CLSA TECHNICAL DOCUMENT

Sampling and Computation of Response Rates and Sample Weights for the Tracking (Telephone Interview) Participants and Comprehensive Participants

CONTENTS

1. BACKGROUND

The Canadian Longitudinal Study on Aging (CLSA) recruited 51,338 Canadian residents aged 45-85 years at baseline to be followed for at least 20 years or until death. There are two components to the sample: (1) CLSA Tracking, whose target was 20,000 people from across the 10 Canadian provinces; and (2) CLSA Comprehensive, which aimed to recruit 30,000 people living within 25-50 km of one of 11 Data Collection Sites (DCS) across 7 Canadian provinces. Since the Canadian Community Health Survey (CCHS) on Healthy Aging was the first source of participants for the CLSA Tracking, the CLSA used the same selection criteria for the recruitment of all participants as the CCHS Healthy Aging Cycle 4.2 survey. The CCHS-Healthy Aging (CCHS-HA) sample is a nationally representative sample of people over the age 45. Excluded from the sampling frame, and consequently the CLSA, are residents in the three territories, persons living on federal First Nations reserves and other First Nations settlements in the provinces, full-time members of the Canadian Armed Forces, and individuals living in institutions. This latter exclusion means that individuals living in long-term care institutions (i.e., those providing 24-hour nursing care) are excluded from the CLSA at baseline. Individuals living in households and transitional housing arrangements (e.g., seniors' residences, in which only minimal care is provided) are included at baseline. Individuals unable to respond in English or French are also excluded from the CLSA. The CLSA team also used three additional sampling sources to select a random sample for both CLSA Tracking and Comprehensive cohorts. The three sampling frames were Provincial Health Registries (HR), Telephone sampling using Random Digit Dialing (RDD), and Quebec Longitudinal Study on Nutrition and Aging (NuAge). The inclusion criteria used to select participants from these three sampling frames were exatcly the same as CCHS-HA.

2. RESPONSE RATES

We defined response rates as the number of participants divided by the estimated number of those sampled who were eligible.

The detailed information about the response rate calculations is given in section 6.

3. SAMPLE WEIGHTS

It is standard practice in surveys to use sampling weights. Each participant in the study is assigned a sample weight constructed based on the inclusion probability. The sampling weights provided with the data aim to provide researchers with an estimate of how many people in each province (and in Canada) are represented by each CLSA participant. The aim is to ensure that, when estimating the mean value of some variable or the proportion with some characteristic, the value obtained is representative of the eligible provincial (and Canadian) population.

The weighting is necessary because the probability of selecting individuals from certain sub-groups of the population varied. As well, the probability that those selected agreed to participate varied within groups. For example, the CLSA ultimately included a much higher proportion of people in PEI than of people in Ontario. If the sample weights are not used, any estimate of Canadian population means or proportions will be skewed toward the mean proportion for PEI. Using the weights in the calculations will remove this skewness. These weights are known as inflation weights.

3.1. Inflation Weights

The CLSA Tracking cohort and Comprehensive cohort inflation weights were constructed to account for sample misrepresentation resulting from unequal sampling probabilities, frame coverage error and non-response, and to improve the precision of estimates through the use of auxiliary information.

First, the basic design weights were computed proportional to the reciprocals of the individual inclusion probabilities; they were then re-calibrated to the sum of the targeted (eligible) Canadian population using the weights given by Statistics Canada.

Inflation weights were also provided for the pooled sample from the two cohorts: based on a combined Tracking and Comprehensive inclusion probability for participants in the DCS areas, and the Tracking cohort inclusion probability for participants in non-DCS areas. In each case the weights were re-calibrated to the sums of the Statistics Canada weights within strata, as had been done for the two cohorts.

For all three sets of weights, in some cases the values of weights obtained in this manner were extremely large. In a small number of cases, the weights with highest values were trimmed, or set equal to second highest values within their provinces, and the calibration was repeated.

From all of these calculations, and for each of the Tracking or Comprehensive or pooled samples, the first wave inflation weight w_i of unit i [WGHTS_TRIMMED_TRM, WGHTS_TRIMMED_COM, WGHTS_TRIMMED_CLSAM] is interpreted as the number of persons in the population that unit represents, and the sum of the w_i over all i in the sample equals the known or assumed population size.

3.2. Analytic Weights

For analytical purposes, some different considerations apply. Regression and logistic regression analyses may be designed to estimate the relationships among variables, not so much for the population at hand as for a hypothetical population of like people, represented by the sample.

The so-called analytic weights supplied with the Tracking cohort data [WGHTS_ANALYTIC_TRM] are proportional to the inflation weights but rescaled to sum to the sample size within each province, so that their mean value is 1 within each province.

The analytic weights supplied with the Comprehensive cohort data [WGHTS_ANALYTIC_COM] are proportional to the inflation weights but rescaled to sum to the sample size within the DCS part of each province, so that their mean value is 1 within that area.

The analytic weights supplied with the pooled data [WGHTS_ANALYTIC_CLSAM] are proportional to the inflation weights for the pooled data but rescaled to sum to the sample size within the DCS and non-DCS part of each province.

3.3. When and How to Use the Weights

For the estimation of a descriptive parameter of the finite study population, the (trimmed) inflation weights should be used. For analyses that examine relationships between variables at the national or provincial level, analytic (rescaled) weights should be used. For analyses of smaller sub-groups, the analytic weights are likely to be appropriate. However, some consultation with a statistician is recommended.

The detailed information about the sample weights calculations is in section 7.

4. PRIMARY SAMPLING UNIT and SAMPLING STRATA VARIABLES

The use of complex survey software, as available for example in SAS, SPSS, Stata and R, is recommended for analyses, so that the sampling design can be accounted for. This will require specification of the appropriate weights variable (given above), and of characteristics of the sampling design, namely strata and primary sampling units or PSUs.

A stratified sampling design involves dividing the population into mutually exclusive (non-overlapping) strata, and sample is taken from every stratum. Within strata, individuals may be selected directly (single stage sampling) with a probability design. Alternatively, the sampling may be done in multiple stages within geographic strata.

If the sampling is done in stages, the PSU is a geographic unit selected by probability sampling at the first stage of sampling, within geographic strata. Within each selected PSU, there is a (possibly multi-stage) design for sampling individuals. This sample structure means that selected individuals are "clustered" into PSUs. Because individuals who live closer together tend to be more alike, a design with this kind of clustering tends to lead to less efficient estimation than does single stage probability sampling.

The samples from the HR and the RDD frame are effectively single stage, with no geographic clustering of respondents. The CCHS-HA design had at least two stages within health regions, but because we have not been provided with PSU information, and because the sample size is small enough that individuals in the same age group are not likely to be clustered together much more than in a single stage design, we take CCHS-HA design also to be single stage. For purposes of specification of the design in complex survey software, the PSU should be taken to be the individual, as represented by the unique ID variable, in the CLSA data for each cohort and for the pooled data.

For the strata variable to be specified in complex survey software, we recommend using the geographic strata variables, namely WGHTS_GEOSTRAT_TRM (10 provinces crossed with DCS/non-DCS with LowED/not LowEd), WGHTS_GEOSTRAT_COM (7 provinces within DCS crossed with LowED/ not LowEd), WGHTS_GEOSTRAT_CLSAM (10 provinces crossed with DCS/non-DCS with LowED/ not LowEd) for the Tracking, Comprehensive and pooled data, respectively. This is essentially the same geographic stratification for all three data sets.

The detailed information about sampling and strata determination is given in section 5.

5. SAMPLE

The sample was obtained via four sources:

1. Canadian Community Health Survey – Healthy Aging (CCHS): only for CLSA Tracking cohort
2. Provincial Health Registries (HR)
3. Telephone Sampling-Random Digit Dial (RDD)
4. Quebec Longitudinal Study on Nutrition and Aging (NuAge): only for CLSA Comprehensive cohort

For each we obtained 'pre-recruits', that is, people expressing preliminary interest in participation. Pre-recruits provided contact information. CLSA staff at Computer Assisted Telephone Interview (CATI) sites

[1] attempted to talk to each person by phone, explain the CLSA in more detail, describe what participation would

entail, and asked if the person would participate. Those who agreed to do so were considered 'recruits'.

Tracking Cohort: The full interview ('60 minute interview') was conducted immediately or later. Once a person completed the interview, s/he was called a 'provisional participant'. Consent forms could have been received by the CLSA before the 60 minute interview; but usually were not sent back until after the interview. People were not considered to be 'participants' until both the interview was completed and written consent was received by the CLSA.

Comprehensive Cohort: A person who completed the in-home interview and consent form (consent form was signed when the interview questionnaire was administered), but had not yet completed the DCS assessment was called a 'partial participant'. People were not considered to be 'participants' until the DCS assessment was completed.

Section 5.2. describes how people were recruited in detail for each source.

5.1. Criteria for the Sample

Tracking Cohort: The total of 20,000 participants was to be divided among the provinces to allow reasonably precise estimates of various parameters or associations to be made within provinces, while obtaining more of the sample from the larger provinces. The minimum target sample in any province was 1,100 in Prince Edward Island, and the maximum was 4,388 in Ontario. As well, the sample was distributed by age and sex within provinces. Eight strata were formed, based on age (45-54, 55-64, 65-74, and 75-85) and sex (male or female). Across the country, the four younger age-sex strata were to include 3,000 each and the four older age-sex strata 2,000 each.

Comprehensive Cohort: The total of 30,000 participants was to be divided among the provinces to allow reasonably precise estimates of various parameters or associations to be made within provinces, while obtaining more of the sample from the larger provinces. The minimum target sample in any province was 3,000 in Alberta, Manitoba, Newfoundland and Labrador, and the maximum was 6,000 in British Columbia, Ontario, and Quebec. Eight strata were formed, based on age (45-54, 55-64, 65-74, and 75-85) and sex (male or female).

Table 1 and Table 2 show the target and actual numbers of participants in each stratum, respectively, for the CLSA Tracking and Comprehensive.

5.2. Sources for the Sample (Sampling Frames)

5.2.1. Canadian Community Health Survey – Healthy Aging (CCHS): Only for Tracking Cohort

The first participants in the CLSA were recruited from participants in the CCHS which was conducted in 2008-2009 by Statistics Canada. As described above, the inclusion and exclusion criteria for the CLSA were adapted from those used in the CCHS. Participants in the CCHS aged 45-85 were asked if they would provide consent to allow Statistics Canada to pass on their names and contact details to the CLSA; they were also asked to give permission for passing on the data they provided as part of the CCHS interview to the CLSA. The disposition of participants in the CCHS who provided contact information and/or the data to the CLSA team is given in Table 3. Those who provided their contact information were approached by the CLSA. They were first sent an information package describing the purpose of the study, the criteria for participation, and a consent form by mail. They were called up

[1] The CATI sites, operated by the CLSA, conducted the 60 minute telephone interviews.

to 10 times[2] to be invited to join the study. Those who completed the CLSA baseline questionnaire[3] and provided written consent were considered CLSA Tracking participants.

5.2.2. Provincial Health Registries (HR)

In the second approach, the provincial government departments or data stewards responsible for housing the healthcare administration databases mailed the information packages directly to the randomly chosen age-eligible persons on behalf of the CLSA.

