Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2003: Challenges in Survey Taking for the Next Decade

2003

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

TRACING AND NON-RESPONSE ADJUSTMENT FOR THE LONGITUDINAL SURVEY OF IMMIGRANTS TO CANADA

M. Simard, T. Leesti and J. Denis[1]

ABSTRACT

The Longitudinal Survey of Immigrants to Canada (LSIC) is a survey designed to study the process by which new immigrants integrate into Canadian society. The survey uses a longitudinal design, with selected immigrants being interviewed at three points in time: approximately six months (Wave 1), two years (Wave 2), and four years (Wave 3) after landing. Given that recent immigrants are characteristically a highly mobile population, tracing for Wave 1 was particularly challenging. As the target population was very new to Canada, conventional administrative sources proved to be of limited use for tracing. For these reasons, the proportion of unresolved units was generally higher for LSIC in Wave 1 than other Statistics Canada surveys (approximately 28%). This paper will discuss the work that has been done on LSIC with regard to tracing respondents, as well as studying and adjusting for non-response. It will highlight challenges faced in locating recent immigrants, as well as the strategies implemented to increase response rates. In addition, the paper presents the chosen method for adjusting for non-response: a model-assisted technique based on the approach proposed by Eltinge-Yanseneh to define adjustment classes. This method incorporates two distinct adjustments - one for untraceable cases and one for non-respondents.

KEYWORDS: Logistic Regression; Model-Assisted; Non-Response; Tracing.

1. INTRODUCTION

The Longitudinal Survey of Immigrants to Canada (LSIC) is designed to study how recent immigrants adjust to life in Canada. New arrivals will be interviewed three times during their first four years in Canada – approximately six months, two years and four years after their arrival - allowing for the creation of a dynamic picture of their experiences. This is the first national survey conducted with the recent immigrant population since 1970.

Recent immigrants are a highly mobile population and, as a result, are difficult to trace and subsequently interview. Hence, finding the appropriate methodology for weighting is a challenge. This paper describes the design of the survey, the challenges of tracing newly-arrived immigrants and the results of the first wave of data collection. It also outlines the strategy used to compensate for non-response: a three-step model-based re-weighting method.

2. SURVEY BACKGROUND

2.1 Survey objectives and content

There exists a growing need for information on recent immigrants to Canada, specifically the adaptation process, the factors that affect their adaptation and the services used by immigrants to facilitate the process. While full adaptation may take several generations to achieve, the LSIC is designed to examine the process during the critical first four years of settlement, during which newcomers establish economic, social and cultural ties to Canadian society.

Respondents are asked questions on various aspects of their life, ranging from their reasons for choosing to relocate to Canada to problems encountered in finding housing, employment and education. It also collects general

[1] M. Simard, T. Leesti and J. Denis, Statistics Canada, Tunney's Pasture, 120 Parkdale Avenue, Ottawa, Ontario, Canada, K1A 0T6, johanne.denis@statcan.ca; michelle.simard@statcan.ca; tracey.leesti@statcan.ca.

information on their children's education and health. Questions pertaining to the ability to access services have been incorporated throughout the questionnaire.

2.2 Survey design

The survey employs a longitudinal design, with recent immigrants being interviewed three times: approximately six months, two years, and four years after landing in Canada[2]. The sample design has been developed using a monotonic or "funnel-shaped" approach. Only immigrants who responded to the previous wave of interviewing will be contacted in the subsequent wave. This approach was chosen because of the nature of the survey and its analytical objectives.

The survey collects information on perceptions, values and attitudes at specific points in time, in order to assess the immigrant's integration during their initial years in Canada. If data were collected only once (i.e., during the fourth year in Canada), significant recall and response errors could be encountered. Furthermore, to facilitate a complete study of the immigrant's adaptation, the full range of longitudinal data must be obtained from each longitudinal respondent.

2.3 Target population and sample

The target population for the survey consists of immigrants who meet the following criteria:

- landed immigrants arriving from outside Canada;
- arrived in Canada between October 1, 2000 and September 30, 2001;
- were age 15 or older at the time of landing.

