

THE INTRICACIES AND IDIOSYNCRASIES OF THE LONGITUDINAL AND INTERNATIONAL STUDY OF ADULTS (LISA)

Angela Quesnel¹, Emmanuel Benhin and Joseph Duggan

ABSTRACT

LISA is a longitudinal survey that collects information from respondents across Canada, within a household construct, on their jobs, education, health and family, and on how changes in these areas affect different aspects of their lives. Statistics Canada collects this information from the same Canadians every two years from 2012 to the present; currently, we are collecting the fourth wave of data. Our unique sample definition and following rules tasked us as survey methodologists into devising an interesting plan for the critical point of weighting for Wave 2 and going forward. All aspects of the following rules and weighting will be discussed, including how the weighting plan evolved from the initial wave to Wave 3, when the pattern is set for perpetuity. Challenges and issues encountered will be discussed and the lessons learned will be shared.

KEY WORDS: Weighting, Longitudinal Survey, Following Rules, Survey Methods.

RÉSUMÉ

L'étude longitudinale et internationale des adultes (ELIA) est une enquête longitudinale qui recueille de l'information auprès des mêmes répondants (des ménages) partout au Canada, sur leur emploi, leur éducation, leur santé et leur famille, et sur la façon dont les changements dans ces domaines ont une incidence sur les différents aspects de leur vie. De 2012 à aujourd'hui, Statistique Canada a recueilli cette information auprès des Canadiens tous les deux ans; actuellement, elle recueille la quatrième vague de données. En tant que spécialistes en méthodologie d'enquête, notre définition de l'échantillon unique et les règles qui l'accompagnent nous ont chargés d'élaborer un plan intéressant de point critique de pondération de la deuxième vague et des vagues suivantes. Nous examinerons tous les aspects des règles et de la pondération suivantes, y compris l'évolution du plan de pondération, de la vague initiale à la troisième vague lorsque le modèle est établi sur une durée indéfinie. Nous traiterons des défis et des problèmes auxquels nous faisons face et communiquerons les leçons que nous avons apprises.

MOTS CLÉS : Pondération; enquête longitudinale; règles de suivi; méthodes d'enquêtes.

1. INTRODUCTION

1.1 What LISA Studies

LISA was sponsored by Employment and Social Development Canada (ESDC) as a long duration survey to examine the social and economic well-being of the adult population of Canada for the advancement of life-course focused research. The study collects information every two years from people across Canada about their jobs, education, health and family to provide evidence that guides decisions about economic, social, health and educational programs. The first wave of LISA collected data from November 2011 to June 2012 and was integrated with the Programme for the International Assessment of Adult Competencies (PIAAC). Data collection for Wave 2 was conducted in the first six months of 2014, and has been repeated every two years thereafter.

1.2 Motivation to Revisit and Revise the Weighting Plan for LISA

The initial plan was to produce six weights for each subsequent wave after the initial wave, and these six were produced for Wave 2. In the course of the LISA post-survey evaluation, it was discovered that the calibration step of the weighting process was incorrectly implemented. It was also determined that there was the need to re-evaluate the implemented wave 2

¹ All authors, Social Statistics Methods Division, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6; angela.quesnel@canada.ca

weighting strategy to reduce its complexity and to facilitate its sustainability and maintenance for future waves. As a result of recommendations from an internal technical committee and several discussions with the Income Statistics Division, a revised weighting strategy for Wave 2 was developed and implemented. The number of weights was reduced from six to four. The simplified strategy meets the survey's core requirements as a longitudinal study and also addresses a large portion of user needs by providing two wave-specific weights that may be used for cross-sectional type analyses.

Section 2 of this paper provides some background information including the sample design, section 3 focuses on the weighting strategy and section 4 discusses some challenges of the weighting and provides some conclusions.

2. BACKGROUND

2.1 Design

The LISA covers the population living in the ten provinces as of the first wave of the survey (November 1, 2011 to June 27, 2012), plus their future descendants (children born or adopted). Excluded from the survey's coverage were those who at the time of Wave 1 were: living on reserves and 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, for example, inmates of correctional facilities and chronic care patients living in hospitals and nursing homes; persons living in other collective dwellings. All together these exclusions represent approximately 2% of the population.

The LISA sample has a stratified multi-stage, multi-phase design. The LISA survey frame was constructed from dwellings containing households that responded to the 2011 Census of Population and that were not selected in the National Household Survey (NHS), which was conducted at the same time as LISA. It was necessary to restrict the frame to Census respondent dwellings because the household composition was required by the sampling plan.