Depending on provincial requirements, the introductory letter included in the information package was signed either jointly by a provincial government representative designated by the province in question and the lead principal investigator (PI) for the CLSA, or separate introductory letters from the CLSA and/or Ministry were included. The letter package contained the introductory letter, a brief explanation of the CLSA, a consent form to be contacted, and a stamped, addressed envelope for potential participants to mail back to the CLSA. A reminder letter was sent twenty days after the initial mail-out. Those who replied and agreed to be contacted were called by the CLSA team. They were given further information and were asked to join the study.

For the Tracking cohort, eight provinces - British Columbia (BC), Manitoba (MB), New Brunswick (NB), Newfoundland and Labrador (NL), Nova Scotia (NS), Ontario (ON), Prince Edward Island (PE) and Saskatchewan (SK) - agreed to use their registry to select a sample and send letters to potential participants.

For the Comprehensive cohort, five provinces - BC[4], MB, NL, NS, and ON - agreed to use their registry to select a sample and send letters to potential participants.

There were two mailings. The first was sent to a stratified random sample of all eligible people in the province.[5] The second was sent to a more targeted sample as described below under 'Targeted Samples'.

For the Tracking cohort, SK and ON sent the first mailing but not the targeted one, while BC sent the targeted but not the first.[6]

For the Comprehensive cohort, only NS sent both first mailing and the targeted one, while BC, MB, NL, and ON sent only the first.

The CLSA team provided the number of people to be sent a letter in each age and sex stratum to the provinces; numbers were calculated taking into account predicted response rates, based on reported response rates of other studies in Canada, and pilot work for the CLSA. The ministries randomly selected people in each stratum, where possible taking account of the exclusion criteria. If more than one individual from a household was chosen, some provinces (MB, NL, SK, ON, PE, and BC) were able to select one person randomly to be sent the CLSA package while the other two provinces (NB and NS) were not able to do so.

---

[2] In practice, for some people, more than 10 attempts were made to contact them by phone.

[3] Because the CLSA received the contact information for CCHS participants prior the launch of the CLSA recruitment, CCHS-based participants first completed a 20-minute pre-recruitment questionnaire prior to completing the 60-minute baseline questionnaire.

[4] For the CLSA comprehensive, only the males aged between 45 and 54 were targeted in BC.

[5] This assumed each registry contained a complete list of people in the province. This will have excluded some eligible people who were not registered.

[6] In BC the process of pre-recruitment was slightly different in that after completing the sample selection, the Ministry provided the contact information to BC-based CLSA researchers who sent invitation letters directly.

The procedure used for sample selection varied by individual province. Although the same exclusion criteria applied to each province, the exact sampling process was based on the province's database/system.[7]

The dates of mail-outs and the numbers of people who were sent letters by province, age and sex are given in Tables 5, 6, 7 and 8.

5.2.3. Telephone Sampling-Random Digit Dialing (RDD)

Random digit dialing is a procedure in which valid telephone numbers[8] are generated randomly to draw a sample of households, which are then called by telephone. The CLSA used RDD as a third sampling approach in all provinces for the Tracking cohort, except NS where the target number of participants was achieved through mail-outs alone; and RDD was used as a second sampling approach in seven provinces which have DCS's for the Comprehensive cohort. Table 9 shows the data collection sites by province. Area codes, specific to provinces except for the code 902 which includes both NS and PE, were used to sample telephone numbers from each province.

Given various difficulties in calling cell phones, only landlines were included in the study. Excluding cell phones could have created an important bias if the study would not reach a large proportion of the eligible population. However, a survey of residential telephone service conducted by Statistics Canada in 2010 showed that very few households with members over the age of 45 did not have a landline. Based on these data we estimated that roughly 5% of potential participants would be excluded from sampling using RDD, and we considered this to be acceptably low.

After a pilot study had shown that RDD was feasible, a well-known Canadian professional polling company, Leger, was hired to conduct the RDD. A script was developed by the CLSA team for the company to use during the initial telephone interview to select and pre-recruit people. The CLSA team provided the company with quotas for the number of people in each stratum to be pre-recruited. The numbers were based on the target sample sizes remaining in those provinces, and on the assumption that 40% of Tracking pre-recruits would later agree to participate in the Tracking cohort and 60% of Comprehensive pre-recruits would later agree to participate in the Comprehensive cohort. The pre-recruitment was spread out over time, via weekly quotas for the polling company, to ensure that CATI sites did not have a backlog of people to call, and thus ensure a relatively short time gap between pre-recruitment by RDD and further contacts with the CLSA. Telephone numbers that were not businesses and appeared to be valid numbers were called up to 10 times before being considered to be non-respondents for pre-recruitment.

During the interview, information on whether there was anyone in the household aged between 45 and 85 and a roster of people in that age range were obtained. One person was randomly selected from the roster as a potential participant and the other eligibility questions were asked of the potential participant. Household members whose age-sex quota had already been filled were excluded from selection. Those who agreed to be contacted by the CLSA team were termed 'pre-recruits'.

[7] For example, NL did not exclude people who lived in long-term care facilities while the other provinces could do this. Of the provinces that did mail-outs, only NB and PE provided the number of packages that were returned as undeliverable, presumably because the selected people had moved or died. More details on the exclusion criteria that the provinces were able to implement are given in Table 4.

[8] Some numbers are known not to be valid, e.g., area codes and exchanges cannot begin with the numerals 0 or 1 and can be excluded from the list of possible numbers to call.

5.2.4. Targeted Samples

Early analyses on initial recruits showed under-representation of people with lower levels of education, a marker for various risk factors (many of which are assessed in the CLSA). This under-representation could potentially lead to low statistical power to identify relationships between these variables and health outcomes. Thus to increase heterogeneity in the independent variables, the CLSA chose to over-sample people from dissemination areas[9] (DAs) with relatively high proportions of people with lower levels of education. We anticipated that a similar problem would also arise later in sampling for the CLSA Comprehensive participants. To ensure we would still have enough participants in the catchment area around the DCS when sampling for the CLSA Comprehensive, we chose to conduct the targeted sampling for the CLSA Tracking in areas outside the catchment area.[10] More detail on the determination of these Dissemination Areas is in Appendix 1. We treated the extra mail-outs from HRs as additional, but different samples from the same sampling frame. We also had random (telephone) sampling from listed telephone numbers (RTS) –which was done by CLSA CATI sites– to increase the proportion of people with lower levels of education as a different sample. To ensure the people called lived in one of the identified DAs, listed telephone numbers, which included addresses, were used as the sampling frame.

This adjustment to the sampling added another stratification variable (see the section below on Strata for more details). As well, the oversampling meant that the target of 50,000 people was exceeded. In total, there were 21,241 participants in the CLSA Tracking, and 30,097 participants in the CLSA Comprehensive.

5.2.5. Quebec Longitudinal Study on Nutrition and Aging (NuAge)[11]: Only for Comprehensive Cohort

The last participants in the CLSA were recruited from the NuAge study, which is a longitudinal study on nutrition as a determinant of successful aging, and includes a cohort of 900 healthy men and 900 healthy women born between 1921 and 1935 to be monitored annually for a period of 5 years. Participants in NuAge were asked by the NuAge investigators if they would provide consent to share their information with the CLSA. Only the participants falling into the age group 75-85 who provided their contact information were approached by the CLSA. They were first sent an information package describing the purpose of the study and the criteria for participation by mail.

5.3. Samples in the CLSA

These approaches led to four samples making up the CLSA:

- CCHS: only for Tracking cohort
- HR
  - HR1 – initial Health Registry mail-outs
  - HR2 – Health Registry mail-outs targeting low-ed areas
- Telephone Sampling (TS)
  - RDD – done by Leger
  - RTS – conducted by CLSA CATI in targeted areas

[9] Statistics Canada states "[a] dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks. It is the smallest standard geographic area for which all census data are disseminated." Each area covers between 400 and 700 people, and this allowed us to identify areas with relatively high proportions of people with lower levels of education.

[10] Except for NS where recruitment had already been done before the decision to oversample exclusively from non-DCS areas.

[11] For the details, please go to http://www.rqrv.com/en/init_NuAge.php.

- NuAge: only for Comprehensive cohort

Tables 10 and 11 give the number of participants in each cohort crossed by sample and province.

5.4. Strata

Strata within a province were defined by the crossing of age group, sex and distance from Data Collection Sites-DCS.

- All 10 provinces are included in the CLSA: 10 provinces in Tracking cohort, and 7 provinces in Comprehensive cohort.
- Age groups were 45-54, 55-64, 65-74 and 75-85.

As noted above, the CLSA Comprehensive component of the study aimed to recruit 30,000 people within a fixed distance of the 11 DCSs.

Geographic areas were split into two groups based on whether they were in the DCS catchment area. We labelled the areas 'DCS' and 'non-DCS'.[12]

These variables created 136 strata[13] for the Tracking cohort and 56 (DCS only) strata for the Comprehensive cohort.

Table 12 and Table 13 show all possible strata and the number of participants within each stratum for Tracking and Comprehensive cohorts, respectively.

6. CALCULATION of RESPONSE RATES

Response rates were computed for each sample separately and the overall response rate was calculated by pooling the estimates for each sample.

CCHS: Statistics Canada provided the CLSA with response rates by province and age group for CCHS participants identified through the CCHS. The rates were not broken down by sex. We calculated the 'contact sharing rate' as the proportion of people in each age and sex group who gave permission to Statistics Canada to pass on their contact information to the CLSA. The CLSA computed recruitment and participation rates among those people who permitted Statistics Canada to pass their contact information to the CLSA. The rates were based on whether people completed the 20 minute interview (recruits) and then completed the 60 minute interview and provided consent forms (participants), respectively. Ineligible people were removed from denominators when calculating the rates; for those who could not be reached/contacted, estimates of the numbers ineligible were used, assuming the proportion of ineligibles in those not contacted was the same as in those who were contacted. Recruitment and participation rates were computed by province, age and sex. Overall response rates for each province, sex, and age group were obtained by multiplying these four rates. The details of the calculations are shown in Appendix 2.

HR: For the health registries, we used the numbers of people who were sampled and were sent letters, replied to the CLSA, were contacted by the CLSA and agreed to join the CLSA in each province, age and sex group. We also had the number of mailed packages returned as undeliverable in some provinces. As well, after contacting some

[12] In QC, ON, and BC, there was more than one DCS, so the 'DCS area' consisted of non-contiguous areas.
[13] In provinces with one or more DCSs there were 16 strata. In NB, PE, and SK, there was no DCS and thus there were 8 strata.

people we discovered they did not meet the CLSA eligibility criteria. We computed the proportion of those sent letters who returned the contact form, and the proportion of those who completed the 60 minute interview and returned their completed consent form for Tracking, or completed the in-home interview, consent form, and DCS assessment for the Comprehensive cohort. Response rates were obtained by multiplying proportions for each province, age and sex group. Calculations adjusted denominators to account for the numbers of returned packages and ineligible people. The formulas are shown in Appendix 3.

