National Population Health Survey

Asthma Supplement

Cycle 2 (1996-1997)

Userguide

Statistics Canada

Table of Contents

Page

Changes To Content

The following changes made to the asthma questionnaire after Quarter 1 collection should be kept in mind during analysis.

1. Questions 4 to 7 were redesigned. Both the question wording and question order were revised. Although data for these questions from Quarter 1 (that is, records with SPA6_CPA = 1) were reformatted during processing to the subsequent quarter variables, differences may remain which are due to the different question design.

2. The category, “a change in temperature or weather” was added to Question 10A which asks for asthma triggers. As a result, the variable CCA6_10Q has been coded as a “9" (not stated) for records with data collected in Quarter 1.

3. Question 10C which asks for the types of fumes or gases which trigger asthma was open-ended in Quarter 1 with responses coded as part of processing. In Quarters 2-4, the question response categories were precoded.

4. Question Proxy2 was omitted in the computer application for Quarter 1 collection. As a result, the variable CCA6_PX2 has been coded as a “9" (not stated) for records in that quarter.

In addition, an error in the computer application for Quarter 4 collection led to the drug code not being saved. A successful paper and pencil follow-up was conducted which recollected that information. However, the drug code variables (CCA6_M1A to CCA6_M7A) have more responses of “don’t know”, “refusal” and “not stated” for records with data collected in Quarter 4 than for those of previous quarters.

1. Sample Design

The sample for the Asthma Supplement of the National Population Health Survey (NPHS) was taken from among respondents to the NPHS. The NPHS is a longitudinal health survey of persons who live in Canada.

The first cycle of the NPHS took place in 1994-1995. During this cycle, a member of each household selected for the survey has been selected for the longitudinal panel. Those longitudinal respondents will be interviewed every two years for a maximum of 20 years. During the second cycle, all longitudinal respondents and their current household members have been interviewed again. As there are no new longitudinal respondents selected during the second cycle, all the longitudinal respondents are two year old or older at the time of the second cycle. The target population of the NPHS consists of residents of households in all provinces except for persons living on Indian reserves, Canadian Forces bases and certain distant regions of Quebec and Ontario. The sample design of the NPHS is described in detail in the *1996/97 NPHS Public Use Microdata Documentation*.

The sample for the Asthma Supplement of the NPHS consists of all individuals who have indicated, in the main survey of 1996-1997, that they had asthma diagnosed by a health professional. This question was asked to every individual 12 year and over and children 2 to 11 year old that were selected as longitudinal respondents by the NPHS in 1994-1995. Among those individuals 12 year old and over who indicated having asthma, some have also been selected as longitudinal respondents for NPHS.

During the NPHS of 1996-1997, Statistics Canada received funds from some provincial governments to increase the sample size in their respective provinces to allow for estimates by health region. The households selected as a result of these funds were not included in the sample for the Asthma Supplement.

The sample size will allow for estimates to be calculated at the national level but not at the provincial level. It is also important to mention that estimates by age group will only be possible if the number of observations in these groups is sufficiently large.

Sample size[1] by province

Province	Sample size
Newfoundland	154
Prince Edward Island	141
Nova Scotia	156
New Brunswick	162
Quebec	442
Ontario	664
Manitoba	194
Saskatchewan	154
Alberta	249
British Columbia	239
Total	2,555

[1] Number of people reporting asthma diagnosed by a health professional during the second cycle of NPHS.

2. Weighting

The principle behind estimation in a probability sample such as the NPHS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population. In the terminology used here, it can be said that each person has a weight of 50.

The weighting phase is a step which calculates, for each person, their associated weight. This weight appears on the file and must be used to derive meaningful estimates from the survey. For example, if the number of individuals in Canada for which asthma is brought on or made worse by dust (see the example in section 4.2) is to be estimated, it is done by selecting the records referring to those individuals in the sample having that characteristic and summing the weights for those records.

Details of the method used to calculate these weights are presented in section 5.

3. Survey Errors

The survey produces estimates based on information collected from and about a sample of individuals. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used in the survey. 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 estimate.

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems with the CAI questionnaire or with the understanding of instructions and procedures to ensure that data collection errors were minimized.

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. In the case of the Asthma supplement, partial non-response was basically non-existent (only 23 cases); once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households who responded to the survey to compensate for those who did not respond.

In most cases, partial non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, could not recall the requested information, or could not provide non-proxy information.

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this file to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (C.V.) of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose that, based upon the survey results, one estimates that approximately 64% of current asthmatics have asthma which is brought on or made worse by dust and that the estimate is found to have standard error of .03. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{.03}{.64}\right) x\ 100\% = 4.69\%$$

4. Guidelines For Tabulation, Analysis And Release

This section of the documentation outlines the guidelines to be adhered to by users tabulating, analyzing, publishing or otherwise releasing any data derived from the survey files. With the aid of these guidelines, users should be able to produce figures that are in close agreement with those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

4.1 Rounding Guidelines

In order that estimates for publication or other release derived from these files correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

4.2 Sample Weighting Guidelines for Tabulation

The sample design used for the NPHS and hence for the Asthma supplement is not self-weighting. That is to say, the sampling weights are not identical for all individuals in the sample. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

If proper weights are not used, the estimates derived from the files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

4.2.1 Definitions of Types of Estimates: Categorical vs. Quantitative

Before discussing how the data from the Asthma supplement can be tabulated and analyzed, it is useful to describe the two main types of point estimates of population characteristics which can be generated from the files.

<u>Categorical Estimates:</u>

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of individuals for which asthma is brought on or made worse by dust is an example of such an estimate. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Example of Categorical Question:

CCA6_10A: Is your asthma brought on or made worse by dust?

__ Yes
__ No

Quantitative Estimates:

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X} / \hat{Y}$ where $\hat{X}$ is an estimate of surveyed population quantity total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of visits to a doctor in the last 12 months among the individuals who do not take their medicines as prescribed. The numerator is an estimate of the total number of visits to a doctor in the last 12 months among the individuals who do not take their medicines as prescribed, and the denominator is an estimate of the number of individuals who do not take their medicines as prescribed.

Example of Quantitative Question:

CCA6_25: In the last 12 months, how many times have you visited any type of doctor for asthma?

|_|_| Number of times

4.2.2 Tabulation of Categorical Estimates

Estimates of the number of people with a certain characteristic can be obtained from the file by summing the final weights of all records possessing the characteristic(s) of interest. Proportions and ratios of the form $\hat{X} / \hat{Y}$ are obtained by:

a) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$),
b) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then
c) dividing the numerator estimate by the denominator estimate.

4.2.3 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of visits to a doctor in the last 12 months among the individuals who do not take their medicines as prescribed, multiply the value reported in question CCA6_25 by the final weight for the record, then sum this value over all records with a response of 'No' to CCA6DMCO.

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of visits to a doctor in the last 12 months among the individuals who do not take their medicines as prescribed,

a) estimate the total number of visits to a doctor in the last 12 months among the individuals who do not take their medicines as prescribed as described above,

b) estimate the number of individuals who do not take their medicines as prescribed by summing the final weights of all records with a response of 'No' to CCA6DMCO, then

c) divide estimate (a) by estimate (b).

4.3 Guidelines for Statistical Analysis

The National Population Health Survey is based upon a complex design, with stratification and multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differs from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, analysis of variance), a method exists which can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is one (1), then the results produced by the standard packages will be more reasonable; they still will not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. The rescaling can be accomplished by using in the analysis a weight which is equal to the original weight divided by the average of the original weights for the sampled units (people) contributing to the estimator in question.

4.4 Release Guidelines

Before releasing and/or publishing any estimate from these files, users should first determine the number of sampled respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines below.

Sampling Variability Guidelines

Type of Estimate	**C.V. (in %)**	**Guidelines**
1. Acceptable	0.0 - 16.5	Estimates can be considered for general unrestricted release. Requires no special notation.
2. Marginal	16.6 - 33.3	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter M (or in some other similar fashion).
3. Unacceptable	greater than 33.3	Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or in some other fashion) and the following warning should accompany the estimates: "The user is advised that . . .(specify the data) . . . do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data."

5. Weighting

Two weights are necessary for the Asthma Supplement. One weight is used for the variables of both the general and asthma questionnaires that are administered to all persons selected for the Asthma Supplement. A second weight is needed for the variables of the health questionnaire administered to persons selected by the NPHS as longitudinal respondents.

The Asthma Supplement uses as a base weight the cross-sectional weights calculated for the NPHS. A description of the process for calculating these weights is found in the *1996/97 NPHS Public Use Microdata Documentation*. The following sections describe the adjustments that must be made to the weight for the Asthma Supplement

5.1 Household member weight

This weight is calculated for all Asthma Supplement respondents.

5.1.1 Basic weight

The households that have been added to augment the sample sizes of the NPHS in certain provinces were not retained for the Asthma Supplement. The adjustment for the integration of the two frames described in the process of the *1996/97 NPHS Public Use Microdata Documentation* is, therefore, not calculated.

Individuals 12 years old and over complete the general questionnaire in the main survey and all those reporting asthma were included in the sample for the Asthma Supplement. The base weight for those individuals is therefore the cross-sectional household member weight after the household member non-response adjustment (Adjustment 4A). Only those 2 to 11 year olds who are longitudinal respondents were asked the chronic conditions section of the general questionnaire in the main survey. Hence, the base weight for these individuals is the selected respondent cross-sectional weight after the interprovincial migration adjustment (Adjustment 2B).

5.1.2 Post-stratification adjustment

This adjustment ensures that the sum of final weights by province, age and sex are equal to the population totals for 1996. The following age groups are used: 2-11, 12-24, 25-44, 45-64, 65 and older. The adjustment is made as follows:

$$\frac{\textit{demographic projections for a province-age-sex group}}{\textit{sum of weights of respondents of an NPHS household for a province-age-sex group}}$$

5.1.3 Asthma Supplement non-response adjustment

In spite of his/her efforts, it is inevitable that the interviewer will not receive a response from certain people for the following reasons: refusal, special circumstances, language problem, temporarily absent or computer problem. One compensates for non-response by proportionately adjusting the weighting factors of responding members. Persons out-of-scope for the survey are considered as respondents. The adjustment is made as follows:

$$\frac{\textit{sum of weights of members sampled for the Asthma Supplement for the age-sex group}}{\textit{sum of weights of members responding to the Asthma Supplement for the age-sex group}}$$

At this point, the persons who were established as being out-of-scope for the survey are no longer considered in the weighting process.

