | 15.7 | 2,675 | 15.5 | 2,063 | 15.2 | 1,710 | 15.0 |
| 45 to 54 | 5,360 | 22.4 | 4,006 | 23.3 | 3,198 | 23.6 | 2,661 | 23.4 |
| 55 to 64 | 4,168 | 17.4 | 3,333 | 19.4 | 2,780 | 20.5 | 2,489 | 21.8 |
| 65 + | 2,932 | 12.3 | 2,436 | 14.2 | 2,167 | 16.0 | 1,968 | 17.3 |
| **Totals** | **23,926** | **100.0** | **17,207** | **100.0** | **13,556** | **100.0** | **11,393** | **100.0** |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

Table 5.1.2 shows the weighted LISA sample composition by age group, using the LISA respondents' person weights (RPW) at wave 1 (2012) and the all-waves longitudinal respondents' person weights (AWRPW) at waves 2, 3, and 4.[17] The weighted sample composition remains essentially the same from wave 1 to wave 4 across all age groups. These results provide strong evidence that for these age groups, the nonresponse weight adjustment and calibration strategy have mitigated the effect of attrition. By applying the adjusted LISA longitudinal weights, estimates for these age groups would be unbiased and would support valid statistical inferences.

In general, it would be a good practice for researchers as a first step in the statistical analysis of LISA data, to verify whether the effect of attrition on the particular analytical variable(s) of interest has been minimized or eliminated by the nonresponse weight adjustment and calibration strategy. This may be done by simply computing and comparing the unweighted and weighted sample compositions similar to the ones shown in Table 5.1.1 and Table 5.1.2.

**Table 5.1.2**
**Weighted respondents' composition in the sample by age group, 2012 to 2018, Canada**

| | Wave 1 (2012) | | Wave 2 (2014) | | Wave 3 (2016) | | Wave 4 (2018) | |
|---|---|---|---|---|---|---|---|---|
| **Age group** | number | percent | number | percent | number | percent | number | percent |
| 15 to 24 | 4,459,690 | 15.8 | 4,480,001 | 15.9 | 4,479,087 | 15.9 | 4,473,804 | 15.9 |
| 25 to 34 | 4,690,854 | 16.6 | 4,638,423 | 16.5 | 4,638,886 | 16.5 | 4,643,656 | 16.5 |
| 35 to 44 | 4,589,204 | 16.3 | 4,565,123 | 16.2 | 4,576,417 | 16.2 | 4,573,927 | 16.2 |
| 45 to 54 | 5,309,417 | 18.8 | 5,306,953 | 18.8 | 5,293,880 | 18.8 | 5,293,295 | 18.8 |
| 55 to 64 | 4,364,251 | 15.5 | 4,366,920 | 15.5 | 4,366,144 | 15.5 | 4,366,652 | 15.5 |
| 65 + | 4,807,699 | 17.0 | 4,816,975 | 17.1 | 4,819,981 | 17.1 | 4,823,063 | 17.1 |
| **Totals** | **28,221,115** | **100.0** | **28,174,395** | **100.0** | **28,174,395** | **100.0** | **28,174,396** | **100.0** |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

## 5.2. Sample composition by aboriginal status from 2012 (wave 1) to 2018 (wave 4)

The unweighted sample compositions for aboriginal status over the four waves from 2012 (wave 1) to 2018 (wave 4) are shown in Table 5.2.1. The results show that the proportion of aboriginals in the sample decreased from 3.8% in 2012 (wave 1) to 3.0% in 2018 (wave 4), whereas the proportion of non-aboriginals slightly increased from 96.0% in 2012 (wave 1) to 96.8% in 2018 (wave 4). This indicates that there may be a potential negative effect of attrition on the aboriginal status as an analytical variable.

**Table 5.2.1**
**Unweighted respondents' composition in the sample by aboriginal status, 2012 to 2018, Canada**

| | Wave 1 (2012) | | Wave 2 (2014) | | Wave 3 (2016) | | Wave 4 (2018) | |
|---|---|---|---|---|---|---|---|---|
| **Aboriginal status** | number | percent | number | percent | number | percent | number | percent |
| Aboriginal | 911 | 3.8 | 562 | 3.3 | 406 | 3.0 | 343 | 3.0 |
| Non-aboriginal | 22,965 | 96.0 | 16,615 | 96.6 | 13,127 | 96.8 | 11,030 | 96.8 |
| Missing | 50 | 0.2 | 30 | 0.2 | 23 | 0.2 | 20 | 0.2 |
| **Totals** | **23,926** | **100.0** | **17,207** | **100.0** | **13,556** | **100.0** | **11,393** | **100.0** |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

17. The estimated total number of all individuals 15 years of age and older living in Canada in 2012 was about 28 million. In theory, this is the population that is followed over time at waves 2, 3 and 4.