Immigrants who applied and landed from within Canada are excluded from the survey. These people may have been in Canada for a considerable length of time before officially "landing" and would therefore likely demonstrate different adaptation characteristics to those recently arrived in Canada. Approximately 21,000 of the 165,000 immigrants aged 15 and over who settled in Canada between October 2000 and September 2001 were selected to participate in the survey.

3. SAMPLE DESIGN

3.1 Sampling frame

The sampling frame for the LSIC is Citizenship and Immigration Canada's Field Operation Support System (FOSS); an administrative database containing information on all persons who have arrived in Canada, including those who have arrived as immigrants. The FOSS is a rich database that includes various characteristics of each immigrant that can be used for survey design purposes, and also provides a very good source of auxiliary information. Some of the key variables of interest found on the FOSS include: age, sex, mother tongue, country of origin, knowledge of English and/or French, class of immigrant, date of landing and intended province of destination in Canada.

Statistics Canada receives detailed information from FOSS on each immigrant landing during the survey reference period (October 2000 to September 2001) two months after the reference month. This allows for the sampling frame to be built month after month by simply adding new monthly landings.

[2] The first cycle of interviews were conducted between April 2001 and March 2002. The second cycle of interviews began in December 2002 and will continue for about a year. The third cycle of interviewing is scheduled to begin in October 2004.

3.2 Sample requirements

The sample can be divided into two components - the core sample and the additional samples. The core sample represents the target population, while the additional samples target specific sub-populations. Based on various requirements of federal and provincial government departments, these specific sub-populations were determined by analyzing the expected sample allocation at wave 3. The following subgroups have been over-sampled:

- government sponsored refugees;
- refugees other than government sponsored;
- contractor and investor immigrants ("Economic-Business");
- family immigrants in British Columbia;
- overall immigrants in Alberta; and
- economic immigrants in Quebec ("Economic-Skilled" and "Economic-Business").

The sample is created using a two-stage stratified sampling method. The first stage involves the selection of Immigrating Units (IUs)[3] using a probability proportional to size (PPS) of the IU method. The second stage involves the selection of one IU member within each selected IU. The selected member of the IU is the longitudinal respondent (LR) who participates in the survey. Only the LR will be followed for the survey and no interviews will be conducted with other members of the IU.

It was determined that for the core sample, 5,000 completed interviews at wave 3 would produce reliable estimates[4] at the national level, the provincial level where the in-flow of immigrants is the most significant (Québec, Ontario and British Columbia) and for certain classes of immigrant (family and economic classes). It would also be possible to obtain reliable estimates for other combinations of provinces and classes of immigrant. Taking into account the requirements for additional samples previously outlined, the expected overall number of completed interviews at wave 3 is 5,755 immigrants. To compute the required sample size in wave 1, a backward method, based on attrition hypotheses coming from different sources, was applied. Using this method, the required sample size for wave 1 was 20,322 units.

4. DATA COLLECTION METHODS

Interviews for wave 1 were conducted between April 2001 and March 2002. Given various operational constraints (field capacity, difficulty to trace the immigrants), each month of sample remained in the field for three months. Interviews averaged approximately 90 minutes and were conducted in one of fifteen different languages, including: English, French, Mandarin, Cantonese, Punjabi, Farsi/Dari, Arabic, Spanish, Russian, Serbo-Croatian, Urdu, Korean, Tamil, Tagalog and Gujarati. Approximately 70% of interviews were conducted in person with the remaining 30% done by telephone. All interviews were conducted with the use of computer assisted interviewing technology. Interviews were not conducted in Yukon, North West Territories and Nunavut given the high collection costs and the fact that very few immigrants established themselves in these areas. Interviews were conducted in all Census Metropolitan Areas and non-remote areas across Canada.

5. TRACING

Tracing and contacting new immigrants is particularly challenging, as recent immigrants tend to be a highly mobile population - a point that was illustrated during the LSIC pilot test, conducted in the Spring of 1997. During the pilot it was found that almost half the population had moved at least once during the first 6 months in Canada.