5.1.4 Data sharing adjustment

Statistics Canada must obtain from respondents permission to provide the information collected to Health Canada. This permission must be given for the information collected in the general questionnaire and for the Asthma Supplement. If the respondent refuses to give either permission, no information concerning that person is disclosed to Health Canada. This holds true for respondents who have only partially answered the Asthma Supplement because the question concerning the sharing of information was not asked of these persons. One must then adjust the weight of respondents who agree to share their information to compensate for those who have refused.

$$\frac{\textit{sum of weights of respondents (complete and partial) for the age-sex group}}{\textit{sum of weights of respondents (complete) for the age-sex group who have agreed to share the information}}$$

The final weight, which we will call the **Household member weight**, is the base weight that is multiplied by each of the adjustments described above. This weight is used to analyze variables from both the general and Asthma Supplement questionnaires.

5.2 Panel member weight

The same adjustments are made as for the household member weight. However, this time, only persons selected as longitudinal respondents for the NPHS who are part of the Asthma Supplement are considered.

5.2.1 Basic weight

The weight used is the persons selected weight. The households added to increase the NPHS sample size in some provinces are not retained for the Asthma Supplement. The adjustment for the integration of the two frames described in the process of the *1996/97 NPHS Public Use Microdata Documentation* is, therefore, not calculated. The base weight used for the panel member is the cross-sectional weight of persons selected after the inter-provincial migration adjustment (Adjustment 2B).

5.2.2 Post-stratification adjustment

This adjustment ensures that the sum of weights by province, age and sex are equal to the population totals in 1996. The age groups used are: 2-11, 12-24, 25-44, 45-64, 65 and older. The adjustment is made as follows:

$$\frac{\textit{demographic projections for a province-age-sex group}}{\textit{sum of weights of panel members of the NPHS for a province-age-sex group}}$$

5.2.3 Asthma Supplement non-response adjustment

As mentioned previously, it is inevitable that the interviewer not obtain a response from some persons. The adjustment is made as follows:

$$\frac{\textit{sum of weights of panel members in the Asthma Supplement for the age-sex group}}{\textit{sum of weights of respondents of the panel in the Asthma Supplement for the age-sex group}}$$

At this point, persons who have been established as out-of-scope for the survey are no longer considered in the weighting process.

5.2.4 Data sharing adjustment

Statistics Canada must obtain from respondents permission to provide information collected to Health Canada. Apart from being granted for information collected on the general questionnaire and for the Asthma Supplement, the respondent must also give his/her permission to share information collected on the health questionnaire. If the respondent refuses the latter, the information on them regarding the health component will not be divulged to Health Canada. An adjustment must be made for these respondents.

$$\frac{\textit{sum of weights of responding panel members (complete and partial) for the age-sex group}}{\textit{sum of weights for responding panel members (complete) for the age-sex group who have agreed to share the information}}$$

The final weight, which we will call the **Panel member weight**, is the base weight that is multiplied by each of the adjustments described above. This weight is used for the analysis of variables on asthma when they also include variables from the health component.

6. Use of weights

In some cases, respondents to the Asthma Supplement of the NPHS answered two, and sometimes three, questionnaires during the survey. All respondents received the general questionnaire (HO5) and the questionnaire on asthma. Among the persons identified for the Asthma Supplement, some were persons selected in the NPHS as panel members. These persons, therefore, received a third questionnaire during the main survey, namely the health component (HO6) questionnaire. Two weights are, therefore, necessary to analyze the data.

6.1 **Household member weight** (WTA6_5S)

This weight is used to analyze the variables of the general component and of the Asthma Supplement if no variables of the health component are included. This is the weight included on the ASMA5S file. This file includes all respondents to the Asthma Supplement who agreed to share the information with Health Canada. The variables of the general questionnaire and of the Asthma Supplement are included on this file.

6.2 **Panel member weight** (WTA6_6S)

This weight is used for all analyses of the data of the Asthma Supplement which also includes variables of the health component. This is the weight that appears on the ASMA6S file. This file consists only of respondents to the Asthma Supplement who are also panel members in the NPHS and who agreed to share the information with Health Canada. The variables of the general questionnaire, of the Asthma Supplement and of the health component are included on this file.

7. Use of data

The microdata files of the Asthma Supplement of the NPHS contain variables from the general and health components of the main survey. They are included to allow these variables to be analyzed in relation to the variables from the Asthma Supplement. **N.B.** Under no circumstances are they to be used to obtain estimates of the general population.

The sample size on the general component file (ASMA5S) is larger than the sample size on the health component file (ASMA6S). Hence, the general component file should be used to calculate estimations whenever it is possible. Estimates calculated using the general component file will have a variance lesser than estimates calculated using the health component file.

It is possible to calculate prevalence estimates at the post-strata level used to do the post-stratification for the Asthma Supplement. The post-stratification has been done at the province-age-sex level. The population sizes for those post-strata are presented in tables 7.1, 7.2, and 7.3. These sample sizes should be used as the denominators for prevalence estimates.

For example, to calculate the prevalence of people with asthma in Canada, one first calculates the number of persons with asthma in Canada by summing the weight of records of individuals who indicated having asthma diagnosed by a physician. This total is 1,715,907. To obtain the prevalence as a percentage, the total number of persons with asthma has to be divided by the total population of Canada, 28,536,553, found in table 7.1. We then obtain the prevalence of asthma in Canada to be (1,715,907/28,536,553)*100=6.01%.

Table 7.1 Population sizes by post-strata

	Age Group					Total
	2-11	12-24	25-44	45-64	65 +	
Newfoundland	71487	115370	184644	120501	57320	549322
Prince Edward Island	19666	25869	41596	28578	16613	132322
Nova Scotia	121356	164383	293311	202916	113948	895914
New Brunswick	96110	138348	240998	161917	90745	728118
Quebec	916374	1240940	2395323	1662263	832628	7047528
Ontario	1516370	1908912	3724577	2386127	1303741	10839727
Manitoba	153910	195117	339286	222292	145433	1056038
Saskatchewan	147396	184996	288714	189943	137462	948511
Alberta	408812	510842	941502	534882	256756	2652794
British Columbia	490183	649376	1259209	826051	461460	3686279
Canada	3941664	5134153	9709160	6335470	3416106	28536553

Table 7.2 Population sizes by post-strata - Females

	Age Group					Total
	2-11	12-24	25-44	45-64	65 +	
Newfoundland	34880	56635	92082	59544	31383	274524
Prince Edward Island	9452	12981	20891	14190	9368	66882
Nova Scotia	58928	81096	150224	101784	65858	457890
New Brunswick	46969	67776	121485	80821	51537	368588
Quebec	447556	608650	1188567	842546	485073	3572392
Ontario	740518	937084	1872135	1205615	737861	5493213
Manitoba	74690	95645	168539	111501	83454	533829
Saskatchewan	72112	90895	143875	94480	76438	477800
Alberta	199294	249554	464665	263276	141787	1318576
British Columbia	240060	319568	632982	408932	254674	1856216
Canada	1924459	2519884	4855445	3182689	1937433	14419910

Table 7.3 Population sizes by post-strata - Males

	Age Groups					Total
	2-11	12-24	25-44	45-64	65 +	
Newfoundland	36607	58735	92562	60957	25937	274798
Prince Edward Island	10214	12888	20705	14388	7245	65440
Nova Scotia	62428	83287	143087	101132	48090	438024
New Brunswick	49141	70572	119513	81096	39208	359530
Quebec	468818	632290	1206756	819717	347555	3475136
Ontario	775852	971828	1852442	1180512	565880	5346514
Manitoba	79220	99472	170747	110791	61979	522209
Saskatchewan	75284	94101	144839	95463	61024	470711
Alberta	209518	261288	476837	271606	114969	1334218
British Columbia	250123	329808	626227	417119	206786	1830063
Canada	2017205	2614269	4853715	3152781	1478673	14116643

8. Variance estimation using boostrap weights

8.1 Bootstrap resampling method

Due to the complexity of the sample design, a resampling technique is used to calculate estimates of the variance. For NPHS as well as the Asthma supplement, the bootstrap resampling method is used.

This technique involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimates from replicate to replicate. The replicates are formed in the following manner: within each stratum, a simple random sample (SRS) is taken, with replacement, of (n-1) of the n clusters in that stratum. For each SRS sample, we recalculate the survey weight for each record in that stratum. This recalculated weight is the first bootstrap weight. We then repeat the process B times (B=1,000 for the Asthma supplement), forming a set of B bootstrap weights for each record on the sample file.

8.2 Bootstrap weights for the Asthma Supplement

Two files (WGT51000 and WGT61000) containing each 1,000 bootstrap weights are included that correspond to each of the data files (A5S and A56S). These weights must be used to calculate an estimation of the variance for a given estimate. For example, if one wants to calculate an estimate for a total, one must:

A) first calculate an estimation of the total using the final weight (WTA6_5S or WTA6_6S);

B) then calculate an estimation of the total needed using each of the 1,000 bootstrap weights as the weighting factor. This gives 1,000 estimates of the total needed;

C) finally, calculate the variance between those 1,000 estimates. This variance is an estimate of the variance of the total calculated in A).

8.3 Program BOOTVAR.SAS

The SAS program BOOTVAR.SAS given with the data files is an example of a program to calculate an estimate of the variance. Users may use this program or they can create their own program. A copy of the program appears at the end of this section.

This program calculates an estimate (for example, a total) and a corresponding estimate of the variance. The user should always calculate the desired estimate (in this case the total) using another program, in order to verify that the program BOOTVAR.SAS has correctly calculated this estimate.

In this section, we describe how to prepare the data for the program BOOTVAR.SAS, and how to use it. Two examples are given as well as the time required to run the program. (Please note that these programs were run on a 266 MHz Pentium Pro, with 64 MB of RAM.)

8.3.1 Definition of variables for the analysis

The program BOOTVAR.SAS permits the calculation of variance estimates for totals, for ratios, for differences of ratios, and for linear and logistic regression parameters. To obtain the desired analyses, the data must first be prepared and saved in a SAS file. For estimates of totals, ratios, and differences of ratios, the estimates are calculated by summing the weights of the records with the characteristic of interest. An indicator variable must be created to identify these records (see example 1A). For regression analyses, the variables to be analysed must be defined according to SAS conventions (see example 1B).