The weighted sample compositions by aboriginal status over the four waves of LISA are shown in Table 5.2.2. The results show that the composition of the weighted samples in each of the four waves is approximately the same and illustrate that the weight adjustment strategy has reduced the effect of attrition on the aboriginal status variable. While the LISA weight adjustment and the calibration strategy does not actually incorporate aboriginal status, it is likely that this variable is correlated with another variable(s) that was used to adjust the weights.[18] This means the effect of attrition would be reduced or eliminated by the nonresponse weight adjustment and calibration strategy for any variable that is highly correlated with those used directly in the weight adjustments.

**Table 5.2.2**
**Weighted respondents' composition in the sample by aboriginal status, 2012 to 2018, Canada**

| | Wave 1 (2012) | | Wave 2 (2014) | | Wave 3 (2016) | | Wave 4 (2018) | |
|---|---|---|---|---|---|---|---|---|
| **Aboriginal status** | number | percent | number | percent | number | percent | number | percent |
| Aboriginal | 786,721 | 2.8 | 750,585 | 2.7 | 700,210 | 2.5 | 727,359 | 2.6 |
| Non-aboriginal | 27,368,876 | 97.0 | 27,372,906 | 97.2 | 27,421,679 | 97.3 | 27,395,970 | 97.2 |
| Missing | 65,516 | 0.2 | 50,903 | 0.2 | 52,507 | 0.2 | 51,066 | 0.2 |
| **Totals** | **28,221,114** | **100.0** | **28,174,395** | **100.0** | **28,174,396** | **100.0** | **28,174,394** | **100.0** |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2012, 2014, 2016, 2018).

As mentioned in section 4, there were a small number of respondents for which data were missing. Researchers are advised to determine the best approach to include the effect of the missing aboriginal status data in their analysis of LISA data.

## 5.3. Attrition ratios for age group from 2014 (wave 2) to 2018 (wave 4)

Another way to quantify or evaluate whether the nonresponse weight adjustment and calibration have reduced the effects of attrition is using attrition ratios. These measure the change in the number of respondents in a category between the first and subsequent waves of a longitudinal survey.

These ratios are defined using the following formula:

$$\mathrm{AR} = \frac{P_{g1}}{P_{gt}}, \quad t = 2,3,4$$

For an analytical variable of more than one category,

$P_{g1}$ = the number of respondents in category $g$ in the 2012 (wave 1) sample.

$P_{gt}$ = the number of respondents in same category in the sample at wave $t = 2,3,4$.

To evaluate the effectiveness of the LISA nonresponse weight adjustment and calibration, attrition ratios were computed with and without the LISA longitudinal weights for comparison purposes. This was done for both the age group and aboriginal status variables. Ratios using weights would indicate the effectiveness of the nonresponse weight adjustment and calibration strategy. If the attrition ratio using weights is one for all categories across all subsequent LISA waves $t = 2,3,4$, then attrition has been mitigated by the remedial measures. If the attrition ratio is greater than one for a given category, this category is under-represented in the sample. If the attrition ratio is less than one, then the category is over-represented in the sample.

The unweighted and weighted attrition ratios by age group from 2014 (wave 2) to 2018 (wave 4) are shown in Table 5.3.1 and Table 5.3.2 respectively. A comparison of the results in the two tables illustrates that nonresponse weight adjustment and calibration have essentially mitigated the impact of attrition on the analytical variable age group.

18. See Benhin, Frolova and Quesnel (2019) for additional details on the weighting strategy for LISA.

**Table 5.3.1**
**Unweighted attrition ratios by age group, 2014 to 2018, Canada**

| Age group | Wave 2 (2014) | Wave 3 (2016) | Wave 4 (2018) |
|---|---|---|---|
| | odds ratio | | |
| 15 to 24 | 1.19 | 1.40 | 1.64 |
| 25 to 34 | 1.14 | 1.19 | 1.21 |
| 35 to 44 | 1.01 | 1.03 | 1.05 |
| 45 to 54 | 0.96 | 0.95 | 0.96 |
| 55 to 64 | 0.90 | 0.85 | 0.80 |
| 65+ | 0.87 | 0.77 | 0.71 |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2014, 2016, 2018).

