Canadian Health Measures Survey

Canadian Health Measures Survey (CHMS)
Data User Guide: Cycle 2

November 2012

Statistics Canada Statistique Canada

Canada

Table of contents

1. Introduction

The Canadian Health Measures Survey (CHMS) is a comprehensive, direct health measures survey, developed to address important data gaps and limitations in existing health information. It is conducted by Statistics Canada in partnership with Health Canada (HC) and the Public Health Agency of Canada (PHAC). The results will provide comprehensive health information to advance health surveillance and research in Canada.

This document will help users work with and understand data from cycle 2 of the CHMS, which was obtained through the collection of directly measured indicators of health and wellness. These measured indicators were taken from August 2009 to November 2011 on a representative sample of 6,395 Canadians aged 3 to 79 years. The survey consisted of an in-home general health interview followed by a visit to a mobile examination centre (MEC), sometimes referred to as a mobile clinic. Reference laboratories and the MEC laboratory analyzed biological specimens for indicators of general health, chronic disease, infectious disease, nutritional status and environmental biomarkers. Indoor air samples were also taken from the home to measure for a number of airborne substances.

This document also provides information to data users about the complexity of the data and any limitations that could affect their use. It explains the methods and concepts used to collect the data at the household, as well as at the MEC. Subsequent sections of the document contain information about data processing and the creation of derived variables[1]. Content regarding sampling and weighting methodology, and guidelines for the creation of tabulations have also been included to assist the data user. Quality assurance and quality control information is provided to describe characteristics of the data which might limit their usefulness or interpretation. The document concludes with a series of appendices which provide supporting information that will be helpful to users of the CHMS data files.

2. Important notes related to this document

2.1 Acronyms and abbreviations

Throughout this document, acronyms and abbreviations for terms associated with the Canadian Health Measures Survey (CHMS) are spelled out the first time they appear in a chapter, with the acronym/abbreviation put in brackets immediately afterwards. The next time that term appears in the chapter, only the acronym/abbreviation is used. A full list of these acronyms and abbreviations can be found in Appendix 1.

2.2 References and end notes

Background information such as references for research articles, definitions and procedural information is cited frequently throughout this document so that the user can find additional information related to the text. When this occurs, a small superscript number is put at the end of the text and users can consult the corresponding number in Chapter 13 – References and end notes for the additional information.

2.3 Survey Documentation

Extensive documentation on the CHMS is available to all data users and the general public. A description of the CHMS, basic methodological information, and links to the household and mobile examination centre (MEC) questionnaires can be accessed through the following link: http://www23.statcan.gc.ca:81/imdb/p2SV.pl?Function=getSurvey&SDDS=5071&lang=en&db=imdb&adm=8&dis=2

In addition, Appendix 2 provides a list of documents that are available upon request. References to the documents in this list are made throughout the User Guide in order to help users identify which documents are related to a particular topic.

Especially important chapters in this User Guide are Chapter 11- File usage, which outlines when each data file and its corresponding documentation are released and describes how to work with the data files, and Chapter 12 – Guidelines for tabulation, analysis and release.

Users wanting to obtain copies of the documents in the list or to obtain further information about the survey can contact Statistics Canada's National Contact Centre (toll-free 1-800-263-1136; 613-951-8116; infostats@statcan.gc.ca; teletypewriter (TTY) 1-800-363-7629).

2.4 Updates to CHMS Documentation

2.4.1 User guide

Background and contextual information related to a particular data file is included in the earliest possible version of this document, even if the corresponding data file has not yet been released. There are two exceptions to this: 1. detailed response rates, which are not permitted to be disseminated prior to the release of the corresponding data file, and 2. information not yet known that is related to a future release. The additional information in these cases will be released at the time the corresponding data file is released. The above approach allows users to not only use the currently released data files, but to also understand and plan for the use of soon to be released files.

There are three different versions of the User Guide, corresponding to the three main data release dates. Version 1 is released on September 20, 2012, Version 2 on November 21, 2012 and Version 3 on April 17, 2013. In order to keep track of content changes between the different versions, a summary of the type of information changed and location where the change was made is provided in the table below. Version 3 is expected to be the final version of the cycle 2 User Guide. If this is not the case, more information will be provided here as it becomes available.