[3] An immigrating unit consists of all individuals who applied to immigrate to Canada on the same Visa form.
[4] By reliable estimates we mean being able to estimate a minimal proportion of 10 % with a coefficient of variation of 16.5 %. A minimum of 450 responding immigrants is necessary to meet this requirement.

5.1 Tracing challenges

One of the most important challenges faced when trying to locate potential respondents for wave 1 was a lack of initial contact information on the survey frame. Upon arrival in Canada, immigrants are only required to provide their intended province of destination. For some immigrants, this was the only contact information available, and as happened, the intended province of destination was not necessarily where the immigrant eventually settled.

In addition to this, at the time of the wave 1 interview, the LSIC target population had only been in Canada for six months. As a result, traditional administrative sources that other surveys may seek permission to access for tracing – such as Revenue Canada's T1 and T4 forms, Child Tax Benefits, Unemployment Insurance or Address Registers – were not available for the first wave of tracing, as new immigrants were not yet included in these files.

5.2 Tracing strategies

In order to address some of the challenges faced in locating respondents, a number of strategies were implemented. For example, from the outset of the survey, the monthly tracing activities within each region were closely monitored. Any potential problems were immediately identified and appropriate solutions were proposed.

In addition, specialized tracing teams dedicated solely to LSIC were put in place in selected regional offices. The co-ordination of the tracing activities, in turn, freed up time for the interviewers to conduct interviews, as well as allowed the tracing unit to develop an expertise with respect to tracing this particular population. And while much of the tracing was being addressed by the tracing unit, it was recognized that some level of tracing was still being sent back to the field to be handled by specific interviewers. In order to compensate for this, the monthly collection period was extended to allow additional time for locating and contacting potential respondents.

Finally, as mentioned earlier, the contact address information on the survey frame was not always complete. A key activity at head office was to update the sample file with as much address information as possible, using timely trace sources that were relevant to this particular population and that might yield some relevant contact information; all before sending the file out to the field for collection. These sources included: the LSIC contact questionnaire; provincial Ministry of Health address files (for all but one province); administrative files from Citizenship and Immigration Canada and telephone files.

The LSIC contact questionnaire was designed to help locate potential respondents after they arrived in Canada. These forms were provided to all immigrants at the time their landing visa was issued overseas – the forms were then collected by Immigration Officers at the Canadian Ports of Entry. On the forms the immigrants provided an intended address in Canada (if known) as well as the address of a contact person in Canada.

During the development of LSIC, it was ascertained that the provincial Ministry of Health Address files would be the best tracing source, as immigrants could apply for a health card within three months of their arrival. Access to these files, however, was only granted with consent from the immigrant. As a result, a consent question was added to the LSIC contact questionnaire which asked immigrants to authorize Statistics Canada to access information held by provincial health organizations for tracing purposes only.

While the consent from immigrants was high (79%), the rate of return of the questionnaires was somewhat low and, as a result, health card addresses were obtained for around 35% of the sample. However, in subsequent analysis of the quality of tracing sources, the health files proved to be one of the most reliable sources, with 77% of cases yielding relevant address information.

Once all potential tracing information had been compiled, each trace source for a selected immigrant was prioritized and the file was sent out to the region for collection. Potential contact information was provided for approximately 75% of the sample, with an average of 5 contacts provided for each immigrant.

6. COLLECTION OUTCOME

6.1 Results and findings

In the end, 59.2% of the original sample was coded as a responding unit, a somewhat better result than the 50% expected at the planning stage. Table 1 presents the possible outcomes of collection as they were expected and as they were observed for wave 1. During collection, there were four possible classifications for a selected immigrant; *respondent, non-respondent, out-of-scope,* and *unresolved*. The first three categories were defined as *resolved* cases as the immigrant had a known status. They resulted in a contact with the immigrant or with someone who was able to confirm their status, i.e. in- or out-of-scope of the survey. Some examples of out-of-scope are: death, wrong birthday on FOSS or units no longer in Canada. The last collection outcome is the *unresolved* or untraceable cases. For these cases, no contact was established and they remained unresolved. No information as to their whereabouts in Canada was available. The table shows that the observed *unresolved* rate (approximately 28%) is higher than expected (23%). Conversely, the observed *resolved out-of-scope* is lower than expected (2.8% versus 7.8%).