There were some people selected by one province's health registry who had moved to a different province. For confidentiality reasons, the CLSA did not have a list of people selected for the mail-outs in each province, and could not necessarily distinguish people who had moved to a different province. Such people, likely a very small proportion of the total,[14] were allocated to the province to which they had moved, rather than the one from which they had been selected. This was done for those who moved to a province that had also conducted a mail-out. For those who moved to a province that had not conducted a mail-out, it was inferred that they had moved from another province. To compute response rates and sample weights, we used data from Statistics Canada on interprovincial migrants (see footnote 12) and weighted random numbers to assign them to a province that had conducted a mail-out.

Telephone Sampling: In RDD, phone numbers were called up to 10 times and each call result was coded by the interviewer (answering machine, line busy, no answer, no one in that age group, refusal, language barrier, etc.). The company that conducted the RDD provided two files to the CLSA for each province; a call history file including the codes for each call attempt and a file including the results of initial interviews (including answers to the eligibility questions). A new file for each province was created by the CLSA team merging these two files; after this, call dates were sorted within phone numbers and the last call was identified for each phone number (household). From these last calls, the numbers of valid phone numbers, answered phone numbers, residential phone numbers, households for which age-eligibility was established, and age-eligible households were obtained. The numbers of rostered households and rostered people (in rostered households) were determined. The file also provided the numbers of selected people, people who agreed to participate in the study, people who were eligible, and people who provided contact information to the CLSA team. These numbers were used to calculate response rates for each province, sex, and age group. Since we did not have age and sex breakdown for household enumeration rate, the same household enumeration rate was used for all age and sex groups within a province.

RTS was conducted by the CLSA CATI sites. The process was a little different. Once a number had been called, the system did not call that number again until all the sampled numbers had been called. As a result, some phone numbers were called only once. In addition, the codes used for call results were different from the ones in RDD. However, the same procedure was used to compute the response rates.

NuAge pre-recruits were treated as RDD pre-recruits and counted in RDD response rates.

The details of the calculations for telephone sampling response rates are shown in Appendix 4.

Response rates are shown in Tables 14 and 15.

[14] The total number of inter-provincial migrants in 2010/11 was roughly 257,000, less than 1% of the Canadian population (http://www.statcan.gc.ca/pub/91-209-x/2013001/article/11786/tbl/tbl3-eng.htm) and concentrated in those younger than the age-eligible range for the CLSA (http://www.statcan.gc.ca/pub/91-209-x/2013001/article/11786/fig/fig2-eng.htm).

7. CALCULATION of SAMPLE WEIGHTS

7.1. Calculation of Inflation Weights

CCHS: Statistics Canada computed the weights for those who allowed Statistics Canada to pass on survey and contact information ('survey+contact' group) or survey data only ('survey only' group) to the CLSA; the weights added up to 13,232,650.77[15] (Table 16).

The CLSA team attempted to contact all 12,269 CCHS participants who provided their contact information. This included 11,742 'survey+contact' CCHS participants and 527 other CCHS participants who allowed Statistics Canada to pass on their contact information, but not their survey data ('contact only' participants). For the contact only participants, the CLSA imputed sampling weights. First, sampling weights for the 'survey+contact' and 'survey only' CCHS participants were grouped by the 136 strata mentioned above. The imputed weight for a 'contact only' participant was the median[16] weight of the relevant stratum. The weights were calibrated within stratum (to stratum totals), and hence across strata they add to the total of Statistics Canada CCHS weights, 13,232,650.77.

HR: In Health Registry mail-out samples, estimates were made of the number of eligible people who had been sent letters. Within each province conducting a mail-out, these estimates were based on the number of packages returned as undeliverable and the 2011 Census Data[17] by sex and age group.[18]

Basic (initial) sample weights for each sex, age and province stratum were calculated by using the estimated number of eligible people in ministry registries, the number of people contacted by the CLSA, the number of ineligible people identified by CLSA and the number of people who agreed to join the CLSA. More detail on the calculation of sample weights is given in Appendix 5. HR participants in the same sex, age group and province stratum were given the same sample weight.

For the Tracking cohort, as a next step, these strata were extended to DCS and education areas based on participants' postal codes.

The two samples, HR1 (initial Health Registry mail-outs) and HR2 (Health Registry mail-outs targeting low-ed areas), were then combined. There were 3,239 participants from HR1 and 571 from HR2, for a total of 3,810.

The full stratification was based on province, age, sex, DCS vs non-DCS, and low-ed vs non-low-ed. This created some small strata (as few as four in one stratum of CCHS participants for whom we had data), so there was a danger that some calibration totals might have fairly large errors. The calibration was thus done in two stages. The first stage used the initial weights for participants in non-DCS areas in each province that did the targeted mail-outs using strata defined by low-ed vs non-low-ed crossed by age (45-64 and 65-85). The aim of this stage was to

[15] For further information, http://www12.statcan.gc.ca/census-recensement/index-eng.cfm, 04.11.2014.
[16] Due to the presence of extreme weights within stratum, median weight was prefered to use instead of mean weight.
[17] We used the census total by province and age groups in Tracking cohort; but in Comprehensive cohort, only the DCS total within province was used for the same age groups.
[18] Since NF was not able to exclude people who lived in long-term care institutions during the sampling process, a further correction was made for NF using Statistics Canada data on Collective Dwellings. Institutional collective dwellings are general hospitals and hospitals with emergency, other hospitals and related institutions, nursing homes and facilities for persons with a disability. For the data, http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/tbt-tt/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=102239&PRID=0&PTYPE=101955&S=0&SHOWALL=0&SUB=0&Temporal=2011&THEME=91&VID=0&VNAMEE=&VNAMEF=, 17.05.2014.

account for the extra sampling in low-ed areas in HR2. The second stage looked at all participants in each province that conducted at least one mail-out, calibrating by DCS vs non-DCS crossed with sex and age decade (45-54, etc.).

For Comprehensive cohort, the strata were extended to education areas based on participants' postal codes.

The two samples, HR1 and HR2 were then combined. There were 3,746 participants from HR1 and 383 from HR2, for a total of 4,129.

The full stratification was based on the DCS part of the province, age, sex, and low-ed vs. non-low-ed. For the same reason mentioned above, the calibration was done in two stages. The first stage used the initial weights for participants in each province using strata defined by low-ed vs non-low-ed crossed by age (45-64 and 65-85). The second stage used the first calibration weights for participants in each province using strata crossed with sex and age decade (45-54, etc.)

Telephone Sampling: In telephone sampling, one age-eligible person per co-operating household was selected randomly to participate in the study. Within a contacted and co-operating household, the probability of selecting an individual is $1/k$ where k is the number of age-eligible people living in the same household. Thus the basic (initial) sample weight for an individual is $w_{ind} = k$, the inverse of the inclusion probability; it is the number of age-eligible people in each contacted and co-operating household, obtained by the interviewers during the initial telephone interview performed to select and pre-recruit people.

After the initial weights for both telephone sampling frames (RDD done by Leger and RTS conducted by CLSA CATI sites) were calculated, the two samples were combined. There were 12,857 participants from RDD and 651 from RTS, for a total of 13,508 in the Tracking cohort; 24,416 from RDD and 1,552 from RTS, for a total of 25,968[19] in the Comprehensive cohort. The same two-stage calibration procedure used for HR participants was applied for telephone recruitment participants.

During the calibration process for each sample in the study, if there was no participant in a sample from a defined stratum, the stratum was merged with the one having the same age group.

7.2. Combining Sample Weights Across All Sample Frames

After weights were calibrated for each sample separately, the weights were combined within strata using the general addition rule of probability (see Appendix 6 for further details).

Tracking Cohort: Combined weights were then re-calibrated within strata to add to the same stratum totals and overall total in the CCHS sample (Statistics Canada weights), 13,232,650.77.

Comprehensive Cohort: Combined weights were then re-calibrated within strata to add to the same stratum totals and overall total in the CCHS sample within the DCS sample, 3,746,315.78.

7.3. Trimming of Weights

Some "combined and re-calibrated" sample weights ('inflated' sample weights, 'inflation' weights), having extreme values relative to most of the other weights, were trimmed to prevent inflated sampling variances of the survey estimates. To identify the extreme values (outliers), box and whisker plots were used. Extreme values were

[19] NuAge participants were treated as RDD participants; so, this number also includes NuAge participants.

trimmed within province. Trimming was done sparingly. For Tracking, the highest weight in BC (which applied to three participants) was trimmed to the second highest value in BC and the highest weight in NS (which applied to two participants) was trimmed to the second highest value in NS (see Figure 1); for Comprehensive, the highest weight in QC (which applied to one participant) was trimmed to the second highest value in QC (see Figure 2). After trimming these extreme weights, in order for the weights of the entire sample to add to the eligible population size, trimmed values were allocated proportionally within the relevant province.

7.4. Calculation of Analytic Weights

Trimmed weights were further rescaled linearly to add to the number of participants within each province in each cohort to provide analytic weights; 21,241 for the Tracking cohort, and 30,097 for the Comprehensive cohort.

7.4.1. Rationale for Rescaling within Provinces

For *analytical purposes*, some different considerations apply. Regression and logistic regression analyses may be designed to estimate the relationships among variables, not so much for the population at hand as for a hypothetical population of like people, represented by the sample. The estimates of the coefficients or other parameters are found by solving estimating equations which are combinations of terms based on model residuals. For example, if we want to study self-reported BMI as a predictor of rheumatoid arthritis, using data from an entire moderate sized population, the *census estimating equation* for the logistic regression would look like this:

$$\sum_{i=1}^{N} X_i \, (y_i - g(X_i^T \beta)) = 0, \qquad (1)$$

where y_i is 1 if i has rheumatoid arthritis and 0 otherwise; X_i is a column vector of k explanatory variables or covariates including a suitable transformation of BMI; *T* denotes transpose; β is a column vector of k regression coefficients; and $g(z) = \frac{e^z}{1+e^z}$. Under the logistic regression model, each term of (1) has expectation 0, and solving (1) yields β_N, the census estimate of β. If the model is true, then because *N* is very large, β_N will be very close to β. If the model is not true, but the correct model is about the same for all provinces, β_N may still be a useful quantity, summarizing the relationship of *X* and *y* in the population.

The *sample-based estimating equation* that is approximately sampling-design-unbiased for the census estimating equation (1) is this:

$$\sum_{i \in s} w_i \, X_i (y_i - g(X_i^T \beta)) = 0, \qquad (2)$$

where *s* is the set of individuals in the sample.

The solution of (2) is a sampling-design-consistent estimator of β_N. Also, if in the sense of the model, the weight w_i given X_i is independent of or uncorrelated with the residual $y_i - g(X_i^T \beta)$, then the solution of (2) is a consistent estimator of β in the sense of the model. But the same would be true for any equation like (2) with a different choice of weights, as long as the expectation of each term is still 0. If the assumed logistic regression model is true, and the weights w_i are highly variable over the population, solving (2) to estimate β will be very inefficient compared to the same equation with a choice of weights which is closer to uniform.

In the CLSA design, the sampling rates are smaller in the larger provinces, and within a province, the sampling rates in the DCS areas are smaller (for the pooled sample) than in the non-DCS areas. Thus the use of the inflation weights in (2) will mean that if the model is correct, the estimation of β will be inefficient. On the other hand, if the model is only an approximation, in the sense that β varies from province to province, then the terms

in the larger provinces being dominant, the estimated value of β would be the value appropriate to the larger provinces. In either case, it may be more satisfactory to use less variable weights for this analytical purpose, and at the same time include province (or province crossed with DCS/non-DCS) in the model, as well as potentially a term for interaction of province or area and BMI.