Type of Change	**Section#**	**Content Description**	**Version #** **(where change first appears)**
Renamed section	2.4 – Updates to CHMS documentation (formerly 2.4 - Updates to User Guide)	Made title more general to be able to include a subsection on changes to the data files; also reworded and reordered text	2
Added section	2.4.2 – Data files	Describes updates made to CHMS data files and refers to more detailed information in Appendix 10	2
Changed content in section	6.3.5.3 – Environmental urine subsampling	Changed the number of respondents required for each sex for the 20 to 79 year old age group to 500 (was incorrectly indicated as 250)	2
Renamed section and added content	8.6 - Analytical range (formerly 8.6 – Limit of detection)	Provides a more complete explanation on the impact of upper analytical ranges and the Limits of Detection on survey data processing and analysis	2
Added section	9.11.3.1 – Weighting for fasted subsample	Describes how data was weighted for the fasted subsample	2
Added section	10.1.2 – Activity monitor response rates	Provides a description of how activity monitor response rates were calculated and refers to a table of the rates in Appendix 8B	2

Type of Change	**Section#**	**Content Description**	**Version #** **(where change first appears)**
Added section	10.1.3 – Indoor air sampler response rates	Provides a description of how indoor air sampler response rates were calculated and refers to a table of the rates in Appendix 8C	2
Added section	10.1.4 – Blood draw and urine response rates	Provides a description of how blood draw and urine response rates were calculated and refers to a table of the rates in Appendix 8D	2
Added section	10.1.5 – Response rates for blood and urine subsample; 10.1.5.1 – Fasted subsample response rates	Provides a description of how blood and urine subsample (fasted subsample) response rates were calculated and refers to a table of the rates in Appendix 8E	2
Added to and changed content in section, including title	10.3.3.16 – Blanks (previously 10.3.3.16 – Field blanks)	New third paragraph provides information on the blanks (cleaning and field) used with the indoor air samplers at the MEC for quality assurance/control; original two paragraphs were modified to improve understanding	2
Added section	10.3.4.2 – Indoor air data	Describes the corrections to the indoor air data made at Statistics Canada's Head Office to improve the quality; also see new paragraph in section 10.3.3.16	2
Added section	10.3.4.3 – Insulin data	Describes procedures and problems encountered and how they are not expected to have an impact on the quality of the insulin data; also see new section 12.2.3.5	2
Added and removed content in section	Table in 11.1	"Insulin" was added as a third test that was done for which data can be found on the fasted subsample file. Reference to "indoor air data" was removed from the explanation for the wave 2 master file since this type of data only applies to the two indoor air subsample files	2

Type of Change	**Section#**	**Content Description**	**Version #** (where change first appears)
Removed content from section	11.1.5 – Master file (wave 2)	Reference to "indoor air data" was removed from the explanation for the wave 2 master file since this type of data only applies to the two indoor air subsample files	2
Added content to section	11.1.6 – Activity monitor subsample file	Provided the exact number of records in file (4,948) and provided information on variable naming	2
Added content to section	11.1.7 and 11.1.8 – Indoor air subsample file – household level and person level	Mentioned that variable names are identical between the two indoor air files	2
Added content to section	11.1.9 – Fasted subsample file	"Insulin" was added as a third test that was done for which data can be found on the fasted subsample file.	2
Added section	12.2.3 – Data comparability over time	Describes differences in methods between cycles for specific types of CHMS data as well as corrections/procedures undertaken to maximize comparability between cycles	2
Moved paragraph to new section	12.2.3.1 - Normative scales (formerly under 12.2.2 – Some recommendations for doing analysis with data from cycle 2 of the CHMS)	Content moved to its own section to emphasize the use of normative scales to make data comparable between CHMS cycles	2
Moved paragraph to new section	12.2.3.2 - Activity monitor data for 3 to 5 year olds (formerly under 12.2.2 – Some recommendations for doing analysis with data from cycle 2 of the CHMS)	Content moved to its own section under data comparability to highlight the difference in methods with the activity monitor for 3 to 5 year olds between cycles	2