Table 1: Wave 1 collection results.

Results	Expected	Observed
Resolved	**77.0**	**71.7**
In-scope responding	65.3	82.7
In-scope non-responding	26.9	14.5
Out-of-scope	7.8	2.8
Unresolved	**23**	**28.3**

This combined result suggests that unresolved cases may have a significantly higher proportion of out-of-scope units. Hence, a full examination of the collection outcomes and exhaustive studies of the response patterns were initiated. Along with close monitoring of the collection outcome codes and review of interviewer notes, discussions with immigration experts took place to better understand the characteristics of this population. While a portion of the unresolved cases could be attributed to a lack of tracing information, some cases remained unresolved despite very solid tracing sources. It was reasonable to believe that a good part of these unresolved cases may have been a result of immigrants leaving the country. During the development process, it became clear that these findings had to be incorporated in the weighting adjustment strategy to appropriately correct the non-response patterns.

6.2 Response mechanisms

Response patterns and mechanisms had to be studied carefully to appropriately correct for non-response. Classic response mechanisms are:

(i) Uniform (or missing completely at random): The probability of response is completely independent of the measurement process and is constant over the population.
(ii) Ignorable dependant (or missing at random): The probability of response is conditionally independent of the unobserved measurements given the observed measurements. In other words, the response mechanism depends on some auxiliary information or variables available for all units being measured.
(iii) Non-ignorable dependant (or not missing at random): The probability of response depends on the variable of interest, thus, not available for all units being measured.

Often, non-response is adjusted through re-weighting within some groups or cells which are defined based on i) the underlying mechanism, and; ii) identifying sample units with similar response probabilities, i.e. Response Homogeneity Group (RHG) theory. Then, within those groups, weights of responding units are inflated to take into account non-respondents. The method used to define the groups is the topic of section 8.2. For LSIC, not only did the response mechanism appear to be non-uniform, but collection results seem to indicate there was more than one mechanism generating the effective sample.

Tables 2 and 3 provide some details about the data collection findings and results. Table 2 presents the sample distribution by response category for each class of immigrant and table 3 for age grouping. These two tables present some preliminary evidence that there were different patterns emerging from the sample. For example, in table 2, economic class had a higher response rate than family but a significantly lower resolved rate; indicating that the economic class was difficult to trace, but once contacted they responded to the survey. Despite having the higher resolved rate, family class had the highest non-response rate. This still results in not having a completed questionnaire for these potential respondents, although generated by a different process than for the economic class. Refugees had the highest response and resolved rates. Finally, out-of-scope rates were different for the three immigrant classes. Similar conclusions can be made when examining the results by age group, as presented in table 3.

Table 2: Wave 1 collection results by class of immigration.

Results	Economic	Family	Refugees	Total
Resolved	**66.5**	**79.9**	**80.0**	**71.7**
In-scope responding	83.2	78.9	88.4	82.7
In-scope non-responding	13.3	18.7	10.6	14.5
Out-of-scope	3.5	2.3	1.0	2.8
Unresolved	**33.5**	**20.1**	**20.0**	**28.3**

Middle age groups, namely 25-34 and 35-44, seemed to be more difficult to resolve, but once traced, they tended to respond more (about 85% for both age groups). There is also a particularly high in-scope non-responding percentage for the 65 and over age group.

Table 3: Wave 1 collection results by age group.