7.4.2. Construction

The so-called analytic weights supplied with the Tracking cohort data [WGHTS_ANALYTIC_TRM] are proportional to the inflation weights but rescaled to sum to sample size within each province, so that their mean value is 1 within each province.

Once this rescaling is done, producing analytic weights $\widetilde{w}_i$, then the sample-based estimating equation

$$\sum_{i \in s} \widetilde{w}_i X_i \left(y_i - g\left(X_i^T \beta\right)\right) = 0 \qquad (3)$$

is approximately unbiased for the new census estimating equation

$$\sum_h \frac{1}{N_h} \sum_{i=1}^{N_h} X_i \left(y_i - g\left(X_i^T \beta\right)\right) = 0 \qquad (4)$$

where N_h is the number in the study population in province h. The solution of (4) would be a new census parameter $\tilde{\beta}_N$, close to an average of the values summarizing the relationships in the provinces. The estimating equation (3) will provide a design-consistent estimator of this new census parameter, which is a model-consistent estimator of β if the model is correct; the estimating equation (3) gives a more efficient estimator than would be provided by estimating equation (2). Here again, the estimating equation (3) is model-unbiased if the weight given X_i is independent of the residual.

The analytic weights supplied with the Comprehensive cohort data [WGHTS_ANALYTIC_COM] are proportional to the inflation weights but rescaled to sum to sample size within the DCS part of each province, so that their mean value is 1 within that area.

The analytic weights supplied with the pooled data [WGHTS_ANALYTIC_CLSAM] are proportional to the inflation weights for the pooled data but rescaled to sum to sample size within the DCS and non-DCS part of each province.

7.4.3. Variables to Include in the Model

A model for something like BMI as a predictor of rheumatoid arthritis is more likely to be correct if sex and age group are included in the model. The practice of including at least province, sex and age group in the model is recommended in general for analytic uses of the data. Within provinces, the weights will vary among age-sex groups, but the loss of efficiency from use of the weights is likely to be small.

For the pooled sample, partly because the analytic weights sum to sample size within DCS and within Non-DCS areas in each province, when researchers fit models to the data, they should consider including a categorical covariate which captures province and DCS or non-DCS within province.

7.4.4. Why Using Weights is Recommended in Modeling

The main reason for the standard advice to use the weights in analysis has to do with possible informativeness of the sampling design and/or the response-nonresponse process. That is, it could be the case that the distribution of the residual $y_i - g(X_i^T\beta)$, conditional on X_i, depends on whether the unit i is sampled, to the point that under the combination of model and design, the term $X_i(y_i - g(X_i^T\beta))$ does not have expectation 0, given that i is in the final sample. For example, suppose marital status has an influence on the relationship between BMI and rheumatoid arthritis, but is not included in the model. Marital status influences the individual's inclusion probability, which for the RDD part of the sample is lower, the more eligible individuals there are in the household. In that case, inclusion in the sample is informative about the relationship of interest. For another example, the design has over-sampled people in low education areas, which might also be areas where environmental factors could contribute to chronic conditions. Here also, being in the sample might alter the statistical relationship between BMI and rheumatoid arthritis.
Under these kinds of circumstances, the terms in (3) will also not have model expectation 0, but because of the presence of the weights, (3) will still be an unbiased estimator of the province population level estimating equation (4), and hence the solution of (3) will be a design-consistent estimator of $\tilde{\beta}_N$.

Software is available to handle most kinds of analyses where the methodology for using weights has been developed and standardized. Packages with survey data features include R, SAS, SPSS, Stata, SUDAAN and Mplus.

7.5. When Not to Use the Weights

For a description of the sample itself, rather than the study population or a hypothetical population behind it, the weights are not used.

The weights might also not be used for an analysis where the object of inference is not readily expressible in terms of estimating equations of the same kind of form as (1) or (4).

An example might be the use of survey data in some kinds of spatial analysis. In such cases it is best to consult with statisticians.

7.6. Sample Weights for the Pooled Data

After getting inflation weights for each cohort , the inflation weights were also provided for the pooled sample from the two cohorts based on a combined Tracking and Comprehensive inclusion probability for participants in the DCS areas, and the Tracking cohort inclusion probability for participants in non-DCS areas. In each case the weights were calibrated to the sums of the Statistics Canada weights within strata, as had been done for the two cohorts. Also, the trimming procedure was applied to the pooled data. The highest weight in AB non-DCS area (which applied to one participant) was trimmed to the second highest value in AB non-DCS area (see Figure 3). Trimmed weights were further rescaled linearly to add to the number of participants (51,338) within DCS and non-DCS areas in each province to provide analytic weights.

TABLES

Table 1. Actual and Target Number of the CLSA Tracking Participants

Province*	Age Group	Sex	Target Number of Participants in Stratum	Actual Number of Participants in Stratum
AB	45-54	Female	306	339
AB	45-54	Male	306	311
AB	55-64	Female	306	348
AB	55-64	Male	306	314
AB	65-74	Female	189	204
AB	65-74	Male	189	205
AB	75-85	Female	189	190
AB	75-85	Male	189	196
BC	45-54	Female	379	407
BC	45-54	Male	379	360
BC	55-64	Female	379	431
BC	55-64	Male	379	403
BC	65-74	Female	234	271
BC	65-74	Male	234	255
BC	75-85	Female	234	255
BC	75-85	Male	234	238
MB	45-54	Female	212	228
MB	45-54	Male	212	224
MB	55-64	Female	212	240
MB	55-64	Male	212	216
MB	65-74	Female	141	141
MB	65-74	Male	141	149
MB	75-85	Female	141	151
MB	75-85	Male	141	135
NB	45-54	Female	190	210
NB	45-54	Male	190	195
NB	55-64	Female	190	212
NB	55-64	Male	190	201
NB	65-74	Female	127	138
NB	65-74	Male	127	143
NB	75-85	Female	127	131
NB	75-85	Male	127	129
NL	45-54	Female	173	190
NL	45-54	Male	173	173
NL	55-64	Female	173	189
NL	55-64	Male	173	196
NL	65-74	Female	125	126
NL	65-74	Male	125	128
NL	75-85	Female	125	123
NL	75-85	Male	125	128
NS	45-54	Female	205	227
NS	45-54	Male	205	223
NS	55-64	Female	205	251
NS	55-64	Male	205	233
NS	65-74	Female	137	167
NS	65-74	Male	137	170
NS	75-85	Female	137	131
NS	75-85	Male	137	151
ON	45-54	Female	658	694
ON	45-54	Male	658	674
ON	55-64	Female	658	755
ON	55-64	Male	658	722
ON	65-74	Female	439	518
ON	65-74	Male	439	460
ON	75-85	Female	439	459
ON	75-85	Male	439	440
PE	45-54	Female	150	165
PE	45-54	Male	150	160
PE	55-64	Female	150	165
PE	55-64	Male	150	151
PE	65-74	Female	125	127
PE	65-74	Male	125	127
PE	75-85	Female	125	121
PE	75-85	Male	125	127
QC	45-54	Female	525	581
QC	45-54	Male	525	526
QC	55-64	Female	525	577
QC	55-64	Male	525	575
QC	65-74	Female	350	349
QC	65-74	Male	350	366
QC	75-85	Female	350	314
QC	75-85	Male	350	320
SK	45-54	Female	202	217
SK	45-54	Male	202	189
SK	55-64	Female	202	221
SK	55-64	Male	202	215
SK	65-74	Female	134	146
SK	65-74	Male	134	144
SK	75-85	Female	134	129
SK	75-85	Male	134	131

* AB=Alberta, BC=British Columbia, MB=Manitoba, NB=New Brunswick, NL=Newfoundland and Labrador, NS= Nova Scotia, ON= Ontario, PE=Prince Edward Island, QC=Quebec, SK=Saskatchewan.

Table 2. Actual and Target Number of the CLSA Comprehensive Participants

Province	Age Group	Sex	Target Number of Participants in Stratum	Actual Number of Participants in Stratum
AB	45-54	Female	450	384
AB	45-54	Male	450	329
AB	55-64	Female	450	509
AB	55-64	Male	450	492
AB	65-74	Female	300	371
AB	65-74	Male	300	375
AB	75-85	Female	300	253
AB	75-85	Male	300	244
BC	45-54	Female	900	831
BC	45-54	Male	900	782
BC	55-64	Female	900	1030
BC	55-64	Male	900	980
BC	65-74	Female	600	724
BC	65-74	Male	600	737
BC	75-85	Female	600	573
BC	75-85	Male	600	597
MB	45-54	Female	450	415
MB	45-54	Male	450	366
MB	55-64	Female	450	527
MB	55-64	Male	450	511
MB	65-74	Female	300	373
MB	65-74	Male	300	367
MB	75-85	Female	300	279
MB	75-85	Male	300	275
NL	45-54	Female	450	309
NL	45-54	Male	450	274
NL	55-64	Female	450	379
NL	55-64	Male	450	343
NL	65-74	Female	300	265
NL	65-74	Male	300	264
NL	75-85	Female	300	179
NL	75-85	Male	300	201
NS	45-54	Female	450	391
NS	45-54	Male	450	378
NS	55-64	Female	450	499
NS	55-64	Male	450	460
NS	65-74	Female	300	389
NS	65-74	Male	300	424
NS	75-85	Female	300	270
NS	75-85	Male	300	267
ON	45-54	Female	900	803
ON	45-54	Male	900	781
ON	55-64	Female	900	1070
ON	55-64	Male	900	1051
ON	65-74	Female	600	780
ON	65-74	Male	600	788
ON	75-85	Female	600	554
ON	75-85	Male	600	591
QC	45-54	Female	900	792
QC	45-54	Male	900	760
QC	55-64	Female	900	1075
QC	55-64	Male	900	930
QC	65-74	Female	600	786
QC	65-74	Male	600	719
QC	75-85	Female	600	510
QC	75-85	Male	600	491

Table 3. Disposition of Participants from Canadian Community Health Survey (CCHS)

People aged between 45-85	Eligible participants who allowed Statistics Canada to pass on their CCHS data to the CLSA	Survey + Contact	**11,742**	26,248
		Survey Only	8,345	
	Contact Only		**527**	
	Neither		5,634	
People aged 85+				4,617
TOTAL				**30,865**

Table 4. Exclusion Criteria Implemented by the Provinces for the Mail-outs from Health Registries

Province*	People living in Long-term Care Facilities and Institutions	People living in First Nations Settlements	Living on Crown Lands	People Who are not Canadian Citizen/ Landed Immigrant/Permanent Resident
BC	Excluded	Excluded**	Excluded	
MB	Excluded	Excluded	Excluded	
NB	Excluded	Excluded**		
NL		Excluded	Excluded	
NS	Excluded	Excluded**		
ON	Excluded	Excluded**		
PE	Excluded	Excluded**		
SK	Excluded	Excluded**		Excluded

* In the Tracking cohort: SK and ON were included only in the first mail-out and BC was included only in the second mail-out; in the Comprehensive cohort: the first mail-out was conducted in MB, NL, NS, ON, and BC, and the second mail-out was conducted in only NS.
** For the second mail-out, the CLSA team provided the provinces with the postal codes excluding First Nations settlements. Only MB and NL excluded the First Nations settlements themselves.

