

## **Estimation: 2001 APS**

In a sample survey, each respondent represents not only himself/herself, but also other persons that were not sampled. Consequently, a weight is associated with each respondent to indicate the number of persons that this respondent represents. This weight must be used for all estimations.

### **Weighting (Phase I – On and Off-community)**

The weights were calculated in a three-stage process. The FIRST stage was the assignment of an initial weight based on the sampling design. The initial weight was simply the inverse of the inclusion probability (probability of falling in the sample).

For the off-community portion of Phase I, the initial weight was the product of two components: the inverse of the primary sampling unit sampling fraction (called the PSU weight) and the Census weight. Three independent frames for North American Indians, Métis and Inuit were developed to select the sample. The PSUs were formed independently on each frame and three independent samples were selected. This means that some individuals who had multiple Aboriginal identities appeared on more than one frame. They thus had more than one chance of being selected. Since a unique survey weight was needed, the selection probabilities were adjusted to take the multiple inclusion probabilities into account.

Following this calculation, individuals selected by mistake<sup>1</sup> and those missed during sample selection were taken into consideration and appropriate weight adjustments were applied to the initial weight.

For the Phase I on-community sample, since sample size determination and sample selection involved manual operations as opposed to computer-based calculations, the initial weights were derived by dividing the weighted number of adults and children in each community (derived from the Census) by the number of such individuals in the sample.

The SECOND stage of the weighting process was the adjustment for non-response. Two adjustments were made to account for the fact that the non-respondents can be classified into two very different categories: the persons not contacted and the persons contacted but who did not respond.

---

<sup>1</sup> Because of the manual listing operation required to select the sample, errors did occur. Sometimes people who should not have been included were put on the list and sometimes people who should have been included were missed. Sample selection procedures were tested using questionnaires from the 1998 National Census Test to identify difficulties that would arise and to minimize errors.

The weights were adjusted first for non-contact and then for non-response. The non-response adjustment was done by forming non-response adjustment classes in such a way that the records in each class had similar response probabilities. The estimated response probabilities were obtained by developing a logistic regression model to predict the response probability using explanatory variables.

Many explanatory variables could be used since all Census long form information was available for each respondent and non-respondent. Separate models were used for children and adults. Approximately 10 classes of roughly the same size were obtained for each logistic regression model. The inverse of the weighted response rate in a class was used as the weighting adjustment factor for that class and the initial weights of the respondents within the class were adjusted accordingly.

The THIRD stage of the weighting adjustment was the post-stratification. This adjustment ensures that the sum of the final weights for the respondents is equal to the population counts from the Census. The adjustment was done for groups (called post-strata) defined by the combination of several variables.

For APS, two consecutive post-stratifications were completed. The first used (among other information) each respondent's answer to Census questions 17, 18, 20 and 21 (Census filter questions) to create post-strata. The weights, which had been corrected for non-response, were adjusted using the ratio of the Census count to the sample count for each post-stratum. This first post-stratification was aimed at ensuring that the sample did not under or over represent Census Aboriginal groups. Since answers to the screening (filter) questions can differ between APS and Census, a second post-stratification was carried out to guarantee that the total Aboriginal population, as estimated from the APS filter questions, matched those estimated from the Census filter questions. Adjustments were not made by Aboriginal group but rather for the total Aboriginal population (identity or origin).

### **Weighting and Non-Response Adjustment (Phase II Off-Community)**

The weighting process used in Phase II was identical to that of Phase I, except on one point. As the selection was made directly from the Census base (i.e. manual listing was not required), no adjustment was necessary for the people falsely selected or for the people who had been missed.

Briefly, the first step consisted of attributing an initial weight based on the sampling plan. The initial weight was simply the product of the Census weight and the subsample weight. The second step consisted of a non-

response adjustment. The method used was identical to the one used in Phase I and consisted in forming the weighting groups using logistic regression models. Models were developed for the non-contacts and the refusals, each separately for adults and children based on the explanatory variables available.

## **POST-STRATIFICATION**

### **First Post-stratification (On-community)**

For the on-community sample, in addition to the Census identity filter questions, three other variables were used to create post-strata: Aboriginal communities, age groups, and sex. For communities where the dominant Aboriginal group represented 90% of the population or more, the adjustment was done at the community level by crossing the age group and sex variables. For all other communities, adjustments were done for the dominant group and for the rest of the community. For both of these portions, post-strata were also defined by crossing the age group and sex variables. When communities were adjusted as a whole and contained 200 respondents or more, five age groups were used (0-4, 5-14, 15-24, 15-39, 40 and over). Otherwise, three age groups were used (0-14, 15-39, 40 and over).