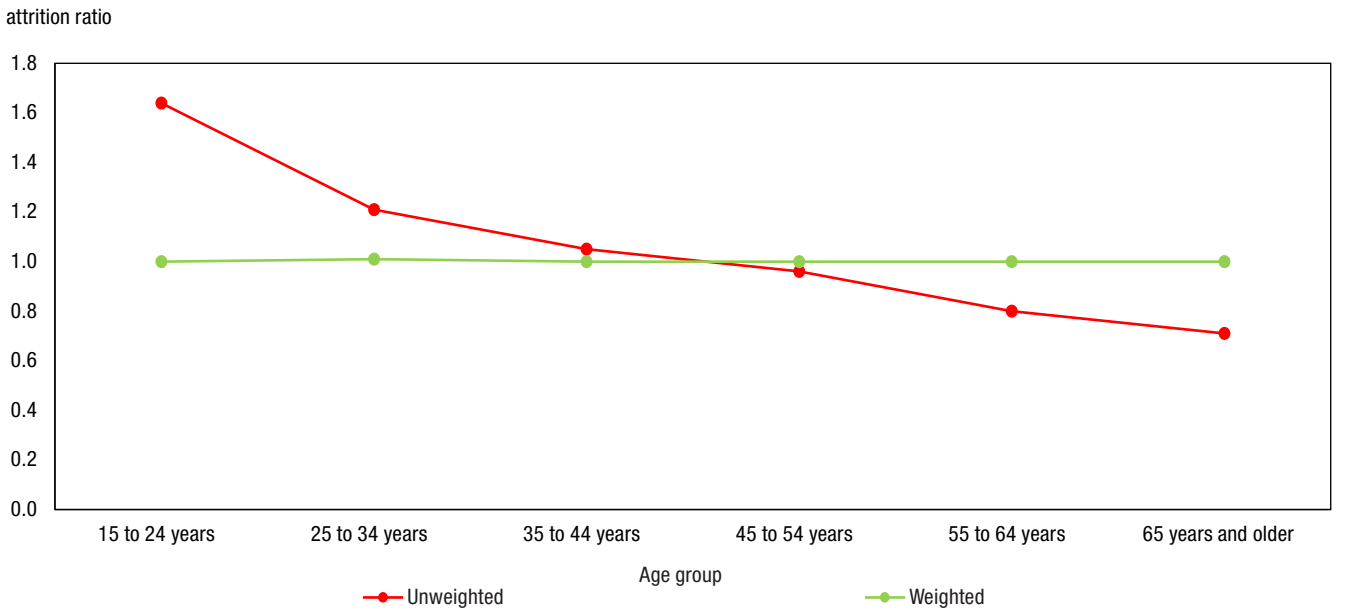
**Table 5.3.2**
**Weighted attrition ratios by age group, 2014 to 2018, Canada**

| Age group | Wave 2 (2014) | Wave 3 (2016) | Wave 4 (2018) |
|---|---|---|---|
| | odds ratio | | |
| 15 to 24 | 0.99 | 0.99 | 1.00 |
| 25 to 34 | 1.01 | 1.01 | 1.01 |
| 35 to 44 | 1.00 | 1.00 | 1.00 |
| 45 to 54 | 1.00 | 1.00 | 1.00 |
| 55 to 64 | 1.00 | 1.00 | 1.00 |
| 65+ | 1.00 | 1.00 | 1.00 |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2014, 2016, 2018).

Chart 5.3.1 shows the weighted and unweighted attrition ratios by age group for 2018 (wave 4). As can be observed, the weighted attrition ratio is a straight line with a value of 1 on the vertical axis. This presentation of the ratios provides a quick way for researchers to determine for a given analytical variable in LISA, the extent to which the nonresponse weight adjustment and calibration strategy has mitigated the effect of attrition.