Table 5. CLSA Tracking HR1-Initial Mail-out by Provincial Health Registries; Number of Letters Sent by Age-Sex Groups

TRACKING COHORT											
Province	**Number of Mail-outs Conducted**	**Year**	**Number of People Mailed Out by Age-Sex Groups**								**TOTAL**
			M 45-54	**F 45-54**	**M 55-64**	**F 55-64**	**M 65-74**	**F 65-74**	**M 75-85**	**F 75-85**	
MB	3	2012	939	835	720	720	410	405	733	700	5,462
SK	1	2012	920	915	780	525	430	405	660	653	5,288
ON	2	2012, 2014	2,354	2,696	2,042	2,743	1,818	2,011	2,200	2,566	18,430
NS	2	2012	1,883	1,173	1,075	763	473	440	845	774	7,426
NB	2	2011, 2012	832	784	666	639	391	357	616	547	4,832
PE	2	2011, 2012	660	645	570	450	470	450	700	687	4,632
NL	2	2012	795	765	635	575	480	435	747	693	5,125
TOTAL											**51,195**

Table 6. CLSA Comprehensive HR1-Initial Mail-out by Provincial Health Registries; Number of Letters Sent by Age-Sex Groups

COMPREHENSIVE COHORT											
Province	**Number of Mail-outs Conducted**	**Year**	**Number of People Mailed Out by Age-Sex Groups**								**TOTAL**
			M 45-54	**F 45-54**	**M 55-64**	**F 55-64**	**M 65-74**	**F 65-74**	**M 75-85**	**F 75-85**	
MB	3	2012	750	750	750	750	500	500	667	667	5,334
NL	2	2012	450	450	449	450	299	300	2448	2096	6,942
NS	5	2012, 2013, 2014	2,606	1,504	1,520	1,125	750	750	1,081	1,333	10,669
ON*	6	2012	1,500	1,500	1,500	1500	1000	1000	1333	1333	10,666
BC	2	2011, 2012	8,250	0	0	0	0	0	0	0	8,250
TOTAL											**41,861**

Table 7. CLSA Tracking HR2-Mail-outs by Health Registries Targeting Low-ed Areas: Number of Letters Sent by Age-Sex Groups

TRACKING COHORT											
Province	**Number of Mail-outs Conducted**	**Year**	**Number of People Mailed Out by Age-Sex Groups**								**TOTAL**
			M 45-54	**F 45-54**	**M 55-64**	**F 55-64**	**M 65-74**	**F 65-74**	**M 75-85**	**F 75-85**	
MB	1	2014	201	99	331	190	69	276	247	251	1,664
BC	1	2014	697	809	637	379	244	325	470	689	4,250
NS	1	2014	1,026	600	517	433	618	464	685	460	4,803
NB	1	2014	834	448	168	134	98	95	331	757	2,865
PE	1	2014	683	385	311	124	376	173	393	497	2,942
NL	1	2014	728	170	286	157	162	173	331	530	2,537
TOTAL											**19,061**

Table 8. CLSA Comprehensive HR2-Mail-outs by Health Registries Targeting Low-ed Areas: Number of Letters Sent by Age-Sex Groups

COMPREHENSIVE COHORT											
Province	**Number of Mail-outs Conducted**	**Year**	**Number of People Mailed Out by Age-Sex Groups**								**TOTAL**
			M 45-54	**F 45-54**	**M 55-64**	**F 55-64**	**M 65-74**	**F 65-74**	**M 75-85**	**F 75-85**	
NS	1	2014	695	439	531	506	1382	513	369	423	4,858
TOTAL											**4,858**

Table 9. Data Collection Sites

Province	Number of DCS's	DCS	Affiliated University/Research Institute
Alberta	1	Calgary Data Collection Site	University of Calgary
British Columbia	3	Victoria Data Collection Site, Vancouver Data Collection Site, Surrey Data Collection Site	University of Victoria (UVIC), The University of British Columbia (UBC), Simon Fraser University (SFU)
Manitoba	1	Winnipeg Data Collection Site	University of Manitoba
Nova Scotia	1	Halifax Data Collection Site	Dalhousie University
Newfoundland and Labrador	1	St. John's Data Collection Site	Memorial University
Ontario	2	Hamilton Data Collection Site, Ottawa Data Collection Site	McMaster University, Bruyère Research Institute
Quebec	2	Montreal Data Collection Site, Sherbrooke Data Collection Site	McGill University, McGill University Health Centre, Université De Sherbrooke

Table 10. Source of Participants for the CLSA Tracking by Sample and Province

Province	Sample			Total Number of participants in each province
	CCHS	HR	TS	
Alberta	352	0	1,755	2,107
British Columbia	425	86	2,109	2,620
Manitoba	350	470	664	1,484
New Brunswick	268	316	775	1,359
Newfoundland and Labrador	209	337	707	1,253
Nova Scotia	310	1,208	35	1,553
Ontario	746	612	3,364	4,722
Prince Edward Island	213	327	603	1,143
Quebec	736	0	2,872	3,608
Saskatchewan	314	454	624	1,392
Total Number of participants in each sample	**3,923**	**3,810**	**13,508**	**TOTAL=21,241**

Table 11. Source of Participants for the CLSA Comprehensive by Sample and Province

Province	Sample		Total Number of participants in each province
	HR	RDD*	
Alberta	0	2957	2,957
British Columbia	133	6121	6,254
Manitoba	496	2617	3,113
Newfoundland and Labrador	443	1771	2,214
Nova Scotia	2093	985	3,078
Ontario	964	5454	6,418
Quebec	0	6063	6,063
Total Number of participants in each sample	**4,129**	**25,968**	**TOTAL=30,097**

* NuAge participants were treated as RDD participants.

Table 12. The 136 Strata and the Number of Participants in Each in the Final CLSA Tracking Group

Stratum Number	Province	Sex	Age Group	DCS	Number of Participants	Stratum Number	Province	Sex	Age Group	DCS	Number of Participants
1	AB	female	45 to 55	DCS	48	**69**	NL	male	65 to 75	DCS	36
2	AB	female	45 to 55	Non_DCS	291	**70**	NL	male	65 to 75	Non_DCS	92
3	AB	female	55 to 65	DCS	65	**71**	NL	male	75 to 96	DCS	38
4	AB	female	55 to 65	Non_DCS	283	**72**	NL	male	75 to 96	Non_DCS	90
5	AB	female	65 to 75	DCS	44	**73**	NS	female	45 to 55	DCS	68
6	AB	female	65 to 75	Non_DCS	160	**74**	NS	female	45 to 55	Non_DCS	159
7	AB	female	75 to 96	DCS	28	**75**	NS	female	55 to 65	DCS	72
8	AB	female	75 to 96	Non_DCS	162	**76**	NS	female	55 to 65	Non_DCS	179
9	AB	male	45 to 55	DCS	36	**77**	NS	female	65 to 75	DCS	45
10	AB	male	45 to 55	Non_DCS	275	**78**	NS	female	65 to 75	Non_DCS	122
11	AB	male	55 to 65	DCS	48	**79**	NS	female	75 to 96	DCS	39
12	AB	male	55 to 65	Non_DCS	266	**80**	NS	female	75 to 96	Non_DCS	92
13	AB	male	65 to 75	DCS	38	**81**	NS	male	45 to 55	DCS	77
14	AB	male	65 to 75	Non_DCS	167	**82**	NS	male	45 to 55	Non_DCS	146
15	AB	male	75 to 96	DCS	17	**83**	NS	male	55 to 65	DCS	65
16	AB	male	75 to 96	Non_DCS	179	**84**	NS	male	55 to 65	Non_DCS	168
17	BC	female	45 to 55	DCS	137	**85**	NS	male	65 to 75	DCS	43
18	BC	female	45 to 55	Non_DCS	270	**86**	NS	male	65 to 75	Non_DCS	127
19	BC	female	55 to 65	DCS	129	**87**	NS	male	75 to 96	DCS	37
20	BC	female	55 to 65	Non_DCS	302	**88**	NS	male	75 to 96	Non_DCS	114
21	BC	female	65 to 75	DCS	72	**89**	ON	female	45 to 55	DCS	72
22	BC	female	65 to 75	Non_DCS	199	**90**	ON	female	45 to 55	Non_DCS	622
23	BC	female	75 to 96	DCS	77	**91**	ON	female	55 to 65	DCS	86
24	BC	female	75 to 96	Non_DCS	178	**92**	ON	female	55 to 65	Non_DCS	669
25	BC	male	45 to 55	DCS	110	**93**	ON	female	65 to 75	DCS	53
26	BC	male	45 to 55	Non_DCS	250	**94**	ON	female	65 to 75	Non_DCS	465
27	BC	male	55 to 65	DCS	135	**95**	ON	female	75 to 96	DCS	36
28	BC	male	55 to 65	Non_DCS	268	**96**	ON	female	75 to 96	Non_DCS	423
29	BC	male	65 to 75	DCS	76	**97**	ON	male	45 to 55	DCS	66
30	BC	male	65 to 75	Non_DCS	179	**98**	ON	male	45 to 55	Non_DCS	608
31	BC	male	75 to 96	DCS	88	**99**	ON	male	55 to 65	DCS	77
32	BC	male	75 to 96	Non_DCS	150	**100**	ON	male	55 to 65	Non_DCS	645
33	MB	female	45 to 55	DCS	120	**101**	ON	male	65 to 75	DCS	59
34	MB	female	45 to 55	Non_DCS	108	**102**	ON	male	65 to 75	Non_DCS	401
35	MB	female	55 to 65	DCS	129	**103**	ON	male	75 to 96	DCS	43
36	MB	female	55 to 65	Non_DCS	111	**104**	ON	male	75 to 96	Non_DCS	397
37	MB	female	65 to 75	DCS	57	**105**	PE	female	45 to 55	Non_DCS	165
38	MB	female	65 to 75	Non_DCS	84	**106**	PE	female	55 to 65	Non_DCS	165
39	MB	female	75 to 96	DCS	64	**107**	PE	female	65 to 75	Non_DCS	127
40	MB	female	75 to 96	Non_DCS	87	**108**	PE	female	75 to 96	Non_DCS	121
41	MB	male	45 to 55	DCS	97	**109**	PE	male	45 to 55	Non_DCS	160
42	MB	male	45 to 55	Non_DCS	127	**110**	PE	male	55 to 65	Non_DCS	151
43	MB	male	55 to 65	DCS	115	**111**	PE	male	65 to 75	Non_DCS	127
44	MB	male	55 to 65	Non_DCS	101	**112**	PE	male	75 to 96	Non_DCS	127
45	MB	male	65 to 75	DCS	63	**113**	QC	female	45 to 55	DCS	98
46	MB	male	65 to 75	Non_DCS	86	**114**	QC	female	45 to 55	Non_DCS	483
47	MB	male	75 to 96	DCS	62	**115**	QC	female	55 to 65	DCS	118
48	MB	male	75 to 96	Non_DCS	73	**116**	QC	female	55 to 65	Non_DCS	459
49	NB	female	45 to 55	Non_DCS	210	**117**	QC	female	65 to 75	DCS	72
50	NB	female	55 to 65	Non_DCS	212	**118**	QC	female	65 to 75	Non_DCS	277
51	NB	female	65 to 75	Non_DCS	138	**119**	QC	female	75 to 96	DCS	65
52	NB	female	75 to 96	Non_DCS	131	**120**	QC	female	75 to 96	Non_DCS	249
53	NB	male	45 to 55	Non_DCS	195	**121**	QC	male	45 to 55	DCS	80
54	NB	male	55 to 65	Non_DCS	201	**122**	QC	male	45 to 55	Non_DCS	446
55	NB	male	65 to 75	Non_DCS	143	**123**	QC	male	55 to 65	DCS	97
56	NB	male	75 to 96	Non_DCS	129	**124**	QC	male	55 to 65	Non_DCS	478
57	NL	female	45 to 55	DCS	59	**125**	QC	male	65 to 75	DCS	76
58	NL	female	45 to 55	Non_DCS	131	**126**	QC	male	65 to 75	Non_DCS	290
59	NL	female	55 to 65	DCS	64	**127**	QC	male	75 to 96	DCS	59
60	NL	female	55 to 65	Non_DCS	125	**128**	QC	male	75 to 96	Non_DCS	261
61	NL	female	65 to 75	DCS	38	**129**	SK	female	45 to 55	Non_DCS	217
62	NL	female	65 to 75	Non_DCS	88	**130**	SK	female	55 to 65	Non_DCS	221
63	NL	female	75 to 96	DCS	36	**131**	SK	female	65 to 75	Non_DCS	146
64	NL	female	75 to 96	Non_DCS	87	**132**	SK	female	75 to 96	Non_DCS	129
65	NL	male	45 to 55	DCS	45	**133**	SK	male	45 to 55	Non_DCS	189
66	NL	male	45 to 55	Non_DCS	128	**134**	SK	male	55 to 65	Non_DCS	215
67	NL	male	55 to 65	DCS	47	**135**	SK	male	65 to 75	Non_DCS	144
68	NL	male	55 to 65	Non_DCS	149	**136**	SK	male	75 to 96	Non_DCS	131