Type of Change	**Section#**	**Content Description**	**Version #** **(where change first appears)**
Moved paragraph to new section	12.2.3.3 – Phthalate data (formerly under 10.3.4.1 – Correcting for bias)	Content moved to its own section to emphasize that adjustments were made to the phthalate data to make data comparable between cycles	2
Added section	12.2.3.4 – Glucose data	Describes data comparability between cycles for the glucose data	2
Added section	12.2.3.5 – Insulin data	Describes data comparability between cycles for the insulin data; also see new section 10.3.4.3	2
Added section	12.2.3.6 – Combining data from cycles 1 and 2	Cautions users to wait until Feb 27, 2013 before combining data from cycles 1 and 2; this is when combined cycle 1,2 weight files and documentation are available	2
Renumbered former section 12.2.3 to 12.2.4	12.2.4 – Software packages	N/A	2
Added Appendix	Appendix 8B	CHMS cycle 2 activity monitor response rates by age group and sex	2
Added Appendix	Appendix 8C	CHMS cycle 2 indoor air sampler (IAS) response rates by household size	2
Added Appendix	Appendix 8D	CHMS cycle 2 blood draw and urine response rates by age group and sex	2
Added Appendix	Appendix 8E	CHMS cycle 2 fasted subsample response rates by age group and sex	2
Added Appendix	Appendix 10 – Changes to wave 1 variables	Listed some of the main changes that were made to the variables on the master and medication files between the cycle 2 wave 1 release (Sept 20, 2012) and the wave 2 release (Nov 21, 2012)	2

2.4.2 Data files

Some errors were discovered in the wave 1 master (household and MEC variables) and medication files that were released on September 20, 2012. These errors were corrected in the wave 2 master and medication files released on November 21, 2012. Most of the errors would have little impact on research done on the files between the two waves, but some may have had a more wide-spread, albeit small, impact. For instance, the weights for the full sample used in the two files were altered after an error was discovered, resulting in many estimates being slightly different (e.g., off by 0.1) when using the new weights vs. the old ones. Documentation of the changes made to the wave 1 variables to address the errors can be found in Appendix 10.

The data dictionaries corresponding to the household and medication files also changed between wave 1 and wave 2 releases, with the main change being the addition of universe statements and notes. Slight changes were also made to the wave 1 version of the derived variables documentation for the household and clinic variables.

Only the most recent master and medication files and their corresponding data dictionaries should be used.

3. Canadian Health Measures Survey (CHMS) background and objectives

3.1 CHMS background

Policy makers, provincial health departments, researchers and health professionals from many fields have expressed a need for an on-going national and comprehensive source of accurate health measures to assist them in addressing the health needs of all Canadians.

In 2003, Health Canada and the Public Health Agency of Canada supported Statistics Canada in obtaining funding for a direct measures health survey to address longstanding limitations within Canada's health information system. This support was announced in the 2003 federal budget as part of an extension of the Health Information Roadmap Initiative and permanent funding was secured in the 2008 Federal Budget.

Collection for cycle 1 of the survey took place from March 2007 to February 2009, with dissemination of the main results occurring from January 2010 until April 2011.

Cycle 2, the focus of this document, contains much of the same content as cycle 1. The main changes between the two cycles are summarized in Chapter 5.

The information collected will create national baseline data on the extent of such major health concerns as obesity, hypertension, cardiovascular disease, exposure to infectious diseases, and exposure to environmental contaminants. In addition, the survey will provide indications about illness and the extent to which many diseases may be undiagnosed among Canadians. Data from the CHMS will enable us to determine relationships between health status and risk factors, and to explore emerging public health issues.

3.2 CHMS objectives

Some of the objectives of the CHMS are to:

- estimate the number of people with selected health conditions, characteristics and environmental exposures based on direct health measures;
- ascertain relationships among risk factors, health behaviours, and health status; and
- maintain a biobank of biospecimens (urine, blood, deoxyribonucleic acid (DNA)) from a representative sample of Canadians, to be used for future research and surveillance.