**Chart 5.3.1**
**Attrition ratios by age group, 2018 (wave 4), Canada**




**Source:** Statistics Canada, Longitudinal and International Study of Adults (2018).

## 5.4. Attrition ratios for aboriginal status from 2014 (wave 2) to 2018 (wave 4)

The unweighted and weighted attrition ratios for aboriginal status for 2014 (wave 2) to 2018 (wave 4) are shown in Table 5.4.1 and Table 5.4.2 respectively. The results show that nonresponse weight adjustment and calibration have reduced the effect of attrition on the aboriginal status variable. The attrition ratios for aboriginals have been reduced in all survey waves. The unweighted and weighted attrition ratios for the 2018 (wave 4) data are also shown in Chart 5.4.1. Even though the variable aboriginal status was not directly incorporated in the nonresponse adjustment and calibration, this strategy still reduced the effect of attrition on this variable. The result suggests that the aboriginal status variable may be correlated with the variables that were used in the weight adjustment strategy. This means that attrition may be reduced or eliminated for any analytical variable that is not directly employed in the weight adjustment strategy but is highly correlated with those that are used.

**Table 5.4.1**
**Unweighted attrition ratios by aboriginal status, 2014 to 2018, Canada**

| | Wave 2 (2014) | Wave 3 (2016) | Wave 4 (2018) |
|---|---|---|---|
| **Aboriginal status** | odds ratio | | |
| Aboriginal | 1.17 | 1.27 | 1.26 |
| Non-aboriginal | 0.99 | 0.99 | 0.99 |
| Missing | 1.20 | 1.23 | 1.19 |

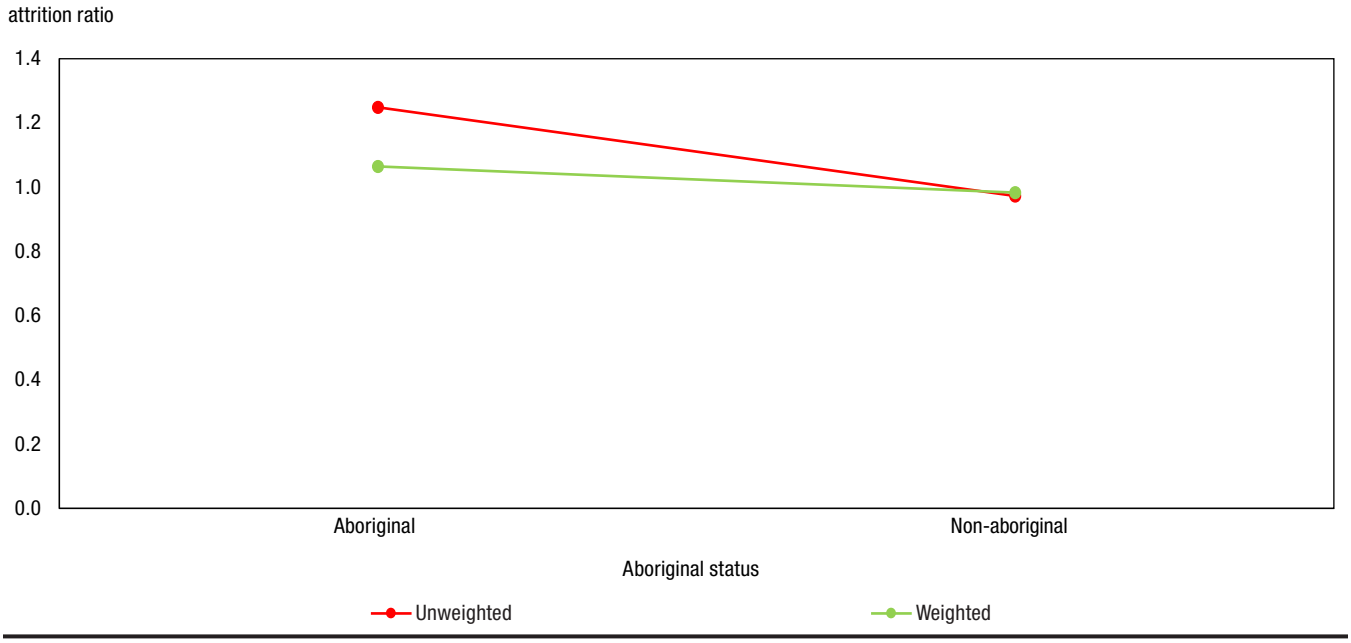
**Source:** Statistics Canada, Longitudinal and International Study of Adults (2014, 2016, 2018).