A supplementary variable (REGION) may be necessary if the analyses are done for geographic subgroups of records on the new datafile created for the program BOOTVAR.SAS. For example, if the new analysis file contains all the records of persons aged twelve and older in Canada, the file represents a group that corresponds to Canada. If the analysis is done at the Canada level, the variable REGION is not necessary. However, if the analysis is done at the province level, then the REGION variable is necessary. This variable must be created by the user and, for each record, will take the value of the geographic region of this record—in this case, the province. As a second example, if the new analysis file contains all the records of persons less than twelve years of age in Ontario, the file represents a group that corresponds to the province of Ontario. If the analysis is done at the Ontario level, the REGION variable is not necessary. In the same way, if the new analysis file contains all the records of persons aged twelve and older in the provinces of Ontario, Manitoba, and Alberta, the REGION variable will be necessary if the analysis is done at the province level.

The following example shows how to define the variables that will subsequently be used to calculate variance estimates with the aid of BOOTVAR.SAS.

Example 1: Definition of variables for the analysis

Example 1A: Totals and ratios

```
/**********************************************************************/
/*  USE THE HOUSEHOLD COMPONENT FILE OF THE ASTHMA SUPPLEMENT
(A5S.TXT)  */
/*  TO CALCULATE THE NUMBER OF PERSONS WHO HAD AN ASTHMA
ATTACK IN THE   */
/*  LAST 12 MONTHS                                    */
/*  FOR THE AGE GROUPS 12-19, 20-34, 35-49, ET 50 ET PLUS           */
/*  BY PROVINCE                                       */
/**********************************************************************/
/*  UTILISE LE FICHIER DE LA COMPOSANTE MÉNAGE DU SUPPLÉMENT
SUR L'ASTHME*/
/*  (A5S.TXT)                                    */
/*  POUR CALCULER LE NOMBRE DE PERSONNES QUI ONT EU DES CRISES
D'ASTHME  */
/*  AU COURS DES 12 DERNIERS MOIS                        */
/*  POUR LES GROUPES D'ÂGES 12-19, 20-34, 35-49, ET 50 ET PLUS         */
/*  PAR PROVINCE                                      */
/*********************************************************************
libname in1 'c:\data';

data in1.asthma_5;
  %let datafid="c:\data\a5s.txt";
  %include "c:\data\a5s_in.sas";

if dhc6_age < 12 then delete;

region=prc6_cur;

/* Had asthma attacks in the last 12 months              */
/* A eu une crise d'asthme au cours des 12 derniers mois */
  if cca6_18=1 then attack=1;
  else if cca6_18=2 then attack=0;
  else attack=.;

/* Medication compliance              */
/* Prend les médicaments comme prescrit */
  if cca6dmco=1 then comp=1;
  else if cca6dmco=2 then comp=0;
```

```
  else comp=.;

/* Age groups   */
/* Groupe d'âge */
  if 12 <= dhc6_age <= 19 then agegrp=1;
  else if 20 <= dhc6_age <= 34 then agegrp=2;
  else if 35 <= dhc6_age <= 49 then agegrp=3;
  else if dhc6_age >= 50 then agegrp=4;

  age1219=0;
  age2034=0;
  age3549=0;
  age50=0;

  if agegrp=1 then age1219=1;
  else if agegrp=2 then age2034=1;
  else if agegrp=3 then age3549=1;
  else if agegrp=4 then age50=1;

/* Cigarettes smoking */
/* Usage du tabac     */
  if cca6_36=1 or cca6_36=2 then smoking=1;
  else if cca6_36=3 then smoking=0;
  else smoking=.;

/* Urban/Rural indicator   */
/* Indicateur Urbain/Rural */
  if ge36durb=1 then urban=1;
  else if ge36durb=2 then urban=0;

/* Asthma attacks*age group    */
/* Crise d'asthme*groupe d'âge */
  if attack=1 then do;
    if agegrp=1 then attage1=1; else attage1=0;
    if agegrp=2 then attage2=1; else attage2=0;
    if agegrp=3 then attage3=1; else attage3=0;
    if agegrp=4 then attage4=1; else attage4=0;
  end;
  else do;
    attage1=0;
    attage2=0;
    attage3=0;
```

```
   attage4=0;
  end;

run;
```

(The execution of this program took less than 1 minute.)

Example 1B: Logistic regression

```
/***********************************************************************/
/*  USE THE HEALTH COMPONENT FILE OF THE ASTHMA SUPPLEMENT
(A56S.TXT)    */
/*  TO STUDY THE RELATION BETWEEN HAVING AN ASTHMA ATTACK IN
THE LAST    */
/*  12 MONTHS AND MEDICATION COMPLIANCE                              */
/***********************************************************************/
/*  UTILISE LE FICHIER DE LA COMPOSANTE SANTÉ DU SUPPLÉMENT SUR
L'ASTHME */
/*  (A56S.TXT)                                        */
/*  POUR ÉTUIDIER LA RELATION ENTRE LE FAIT D'AVOIR EU DES CRISES
    */
/*  D'ASTHME AU COURS DES 12 DERNIERS MOIS ET LA PRISE DE
MÉDICAMENTS    */
/*  COMME PRESCRIT                                   */
//***********************************************************************
libname in1 'c:\data';

data in1.asthma_6;
  %let datafid="c:\data\a56s.txt";
  %include "c:\data\a56s_in.sas";

if dhc6_age < 12 then delete;

/* Had asthma attacks in the last 12 months            */
/* A eu une crise d'asthem au cours des 12 derniers mois */
  if cca6_18=1 then attack=1;
  else if cca6_18=2 then attack=0;
  else attack=.;

/* Medication compliance               */
/* Prend les médicaments comme prescrit */
  if cca6dmco=1 then comp=1;
```

```
  else if cca6dmco=2 then comp=0;
  else comp=.;

/* Age groups   */
/* Groupe d'âge */
  if 12 <= dhc6_age <= 19 then age19=1;
  else age19=0;

  if 20 <= dhc6_age <= 34 then age2034=1;
  else age2034=0;

  if 35 <= dhc6_age <= 49 then age3549=1;
  else age3549=0;

  if dhc6_age >= 50 then age50=1;
  else age50=0;

/* Cigarettes smoking */
/* Usage du tabac     */
  if cca6_36=1 or cca6_36=2 then smoking=1;
  else if cca6_36=3 then smoking=2;
  else smoking=.;

/* Urban/Rural indicator   */
/* Indicateur Urbain/Rural */
  if ge36durb=1 then urban=1;
  else if ge36durb=2 then urban=0;

/* Health status */
/* État de santé */
  if ghc6_1=1 or ghc6_1=2 or ghc6_1=3 then health=1;
  else if ghc6_1=4 or ghc6_1=5 then health=0;
  else health=.;
run;

proc logistic data=in1.asthma_6 descending;
  model attack=comp age2034 age3549 age50 smoking urban health;
  weight wta6_6s;
  title 'Asthma attacks';
  title2 'Crises d''asthme';
run;
```

(The execution of this program took less than 1 minute.)

8.3.2 Use of the program BOOTVAR.SAS

Once the new analysis file has been created, BOOTVAR.SAS can be run. The user must however complete the program by specifying the parameters and by indicating the list of variables to analyse. BOOTVAR.SAS is divided into four sections.

In **section 1**, the users must specify the following:

- the number of bootstrap weights used (1000 for AS5 and AS56),
- the subdirectory containing the new analysis file (created by the user as in example 1), the bootstrap weight file, and the file that will contain the results,
- the name of the new analysis file,
- the name of the bootstrap weight file, and
- the parameters that indicate if the analysis is done at a geographic level, using the REGION variable.

In **section 2,** the only changes to make are to the macro for the general linear model. This macro must be modified by the user according to the analysis required.

In **section 3**, the user must specify the list of variables to keep for the analysis. BOOTVAR.SAS runs quickly for variance estimates of totals, ratios, and differences of ratios. However, more time is required for variance estimates of regression parameters. To reduce the run time, it is therefore recommended to keep only those variables necessary for the desired analysis.

In addition to the analysis variables, two variables must be kept on the file, REALUKEY and PERSONID. These variables are needed to match the survey datafile and the bootstrap weight file during the running of the program. Please note that BOOTVAR.SAS reads two files: the survey data and the bootstrap weights, and that these two files must be SAS files.

Finally, in **section 4**, the user must indicate which analyses will be run and for which variables.

The following examples indicate in ***italics*** those parts of the program that the user must specify for variance estimates of totals, ratios, and logistic regression parameters. Please note that only those parts of the program where changes are necessary are presented. It is necessary, however, to submit the entire program.