In Wave 1, LISA was integrated with the Programme for the International Assessment of Adult Competencies (PIAAC). The two surveys shared a portion of their samples and the data collection activities were integrated. It is important to note that while sampling and collection activities were shared by the two surveys, PIAAC and LISA, each survey team processed their own data separately. The PIAAC was a cross-sectional survey and its target population differed slightly from LISA in that PIAAC included the territories and only considered people aged 16 to 65 years. The portion of the LISA sample that shared units with PIAAC is known as the ISA portion (for the International Study of Adults).

At any given wave, LISA only interviews respondents 15 years of age and older, but asks certain health and childcare related questions to a person most knowledgeable (PMK) about the younger children and asks certain financial questions to a PMK about the household's finances. The ISA interview targeted the selected person and administered tests in aptitude only in the first wave of the survey, but these respondents are interviewed using the LISA questionnaire in subsequent waves. The ISA provided valuable additional information from literacy, numeracy and problem solving assessments for some LISA respondents in the initial wave of collection, and as such at every wave a separate ISA longitudinal weight will be produced for ISA sample members along with a longitudinal weight for the LISA sample members.

2.2 Sample Membership Definitions and Following Rules

The longitudinal sample for LISA consists of all individuals in responding households at Wave 1. A responding household refers to a household that has at least one person responding to the survey. All persons in the responding households in Wave 1 are referred to as (original) permanent sample members, denoted by PSM1. In subsequent waves, children born to or adopted by a PSM1 since Wave 1 also become permanent sample members, denoted by PSM2. The PSM2s are given initial weights in the wave they are first identified, but are not interviewed until they turn 15 years of age. Also at any given wave, joiners to a household who are not classified as PSM2s are referred to as temporary sample members, denoted by TSM. Data are collected from TSMs who are 15 years of age or older at the time of the interview. Once a person has been identified as a PSM or TSM, they are assigned a unique identifier that follows them for the life of the survey and they cannot change status.

The PSM1s and the PSM2s are followed over the life of the survey for data collection. Data are collected only from those TSMs who are at least 15 years of age and live with a PSM at the time of the survey.

2.3 Types of Populations

In longitudinal surveys, it is a challenge to clearly define the study population at wave t . In general, at wave t there are a total of 2^t-1 potential populations, t potential cross-sectional populations and 2^t-t-1 potential longitudinal populations. For example at Wave 3, there are a total of seven potential populations: three potential cross-sectional populations at Wave 1, Wave 2 and Wave 3 respectively, and four potential longitudinal populations including all-waves (Waves 1, 2 and 3; longitudinal population with reference at Wave 1), Waves 1 and 2, Waves 1 and 3, and Waves 2 and 3. The number of potential populations increases exponentially as the number of waves in a longitudinal survey increases. In the context of survey sampling, unique sampling weights are created for each of the possible potential populations and so the number of sets of sampling weights also increases exponentially.

3. WEIGHTING

3.1 Types of Weights Produced

LISA is a sample survey and so the data collected are associated with survey sampling weights. These weights are required in order to produce estimates that may be used to make statistical inferences about characteristics of a well-defined population. The LISA longitudinal sample size was determined in Wave 1 and consists of all individuals living in the 11,458 responding households. These 32,133 individuals include responding adults, nonresponding adults and children less than 15 years of age. These individuals are considered in scope for the lifetime of the survey and constitute the LISA longitudinal sample.

For each of the populations described in section 2.3, there is a corresponding set of sampling weights (and associated coordinated bootstrap weights) which are produced to enable analysts to perform reliable statistical inferences concerning these populations. The all-waves weights are produced to allow for inferences on characteristics of the all-waves (longitudinal) populations. These types of weights may be useful for inferences about changes in people's characteristics from Wave 1 through to all subsequent waves. One can also estimate the number of people in the cohort that are deceased, moved outside of Canada or have been institutionalized.

The wave t weights are produced to allow for reliable statistical inferences on characteristics of the wave t populations. These types of weights may be useful for inferences about characteristics of the population at a specific wave.

The all-waves weights are calibrated to always represent the target population at Wave 1. The wave t weights are calibrated to represent the Wave 1 population plus the population of new children born or adopted since Wave 1. The wave t weights allow users to approximate estimates of change between any two waves, but will not indicate changes in the overall Canadian population. This is because certain elements, most notably new immigrants, are not accounted for in subsequent waves.