**Table 5.4.2**
**Weighted attrition ratios by aboriginal status, 2014 to 2018 , Canada**

| | Wave 2 (2014) | Wave 3 (2016) | Wave 4 (2018) |
|---|---|---|---|
| **Aboriginal status** | odds ratio | | |
| Aboriginal | 1.05 | 1.12 | 1.08 |
| Non-aboriginal | 1.00 | 1.00 | 1.00 |
| Missing | 1.28 | 1.25 | 1.28 |

**Source:** Statistics Canada, Longitudinal and International Study of Adults (2014, 2016, 2018).

**Chart 5.4.1**
**Attrition ratio by aboriginal status, 2018 (wave 4), Canada**




**Source:** Statistics Canada, Longitudinal and International Study of Adults (2018).

## 6. Limitations

Given the number of ways that attrition rates can be defined, caution should be used when comparing findings in this article to those identified in other studies of attrition. As referenced in section 3.2, the attrition rate primarily used in this study is the cumulative attrition rate which is calculated based on respondents who participated in all survey waves. This contrasts with how attrition rates are defined in other studies,[19] whereby the ratio numerator includes respondents who did not participate in each wave of the survey.

Due to differences in collection methodologies, comparisons of attrition rates between LISA and other longitudinal surveys were not explored extensively in this study. Data collection for LISA occurs every two years. Evaluating differences in attrition by wave can be problematic when comparisons are made to panel surveys that collect data on an annual basis, where respondent burden may be greater.[20] Furthermore, differences based on sampling design and population of interest can also complicate comparisons of attrition rates among longitudinal surveys.[20]

As mentioned in section 2.6, it has been argued that attrition in panel surveys has worsened over time.[20] This makes LISA distinctive from panel surveys such as the GSOEP and the BHPS, as both began collection before 1995. The LISA sample was not refreshed or topped-up over time. This has been done for other longitudinal surveys such as the German Socio-Economic Panel (SOEP)[21] and the Panel Study for Income Dynamics (PSID)[22] in order to increase the number of respondents.

While the LISA nonresponse weight adjustment and calibration strategy has been able to minimize the effect of attrition on the estimates as illustrated for the age group, aboriginal status and related variables, its effectiveness may vary depending on the variable used for analysis.

## 7. Conclusion

Attrition is one of the major challenges for the use of longitudinal survey data in research studies. It can lead to biased estimates when the characteristics of the respondents who drop out of the survey are significantly different from those who remain in the survey. It may also lead to reduced precision of estimates due to the reduction in the sample size over time.[20] Under the assumption that sample attrition in longitudinal surveys has become worse over time, there is an increased need for statistical techniques that can correct for the bias resulting from attrition.

In this report, some simple tools have been presented to help researchers to measure and evaluate the impact of attrition on their particular LISA analytical research variable(s). Attrition rates are commonly used to quantify attrition for any analytical variable of interest. The formulas to compute the attrition rates are provided in equations (1) and (2) of section 2.6.2 of this report. Equation (1) computes these rates at the sample level and equation (2) estimates these rates at the population level. Attrition rates at the sample level measure the proportion of the initial respondents' lost from the survey, while attrition rates at the population level measure the proportion of people in the pre-defined population at the start of the longitudinal survey who left the survey. In addition to overall survey attrition, comparisons of attrition for subsets of the population based on variables of interest can help in identifying the sources of attrition and in evaluating their potential impacts on survey estimates. As described in section 5 of this report, a logistic regression model can be used to investigate whether the difference in attrition rates between two or more categories for a variable of interest is statistically significant.

One step to address the potential negative impacts of attrition on survey estimates is to use a nonresponse weight adjustment and calibration strategy. Applying the final LISA survey weights generated using these methods in estimation can help ensure that survey estimates are unbiased and reliable. The effectiveness of the nonresponse weight adjustment and calibration methods can be measured based on comparisons of the sample composition and attrition ratios for a given analytical variable(s) of interest before and after weight adjustments.

19. See Statistics Canada (2010), Taylor (2018) or National Centre for Longitudinal Data (2018).
20. See Watson and Wooden (2009).
21. See Siegers et al. (2020).
22. See University of Michigan (2019).