4. Survey approval

4.1 Authority

The 2008 federal budget provided on-going funding for the Canadian Health Measures Survey (CHMS) to be conducted by Statistics Canada in partnership with Health Canada (HC) and the Public Health Agency of Canada (PHAC).

The survey falls under the authority of the federal *Statistics Act*. Statistics Canada may only collect health information for statistical research purposes and may not use or disclose individual participant information for any other purpose without the written consent of participants.

4.2 Ethical protocols and privacy standards

The CHMS was conducted in cooperation with provincial and municipal officials, with the support of health professional associations, and with the highest regard to Canadians' health and safety.

All processes of the CHMS were reviewed and approved by the HC and PHAC Research Ethics Board to ensure that internationally recognized ethical standards for human research were met and maintained. In addition, protocols were developed through extensive consultation with recognized experts and were performed by accredited health professionals in conformance with universal precautions.

Several meetings were also held with the Office of the Privacy Commissioner of Canada and with provincial privacy commissioners regarding CHMS protocols to ensure that participants' privacy rights were protected. A full Privacy Impact Assessment was completed for the CHMS and reviewed through the Office of the Privacy Commissioner of Canada—the authority that continues to provide oversight to the CHMS as well as a complaint route and redress mechanism for CHMS participants.

Participation in this survey was voluntary. The voluntary nature of the survey was stated in the introductory letter, brochure, video, *Information and Consent Booklet*, and on the Statistics Canada website (www.statcan.gc.ca\chms). The documents also emphasized the safety and standards used in all tests. CHMS staff answered any questions respondents had regarding the risks of participating in the tests and the use of their data in an interactive consent process throughout the household health interview and the visit to the mobile examination centre.

5. Survey content

Consultation on proposed content, data requirements and operational considerations has been on-going with many groups, individuals and agencies. Some of the groups involved include Health Canada, the Public Health Agency of Canada, and several expert advisory groups and committees, both internal and external to Statistics Canada.

Cycle 2 of the CHMS is, for the most part, unchanged from cycle 1. Both cycles were made up of a household interview and a visit to a mobile examination centre (MEC). The household interview included general demographic information and an in-depth health questionnaire. The MEC visit included not only physical measure tests but also the collection of blood and urine samples from respondents. Some samples were analyzed in a laboratory at the MEC, such as the complete blood count (CBC), which includes platelets, red blood count and white blood count. The remaining samples were analyzed at three external reference laboratories. Respondents were also asked to wear an activity monitor for the seven days following their visit to the MEC.

One of the main changes between cycle 1 and 2 was the addition of an indoor air measure. Respondents were asked to place an indoor air sampler in their home for seven days after their MEC visit to measure airborne substances and then send the sampler to a fourth reference laboratory for analysis.

Other changes to the survey between the two cycles included the addition of respondents aged three to five years old, the removal of oral health content, the addition of questions related to the indoor air measure, the modification of the fish and shellfish questions, the addition of human papillomavirus (HPV) and osteoporosis questions, and the addition of a neck circumference measure. There were also a large number of new and changed laboratory measures, particularly in relation to infectious disease markers and environmental exposures.

A content summary document for all cycles of the CHMS is available upon request.

6. Sample design

6.1 Target population

Cycle 2 of the Canadian Health Measures Survey (CHMS) covers the population 3 to 79 years of age living in the ten provinces and the three territories. Excluded from the survey's coverage are: persons living on reserves and other Aboriginal settlements in the provinces; full-time members of the Canadian Forces; the institutionalized population and residents of certain remote regions. Altogether these exclusions represent less than 4% of the target population.

6.2 Sample size and allocation

To produce reliable estimates at the national level by age group and sex, it was determined that this survey must be carried out on a sample of at least 5,700 persons over a two-year period to allow estimates for the following 11 groups: ages 3 to 5 (both sexes combined), ages 6 to 11 (males and females), 12 to 19 (males and females), 20 to 39 (males and females), 40 to 59 (males and females) and 60 to 79 (males and females).