3.2 Wave 1 Cross-sectional Weights

In the initial wave, the sampled population represents Canadians living in private households in the 10 provinces, with the few exceptions mentioned above. Three sets of weights were produced to facilitate analysis of this population, corresponding to data collected for LISA and ISA and their intended target populations.

Enumerated Person Weight (EPW): This weight was created as an initial weight for each and every PSM1. It was also instrumental for moving from the initial sample weights which were at the household level to create person level weights. This weight applies to all enumerated persons in the longitudinal sample. In short, the design weight was first adjusted for nonresponse at the dwelling level for each sample type separately. Next, the person level weight was constructed by assigning the same nonresponse-adjusted dwelling weight to every member of a household. The EPW resulted from calibrating the person level weight on regions and classes based on age, sex and household size.

Responding Person Weight (RPW): Constructed for all PSM1s aged 15 and over who responded to the initial LISA wave, these sampled units form the basis of the longitudinal (all waves) sample. The weight used for this calculation was the pre-calibration weight of an individual adjusted for household nonresponse. The person level nonresponse adjustment was simply the weighted inverse response rate of adults within a group, where the group is a cross-classification of sex, marital

status and age category. The RPW resulted from calibrating the person level weight on regions and classes based on age, sex and household size.

ISA Responding Person Weights (IRPW): Constructed for all ISA respondents, these sampled units form the basis of the longitudinal (all waves) ISA sample. The person level nonresponse adjustment was simply the weighted inverse response rate of adults within a group, where the group is a cross-classification of sex, marital status and age category. The weight used for this calculation was the weight of an individual adjusted for household nonresponse. The IRPW resulted from calibrating the person level weight on regions and classes based on age, sex, household size and education level.

3.3 Wave 2 Proposed Weights

The weights for Wave 1 of the survey were quite clear as they were cross-sectional. Weights were needed to represent the population at that wave for the sample that responded to the LISA portion of the survey and also weights were needed for the sample that responded to the ISA portion. In order to create person level weights it is always necessary to get appropriate household level weights because the frame was at that level, and respondents are contacted and interviewed as part of a household construct. Because the plan was to follow this population and their descendants, weights for everyone in those households were required, which became enumerated person weights and the initial weights for all people in LISA.

Thus, for Wave 2 there were many potential populations and combinations of waves for which weights could be produced. One of the key initiatives undertaken as part of the evaluation was to review the procedures of longstanding longitudinal surveys of other countries and compare them to LISA objectives (not documented here due to page limits). It was quickly confirmed that there were no universally accepted methods for defining the PSMs, determining the following rules or tackling these weights.

At the outset the intent was to produce six sets of weights for each subsequent wave of LISA; three of these would correspond to the cross-sectional weights from Wave 1 and the other three would be longitudinal versions of the first three weights. However, through extensive evaluation and internal consultation, it was decided not to produce a longitudinal all waves enumerated person weight due to its limited utility and the lack of a longitudinal construct at the household level, nor to produce a wave t “cross-sectional like” ISA responding person weight, as the competency assessments were only collected in the first wave. These choices resulted in a more manageable and defensible four sets of weights.

3.4 Wave 2 Final Production Weights

The pattern for each new wave is now set using these four weights for Wave 2 and all subsequent waves. The details for the longitudinal and wave t weights are outlined below.

3.4.1 LISA All Waves (Longitudinal) Weights

All Waves Responding Person Weight (AWRPW): This weight is constructed for all respondents aged 15 and older, as of the initial survey cycle, from responding households that responded in every wave. Only PSM1s that were at least 15 years old in Wave 1 and who responded in Wave 1 and in each following wave are included. The longitudinal weight accounts for people who have died, moved outside of Canada or who have been institutionalized. This weight is referred to as an “all-waves” weight since the weighted sample has been accounted for in each subsequent wave since Wave 1.

All Waves ISA Responding Person Weight (AWIRPW): This weight is created for a subset of the above population who were 16 to 65 years of age as of the initial survey cycle, and who answered the PIAAC component test questionnaire on literacy, numeracy or other functional competencies. Note that the ISA longitudinal weights are calculated in much the same way as the AWRPW, but only for those individuals in the ISA portion in Wave 1.