## References


1. Benhin, E. (2018). Weighting for the Longitudinal and International Study of Adults (LISA), wave 3 – excluding bootstrap weights. *Statistics Canada, Internal document*.
2. Benhin, E., Olena Frolova and Angela Quesnel. (2019). Weighting for the Longitudinal and International Study of Adults (LISA), Wave 4 – excluding bootstrap weights. *Statistics Canada, Internal document*.
3. Binder, D. A. (1981). On the Variances of Asymptotically Normal Estimators from Complex Surveys. Survey Methodology 7:157–170.
4. Binder, D. A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. *International Statistical Review* 51:279–292.
5. Bocci, C. (2014). Weighting for the Longitudinal and International Survey of Adults (LISA), wave 1 (2011) - excluding bootstrap weights. *Statistics Canada, Internal document*.
6. Gossen, M. and Levesque, A. (2013). Longitudinal and International Study of Adults, Sample Design, Allocation and Selection. *Statistics Canada, Internal document*.
7. Lehtonen, R., and Pahkinen, E. (1995). *Practical Methods for Design and Analysis of Complex Surveys*. Chichester, UK: John Wiley & Sons.
8. Morel, J. G. (1989). Logistic Regression under Complex Survey Designs. *Survey Methodology* 15:203–223.
9. National Centre for Longitudinal Data (2018). Sample sizes and response rates for the Centre studies, Household, Income and Labour Dynamics in Australia (HILDA). https://www.dss.gov.au/about-the-department/national-centre-for-longitudinal-studies/growing-up-in-australia-the-longitudinal-study-of-australian-children-lsac/sample-sizes-and-response-rates-for-the-centre-studies
10. Rao, J. N. K., Wu, C. F. J., and Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. *Survey Methodology* 18:209–217.
11. Roberts, G., Rao, J. N. K., and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. *Biometrika* 74:1–12.
12. Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
13. Siegers, Rainer., Veronika Belcheva and Tobias Sibermann. (2020). SOEP-Core v35 – Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2018). SOEP Survey Papers 826: Series C. Berlin: DIW/SOEP.
14. Skinner, C. J., Holt, D., and Smith, T. M. F. (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons.
15. Statistics Canada (2010). National Population Health Survey, Household Component, Cycles 1 to 8 (1994/1995 to 2008/2009). Longitudinal Documentation.
16. Statistics Canada (2014). Longitudinal and International Study of Adults 2012 (Wave 1) User Guide. Longitudinal Documentation.
17. Statistics Canada (2020). Longitudinal and International Study of Adults 2018 (Wave 4) Data Quality Report. Longitudinal Documentation.
18. Statistics Canada (2020). Longitudinal and International Study of Adults 2018 (Wave 4) User Guide. Longitudinal Documentation.
19. Taylor, Marci Freed (ed.) with John Brice, Nick Buck and Elaine Prentice-Lane (2018). British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester, University of Essex.
20. Thoemmes, F. and Mohan, K. (2015). Graphical Representation of Missing Data Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 22: 631–642. Taylor & Francis Group, LLC.
21. University of Michigan. (2019). PSID Main Interview User Manual. Institute for Social Research. *Release 2019*.

22. Watson, N and Wooden, M. (2009). *Identifying Factors Affecting Longitudinal Survey Response*. Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

23. Wolter, K. M. (2007). *Introduction to Variance Estimation*. 2nd ed. New York: Springer.

24. Wooldridge, Jeffrey M. (2010). Econometric analysis of cross section and panel data. MIT Press.

## Appendix A: Logistic Regression for the analyses of attrition

Logistic regression analysis is used to investigate the relationship between a response outcome with categorical data and a set of explanatory variables. The discrete responses may be of three types: binary, ordinal or nominal. A binary response is one where there are only two possible response outcomes. For example, at a given wave/cycle a sample unit (person) could leave a survey or remain in a survey. An ordinal response is one where there are more than two possible response outcomes, as well as a measure of the degree of relative significance of these outcomes. For example, a sample unit may report having mild, moderate or severe pain. Finally, a nominal response is one where there are more than two possible outcomes but there is no measure of degree of importance or ranking among the response outcomes. For example, a sample unit may report being Black African, East Asian, Southeast Asian or white.

For the analysis of the effect of attrition on the data from Longitudinal and International Study of Adults (LISA), the response outcome of interest is binary – whether for example, at wave 4 (2018), a sample unit left the survey or remained in the survey. Suppose $x$ represents an explanatory variable (e.g., aboriginal status) and $\pi$ is the probability of a sample unit leaving the study at wave 4, then the linear logistic model is:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + x\beta$$

where $\alpha$ is the intercept parameter and $\beta$ is the regression (slope) parameter. The latter measures the corresponding increase in the logarithm of the outcome variable for every unit of increase in the value of the independent variable $x$.