### **First Post-Stratification (Off-Community, Phase I and II Combined)**

For the off-community component, the post-strata were defined using the subprovincial geography (closely corresponding to the off-community strata - Phase I), the child/adult variable as well as the Aboriginal identity and ancestry according to the Census. The post-stratifications of the “identity” and “ancestry only” components were made independently. At first, the same geography was used for both components in the post-strata definition. At the Aboriginal identity or ancestry level, the NAI groups, the Métis, the Inuit, the multiple NAI and Métis and the other multiples (all multiples types combined excluding NAI and Métis) were used to define the post-strata. In some cases, some aggregations were required. For example, in Newfoundland and Labrador, the post-stratification for the identity adjustment was done for the whole province instead of the subprovincial region (rural/urban). Also, in some domains, children and adults were combined when forming the post-strata.

### **Second Post-Stratification**

A second post-stratification was carried out to adjust the counts obtained from the APS screening questions to the Census counts. This post-

stratification was done according to the total count of Aboriginal people, and not according to each Aboriginal group, in order not to hide the transitions observed between the survey and the Census. In fact, these transitions reflect the Aboriginal status as a concept that can be affected by the context of the survey, the data collection method and the answers obtained by proxy.

This post-stratification was done for all Phase I and Phase II respondents combined. This weight adjustment was done such that the total Aboriginal population (identity and/or ancestry) as estimated from the survey would match the Census total Aboriginal population. This adjustment was done for each geographic domain (subprovincial: urban, rural or CMA) and age (child or adult) combination for the off-community component. For the on-community component, the adjustment was made separately for adults and children and was done for each community for which the publication of a profile was planned. Other communities for which no profile was planned have been post-stratified together.

### **Variance estimation**

The estimates that can be derived from this survey are based on a sample of individuals. Somewhat different estimates might be obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions is called the sampling error of the estimates.

In order to provide estimates of sampling error for statistics produced in APS, the bootstrap method was used. This method, which is a resampling method, consists of selecting  $M$  subsamples from the main sample and producing estimates for each subsample.

The bootstrap variance estimate (the variance is a particular measure of sampling error) was calculated as the empirical variance of the  $M$  estimates. For each subsample, the initial weights first had to be adjusted for bootstrap subsampling which produces what is called “initial bootstrap weights”. These initial bootstrap weights were then adjusted as described previously to obtain the final weights for each subsample. In other words, adjustments for individuals either missed or sampled by mistake, non-response and post-stratification were done for each subsample in almost exactly the same way as the full sample. For APS, 500 bootstrap samples were selected.

Each subsample was selected to reflect the sampling design of the full sample. However, units were selected *with replacement* as opposed to *without replacement*. For the off-community portion of the survey, each stratum was subdivided into two components according to whether or not the PSUs were selected with certainty (take-all PSUs) or not (take-some PSUs). In the take-some component of each stratum, a *with replacement* simple random sample of  $n-1$  PSUs within  $n$  PSUs was selected for each bootstrap sample. Here,  $n$  is the number of take-some PSUs in the original sample for that stratum. The particular choice of  $n-1$  has the advantage of simplifying the formula for the bootstrap weights.

In the take-all component of each stratum, as well as in the on-community component, a *with replacement* simple random sample of  $m-1$  individuals within  $m$  individuals was selected. Here,  $m$  is the number of individuals in the initial sample falling in the take-all component of the stratum or the sample size in each on-community stratum. This reflects the fact that, for the take-all component of the stratum, the initial sample can be seen as a one stage of individuals as opposed to a two-stage sample of the PSUs.

For the off-community portion of the survey, three independent frames of North American Indian, Métis and Inuit were developed to select the main sample. The PSUs were formed independently on each frame and three independent samples were selected. For the bootstrap method, the same sampling design was used. Therefore, 500 bootstrap samples were selected in each of the three frames. As for the survey sample, bootstrap weights were adjusted so that individuals with multiple identities/origins have a single bootstrap weight in each of the 500 iterations.