Example 2: BOOTVAR.SAS

Example 2A: Totals and ratios

.
.
.

```
/*************************************************************************/
*** Section 1: Declaration of the macro variables                    ***/
/*************************************************************************/

%let R=01;                /* bootstrap averaging parameter (fixed) */

/*************************************************************************/
/*** USER SPECIFIES THE NUMBER OF BOOTSTRAP WEIGHTS USED,
DIVIDED BY 100   ***/
/*** %let B=xx;                              ***/
/*************************************************************************/

%let B=10;                    /* bootstrap Bx100 times */
/*************************************************************************/
/*** USER SPECIFIES THE DIRECTORY CONTAINING THE ANALYSIS FILE
(in1)  ***/
/*** AND THE BOOTSTRAP WEIGHTS FILE (in2) SPECIFIED BELOW.          /
/*** THE OUTPUT FILES WILL BE SAVED IN THE DIRECTORY SPECIFIED
FOR in3 **/
/*************************************************************************/

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";

/*************************************************************************/
/*** USER SPECIFIES THE NAME OF THE SAS FILE CONTAINING THE
ANALYSIS VARIABLES.  */
/*** %let Mfile=xxx;                               */
/*************************************************************************/

%let Mfile=in1.asthma_5;                  /* Main SAS file */
```

```
/************************************************************************/
/*** USER SPECIFIES THE NAME OF THE SAS FILE CONTAINING THE
BOOTSTRAP WEIGHTS  ***/
/*** %let bsamp=wgt51000; (to be used with the file A5S)   OR            ***/
/*** %let bsamp=wgt61000; (to be used with the file A56S)               ***/
/************************************************************************/

%let bsamp=in2.wgt51000;            /* bootstrap weights SAS file */

/*********************************************************************/
/*** USER SPECIFIES THE PARAMETERS by, wo AND kp, WHICH INDICATE /
/*** IF THE ANALYSIS IS DONE AT THE REGIONAL LEVEL OR FOR THE
***/
/*** WHOLE DATASET.                              ***/
/***                                  ***/
/*** IF THE ANALYSIS IS DONE AT THE CANADA LEVEL, THEN          ***/
/***    %let by=*;                         ***/
/***    %let wo=;                          ***/
/***    %let kp=;                          ***/
/***                                  ***/
/*** ELSE IF THE ANALYSIS IS DONE AT THE PROVINCIAL LEVEL, THEN
***/
/***    %let by=;                          ***/
/***    %let wo=*;                          ***/
/***    %let kp=region;                      ***/
/*********************************************************************/

%let by=;
%let wo=*;
%let kp=region;

data time;
  format start datetime16.;
  start=datetime();
  output;
run;
```

```
/*************************************************************************/
/*** Section 2: Declaration of the macros                         ***/
/*************************************************************************/

.
.
.

*************************************************************************;
*  Section 3: MAIN PROGRAM                          *;
*************************************************************************;

/* Read in Main file */

/***********************************************************************
**/
/*** DUE TO THE LARGE NUMBER OF OBSERVATIONS, ONLY THE
VARIABLES      ***/
/*** THAT ARE NEEDED TO DO THE ANALYSES AND THE VARIABLES
"PROVINCE", ***/
/*** "REALUKEY", AND "PERSONID" SHOULD BE KEPT AT THIS POINT.        /
/*************************************************************************/

  data nphs (index=(id=(realukey personid)));
   set &Mfile (keep= realukey personid region attack attage1 attage2 age1219
        age2034);
  run;

/* Read in Bootstrap Weights */
/* FWGT corresponds to WTA6_5S or WTA6_6S, depending on which file is analysed
*/


  data bsamp (index=(id=(realukey personid)));
   set &bsamp;
   keep fwgt realukey personid bsw1-bsw&B.00;
  run;


/* Merge Main file and Bootstrap Weights */

&by data bs_nphs (index=(region));
```

```
&wo data bs_nphs;
     merge nphs (in=in1) bsamp (in=in2);
     by realukey personid;
     if in1;
   run;

  proc datasets library=work;
    delete nphs bsamp;
  run;


/**********************************************************************/
/*** Section 4: Macro calls                                        ***/
/**********************************************************************/

/**********************************************************************/
/*** The variance estimates are calculated using the SAS macros that have been  ***/
/*** defined earlier in the program.  Those SAS macros can be called in the     ***/
/*** program as needed.  A macro call is a SAS statement that specifies the name **/
/*** of the macro called and some parameters.  In this program, the parameters  ***/
/*** indicate which file has to be read, which variables will be used, how many ***/
/*** bootstrap weights are used, and how many bootstrap samples were selected.  ***/
/*** The name of the file to read is always the same (BS_NPHS) and the last two ***/
/*** parameters have already been defined in the first section of the program.  ***/
/*** Hence, the user only has to specify the variables to be used.            ***/
/***                                                              ***/
/*** Each macro call will result in a variance estimate of ONE estimator.  If   ***/
/*** more than one estimator and its variance estimate have to be calculated,   ***/
/*** the macro has to be called multiple times.                       ***/
/***                                                              ***/
/*** A commented statement for the macro call appears where the user's statement***/
/*** should be placed in the program.  The user's statement will              ***/
/*** be identical to the commented one, except for the names of the analysis    ***/
/*** variables.                                                   ***/
/**********************************************************************/

/*********/
/*       */
/* Total */
/*       */
/*********/
  proc datasets library=work; /* initial alltots */
```

```
  delete alltots;
 run;

/**********************************************************************/
/*** To call the macro total, the following statement is used:     ***/
/*** %total(infile,var,bssz=,multi=);                               ***/
/***     where infile : bs_nphs                                      ***/
/***           var    : the variable for which a total is calculated ***/
/***           bssz=  : bssz=&B.00                                   ***/
/***           multi= : multi=&R                                     ***/
/***                                                                 ***/
/**********************************************************************/

 *  %total(bs_nphs,VAR,bssz=&B.00,multi=&R);
   %total(bs_nphs,attack,bssz=&B.00,multi=&R);

/* Delete sas files in total macro */

 proc datasets library=work;
  delete ytot est;
 run;


/*********/
/*      */
/* Ratio */
/*      */
/*********/

/**********************************************************************/
/*** To call the macro ratio, the following statement is used:     ***/
/*** %ratio(infile,var1,var2,bssz=,multi=);                         ***/
/***     where infile : bs_nphs                                      ***/
/***           var1   : the variable of the numerator of the ratio   ***/
/***           var2   : the variable of the denominator of the ratio ***/
/***           bssz=  : bssz=&B.00                                   ***/
/***           multi= : multi=&R                                     ***/
/***                                                                 ***/
/**********************************************************************/


 *  %ratio(bs_nphs,VAR1,VAR2,bssz=&B.00,multi=&R);
```

```
  %ratio(bs_nphs,attage1,age1219,bssz=&B.00,multi=&R);
  %ratio(bs_nphs,attage2,age2034,bssz=&B.00,multi=&R);

/* Delete sas files in ratio macro */
  proc datasets library=work;
   delete ytot xtot est;
  run;

%prnttot;

/*******************************************/
/*** The results of total and ratio macro ***/
/*** can be saved in a permanent file.    ***/
/*******************************************/

  data in3.ZZZ;
   set alltots;
  run;

.
.
.

data time;
 set time;
 format stop datetime16.;
 stop=datetime();
 output;
run;

proc print data=time;
 title 'Time Taken to Run Program';
run;

/* End of SAS program BootVar */
```

(The execution of this program took about 1 minute.)

Example 2B: Logistic regression

.
.
.

```
/**********************************************************************/
/*** Section 1: Declaration of the macro variables                 ***/
/**********************************************************************/

%let R=01;                    /* bootstrap averaging parameter (fixed)*/

/**********************************************************************/
/*** USER SPECIFIES THE NUMBER OF BOOTSTRAP WEIGHTS USED,
DIVIDED BY 100   ***/
/*** %let B=xx;                              ***/
/**********************************************************************/

%let B=10;                    /* bootstrap Bx100 times */


/**********************************************************************/
/*** USER SPECIFIES THE DIRECTORY CONTAINING THE ANALYSIS FILE
(in1)  ***/
/*** AND THE BOOTSTRAP WEIGHTS FILE (in2) SPECIFIED BELOW.
***/
/*** THE OUTPUT FILES WILL BE SAVED IN THE DIRECTORY SPECIFIED
FOR in3 **/
/**********************************************************************/

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";


/**********************************************************************/
/*** USER SPECIFIES THE NAME OF THE SAS FILE CONTAINING THE
ANALYSIS VARIABLES.  */
/*** %let Mfile=xxx;                              */
/**********************************************************************/

%let Mfile=in1.asthma_6;               /* Main SAS file name */
```

```
/**********************************************************************/
/*** USER SPECIFIES THE NAME OF THE SAS FILE CONTAINING THE
BOOTSTRAP WEIGHTS  ***/
/*** %let bsamp=b1h35;   (to be used with the file H35.TXT)    OR            ***/
/*** %let bsamp=b5h356;  (to be used with the file H356.TXT)   OR            ***/
/*** %let bsamp=b5h356a; (to be used with the file H356.TXT)   OR            ***/
/*** %let bsamp=b5h356c; (to be used with the file H356.TXT)   OR            ***/
/*** %let bsamp=b5longf; (to be used with the file LONG.TXT)                 ***/
/**********************************************************************/

%let bsamp=in2.wgt61000;            /* bootstrap weights SAS file */


/******************************************************************/
/*** USER SPECIFIES THE PARAMETERS by, wo AND kp, WHICH INDICATE /
/*** IF THE ANALYSIS IS DONE AT THE REGIONAL LEVEL OR FOR THE
***/
/*** WHOLE DATASET.                              ***/
/***                                ***/
/*** IF THE ANALYSIS IS DONE AT THE CANADA LEVEL, THEN          ***/
/***    %let by=*;                             ***/
/***    %let wo=;                              ***/
/***    %let kp=;                             ***/
/***                                ***/
/*** ELSE IF THE ANALYSIS IS DONE AT THE PROVINCIAL LEVEL, THEN
***/
/***    %let by=;                             ***/
/***    %let wo=*;                              ***/
/***    %let kp=region;                          ***/
/******************************************************************/

%let by=*;
%let wo=;
%let kp=;

data time;
  format start datetime16.;
  start=datetime();
  output;
run;


/**********************************************************************/
```

```
/*** Section 2: Declaration of the macros                              ***/
/**********************************************************************/

.
.
.

********************************************************************;
*  Section 3: MAIN PROGRAM                                    *;
********************************************************************;

/* Read in Main file */

/**********************************************************************/
/*** DUE TO THE LARGE NUMBER OF OBSERVATIONS, ONLY THE
VARIABLES      ***/
/*** THAT ARE NEEDED TO DO THE ANALYSES AND THE VARIABLES
"PROVINCE", ***/
/*** "REALUKEY", AND "PERSONID" SHOULD BE KEPT AT THIS POINT.       /
/**********************************************************************/

  data nphs (index=(id=(realukey personid)));
   set &Mfile (keep= realukey personid attack comp age2034 age3549 age50 smoking
                urban health);
  run;

/* Read in Bootstrap Weights */
/* FWGT corresponds to WT56_S, WT66_S or WT66_SN, depending on which file is
analysed */

  data bsamp (index=(id=(realukey personid)));
   set &bsamp;
   keep fwgt realukey personid bsw1-bsw&B.00;
  run;


/* Merge Main file and Bootstrap Weights */
```

```
&by data bs_nphs (index=(region));
&wo data bs_nphs;
      merge nphs (in=in1) bsamp (in=in2);
      by realukey personid;
      if in1;
    run;

   proc datasets library=work;
     delete nphs bsamp;
   run;


/***************************************************************************/
/*** Section 4: Macro calls                                            ***/
/***************************************************************************/

.
.
.

/************************/
/*                    */
/* Logistic Regression */
/*                    */
/************************/

   proc datasets library=work; /* initial bs_reglg */
     delete bs_reglg;
   run;
```

```
/*************************************************************************/
/*** To call the macro logreg, the following statement is used:        ***/
/*** %logreg(infile,yvar,xvar,bssz=,multi=);                           ***/
/***     where infile : bs_nphs                                        ***/
/***        yvar   : the dependent variable                            ***/
/***        xvar   : the independent variables.  The variables should  ***/
/***               be listed as they would be in a model statement     ***/
/***               (no commas in between).                             ***/
/***        bssz=  : bssz=&B.00                                        ***/
/***        multi= : multi=&R                                          ***/
/***                                                                   ***/
/*************************************************************************/
   title "Bootstrap &B.00:&R Variance Estimate for";
   title2 "Logistic Regressions";

 *  %logreg(bs_nphs,YVAR,XVAR,bssz=&B.00,multi=&R);
  %logreg(bs_nphs,attack,comp age2034 age3549 age50 smoking urban health,
bssz=&B.00,multi=&R);

/* Delete sas files in logreg macro */
  proc datasets library=work;
   delete betas betat bsbeta origest;

  run;

.
.
.

data time;
  set time;
  format stop datetime16.;
  stop=datetime();
  output;
run;

proc print data=time;
  title 'Time Taken to Run Program';
run;

/* End of SAS program BootVar */
(The execution of this program took about 10 minutes.)
```

The following examples present the results of the analyses done using the programs from the example 2. Results for the totals and ratios are presented in the example 3A. For example, if we want the ratio of the number of persons aged 12 to 19 years old who had an asthma attack in the last 12 month to the number of persons aged 12 to 19 years old who have asthma, in Prince-Edward-Island, we look at observation 5. The region 11 corresponds to the province of Prince Edward Island (see the data dictionnary of the *NPHS Asthma Supplement Documentation* for the codes associated with each province) and the variable Type indicates the type of analysis, in this case a ratio. We find the variables attage1 (VAR1) at the numerator of the ratio and age1219 (VAR2) at the denominator. The estimate of the ratio is 39.18% (YHAT) with a standard deviation of 10.51 (BS_SD) and a coefficient of variation of 26.82 (BS_CV). The 95% confidence interval for this estimate is (18.58%, 59.78%) (CIL95, CIU95). Following the results, we can see that the execution of this program started at 11:20 and finished at 11:21.