Results	15 - 24	25 - 34	35 - 44	45 - 64	65 +	Total
Resolved	**74.7**	**67.9**	**70.6**	**77.5**	**85.7**	**71.7**
In-scope responding	82.8	84.9	84.0	77.9	69.2	82.6
In-scope non-responding	14.6	12.5	13.1	19.0	27.2	14.5
Out-of-scope	2.6	2.6	3.0	3.1	3.6	2.8
Unresolved	**25.3**	**32.1**	**29.4**	**22.5**	**14.3**	**28.3**

The wealth of information available on the population frame allowed for comparable studies with other variables - all indicating similar findings. The study of the sample distributions was revealing: non-responding and responding units were different; unresolved and non-responding units were different and finally unresolved units seemed to have similar patterns as the resolved out-of-scope units. Based on these findings, some assumptions were made. First, there are different non-uniform mechanisms generating the resulting sample. Second, non-response is comprised of two components - the in-scope non-responding units from the resolved cases and the unresolved units. Given that it is legitimate to believe that a good portion of the unresolved units are a result of immigrants who have left Canada, how will these be incorporated into the weighting strategy? This question relates directly to the issue of who the final weights represent, which is the topic of the next section.

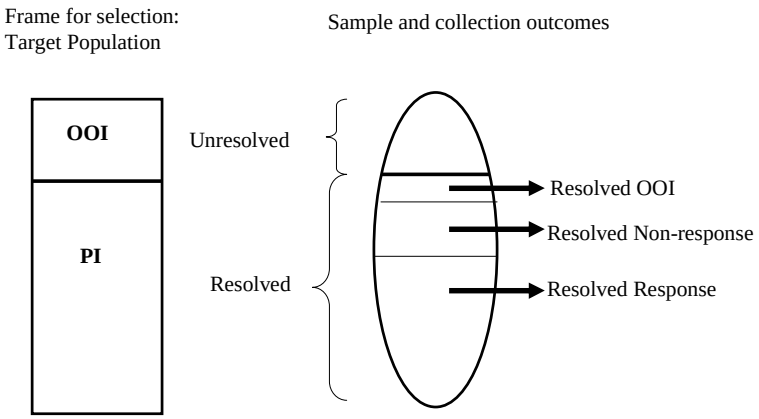
7. WEIGHTING ISSUE - THE FINAL WEIGHT

As is the case with any probability survey, the LSIC sample is selected to represent a reference population - the immigrant population - at a specific point in time. Each unit in the sample must therefore represent a certain number of units in the population. If all selected units were traced, contacted and an interview was completed, and if the frame used was perfect (i.e. covering exactly the population of interest), then the design weight assigned to each unit

would represent accurately and exactly the number of immigrants in the target population. In this situation, using this weight would yield unbiased estimates. However, this is not the case when surveys are faced with non-response, unresolved/untraceable units and imperfect frames.

For most surveys, the sum of the final weights represents the estimated target population counts, which usually coincides with the population of interest. Recall that the survey frame covers the target population - immigrants who meet all of the three criteria as described in section 2.3. However, it was confirmed that some of these immigrants landed in Canada but, resided only for a short period of time before returning to their original country or migrating to another country. Since the survey's main objective is to understand the integration process of immigrants who have recently arrived in Canada; those who have left do not have similar adaptation characteristics as the ones who are permanently residing in Canada. Bias could be introduced in the weight adjustment if one includes the immigrants who moved out of Canada with those who still reside in Canada. Therefore, the concepts of population of interest (PI) and population out of interest (OOI) were established. The **population of interest** consists of immigrants who meet all three criteria described in the definition of the target population (section 2.3) AND have lived in Canada for more than 6 months in a given collection year. The **out-of-interest population** consists of immigrants who no longer live in Canada, i.e., who have left since landing in Canada. Diagram 1 visually presents the concepts related to the population, the sample and the collection outcome.