3.4.2 LISA Wave t Weights

Wave t weights reflect the population of PSM1s and PSM2s. It is important to note that these weights are not at all cross-sectional in nature. They reflect the evolution of the initial LISA population defined in Wave 1 through time. People who have died, moved outside of Canada or who have been institutionalized are considered out-of-scope for wave t . TSMs are again not weighted but provide contextual information to the household.

Wave t Enumerated Person Weight (EPW- t): These represent the evolution of the LISA population since the initial wave, at wave t , which includes children born or adopted since Wave 1 to the LISA longitudinal population. This weight is the key to getting initial weights for the PSM2s. Enumerated or rostered person weights are given to all PSM1s and PSM2s in responding households at the time of the current wave. This includes assigning an initial weight to PSM2s (this is the only way they receive an initial weight for weighting in future waves). The EPW- t represents the initial population's evolution at wave t , and is not a cross-sectional weight at wave t , so for example excludes new immigrants since Wave 1 by definition.

Wave t Responding Person Weight (RPW- t): These are applied to respondents to the LISA questionnaire at wave t . Similar to the EPW- t , wave t responding person weights are given to all responding PSMs (aged 15+) in responding households at the time of the current wave. The RPW- t represents the initial population's evolution at wave t , and is not a cross-sectional weight at wave t , so for example excludes new immigrants since Wave 1 by definition. These weights maximize the analytical potential for the data collected and facilitate analysis between waves.

4. CHALLENGES AND CONCLUSIONS

4.1 Populations of Study

The only longitudinal populations for which LISA data may be used for reliable statistical inferences are the populations in Wave 1 that are followed every two years (without a break) over the life of the survey. Furthermore, the LISA data may not be used to make statistical inferences about the Canadian population in Wave 2 (2014) or Wave 3 (2016). This is because the LISA data for 2014 or 2016 does not capture all the changes in the Canadian population since 2012 (the start of the LISA survey); in particular, new immigrants since 2012. However, the wave t weights can be used as an approximation to study the changes between any two waves. It is important to examine all aspects of the survey completely to implement a clear and sound weighting strategy for both the longitudinal and the wave t weights, from Wave 2 onwards. As LISA is longitudinal, not only can its data influence new programs, but it can also track how events from the past are impacting Canadians today and into the future.

4.2 Other Longitudinal Surveys

At several points in the development of the survey the directions other countries had taken for their longitudinal panel surveys was reviewed and evaluated those options against the LISA survey goals. Through these evaluations it was discovered that most other household surveys of this type face similar issues that could lead to inconsistencies, inadequacies and complexities. The subsequent survey review, specifically the initially proposed Wave 2 weights, led to the creation of a revised weighting plan. This streamlined the proposed weights, clarified their use and still enabled use of all collected data, for a reduced and rationalized set of weights.

4.3 Future Work

Collection of LISA wave 4 has been completed and processing of the data has begun. An assessment is being completed to determine the feasibility of more cycles after Wave 5 and potentially planning a top-up sample in Wave 6. Similar to other panel surveys, a top-up sample is being considered for LISA; however, there are many factors to consider. The way the survey is designed and executed means carefully evaluating the need of such a sample in the context of the estimates. Although you can top-up the sample to increase it, it does little to help the power of the estimates longitudinally, as those units are effectively a brand new longitudinal cohort. With the top-up, it is only at the overlapping wave that the samples can be combined to produce a cross-sectional estimate.

4.4 Important Analytical Uses

LISA has immense analytical potential and analysts have been very excited with the prospect of these data since the survey's inception. There is content that is collected across all waves and content that has been collected only for specific wave(s). To date, over 85 research projects are using LISA data to look at: the factors influencing employment insurance utilization; impacts of a growing number of older workers on job availability for younger workers; what can be done to reduce the gender wage gap; and, how to encourage girls early in their school careers to pursue careers in typically male-dominated fields, to name a few.

ACKNOWLEDGMENTS

The authors would like to thank Olena Frolova and the Income Statistics Division team with whom we closely worked to unravel the challenges of the intricacies and idiosyncrasies of LISA.

REFERENCES

Statistics Canada. (2018). *Revised Data Quality Report: Longitudinal and International Study of Adults, Wave 2 re-release. Internal Document.*

Statistics Canada. (2017). *Revised Weighting Strategy. Internal Document.*

Statistics Canada. (2014). *Data Quality Report: Longitudinal and International Study of Adults, Wave 1. Internal Document.*

Statistics Canada. (2011). *Weighting for the Longitudinal International Survey of Adults (LISA) Wave 1. Internal Document.*