Results from the logistic regression are shown in example 3B. For example, the estimate of the parameter for the variable HEALTH is -0.86763 (BETA) and the odds ratio is 0.41994 (ODDS). The Wald's statistic for this parameter and its p-value associated are 6.25034 (WALD) and p=0.01242 (PVALUE) respectively. The variance estimate and the standard deviation for the parameter estimate are 0.12044 (BS_VAR) and 0.34704 (BS_SD) and the coefficient of variation is 40.00 (BS_CV). Finally, the confidence interval for the odds ratio is (0.21271, 0.8291) (CIL95,CIU95). The execution of this program started at 11:35 and finished at 11:45.

Example 3: Results of BOOTVAR.SAS and the time required to run the program

Example 3A: Totals and ratios

Bootstrap 1000:01 Variance Estimate for
Totals and Ratios

OBS	REGION	TYPE	VAR1	VAR2	YHAT	BS_SD	BS_CV	CIL95	CIU95
1	10	Total	attack	Aucune	11571.41	1818.68	15.72	8006.80	15136.02
2	10	Ratio	attage1	age1219	60.08	10.38	17.28	39.73	80.43
3	10	Ratio	attage2	age2034	45.20	13.46	29.78	18.81	71.59
4	11	Total	attack	Aucune	2707.89	448.34	16.56	1829.15	3586.63
5	11	Ratio	attage1	age1219	39.18	10.51	26.82	18.58	59.78
6	11	Ratio	attage2	age2034	51.18	16.64	32.51	18.57	83.79
7	12	Total	attack	Aucune	21703.58	2797.00	12.89	16221.45	27185.71
8	12	Ratio	attage1	age1219	61.35	10.37	16.90	41.03	81.67
9	12	Ratio	attage2	age2034	67.28	10.36	15.40	46.97	87.59
10	13	Total	attack	Aucune	19653.20	2872.28	14.61	14023.53	25282.87
11	13	Ratio	attage1	age1219	56.60	11.15	19.70	34.74	78.46

12	13	Ratio attage2	age2034	58.07	11.15	19.21	36.21	79.93
13	24	Total attack	Aucune	150456.35	13942.21	9.27	123129.61	177783.09
14	24	Ratio attage1	age1219	44.70	5.95	13.31	33.04	56.36
15	24	Ratio attage2	age2034	66.08	6.85	10.36	52.66	79.50
16	35	Total attack	Aucune	338993.45	23653.79	6.98	292632.03	385354.87
17	35	Ratio attage1	age1219	49.49	5.85	11.82	38.02	60.96
18	35	Ratio attage2	age2034	62.60	4.99	7.97	52.82	72.38
19	46	Total attack	Aucune	35455.99	5299.92	14.95	25068.14	45843.84
20	46	Ratio attage1	age1219	61.57	9.22	14.98	43.49	79.65
21	46	Ratio attage2	age2034	75.34	6.91	9.17	61.80	88.88
22	47	Total attack	Aucune	22663.24	3172.95	14.00	16444.27	28882.21
23	47	Ratio attage1	age1219	42.53	13.08	30.75	16.90	68.16
24	47	Ratio attage2	age2034	56.99	9.27	16.26	38.82	75.16
25	48	Total attack	Aucune	81139.41	9989.28	12.31	61560.41	100718.41
26	48	Ratio attage1	age1219	76.22	7.54	9.90	61.44	91.00
27	48	Ratio attage2	age2034	68.56	7.54	10.99	53.79	83.33
28	59	Total attack	Aucune	79525.61	9148.32	11.50	61594.91	97456.31
29	59	Ratio attage1	age1219	53.69	9.03	16.83	35.98	71.40
30	59	Ratio attage2	age2034	55.22	9.42	17.06	36.76	73.68

Time Taken to Run Program

OBS	START	STOP
1	28OCT98:11:20:16	28OCT98:11:21:03

Example 3B: Logistic regression

Bootstrap 1000:01 Variance Estimate for
Logistic Regressions
Dependent variable: attack

OBS	BETA	BHAT	ODDS	WALD	PVALUE
1	INTERCEP	1.56001	4.75886	5.87215	0.01538
2	COMP	-0.28298	0.75353	0.88929	0.34567
3	AGE2034	0.26398	1.30210	0.68375	0.40830
4	AGE3549	0.37246	1.45131	1.30213	0.25383
5	AGE50	-0.38554	0.68009	1.49725	0.22109
6	SMOKING	-0.13746	0.87157	0.28890	0.59093
7	URBAN	-0.16283	0.84973	0.29619	0.58628
8	HEALTH	-0.86763	0.41994	6.25034	0.01242

OBS	BS_VAR	BS_SD	BS_CV	CIL95	CIU95
1	0.41444	0.64377	41.27	1.34747	16.8069
2	0.09005	0.30008	106.04	0.41848	1.3569
3	0.10191	0.31924	120.93	0.69647	2.4344
4	0.10654	0.32641	87.63	0.76545	2.7517
5	0.09927	0.31508	81.72	0.36674	1.2611
6	0.06540	0.25574	186.05	0.52798	1.4388
7	0.08952	0.29919	183.74	0.47272	1.5274
8	0.12044	0.34704	40.00	0.21271	0.8291

Time Taken to Run Program

OBS	START	STOP
1	28OCT98:11:35:07	28OCT98:11:45:54

8.3.3 Modifications to the program for testing purposes

As mentioned earlier, program run time can be lengthy for regression analyses. It is possible to reduce the number of bootstrap weights used in order to test the program (**However, to obtain variance estimations, it is necessary to use the full set of bootstrap weights provided, i.e. 1000 weights**). For example, in the case of logistic regression, the user can specify the number of times that the regression will be calculated using the bootstrap weights (for example, 2 instead of 1000) in the routine "logreg" in section 2 of the program. The user has to add three lines as shown in *italics* in example 4. The value kL corresponds to the number of times that the regression will be calculated (in this case, 2 times). The two other parameters, k and j, should equal respectively 10 and 1 for testing purposes. Please note that those modifications have to be made in conjunction with specifying parameters as shown in example 2B. Again, please note that variance estimates calculated this way will not be valid.

One must also specify the number of bootstrap weights to be read in section 3. The following example shows where the change must be made. The part in italics must replace the part in parentheses. The same change can be made for the routines "regress" and "regglm".

(The program in Example 2B was tested in this fashion and the execution took less than 1 minute.)

Example 4: Changing the number of iterations for testing

```
.
.
.

/**********************************************************************/
/*** Section 2: Declaration of the macros                          ***/
/**********************************************************************/
.
.
.

********************************************;

%macro logreg(infile,yvar,xvar,bssz=,multi=);

********************************************;
proc logistic data=&infile outest=orig(keep=&kp intercep &xvar) descending noprint;
  model &yvar=&xvar;
  &by by region;
  weight fwgt;
run;

proc transpose data=orig out=origest prefix=bhat name=beta;
  var intercep &xvar;
  &by by region;
run;

%let L=%eval(&bssz/10);

%do k=1 %to 10;
  %let j=%eval(1+((&k-1)*&L));
  %let kL=%eval(&k*&L);


%let k=10;
%let j=1;
%let kL=2;

  data poids;
    set &infile;
    keep bsw&j-bsw&kL &yvar &xvar &kp;
```

```
  run;

  %do i=&j %to &kL;

    proc logistic data=poids outest=betas(keep=&kp intercep &xvar) noprint descending;
      model &yvar=&xvar;
      &by by region;
      weight bsw&i;
    run;

    proc transpose data=betas out=betat prefix=best name=beta;
      var intercep &xvar;
      &by by region;
    run;

    data betat;
      set betat;
      rename best1=best&i;
    run;

    %if (&i =1) %then %do;

      data bsbeta;
        set betat;
      run;

    %end;
    %else %do;

      data bsbeta;
        merge bsbeta betat;
        &by by region;
      run;

    %end;
  %end;
%end;

data bs_reglg;
  merge origest bsbeta;
  rename bhat1=bhat;
  bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
```

```
  bs_sd=sqrt(bs_var);
  bs_cv=abs(round((bs_sd/bhat1)*100,.01));
  wald=(bhat1/bs_sd)*(bhat1/bs_sd);
  pvalue=1-probchi(wald,1);
  lo95=bhat1-1.96*bs_sd;
  hi95=bhat1+1.96*bs_sd;
  odds=exp(bhat1);
  cil95=exp(lo95);
  ciu95=exp(hi95);
  ydep="&yvar";
  drop best1-best&bssz;
run;
%let printlog=1;
%let dep=&yvar;

%mend logreg;
.
.
.
*********************************************************************;
*  Section 3: MAIN PROGRAM                                  *;
*********************************************************************;

/* Read in Main file */

/*********************************************************************/
/*** DUE TO THE LARGE NUMBER OF OBSERVATIONS, ONLY THE
VARIABLES      ***/
/*** THAT ARE NEEDED TO DO THE ANALYSES AND THE VARIABLES
"PROVINCE", ***/
/*** "REALUKEY", AND "PERSONID" SHOULD BE KEPT AT THIS POINT.        /
/*********************************************************************/
   data nphs (index=(id=(realukey personid)));

     set &Mfile (keep= realukey personid attack comp age2034 age3549 age50 smoking
                  urban health);
   run;

/* Read in Bootstrap Weights */
/* FWGT corresponds to WT56_S, WT66_S or WT66_SN, depending on which file is
analysed */
```

```
data bsamp (index=(id=(realukey personid)));
  set &bsamp;
  keep fwgt realukey personid bsw1-bsw2; (instead of bsw1-bsw&B.00;)
run;
.
.
.
```