Table 13. The 56 Strata and the Number of Participants in Each in the Final CLSA Comprehensive Group

Stratum Number	Province	Sex	Age Group	DCS	Number of Participants	Stratum Number	Province	Sex	Age Group	DCS	Number of Participants
1	AB	female	45 to 55	DCS	384	29	NL	male	45 to 55	DCS	274
2	AB	female	55 to 65	DCS	509	30	NL	male	55 to 65	DCS	343
3	AB	female	65 to 75	DCS	371	31	NL	male	65 to 75	DCS	264
4	AB	female	75 to 96	DCS	253	32	NL	male	75 to 96	DCS	201
5	AB	male	45 to 55	DCS	329	33	NS	female	45 to 55	DCS	391
6	AB	male	55 to 65	DCS	492	34	NS	female	55 to 65	DCS	499
7	AB	male	65 to 75	DCS	375	**35**	NS	female	65 to 75	DCS	389
8	AB	male	75 to 96	DCS	244	**36**	NS	female	75 to 96	DCS	270
9	BC	female	45 to 55	DCS	831	**37**	NS	male	45 to 55	DCS	378
10	BC	female	55 to 65	DCS	1030	**38**	NS	male	55 to 65	DCS	460
11	BC	female	65 to 75	DCS	724	**39**	NS	male	65 to 75	DCS	424
12	BC	female	75 to 96	DCS	573	**40**	NS	male	75 to 96	DCS	267
13	BC	male	45 to 55	DCS	782	**41**	ON	female	45 to 55	DCS	803
14	BC	male	55 to 65	DCS	980	**42**	ON	female	55 to 65	DCS	1070
15	BC	male	65 to 75	DCS	737	**43**	ON	female	65 to 75	DCS	780
16	BC	male	75 to 96	DCS	597	**44**	ON	female	75 to 96	DCS	554
17	MB	female	45 to 55	DCS	415	**45**	ON	male	45 to 55	DCS	781
18	MB	female	55 to 65	DCS	527	**46**	ON	male	55 to 65	DCS	1051
19	MB	female	65 to 75	DCS	373	**47**	ON	male	65 to 75	DCS	788
20	MB	female	75 to 96	DCS	279	**48**	ON	male	75 to 96	DCS	591
21	MB	male	45 to 55	DCS	366	**49**	QC	female	45 to 55	DCS	792
22	MB	male	55 to 65	DCS	511	**50**	QC	female	55 to 65	DCS	1075
23	MB	male	65 to 75	DCS	367	**51**	QC	female	65 to 75	DCS	786
24	MB	male	75 to 96	DCS	275	**52**	QC	female	75 to 96	DCS	510
25	NL	female	45 to 55	DCS	309	**53**	QC	male	45 to 55	DCS	760
26	NL	female	55 to 65	DCS	379	**54**	QC	male	55 to 65	DCS	930
27	NL	female	65 to 75	DCS	265	**55**	QC	male	65 to 75	DCS	719
28	NL	female	75 to 96	DCS	179	**56**	QC	male	75 to 96	DCS	491

Table 14. CLSA Tracking Cohort Response Rates by Province and Sample and Overall

	AB	BC	MB	NB	NL	NS	ON	PE	QC	SK	CANADA
CCHS	**0.12**	**0.11**	**0.15**	**0.12**	**0.11**	**0.13**	**0.11**	**0.13**	**0.13**	**0.14**	**0.12**
RDD	0.09	0.11	0.10	0.13	0.09	-	0.10	0.13	0.15	0.09	0.11
RTS	0.01	0.01	0.01	0.01	0.01	0.02	0.01	-	0.02	0.01	0.01
TS	**0.07**	**0.10**	**0.09**	**0.10**	**0.08**	**0.02**	**0.09**	**0.13**	**0.13**	**0.07**	**0.10**
HR1	-	-	0.08	0.07	0.06	0.12	0.04	0.06	-	0.09	0.07
HR2	-	0.02	0.03	0.02	0.01	0.08	-	0.02	-	-	0.03
HR	**-**	**0.02**	**0.07**	**0.05**	**0.05**	**0.10**	**0.04**	**0.05**	**-**	**0.09**	**0.06**
OVERALL	**0.08**	**0.09**	**0.09**	**0.08**	**0.07**	**0.10**	**0.08**	**0.09**	**0.13**	**0.08**	**0.09**

Table 15. CLSA Comprehensive Cohort Response Rates by Province and Sample and Overall

	AB	BC	MB	NL	NS	ON	QC	CANADA
RDD	0.11	0.10	0.13	0.19	0.16	0.10	0.12	0.11
RTS	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.02
TS	**0.11**	**0.10**	**0.10**	**0.15**	**0.12**	**0.09**	**0.10**	**0.10**
HR1	-	0.02	0.09	0.06	0.16	0.09	-	0.09
HR2	-	-	-	-	0.08	-	-	0.08
HR	**-**	**0.02**	**0.09**	**0.06**	**0.14**	**0.09**	**-**	**0.09**
OVERALL	**0.11**	**0.09**	**0.10**	**0.12**	**0.13**	**0.09**	**0.10**	**0.10**

Table 16. Total Weights Computed by Statistics Canada*

	Number of Eligible Participants Who Allowed Statistics Canada to Pass on Their CCHS Data to the CLSA	Weights		
		Minimum	Maximum	Sum
Survey+Contact	11,742	3.07	22,374.19	7,678,625.69
Survey Only	8,345	3.25	22,317.57	5,554,025.08
TOTAL	**20,087**			**13,232,650.77****

*Statistics Canada re-calculated the CCHS weights for these 20,087 people to add to the total eligible population.
**The total in DCS areas was 3,746,315.46.

Table 17.1 Number of Postal Codes per Province by % of People with Lower Levels of Education*

	≥ 40		≥ 45		≥ 50		≥ 55		≥ 60		≥ 65		TOTAL
Province	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**
AB	30648	38.2	21416	26.7	13961	17.4	7685	9.6	**3494**	4.4	1557	1.9	80235
BC	43558	38.3	28648	25.2	17823	15.7	8743	7.7	**3944**	3.5	1703	1.5	113732
MB	12984	53.7	9868	40.8	7097	29.3	3981	16.5	2353	9.7	**1314**	5.4	24197
NB	41711	70.2	33587	56.5	25667	43.2	17477	29.4	**11157**	18.8	6197	10.4	59407
NL	3705	34.1	2303	21.2	1524	14.0	862	7.9	**553**	5.1	264	2.4	10872
NS	12356	43.9	8277	29.4	5537	19.7	2759	9.8	**1461**	5.2	464	1.7	28117
ON	120593	43.7	85631	31.0	60167	21.8	35175	12.7	18861	6.8	**8862**	3.2	276083
PE	1077	32.6	681	20.6	374	11.3	**243**	7.3	210	6.4	137	4.1	3307
QC	86713	41.0	57404	27.2	37366	17.7	19204	9.1	9208	4.4	**3840**	1.8	211384
SK	11932	54.0	9049	41.0	7001	31.7	4204	19.0	**2245**	10.2	914	4.1	22087
TOTAL	**365277**	**44.0**	**256864**	**31.0**	**176517**	**21.3**	**100333**	**12.1**	**53486**	**6.4**	**25252**	**3.0**	**829421**

* Percentage of people aged 25-64 and 65+ in DAs (See Appendix 1)

Table 17.2 Number of Postal Codes per Province

Province	Cut Point	Number of Postal Codes	%
AB	60	3494	4.4
BC	60	3944	3.5
MB	65	1314	5.4
NB	60	11157	18.8
NL	60	553	5.1

Province	Cut Point	Number of Postal Codes	%
NS	60	1461	5.2
ON	65	8862	3.2
PE	55	243	7.3
QC	65	3840	1.8
SK	60	2245	10.2

APPENDICES

Appendix 1. Determination of the 'Low-Education' Dissemination Areas

As early indications showed that the proportion of participants with lower levels of education (defined as having no education beyond high school) was below that in the population, it was decided to try to oversample people with lower levels of education. The first step was to estimate how many people had to be sampled in each province to achieve the desired number of participants with high school or lower education.

The next step was to identify 'low education (low-ed)' areas - areas in which the proportion of people with high school or less education was relatively high. The CLSA used data from the 2006 Census. These data were the latest available that provided the required information. The aim was to restrict sampling to those areas, so that the 'yield' of people with lower levels of education would be relatively high, and the sampling process would thus be more efficient.

For confidentiality reasons, the CLSA did not have access to individual level data. Rather the data were grouped into age groups 25-64 and 65+, and according to 'Dissemination areas (DAs)' - see footnote 5. The CLSA identified DAs in which the proportion of people with high school or lower education in both age groups exceeded various percentages. The number of postal codes (PCs) in those DAs was determined (see Table 17.1). For each province, an estimate was made of the number of people who had to be sampled to achieve the target number of participants.