### Estimation of the parameters

Using LISA wave 1 (2012) and wave 4 (2018) data, the LISA respondent person-level (RPW) weights at wave 1 and the corresponding bootstrap weights, estimation and variance estimation of the regression parameter and intercept parameter were done using the maximum likelihood estimation method and employing the SAS software procedure SURVEYLOGISTIC.

### Interpretation of the Estimates via odds ratio

Consider a dichotomous explanatory variable $x$ which takes the value 1 for if the sample unit is an aboriginal and value 0 if the sample unit is not an aboriginal.

The odds of a sample unit leaving the survey if the unit is an aboriginal is given by

$$\text{odds}(1) = \text{odds}(\text{aboriginal}) = \frac{\text{P}(\text{attrite} / x = 1(\text{aboriginal}))}{\text{P}(\text{does not attrite} / x = 1(\text{aboriginal}))}$$

The odds of a sample unit leaving the survey if the unit is a not an aboriginal is given by

$$\text{odds}(0) = \text{odds}(\text{not an aboriginal}) = \frac{\text{P}(\text{attrite} / x = 0(\text{not an aboriginal}))}{\text{P}(\text{does not attrite} / x = 0(\text{not an aboriginal}))}$$

With respect to this example, the odds ratio (OR) is defined as the ratio of the odds of a sample unit leaving the survey given that the unit is an aboriginal to the odds of a sample unit leaving the survey given that the unit is a not an aboriginal. This ratio is represented by the following formula:

$$\mathrm{OR}(x=1)=\mathrm{OR}(\text{aboriginal})=\frac{\text{odds}(x=1(\text{aboriginal}))}{\text{odds}(x=0(\text{not an aboriginal}))}=\begin{cases}>1\\=1\\<1\end{cases}$$

For example if the OR = 2, this means that the odds of a sample unit leaving the study given that the unit is an aboriginal is twice the odds of the sample unit leaving the study if he/she was not an aboriginal. In other words, this result indicates that aboriginals are two times more likely to leave the study than non-aboriginals.

There are other options to consider for multi-category explanatory variables, where $x$ takes the values 1, 2, 3 or 4. However, these are not discussed in this report.

## Appendix B: Survey weighting methodology[23]

LISA produces four sets of survey weights: 1) the responding person weight (AWRPW) for longitudinal analyses, 2) the Programme for the International Assessment of Adult Competencies (PIAAC) subsample person weight (AWIRPW) for longitudinal analyses, 3) the enumerated persons weight (EPW) for wave-t analyses, and 4) the responding person weight (RPW) for wave-t analyses. Both the EPW and the RPW are wave-specific weights that can be used to perform cross-sectional analysis. If multiple waves are used for analysis, it is recommended that the AWRPW or the AWIRPW waves be used.

### All waves responding person weights

The AWRPW are calculated only for individuals in the original LISA sample who responded in each wave of the survey up to and including the most recent wave. Children less than 15 years of age and non-responding adults are not assigned an AWRPW. The weighted responding individuals represent the 2012 Census population living in the 10 provinces aged 15 years and older at the time of LISA 2012 (Wave 1).

### All waves PIAAC responding person weights

The AWIRPW are calculated for all individuals within the responding LISA 2012 (Wave 1) households who were selected for, and responded to, the PIAAC survey. The PIAAC weighted individuals represent the 2012 Census population living in the 10 provinces aged 16 to 65 years at the time of LISA 2012 (Wave 1).

### Enumerated person weights

The EPW are calculated for all individuals in responding households at wave-t, including children and non-responding individuals. Weighted individuals, using the EPW, will represent the 2012 Census population living in the 10 provinces at 2012 (Wave 1), plus their descendants by birth or adoption.

### Responding person weights

The RPW are calculated only for respondents in wave t of the LISA survey. Children less than 15 years of age and non-responding adults are not assigned an RPW. Temporary subsample members are also not assigned an RPW, even if they responded to the survey.

Weighted individuals with an RPW will represent the evolution of the 2012 Census population living in the 10 provinces at 2012 (Wave 1), which includes individuals who were less than 15 years of age in 2012 but turned 15 in wave-t.

23. See Statistics Canada (2020), LISA 2018 user guide.