Program BOOTVAR.SAS

```
*                          WARNING

* The Government of Canada (Statistics Canada) is the owner of all intellectual
* property rights (including copyright) in this software.  Subject to the terms below,
* you are granted a non-exclusive and non-transferable licence to use this software.
*
* This software is provided "as-is", and the owner makes no warranty, either express
* or implied, including but not limited to, warranties of merchantability and fitness
* for any particular purpose.  In no event will the owner be liable for any indirect,
* special, consequential or other similar damages.  This agreement will terminate
* automatically without notice to you if you fail to comply with any term of this
* agreement.;

/*******************************************************************/
/* Date: December 01, 1997                                        */
/* Modified: September 21, 1998                                    */
/*******************************************************************/

/*********************************************************************/
/***                                                            ***/
/***                  SAS Program BOOTVAR                       ***/
/***                                                            ***/
/*** This program calculates variance estimates using the bootstrap weights    ***/
/*** for different types of estimators.  Using SAS Macros, this program can    ***/
/*** calculate variance estimates for totals, ratios and differences between   ***/
/*** ratios.  It can also calculate variance estimates for the parameters      ***/
/*** of a regression, logistic regression, or generalized linear model.        ***/
/*** This program can also be customized for other types of analyses.          ***/
/***                                                            ***/
/*** The program is divided in 4 sections, described below. Throughout the     ***/
/*** program, it is indicated where the user has to make changes.              ***/
/***                                                            ***/
/*** Section 1: Declaration of the macro variables                  ***/
/***          This is where variables that are going to be used throughout   ***/
/***          the program have to be defined.  Some changes must be made     ***/
/***          by the user.                                      ***/
/*** Section 2: Declaration of the macros                      ***/
/***          This is where the portion of the program to calculate variance ***/
/***          estimates is defined.  No changes have to be made by the user  ***/
/***          in this section (excepted for the macro REGGLM, which has to be **/
```

```
/***        customized).                                    ***/
/*** Section 3: Main program                                ***/
/***        This is where the analysis and the bootstrap weights data set  ***/
/***        are read.  The variables to be used in the analyses have to be ***/
/***        defined by the user prior to using this program.          ***/
/*** Section 4: Macro calls                                 ***/
/***        This is where the user will call the macro(s) that he/she is   ***/
/***        interested in submitting.  At that point, the user can call    ***/
/***        one macro once, one macro multiple times, or multiple macros.  ***/
/***        Due to the large number of observations to analyse, caution    ***/
/***        should be used however in the number of macro calls requested, ***/
/***        particularly in the case of regressions.                 ***/
/***                                                        ***/
/*** Before using this program, the user has to prepare an analysis file that  ***/
/*** will contain the variables to analyse.  This file should contain only the ***/
/*** records for which the analysis is required (e.g., if the analysis is done ***/
/*** for the agegroup 12 years old and over, the file should contain only the  ***/
/*** records of those 12 years old and over).  A variable called REGION should ***/
/*** also be created if the analysis are done at a regional level. For example,***/
/*** if the analysis is done at the provincial level, the variable REGION will ***/
/*** take the value of the variable containing the province on the microdata   ***/
/*** file (e.g. PRC6_CUR).                                  ***/
/***                                                        ***/
/*** N.B. When calculating a total, a ratio, or a difference of ratios,       ***/
/*** the estimates are obtained by summing the weight of the records that     ***/
/*** have the characteristic of interest.  Hence, a dummy variable must be    ***/
/*** created for each of the variables to be analysed, the variable taking a  ***/
/*** value of 1 when a record has the characteristic of interest and 0 otherwise.*/
/****************************************************************************/

options ps=62 ls=80 ;

/****************************************************************************/
/*** Section 1: Declaration of the macro variables                     ***/
/****************************************************************************/

%let R=01;                  /* bootstrap averaging parameter (fixed) */

/****************************************************************************
*******/
/*** USER SPECIFIES THE NUMBER OF BOOTSTRAP WEIGHTS USED,
DIVIDED BY 100   ***/
```

```
/*** %let B=xx;                                    ***/
/**************************************************************************/
%let B=10;                  /* bootstrap Bx100 times */


/**************************************************************************/
/*** USER SPECIFIES THE DIRECTORY CONTAINING THE ANALYSIS FILE
(in1)  ***/
/*** AND THE BOOTSTRAP WEIGHTS FILE (in2) SPECIFIED BELOW.
***/
/*** THE OUTPUT FILES WILL BE SAVED IN THE DIRECTORY SPECIFIED
FOR in3 **/
/**************************************************************************/

libname in1 "c:\data";
libname in2 "c:\bootstrp";
libname in3 "c:\output";


/**************************************************************************/
/*** USER SPECIFIES THE NAME OF THE SAS FILE CONTAINING THE
ANALYSIS VARIABLES.  */
/*** %let Mfile=xxx;                                  */
/**************************************************************************/

%let Mfile=in1.a5s;                 /* Main SAS file name */


/**************************************************************************/
/*** USER SPECIFIES THE NAME OF THE SAS FILE CONTAINING THE
BOOTSTRAP WEIGHTS  ***/
/*** %let bsamp=wgt51000; (to be used with the file A5S)   OR            ***/
/*** %let bsamp=wgt61000; (to be used with the file A56S)              ***/
/**************************************************************************/

%let bsamp=in2.wgt51000;                /* bootstrap weights SAS file */

/************************************************************************/
/*** USER SPECIFIES THE PARAMETERS by, wo AND kp, WHICH INDICATE
***/
/*** IF THE ANALYSIS IS DONE AT THE REGIONAL LEVEL OR FOR THE
***/
/*** WHOLE DATASET.                                ***/
```

```
/***                                          ***/
/*** IF THE ANALYSIS IS DONE FOR THE WHOLE DATASET, THEN        ***/
/***    %let by=*;                                ***/
/***    %let wo=;                                 ***/
/***    %let kp=;                                ***/
/***                                          ***/
/*** ELSE IF THE ANALYSIS IS DONE AT THE REGIONAL LEVEL, THEN    ***/
/***    %let by=;                                ***/
/***    %let wo=*;                                ***/
/***    %let kp=region;                            ***/
/*********************************************************************/

%let by=*;
%let wo=;
%let kp=;

data time;
  format start datetime16.;
  start=datetime();
  output;
run;


/**********************************************************************/
/*** Section 2: Declaration of the macros                   ***/
/**********************************************************************/

%let printtot=0;
%let printdif=0;
%let printreg=0;
%let printlog=0;
%global dep;

*************************************;

%macro total(infile,var,bssz=,multi=);

*************************************;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var;
```

```
&by  by region;
  output out=ytot
        sum=yhat ybs1-ybs&bssz;
run;

data est;
  set ytot;
  length var1 $ 8;
  length var2 $ 8;
  length type $ 8;
  bs_var=&multi*((&bssz-1)*(var(of ybs1-ybs&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=round((bs_sd/yhat)*100,.01);
  cil95=yhat-1.96*bs_sd;
  ciu95=yhat+1.96*bs_sd;
  var1="&var";
  var2="None";
  type="Total";
  drop ybs1-ybs&bssz _type_ _freq_;
run;

proc append data=est base=alltots;
run;

%let printtot=1;

%mend total;


*******************************************;

%macro ratio(infile,var1,var2,bssz=,multi=);

*******************************************;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var1;
&by  by region;
  output out=ytot
        sum=yhat ybs1-ybs&bssz;
run;
```

```
proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var2;
&by  by region;
  output out=xtot
        sum=xhat xbs1-xbs&bssz;
run;

data est;
  merge ytot xtot;
  array ybs{&bssz};
  array xbs{&bssz};
  array rbs{&bssz};
  length var1 $ 8;
  length var2 $ 8;
  length type $ 8;
  yhat=round((yhat/xhat)*100,.01);
  do i=1 to &bssz;

    rbs{i}=round((ybs{i}/xbs{i})*100,.01);
  end;
  bs_var=&multi*((&bssz-1)*(var(of rbs1-rbs&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=round((bs_sd/yhat)*100,.01);
  cil95=yhat-1.96*bs_sd;
  ciu95=yhat+1.96*bs_sd;
  var1="&var1";
  var2="&var2";
  type="Ratio";
  drop ybs1-ybs&bssz xbs1-xbs&bssz rbs1-rbs&bssz xhat i _type_ _freq_;
run;

proc append data=est base=alltots;
run;

%let printtot=1;

%mend ratio;


*******************************************************;
```

```
%macro diff_rat(infile,var1,var2,var3,var4,bssz=,multi=);

*******************************************************;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var1;
&by  by region;
  output out=ytot
        sum=yhat ybs1-ybs&bssz;
run;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var2;
&by  by region;
  output out=xtot
        sum=xhat xbs1-xbs&bssz;
run;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var3;
&by  by region;
  output out=yytot
        sum=yyhat yybs1-yybs&bssz;
run;

proc means data=&infile noprint;
  var fwgt bsw1-bsw&bssz;
  weight &var4;
&by  by region;
  output out=xxtot
        sum=xxhat xxbs1-xxbs&bssz;
run;

data est;
  merge ytot xtot yytot xxtot;
  array ybs{&bssz};
  array xbs{&bssz};
  array yybs{&bssz};
  array xxbs{&bssz};
```

```
  array drbs{&bssz};
  length var1 $ 8;
  length var2 $ 8;
  length var3 $ 8;
  length var4 $ 8;
  length type $ 10;
  yhat=round(((yhat/xhat)-(yyhat/xxhat))*100,.01);
  do i=1 to &bssz;
    drbs{i}=round(((ybs{i}/xbs{i})-(yybs{i}/xxbs{i}))*100,.01);
  end;
  bs_var=(&multi*((&bssz-1)*(var(of drbs1-drbs&bssz)))/&bssz);
  bs_sd=sqrt(bs_var);
  bs_cv=abs(round((bs_sd/yhat)*100,.01));
  cil95=yhat-1.96*bs_sd;
  ciu95=yhat+1.96*bs_sd;
  var1="&var1";
  var2="&var2";
  var3="&var3";
  var4="&var4";
  type="Diff_Ratio";
  drop ybs1-ybs&bssz xbs1-xbs&bssz yybs1-yybs&bssz xxbs1-xxbs&bssz
drbs1-drbs&bssz
       xhat xxhat yyhat i _type_ _freq_;
run;