Using this figure and Table 17.1, an appropriate cut-point was chosen for the proportion of people with 'low-ed'. The cut-point was specific to the individual province, and was intended to balance the efficiency of sampling with the need to obtain the target number of participants. As well, an average of no more than one person in any age-sex group was sampled per PC to ensure that the average total number of people sampled per PC was not so high that the sampling would be too concentrated in these PCs. The DAs within each province were divided into those within the catchment area of the Data Collection Sites, and those outside that area. Sampling for CLSA Tracking was then restricted to areas outside the catchment areas. In some provinces - PE, NB, SK - there was no DCS, so this included the whole province.

For example, Table 17.2 shows that in SK, there were 2245 Postal Codes covered by the Dissemination areas in which at least 60% of people had 'low education'. Targeted sampling for the 'low-ed' mail-out (HR2) and for the second phase of telephone sampling (RTS) was restricted to these postal codes.

Appendix 2. CCHS Response Rate Calculation

CCHS response rates by province and age groups (A_h) were provided by Statistics Canada. The same rates were used for the male and female people in the same age groups within province.

Within province:

Proportion of people who provided their contact information to the CLSA in sex*age group h (B_h): (Number of pre-recruits in sex*age group h)/Number of people aged 45-85 from CCHS in sex*age group h

Recruitment Rate in sex*age group h (C_h): Number of recruits in sex*age group h /(Number of pre-recruits in sex*age group h- Number of ineligible people because of their proxy mode and language in sex*age group h -Number of people contacted by the CLSA team and found ineligible in sex*age group h -Number of people not contacted by the CLSA team and estimated ineligible in sex*age group h)

Participation Rate in sex*age group h (D_h): Number of participants in sex*age group h /(Number of recruits in sex*age group h -Number of people unreachable and estimated ineligible in sex*age group h-Number of pre-recruits contacted but found ineligible in sex*age group h)

Response Rate 1= $B_h \times C_h \times D_h$

Response Rate 2=$A_h \times B_h \times C_h \times D_h$

Overall rates for provinces and Canada can be calculated by modifiying the same formulas above.

Appendix 3. Provincial Health Registry Mail-outs Response Rate Calculation

1. Rates for each age and sex group within province

Pre-Recruitment Rate (A_h)= $r_h/(n_h - x_h)$,

where n_h is number of people sampled and sent letters in each age and sex group within province, r_h is number of people who replied to the CLSA in each age and sex group within province, x_h is number of returned mail-outs in each age and sex group within province.

Contact Rate (B_h)= s_h/r_h

where s_h is number of people contacted by the CLSA in each age and sex group within province

Full Participation Rate (C_h)=$t_h/(s_h - y_h)$

where t_h is number of peple who both completed baseline interview and provided consent form (participants) in each age and sex group within province, y_h is number of other ineligible in each age and sex group within province.

Response Rate for each age and sex group within province =$A_h \times B_h \times C_h$

2. Overall Response Rates by Provinces:

Pre-Recruitment Rate (A)= $r/(n - x)$

Contact Rate (B)= s/r

Full Participation Rate (C)= $t/(s - y)$

Response Rate =$A \times B \times C$

Appendix 4. Telephone Sampling Response Rate Calculation

1. N: Numbers of TNs in ASDE data set
2. n: Numbers of TNs called by Leger/CLSA CATI
3. $P(TNs\ called) = P_c = \frac{n}{N}$
4. $P(TNs\ valid|called) = P_{val} = \frac{n_{val}}{n}$
5. $P(TNs\ answered|valid) = P_{ans} = \frac{n_{ans}}{n_{val}}$
6. $P(TNs\ residence|answered) = P_{res} = \frac{n_{res}}{n_{ans}}$
7. $P(age\ eligible\ established|residence) = P_{age.est} = \frac{n_{age.est}}{n_{res}}$
8. $P(age\ eligible\ HHs|age\ eligible\ established) = P_{age} = \frac{n_{age}}{n_{age.est}}$
9. $P(rostered\ HHs|age\ eligible\ HHs) = P_{rost} = \frac{n_{rost}}{n_{age}}$
10. $'P'(individual\ selection) = P_{ind} = \frac{1}{k}$
    $k =$ the number of age-eligible people in contacted and co-operating household

10'. $P(selected\ people|rostered\ HHs) = P_{sel} = \frac{\#\ of\ people\ selected}{\#\ of\ people\ rostered*} = \frac{n_{sel}}{n_{rost}}$

11. $P(individual\ agrees|selected\ people) = P_{agr} = \frac{\#\ of\ people\ agreeing}{\#\ of\ people\ selected} = \frac{n_{agr}}{n_{sel}}$
12. $P(full\ eligible|individual\ agrees) = P_{fe} = \frac{n_{fe}}{n_{agr}}$
13. $P(coord|agreeing, full\ eligible) = P_{coord} = \frac{n_{coord}}{n_{fe}}$ Provides contact information ('co-ordinates')
14. $P(CLSA\ contact|coord) = Contact\ Rate = P_{CLSA} = \frac{n_{CLSA}}{n_{coord}}$
15. $P(agree\ to\ participate|CLSA\ contact) = P_{agree} = \frac{n_{agree}}{n_{CLSA}}$
16. $P(interview|agree\ to\ participate) = P_{int} = \frac{n_{int}}{n_{agree}}$
17. $P(consent\ form\ returned|interview) = P_{cons} = \frac{n_{cons}}{n_{int}}$
18. $P(full\ participant|prerecruit) = Full\ Participation\ Rate = P_{fullpart} = \quad P_{agree} \times P_{int} \times P_{cons} = \frac{n_{cons}}{n_{coord}}$

Estimated number of eligible people=$N \times P_{val} \times P_{res} \times P_{age} \times \frac{n_{age/sex}}{n_{rost}} \times P_{fe}$

Estimated number of eligible HHs=$N \times P_{val} \times P_{res} \times P_{age}$

HH Enumeration Rate=$P_{ans} \times P_{age.est} \times P_{rost}$

Pre-Recruitment Rate= HH Enumeration Rate$\times P_{agr} \times P_{coord}$

Conversion Rate=P_{conv}=$P_{CLSA} \times P_{fullpart}$

Response Rate=Pre-Recruitment Rate$\times P_{conv}$= $P_{ans} \times P_{age.est} \times P_{rost} \times P_{agr} \times P_{coord} \times P_{conv}$

Note: Multiplication of probabilities assumes Markov dependence of the separate components. To the extent this assumption is wrong, the final rates are not necessarily interpretable as probabilities.

* Number of people in the rostered households.

Appendix 5. Provincial Health Registry Sample Weight Calculation

The numbers given in Appendix 2 and the response rates for each sex*age group in province are used to get the basic sample weights.

I. way:

$$\boldsymbol{Wts} = \left[\frac{n_h - x_h}{N_h - \hat{X}_h} \times A_h \times B_h \times C_h\right]^{-1} = \left[\frac{n_h - x_h}{N_h - \hat{X}_h} \times \frac{r_h}{n_h - x_h} \times \frac{s_h}{r_h} \times \frac{t_h}{s_h - Y_h}\right]^{-1} = \frac{(s_h - Y_h) \times (N_h - \hat{X}_h)}{s_h \times t_h}$$

where $N_h - \hat{X}_h = \frac{N_h}{n_h} \times (n_h - x_h) = E_h$estimate of 'eligible' at Ministry.

II. way:

Number of people contacted are s_h and $s_h - Y_h$ is number of eligible people...

New estimate of total number of eligible people in the population $E_h^l = E_h \times \frac{(s_h - Y_h)}{s_h}$

$\boldsymbol{Wts} = \frac{E_h^l}{t_h} = \frac{(s_h - Y_h) \times (N_h - \hat{X}_h)}{s_h \times t_h}$

Appendix 6. Combining Sample Frames

Calibrated weights for each sample were combined within stratum by using extended form of general addition rule of probability.

General rule of probability: P(AUB)=P(A)+P(B)-P(A∩B).

The extended form for three samples in Tracking cohort:

P(AUBUC) =P(A)+P(B)+P(C)-P(A∩B)-P(A∩C)-P(B∩C)+P(A∩B∩C), where

P(A)=P(CCHS), inverse of the weight of a participant coming from the CCHS sample in a stratum

P(B)=P(HR), inverse of the weight of a participant coming from the HR sample in a stratum

P(C)=P(TS), inverse of the weight of a participant coming from the (combined) TS sample in a stratum

Multipication rule of probability P(A∩B)=P(A)×P(B) is only valid for independent events and can be extended to 3 samples.

Table 10 shows the methods applied in each province to recruit Tracking cohort participants. For a sample not used in a province, probability of being in the sample becomes zero.

P($\tilde{A}$) =P($\widetilde{CCHS}$) : Inverse of average weights of participants coming from the CCHS sample in the same stratum.

P($\tilde{B}$) =P($\widetilde{HR}$) : Inverse of average weights of participants coming from the HR sample in the same stratum.

P($\tilde{C}$) =P($\widetilde{TS}$) : Inverse of average weights of participants coming from the TS sample in the same stratum.

Let W_{Tr} be the combined weight for a participant in the CLSA tracking cohort.

- For a participant coming from the CCHS sample,

W_{Tr}=(P(A)+P($\tilde{B}$)+P($\tilde{C}$)-P(A∩$\tilde{B}$)-P(A∩$\tilde{C}$)-P($\tilde{B}$∩$\tilde{C}$) +P(A∩$\tilde{B}$∩$\tilde{C}$))$^{-1}$

- For a participant coming from the HR sample,

W_{Tr}=(P($\tilde{A}$)+P(B)+P($\tilde{C}$)-P($\tilde{A}$∩B)-P($\tilde{A}$∩$\tilde{C}$)-P(B∩$\tilde{C}$)+P($\tilde{A}$∩B∩$\tilde{C}$))$^{-1}$

- For a participant coming from the TS sample,

W_{Tr}=(P($\tilde{A}$)+P($\tilde{B}$)+P(C)-P($\tilde{A}$∩$\tilde{B}$)-P($\tilde{A}$∩C)-P($\tilde{B}$∩C)+P($\tilde{A}$∩$\tilde{B}$∩C))$^{-1}$

For the Comprehensive cohort, we applied the same procedure to the two sampling frames; HR and TS.

FIGURES

Figure 1.