Diagram 1: Overview of the design concepts as they relate to weighting



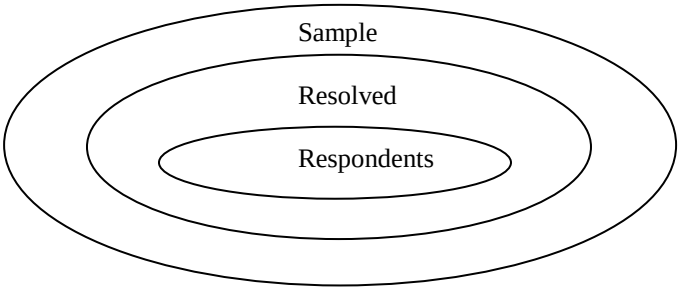
8. THREE-STEP MODEL-ASSISTED ADJUSTMENT

8.1 General principle

The LSIC weighting strategy is based on a series of cascading adjustments incorporating characteristics of all units identified as part of the *population of interest* only.[5] Based on the two distinct non-response mechanisms (i.e. the non-response within the resolved and the unresolved portions of the sample), two corrective adjustments are applied consecutively to the design weight of responding units in PI. First, there is an adjustment to redistribute the weights of the non-responding units in the resolved portion of the sample to the weights of the responding units within the resolved. The result of this first step is an unbiased estimate of the total resolved units in the PI. Second, the weights of the resolved units in the PI are adjusted to incorporate the weights of unresolved units. As shown in diagram 2, this is similar to a two-phase approach. However, recall that there is no information about the status, i.e., in PI or in OOI available for these unresolved units, therefore one will be predicted. The second adjustment actually consists of incorporating the predicted-in-the-PI units in the unresolved portion of the sample.

[5] This paper covers only the final weighting for the units in the PI. There is also a final weight attributed to the unit in OOI, but it is not cover in this paper. Please refer to LSIC user guide (2003) for details on the OOI weights.

Diagram 2: LSIC 2-phase adjustments for the units in the population of interest



8.2 Prediction of status for the unresolved units

Let's first introduce some notation. Following data collection, each unit can be assigned to either the resolved or unresolved portion of the sample.

Sample: $S = U + R$ where U is the set of unresolved units and R the set of resolved units. The resolved portion can further be broken down as:

$$R = RR + RN + RO$$

where *RR* are the responding units; *RN* the non-responding units; *RO* the units not in the population of interest referred to as OOI[6]and $RPI = RR + RN$ the resolved units in the PI.

Conceptually, the set of unresolved units (U) is composed of units in the PI and in the OOI population. The challenge is to predict a split between these two sub-populations. One option could be to assume the same split as within the resolved portion of the sample. However, there is a strong indication that the proportion of OOI units in the unresolved portion is higher that the one in the resolved portion of the sample. Moreover, the unresolved rates differ by some characteristics, suggesting a non-uniform response mechanism[7]. The solution was therefore to predict the probability of being in the PI based on different socio-demographic characteristics.

Let X'_k be the set of auxiliary variables available for all *resolved units*. X'_k is used to fit the following logistic regression model to obtain the estimated regression parameter $\hat{\alpha}$:

$$P\left[RPI_k = 1 \middle| X'_k\right] = \left(1 + \exp\left[-\left(\alpha_0 + \alpha_k X'_k\right)\right]\right)^{-1}$$

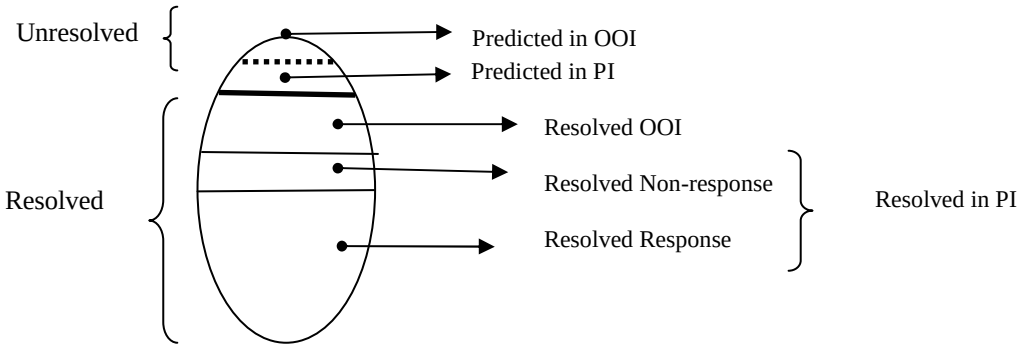
where $\mathrm{RPI_k} = \begin{cases} 1 \text{ if } k \in RPI \\ 0 \text{ if } k \in RO. \end{cases}$