proc append data=est base=diffrat;
run;

%let printdif=1;

%mend diff_rat;


********************************************;

%macro regress(infile,yvar,xvar,bssz=,multi=);

********************************************;

proc reg data=&infile outest=orig(keep=&kp intercep &xvar) noprint;
  model &yvar=&xvar;
  weight fwgt;
```

```
&by  by region;
run;

proc transpose data=orig out=origest(drop=_label_) prefix=bhat name=beta;
  var intercep &xvar;
&by  by region;
run;

%let L=%eval(&bssz/10);

%do k=1 %to 10;
  %let j=%eval(1+((&k-1)*&L));
  %let kL=%eval(&k*&L);

  data poids;
    set &infile;
    keep bsw&j-bsw&kL &yvar &xvar &kp;
  run;

  %do i=&j %to &kL;

   proc reg data=poids outest=betas(keep=&kp intercep &xvar) noprint;
     model &yvar=&xvar;
     weight bsw&i;
&by  by region;
   run;

   proc transpose data=betas out=betat prefix=best name=beta;
     var intercep &xvar;
&by  by region;
   run;

   data betat;
     set betat;
     drop _label_;
     rename best1=best&i;
   run;

   %if (&i =1) %then %do;

     data bsbeta;
       set betat;
```

```
     run;

   %end;
   %else %do;

     data bsbeta;
       merge bsbeta betat;
&by    by region;
     run;

   %end;
  %end;
%end;

data est;
  merge origest bsbeta;
  rename bhat1=bhat;
  bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=abs(round((bs_sd/bhat1)*100,.01));
  cil95=bhat1-1.96*bs_sd;
  ciu95=bhat1+1.96*bs_sd;
  ydep="&yvar";
  drop best1-best&bssz;
run;

proc append data=est base=bs_reg;
run;

%let printreg=1;
%let dep=&yvar;

%mend regress;


*********************************************;

%macro logreg(infile,yvar,xvar,bssz=,multi=);

*********************************************;
proc logistic data=&infile outest=orig(keep=&kp intercep &xvar) descending noprint;
  model &yvar=&xvar;
```

```
  &by by region;
  weight fwgt;
run;

proc transpose data=orig out=origest prefix=bhat name=beta;
  var intercep &xvar;
  &by by region;
run;

%let L=%eval(&bssz/10);

%do k=1 %to 10;
  %let j=%eval(1+((&k-1)*&L));
  %let kL=%eval(&k*&L);

  data poids;
    set &infile;
    keep bsw&j-bsw&kL &yvar &xvar &kp;
  run;

  %do i=&j %to &kL;

    proc logistic data=poids outest=betas(keep=&kp intercep &xvar) noprint descending;
      model &yvar=&xvar;
      &by by region;
      weight bsw&i;
    run;

    proc transpose data=betas out=betat prefix=best name=beta;
      var intercep &xvar;
      &by by region;
    run;

    data betat;
      set betat;
      rename best1=best&i;
    run;

    %if (&i =1) %then %do;

      data bsbeta;
        set betat;
```

```
      run;

    %end;
    %else %do;

      data bsbeta;
        merge bsbeta betat;
        &by by region;
      run;

    %end;
  %end;
%end;

data est;
  merge origest bsbeta;
  rename bhat1=bhat;
  bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=abs(round((bs_sd/bhat1)*100,.01));
  wald=(bhat1/bs_sd)*(bhat1/bs_sd);
  pvalue=1-probchi(wald,1);
  lo95=bhat1-1.96*bs_sd;
  hi95=bhat1+1.96*bs_sd;
  odds=exp(bhat1);
  cil95=exp(lo95);
  ciu95=exp(hi95);
  ydep="&yvar";
  drop best1-best&bssz;
run;

proc append data=est base=bs_reglg;
run;

%let printlog=1;
%let dep=&yvar;

%mend logreg;


********************************************;
```

```
%macro regglm(infile,yvar,xvar,bssz=,multi=);

*******************************************;
proc glm data=&infile noprint;
     class &xvar;
     model &yvar= &xvar;
     LSMEANS &xvar / out=orig;
     weight fwgt;
&by  by region;
run;

data orig2;
  set orig;
  drop _name_ stderr;
  array meanss{3};
  do i=1 to 3;
    if _n_ = i then do;
      meanss(i)=lsmean;
      retain;
    end;
  end;

  if _n_=3 then do;
    diff1_2= meanss1-meanss2;
    diff1_3= meanss1-meanss3;
    diff2_3= meanss2-meanss3;
    drop  smc6_2 lsmean i meanss1- meanss3;
    output;
  end;
run;

proc transpose data=orig2 out=origest prefix=bhat name=beta;
var diff1_2 diff1_3 diff2_3;
run;

%let L=%eval(&bssz/10);

%do k=1 %to 10;
  %let j=%eval(1+((&k-1)*&L));
  %let kL=%eval(&k*&L);

  data poids;
```

```
  set &infile;
  keep bsw&j-bsw&kL &yvar &xvar &kp;
run;

%do i=&j %to &kL;

proc glm data=poids noprint;
  class &xvar;
  model &yvar= &xvar;
  LSMEANS &xvar / out=betas;
  weight bsw&i;
  &by by region;
run;

data betas2;
  set betas;
  drop _name_ stderr;
  array meanss{3};
  do i=1 to 3;
    if _n_ = i then do;
      meanss(i)=lsmean;
      retain;
    end;
  end;

  if _n_=3 then do;
    diff1_2= meanss1-meanss2;
    diff1_3= meanss1-meanss3;
    diff2_3= meanss2-meanss3;
    drop smc6_2 lsmean i meanss1- meanss3;
    output;
  end;
run;

proc transpose data=betas2 out=betat prefix=best name=beta;
 var diff1_2 diff1_3 diff2_3;
run;

 data betat;
  set betat;
  rename best1=best&i;
 run;
```

```
  %if (&i =1) %then %do;

    data bsbeta;
      set betat;
    run;

  %end;
  %else %do;

     data bsbeta;
       merge bsbeta betat;
       &by by region;
     run;

  %end;
 %end;
%end;

data est;
  merge origest bsbeta;
  rename bhat1=bhat;
  bs_var=&multi*((&bssz-1)*(var(of best1-best&bssz)))/&bssz;
  bs_sd=sqrt(bs_var);
  bs_cv=abs(round((bs_sd/bhat1)*100,.01));
  cil95=bhat1-2.39406*bs_sd;
  ciu95=bhat1+2.39406*bs_sd;
  drop best1-best&bssz;
run;

proc append data=est base=bs_reg;
run;

%let printreg=1;
%let dep=&yvar;

%mend regglm;


****************;

%macro prnttot;
```

```
****************;
%if &printtot=1 %then %do;

&by   proc sort data=alltots;
&by     by region;
&by   run;

/*********************************************/
/* Print the result of total and ratio macro */
/*********************************************/


  proc print data=alltots;
    title "Bootstrap &B.00:&R Variance Estimate for";
    title2 "Totals and Ratios";
    var &kp type var1 var2 yhat bs_sd bs_cv cil95 ciu95;
  run;
%end;

      /********************************************************/
      /*** Where:                                          ***/
      /*** type        : type of estimator (Total or Ratio)  ***/
      /*** var1 and var2: the variables used to calculate the ***/
      /***              estimate.  For a Total, var2=None   ***/
      /*** yhat        : the estimate (in % for a Ratio)     ***/
      /*** bs_sd        : the standard deviation             ***/
      /*** bs_cv        : the coefficient of variation        ***/
      /*** cil95        : the lower 95% confidence limit      ***/
      /*** ciu95        : the upper 95% confidence limit      ***/
      /********************************************************/


%mend prnttot;


****************;

%macro prntdiff;

****************;
%if &printdif=1 %then %do;
```

```
&by   proc sort data=diffrat;
&by     by region;
&by   run;

/*************************************/
/* Print the result of diff_rat macro */
/*************************************/

  proc print data=diffrat;
    title "Bootstrap &B.00:&R Variance Estimate for";
    title2 "Difference between Ratios";
    var &kp type var1 var2 var3 var4 yhat bs_sd bs_cv cil95 ciu95;
  run;
%end;

      /*******************************************************/
      /*** Where:                                    ***/
      /*** type        : type of estimator (Diff_Ratio)      ***/
      /*** var1,var2,                                ***/
      /*** var3 and var4: the variables used to calculate the ***/
      /***              estimate.                    ***/
      /*** yhat        : the estimate (difference in %)      ***/
      /*** bs_sd        : the standard deviation           ***/
      /*** bs_cv        : the coefficient of variation       ***/
      /*** cil95       : the lower 95% confidence limit      ***/
      /*** ciu95        : the upper 95% confidence limit      ***/
      /*******************************************************/
%mend prntdiff;


****************;

%macro prntreg;

****************;
%if &printreg=1 %then %do;

/**********************************************/
/* Print the result of regress and regglm macro */
/**********************************************/

  proc print data=bs_reg;
```

```
    title "Bootstrap &B.00:&R Variance Estimate for";
    title2 "Regressions";
    title3 "Dependent variable: &yvar";
    var &kp beta bhat bs_var bs_sd bs_cv cil95 ciu95;
  run;
%end;

     /**********************************************************/
     /*** Where:                                           ***/
     /***   beta      : parameter to estimate              ***/
     /***   bhat      : parameter estimator                ***/
     /***   bsvar     : variance of the parameter estimator ***/
     /***   bs_sd     : the standard deviation of the para- ***/
     /***               meter estimator                    ***/
     /***   bs_cv     : the coefficient of variation of the ***/
     /***               parameter estimator                ***/
     /***   cil95     : the lower 95% confidence limit     ***/
     /***   ciu95     : the upper 95% confidence limit     ***/
     /**********************************************************/


%mend prntreg;


****************;

%macro prntlog;
****************;
%if &printlog=1 %then %do;

/************************************/
/* Print the result of logreg macro */
/************************************/

  proc print data=bs_reglg;
    title "Bootstrap &B.00:&R Variance Estimate for";
    title2 "Logistic Regressions";
    title3 "Dependent variable: &yvar";
    var &kp beta bhat odds wald pvalue bs_var bs_sd bs_cv cil95 ciu95;
  run;
%end;
```

```
/*****************************************************/
/*** Where:                                       ***/
/***   beta      : parameter to estimate           ***/
/***   bhat      : parameter estimator             ***/
/***   odds      : odds ratio                      ***/
/***   wald      : Wald statistic                  ***/
/***   pvalue    : p-value of the Wald statistic   ***/
/***   bsvar     : variance of the parameter estimator ***/
/***   bs_sd     : the standard deviation of the para- ***/
/***              meter estimator                  ***/
/***   bs_cv     : the coefficient of variation of the ***/
/***              parameter estimator              ***/
/***   cil95     : the lower 95% confidence limit  ***/
/***              for the odds ratio               ***/
/***   ciu95     : the upper 95% confidence limit  ***/
/***              for the odds ratio               ***/
/*****************************************************/


%mend prntlog;


*****************************************************************;
*  Section 3: MAIN PROGRAM                                      *;
*****************************************************************;

/* Read in Main file */

/*****************************************************************/
/*** DUE TO THE LARGE NUMBER OF OBSERVATIONS, ONLY THE
VARIABLES      ***/
/*** THAT ARE NEEDED TO DO THE ANALYSES AND THE VARIABLES
 ***/
/*** "REALUKEY", AND "PERSONID" SHOULD BE KEPT AT THIS POINT.       /
/*****************************************************************/

  data nphs (index=(id=(realukey personid)));
   set &Mfile (keep= realukey personid + VARIABLES TO ANALYZE);
  run;

/* Read in Bootstrap Weights */
```

```
/* FWGT corresponds to WTA6_5S or WTA6_6S, depending on which file is analysed
*/

  data bsamp (index=(id=(realukey personid)));
   set &bsamp;
   keep fwgt realukey personid bsw1-bsw&B.00;
  run;


/* Merge Main file and Bootstrap Weights */

&by data bs_nphs (index=(region));
&wo data bs_nphs;
     merge nphs (in=in1) bsamp (in=in2);
     by realukey personid;
     if in1;