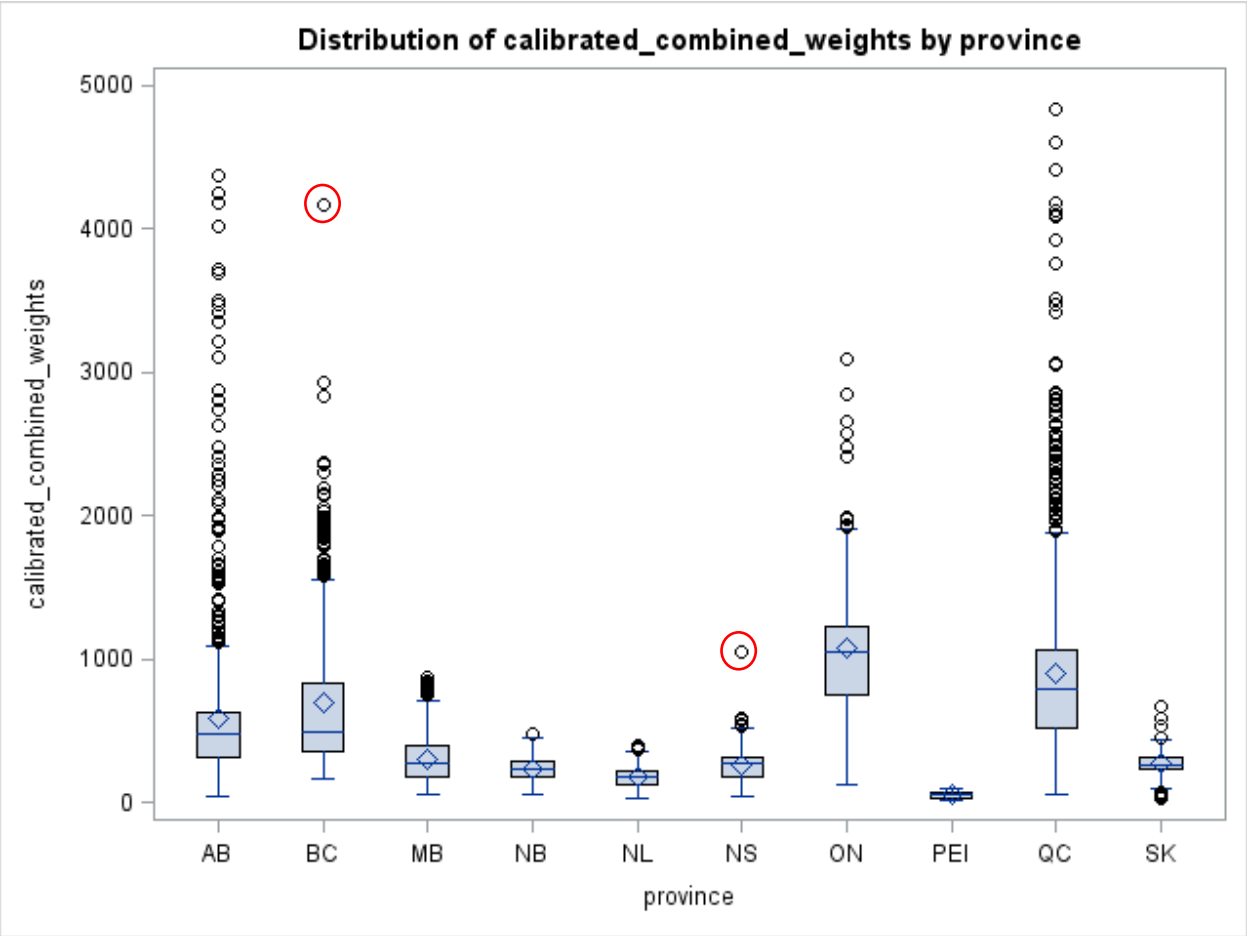

Figure 1. Box and Whisker Plot of the weights within province in the Tracking cohort. The outlier weights (three in BC and two in NS) are shown in red circles.

Figure 2.

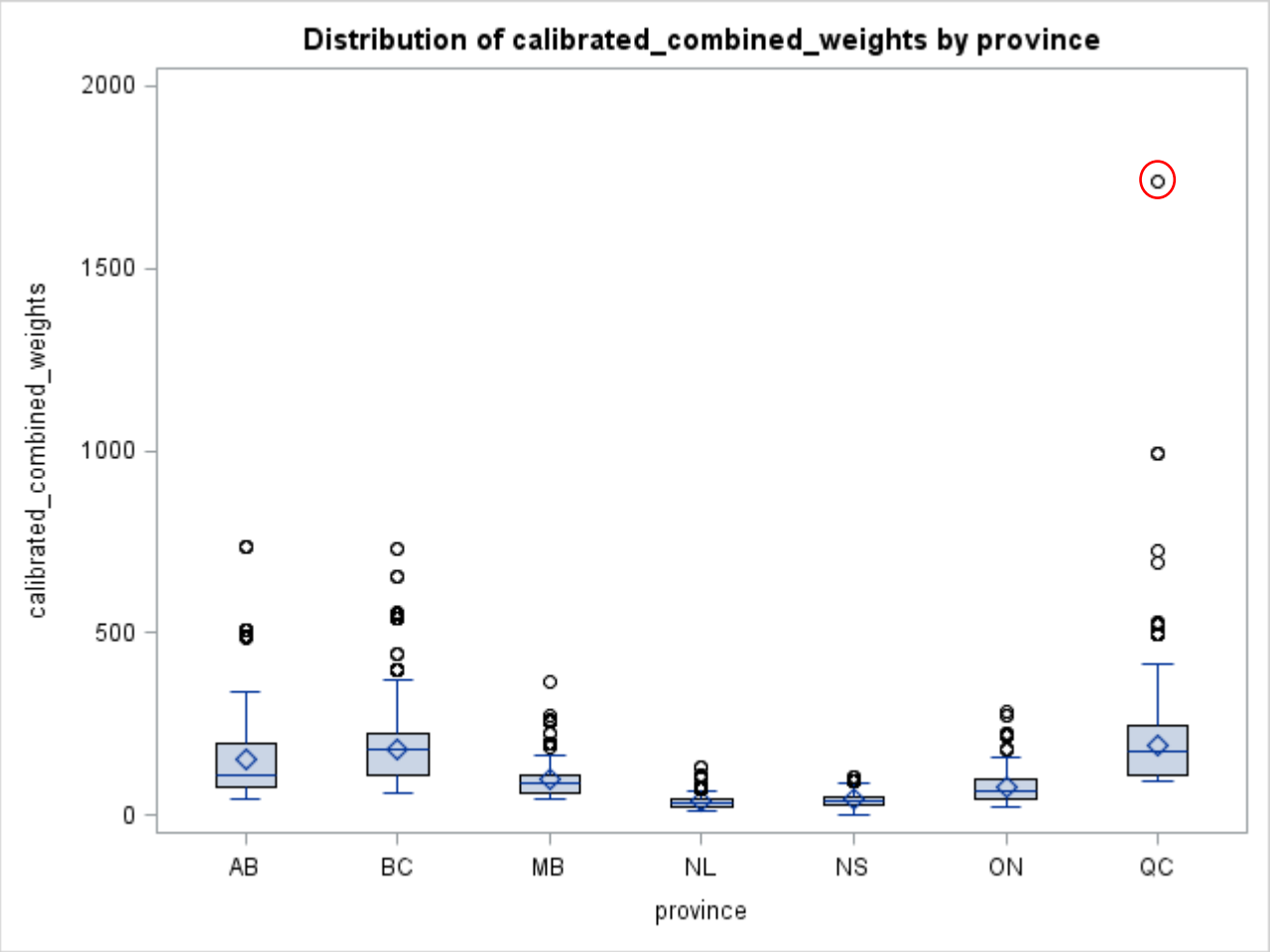

Figure 2. Box and Whisker Plot of the weights within province in the Comprehensive cohort. The outlier weight (one in QC) is shown in the red circle.

Figure 3.

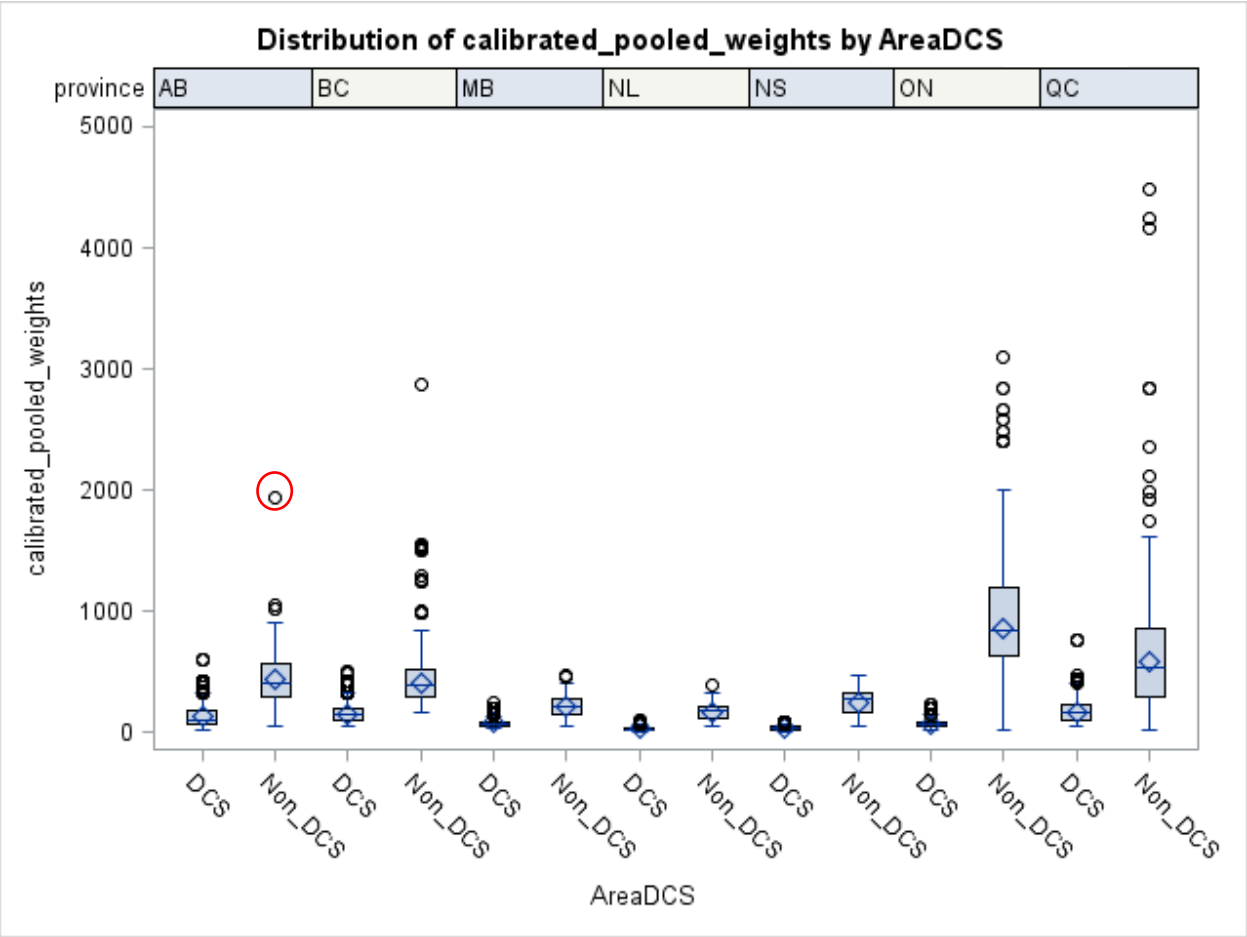

Figure 3. Box and Whisker Plot of the weights by DCS and non-DCS within province for the pooled CLSA data. The outlier weight (one in AB non-DCS) is shown in the red circle.

REFERENCES

Statistics Canada 2011. Census Dictionary, Census Year 2011. Catalogue no. 98-301-X2011001.

Statistics Canada [2010]. Residential Telephone Service Survey (RTSS)

Raina P, et al. Canadian Longitudinal Study on Aging (CLSA) Protocol. https://clsa-elcv.ca/doc/511. Accessed 03.11.2016.