The next step consists of estimating $\hat{\eta}'_k = \left(1 + \exp\left[-\left(\hat{\alpha}_0 + \hat{\alpha}_k X'_k\right)\right]\right)^{-1}$, for $k \in U$, i.e. the probability of being in the PI for each unit in the unresolved portion of the sample. The predicted status, i.e. in PI or in OOI is then obtained from a Bernoulli ($\hat{\eta}'_k$) trial. As shown in diagram 3, after this first step, all units received a status indicating if it is part of PI or OOI. Note that the most significant predictors for being an immigrant in the population residing in Canada are, in order of significance: *level of education, age groups and class of immigrant.*

[6] For simplicity, we defined them out of interest as opposed to the population of interest. Traditional out of scope is one component of "not in the population of interest" portion.

[7] Actually, there was a strong suspicion that the immigrants remained untraced because they had left Canada, similar to a non-ignorable dependent mechanism. However, the few available cases made it almost impossible to robustly estimate the "non-ignorability" parameters in the joint model used in such a situation.

Diagram 3: Final status for all sample units.



8.3 Creation of classes based on predictive models

As mentioned in section 6.2, weight adjustment classes are constructed under the assumption that they must be homogeneous - related to the correction being made (referred to as Homogeneity Group – HG). LSIC non-response adjustment classes are constructed based on units with similar propensity to respond. As for the unresolved adjustment classes, they are constructed based upon the same homogeneity property except that it involves units with "similar propensity of being resolved". Several methods can be used for the formation of the HG. For example, decision tree algorithms, such as CHAID in the software Knowledge Seeker (Kass, 1980; Angoss Software, 1995) and logistic regression models have been extensively used to create HGs. Another method, recently proposed by Eltinge and Yansaneh (1997), creates classes with similar estimated probabilities of responding. These probabilities are predicted values from a logistic regression model. For LSIC, the non-response and the unresolved adjustments' HGs are derived based on the Eltinge-Yanseneh method.

Non-response adjustment classes

Let X''_j be the set of auxiliary variables available for all *resolved units in PI* (i.e. respondents and non-respondents). X''_j are used to fit the following logistic regression model:

$$P\left[RR_j = 1 \middle| X''_j\right] = \left(1 + \exp\left[-\left(\alpha_0 + \alpha_j X''_j\right)\right]\right)^{-1}$$

where $\mathrm{RR}_j = \begin{cases} 1 & \text{if } j \in \mathrm{RR} \\ 0 & \text{if } j \in RN. \end{cases}$

From the model, we obtain $\hat{\eta}''_j = \left(1 + \exp\left[-\left(\hat{\alpha}_0 + \hat{\alpha}_j X''_j\right)\right]\right)^{-1}$, for $j \in RPI$. In other words, we obtain the probability of being a respondent for all resolved units in the population of interest. We sort units $j \in RPI$ based on their associated $\hat{\eta}''_j$. Similar steps, as described in the Eltinge-Yanseneh method, are then completed:

> *Step1:* Construct C adjustment classes with equal number of units within classes. We choose the equal-quantile method because of the control over the expected number of respondents in each cell. Compute adjusted-mean estimates based on the C classes. Six variables were used for LSIC.
> *Step 2:* Repeat step 1 with C+1 classes, until all estimates appear constant.

The estimated adjusted-means of the 6 variables used for diagnostics were converging as soon as the number of classes, C, was greater than 11. Note that the most significant predictors of being a respondent are, in order of significance: *class of immigrant, age group, level of education, knowledge of official language, and mother tongue*.