6.3 Sampling frames and sampling strategy

To meet the requirements of the CHMS, a multistage sampling strategy was used. An overview of the sampling strategy is provided below; for further information about the sampling strategy refer to Giroux, Labrecque and Quigley (2012).[2]

6.3.1 Sampling of collection sites

Because the CHMS requires that participants report to a mobile examination centre (MEC), they should be able to travel to the MEC within a reasonable period of time. To accommodate this requirement the sample of dwellings is selected within collection sites. The Labour Force Survey (LFS) sampling frame was used to create the collection sites and control their size. The LFS sampling frame uses small geographic units that contain approximately 200 dwellings called clusters. These clusters were combined together to create the 257 potential collection sites for the CHMS. A collection site is a geographic area with a population of at least 10,000 and a maximum respondent travel distance of 50 kilometres in urban areas and 100 kilometres in rural areas. Areas not meeting these criteria were excluded.

Although only national estimates were required, the collection sites were stratified into five regions to ensure that the allocation of the sample was representative of the national population. The regions identified, based on Statistics Canada's standard regional boundaries, were British Columbia (including Whitehorse, Yukon), the Prairies (Alberta, Manitoba, Saskatchewan and Yellowknife, Northwest Territories), Ontario, Quebec and the Atlantic provinces (Newfoundland and Labrador, Prince Edward Island, Nova Scotia and New Brunswick).

A large number of collection sites (with few respondents) would be the recommended sampling strategy because it would help to optimize the precision of the estimates. However, the logistical and cost constraints associated with the use of a MEC restricted the number of collection sites to 18. Sixteen collection sites were originally allocated to the regions in proportion to the size of the population and to ensure a minimum of two sites per region for variance calculation. Later, it was determined that an additional two sites could be added. These sites were added to the British Columbia and Prairies regions following a study on the effect of cluster sampling on the variance estimates.[3] The number of sites selected by region is provided in Table 6.1.

Table 6.1 Selection of collection sites for the CHMS, by region – Cycle 2

Region	Estimated target population, ages 3 to 79, 2006 Census	Number of sites in region	Number of sites allocated
Atlantic	2,130,890	36	2
Quebec	7,044,190	50	4
Ontario	11,317,075	61	6
Prairies (including Yellowknife)	5,034,985	77	3
British Columbia (including Whitehorse)	3,848,985	33	3
Total	**29,431,725**	**257**	**18**

Within each region, the collection sites were sorted according to whether they belonged to a census metropolitan area (CMA), and then by the size of the population before the selection took place. A CMA is an area consisting of one or more adjacent municipalities centering on a large urban area (known as an urban core). The urban core must have a population of at least 100,000 to form a CMA. The collection sites were then sampled systematically with a probability of selection proportional to the size of their population. This selection method, combined with the sorting of sites by CMA and non-CMA and by population size, ensured that the sites selected would be allocated among CMA and non-CMA areas and among areas with larger and smaller populations. While not every province/territory would have a collection site, the CHMS sites were chosen to represent the Canadian population, east to west, with larger and smaller population densities.

Data collection at the 18 sites was carried out sequentially over two years. The sites were classified to take into account seasonality and the temporal effect, subject to operational and logistical constraints. The temporal effect was corrected by distributing uniformly the number of sites per region between the first year and the second year. The seasonality of the sites collected in each region during cycle 1 was also considered when determining the sequential order of the sites for cycle 2, given that researchers will have the opportunity to combine the data from the two cycles at a later date. See section 12.2.3.6 for more information.

The specific collection sites by city/municipality are listed in Appendix 3. Note that as the CHMS was designed to produce national estimates only, it is not recommended to do analysis at lower geographic levels as it could result in either extreme sampling variability or unstable estimates of the sampling variability.

6.3.2 Dwelling sampling

Several options were examined to determine how best to obtain the required number of participants by age group. The option chosen uses the 2006 Census as a sampling frame. The household composition of dwellings as of May 2006 was available and could be used to develop a design to meet the sample requirements in each age group. Prior to selecting the sample, the household composition from the Census was updated with more recent information from administrative files. To reduce under-coverage, new dwellings constructed since the 2006 Census or dwellings that were missed during the 2006 Census were added to the frame from the Address Register, a list of dwelling addresses across Canada compiled by Statistics Canada.