   run;

  proc datasets library=work;
   delete nphs bsamp;
  run;

/***********************************************************************/
/*** Section 4: Macro calls                                         ***/
/***********************************************************************/

/***********************************************************************/
/*** The variance estimates are calculated using the SAS macros that have been  ***/
/*** defined earlier in the program.  Those SAS macros can be called in the     ***/
/*** program as needed.  A macro call is a SAS statement that specifies the name **/
/*** of the macro called and some parameters.  In this program, the parameters  ***/
/*** indicate which file has to be read, which variables will be used, how many ***/
/*** bootstrap weights are used, and how many bootstrap samples were selected.  ***/
/*** The name of the file to read is always the same (BS_NPHS) and the last two ***/
/*** parameters have already been defined in the first section of the program.  ***/
/*** Hence, the user only has to specify the variables to be used.            ***/
/***                                                                ***/
/*** Each macro call will result in a variance estimate of ONE estimator.  If   ***/
/*** more than one estimator and its variance estimate have to be calculated,   ***/
/*** the macro has to be called multiple times.                          ***/
/***                                                                ***/
/*** A commented statement for the macro call appears where the user's statement***/
```

```
/*** should be placed in the program.  The user's statement will              ***/
/*** be identical to the commented one, except for the names of the analysis    ***/
/*** variables.                                               ***/
/**********************************************************************/

/*********/
/*       */
/* Total */
/*       */
/*********/

  proc datasets library=work; /* initial alltots */
   delete alltots;
  run;

/**********************************************************************/
/*** To call the macro total, the following statement is used:            ***/
/*** %total(infile,var,bssz=,multi=);                              ***/
/***       where infile : bs_nphs                              ***/
/***            var    : the variable for which a total is calculated    ***/
/***            bssz=  : bssz=&B.00                              ***/
/***            multi= : multi=&R                                ***/
/***                                                       ***/
/**********************************************************************/

 *  %total(bs_nphs,VAR,bssz=&B.00,multi=&R);

/* Delete sas files in total macro */

  proc datasets library=work;
   delete ytot est;
  run;


/*********/
/*       */
/* Ratio */
/*       */
/*********/

/**********************************************************************/
/*** To call the macro ratio, the following statement is used:            ***/
```

```
/*** %ratio(infile,var1,var2,bssz=,multi=);                    ***/
/***      where infile : bs_nphs                           ***/
/***            var1   : the variable of the numerator of the ratio        ***/
/***            var2   : the variable of the denominator of the ratio      ***/
/***            bssz=  : bssz=&B.00                          ***/
/***            multi= : multi=&R                           ***/
/***                                                ***/
/*********************************************************************/


 *  %ratio(bs_nphs,VAR1,VAR2,bssz=&B.00,multi=&R);

/* Delete sas files in ratio macro */
  proc datasets library=work;
    delete ytot xtot est;
  run;

%prnttot;

/*********************************************/
/*** The results of total and ratio macro ***/
/*** can be saved in a permanent file.    ***/
/*********************************************/

*   data in3.ZZZ;
*     set alltots;
*   run;


/*****************************/
/*                         */
/* Difference between Ratios */
/*                         */
/*****************************/

  proc datasets library=work; /* initial diffrat */
    delete diffrat;
  run;

/*********************************************************************/
/*** To call the macro diff_rat, the following statement is used:         ***/
/*** %diff_rat(infile,var1,var2,var3,var4,bssz=,multi=);              ***/
```

```
/***      where infile : bs_nphs                                    ***/
/***           var1   : the variable of the numerator of the 1st ratio    ***/
/***           var2   : the variable of the denominator of the 1st ratio  ***/
/***           var3   : the variable of the numerator of the 2nd ratio    ***/
/***           var4   : the variable of the denominator of the 2nd ratio  ***/
/***           bssz=  : bssz=&B.00                                  ***/
/***           multi= : multi=&R                                    ***/
/***                                                           ***/
/*** Note: The confidence interval is calculated for analysis doing only one **/
/*** comparison of ratios.  If multiple comparisons are done, the calculation */
/*** of confidence intervals must take into account this situation.  For    ***/
/*** this reason, in the case of multiple comparisons, the value of the     ***/
/*** normal distribution (Z) used in the calculation of the confidence      ***/
/*** interval must be corrected in the macro declaration in Section 2 ,     ***/
/*** using, for example, the Bonferroni approach for multiple comparisons.  ***/
/***                                                           ***/
/**********************************************************************/


 *  %diff_rat(bs_nphs,VAR1,VAR2,VAR3,VAR4,bssz=&B.00,multi=&R);

/* Delete sas files in diff_rat macro */
  proc datasets library=work;
    delete ytot xtot yytot xxtot est;
  run;

%prntdiff;

/*****************************************/
/*** The results of diff_rat macro      ***/
/*** can be saved in a permanent file.  ***/
/*****************************************/

*   data in3.ZZZ;
*     set diffrat;
*   run;


/**************/
/*          */
/* Regression */
/*          */
```

```
/**************/

  proc datasets library=work; /* initial bs_reg */
   delete bs_reg;
  run;

/**********************************************************************/
/*** To call the macro regress, the following statement is used:          ***/
/*** %regress(infile,yvar,xvar,bssz=,multi=);                             ***/
/***      where infile : bs_nphs                                          ***/
/***            yvar   : the dependent variable                           ***/
/***            xvar   : the independent variables.  The variables should  ***/
/***                   be listed as they would be in a model statement   ***/
/***                   (no commas in between).                            ***/
/***            bssz=  : bssz=&B.00                                       ***/
/***            multi= : multi=&R                                         ***/
/***                                                                      ***/
/**********************************************************************/


 *  %regress(bs_nphs,YVAR,XVAR,bssz=&B.00,multi=&R);

/* Delete sas files in regress macro */
  proc datasets library=work;
   delete betas betat bsbeta origest;
  run;


/**********************/
/*                   */
/* Logistic Regression */
/*                   */
/**********************/

  proc datasets library=work; /* initial bs_reglg */
   delete bs_reglg;
  run;

/**********************************************************************/
/*** To call the macro logreg, the following statement is used:           ***/
/*** %logreg(infile,yvar,xvar,bssz=,multi=);                              ***/
/***      where infile : bs_nphs                                          ***/
```

```
/***          yvar   : the dependent variable                    ***/
/***          xvar   : the independent variables.  The variables should  ***/
/***                be listed as they would be in a model statement   ***/
/***                (no commas in between).                    ***/
/***          bssz=  : bssz=&B.00                              ***/
/***          multi= : multi=&R                                ***/
/***                                                   ***/
/*************************************************************************/


 *  %logreg(bs_nphs,YVAR,XVAR,bssz=&B.00,multi=&R);

/* Delete sas files in logreg macro */
  proc datasets library=work;
    delete betas betat bsbeta origest;
  run;


/**********************************/
/*                               */
/* Generalized Linear Model (GLM) */
/*                               */
/**********************************/

/*************************************************************************/
/*** N.B. The macro regglm has to be customized.  The macro definition that ***/
/***      appears in this program is an example of one analysis that has    ***/
/***      been done using this program.                          ***/
/***                                                   ***/
/*** To call the macro regglm, the following statement is used:           ***/
/*** %regglm(infile,yvar,xvar,bssz=,multi=);                       ***/
/***       where infile : bs_nphs                        ***/
/***          yvar   : the dependent variable                    ***/
/***          xvar   : the independent variables.  The variables should  ***/
/***                be listed as they would be in a model statement   ***/
/***                (no commas in between).                    ***/
/***          bssz=  : bssz=&B.00                              ***/
/***          multi= : multi=&R                                ***/
/***                                                   ***/
/*** Note: If the regglm macro is adapted to calculate differences between  ***/
/***       means, confidence interval limits should be calculated to take   ***/
/***       into account multiple comparisons.  The value of the normal      ***/
```

```
/***    distribution (Z) used in the calculation of the confidence       ***/
/***    interval must be corrected in the macro declaration in Section 2,***/
/***    using, for example, the Bonferroni approach for multiple         ***/
/***    comparisons.                                                     ***/
/************************************************************************/


 *  %regglm(bs_nphs,YVAR,XVAR,bssz=&B.00,multi=&R);

/* Delete sas files in regglm macro */
  proc datasets library=work;
    delete betas betat bsbeta origest;
  run;

%prntreg;

/**********************************************/
/*** The results of regress and regglm macro ***/
/*** can be saved in a permanent file.       ***/
/**********************************************/

*   data in3.ZZZ;
*     set bs_reg;
*   run;


%prntlog;

/*****************************************/
/*** The results of logreg macro       ***/
/*** can be saved in a permanent file. ***/
/*****************************************/

*   data in3.ZZZ;
*     set bs_reglg;
*   run;


data time;
  set time;
  format stop datetime16.;
  stop=datetime();
```

```
  output;
run;

proc print data=time;
  title 'Time Taken to Run Program';
run;

/* End of SAS program BootVar */
```