Unresolved adjustment classes

A similar approach is used to derive the adjustment classes for the unresolved portion of the sample. Let X_i''' be the set of auxiliary variables available for all *units in the population of interest* (i.e. resolved-PI and predicted-PI). X_i''' is then used to fit the following logistic regression model[8]:

$$P\left[RES_i = 1 \middle| X_i'''\right] = \left(1 + \exp\left[-\left(\alpha_0 + \alpha_i X_i'''\right)\right]\right)^{-1}$$

where $\mathrm{RES_i} = \begin{cases} 1 & \text{if } i \in \mathrm{RPI} \\ 0 & \text{if } i \in UPI. \end{cases}$

From the model, we obtain $\hat{\eta}_i''' = \left(1 + \exp\left[-\left(\hat{\alpha}_0 + \hat{\alpha}_i X_i'''\right)\right]\right)^{-1}$ for $i \in PI$, where $\mathrm{PI} = UPI + RPI$. In other words; we obtain the probability of being resolved for all units in the population of interest. We sort units $i \in PI$ based on their associated $\hat{\eta}_i'''$. The same steps, 1 and 2, are completed as described above. The convergence of the estimated adjusted means were established when C was greater than 12. Note that the most significant predictors of being a resolved unit are, in order of significance: *quality of tracing source, reference month,* and *number of years of school*. *Level of education, age group and class of immigrant* were also included to implicitly account for the predicted PI from the unresolved portion of the sample.

8.4 Non-response and unresolved adjustment weights

The weight of each responding unit in the PI resolved portion of the sample is finally adjusted using the following procedure:

i) Adjustment within each of the 13 non-response adjustment class:

$$\frac{\text{weighted sum of non-respondents} + \text{weighted sum of respondents}}{\text{weighted sum of respondents}}$$

ii) Adjustment within each of the 12 unresolved adjustment class:

$$\frac{\text{weighted sum of predicted unresolved} + \text{weighted sum of resolved}}{\text{weighted sum of resolved}}$$

9. CONCLUSION

Throughout the development of the survey methods, it became evident that LSIC weight adjustment would be a challenge, requiring a non-traditional adjustment. However, with no empirical evidence, it was a not a trivial exercise to come up with a new and unbiased strategy. It had to incorporate as much information as possible about the response mechanisms without creating any artificial patterns in the data or too many adjustment cells. The use of the Eltinge-Yanseneh method not only ensures equal-size class and homogeneous classes but, also allows the use of all possible explanatory variables without introducing any impact on the analysis. LSIC is a very good example of a model-based technique using all relevant auxiliary information. It is also a good example of careful and meticulous studies of response patterns in order to clearly understand the underlying mechanisms.

For future waves, collection outcomes will still be closely monitored. Because LSIC response rates were expected to be lower than for other surveys, at the outset of the project, the wave 1 sample size was over-sample accordingly. This ensures that reliable estimates would be possible at the end of wave 2 and wave 3. If everything goes to plan, the effective sample size will be greater than expected. New tracing sources are continuously being investigated and some new ones will be introduced at wave 3.

[8] Note that since this model includes implicitly the predicted PI, we also included in the model, auxiliary information coming from the first model described in 8.2.

REFERENCES

Angoss Software (1995), "Knowledge Seeker IV for Windows – User's Guide." Angoss Software International Limited.

Kass, G. V. (1980), "An explanatory technique for investigating large quantities of categorical data." *Applied Statistics*, Vol. 29, No.2, pp.119-127.

Eltinge, J.L. and Yansaneh, I.S. (1997), "Diagnostics for formation of non-response adjustment cells with an application to income non-response in the U.S. Consumer Expenditure Survey." *Survey Methodology*, June 1997, Vol. 23, No. 1, pp. 33-40.

Little, R.J.A. (1982), "Models for nonresponse in sample surveys." *Journal of the American Statistical Association,* Vol. 77, pp. 237-250.

LSIC microdata user guide (2003), *available upon request.*