Within each collection site, dwellings with known household composition at the time of sample selection were stratified by the occupants' age at the time of the survey. Age was determined based on the starting date of data collection at each site. Six age-group strata were created, corresponding to the six CHMS age groups (3 to 5, 6 to 11, 12 to 19, 20 to 39, 40 to 59 and 60 to 79 years) as follows:

- 3 to 5 stratum[4]: dwellings where at least one 3-to-5-year-old is present, else,
- 6 to 11 stratum: dwellings where at least one 6-to-11-year-old is present, else,
- 12 to 19 stratum: dwellings where at least one 12-to-19-year-old is present, else,
- 60 to 79 stratum: dwellings where at least one 60-to-79-year-old is present, else,
- 20 to 39 stratum: dwellings where at least one 20-to-39-year-old is present, else,
- 40 to 59 stratum: dwellings where at least one 40-to-59-year-old is present, else,
- "Other" stratum: dwellings not included in the above-mentioned strata, such as vacant dwellings at the time of sampling or dwellings with people outside the CHMS age range based on household composition at the time sampling.

Each stratum had a high probability of having dwellings inhabited by persons in the desired age groups, whether they were the same occupants or were replaced by a similar household. Within each site, a simple random sample of dwellings was selected in each stratum. The sample size was allocated in each stratum so that, combined with the strategy for sampling participants in the survey, an equal number of respondents by age group would be obtained. Each selected dwelling was contacted to draw up a current list of the members of the household, and this list was then used to select survey respondents.

Table 6.2 shows the distribution of the number of dwellings selected per site. In all, 9,259 dwellings were selected, with an average of 514 dwellings per site.

Table 6.2 CHMS cycle 2 — Number of dwellings selected per site

Site	Number of dwellings
1	525
2	525
3	574
4	526
5	524
6	525
7	525
8	525
9	520
10	500
11	455
12	525
13	480
14	507
15	575
16	410
17	490
18	548
TOTAL	**9,259**

6.3.3 Respondent sampling

Different selection probabilities by age group within each stratum were used to ensure that the sampling targets were attained. The sampling factors shown in Table 6.3 were used for the majority of sites in cycle 2. Small increases were made to the sampling factors for the 12 to 19 and 60 to 79 age groups for the last three sites of cycle 2 and for the 3 to 5 age group for the final site in order to increase the number of people selected in these groups. The dwellings selected were contacted to obtain a list of current household members. In each dwelling, one or two people were selected, depending on the household composition. Two people were selected from households with children aged 3 to 11: one child randomly selected from those aged 3 to 11, and a second person aged 12 to 79. If no 3 to 11 year olds were living in the household, only one person was selected from the household members aged 12 to 79. The weight vector for the selection of people aged 12 to 79 was designed to avoid large person sampling weights. Since some age groups have a weight that is up to five times higher than that of other age groups, it is possible that a selected person would have a very high sampling weight when there are many household members in a dwelling. Therefore, when a specified minimum number of people aged 12 to 79 are living in a household, the weight for each person is reset to 1. In such cases, each household member has an equal chance of being selected. A careful balance of the parameters required for each of the measures put in place was obtained through studies and simulations.

Table 6.3 CHMS cycle 2 – Selection weight multiplicative factors for the person-level sampling strategy, by stratum and age group

	Age group					
Stratum[5]	**3 to 5**	**6 to 11**	**12 to 19**	**20 to 39**	**40 to 59**	**60 to 79**
3 to 5 stratum	3	2	5	1	1	1
6 to 11 stratum	3	2	5	1	1	1
12 to 19 stratum	3	2	5	1	1	1
20 to 39 stratum	3	2	5	2	1	1
40 to 59 stratum	3	2	5	1	2	1
60 to 79 stratum	3	2	5	1	1	2
"Other" stratum	3	2	5	1	1	1

6.3.4 Activity monitor and indoor air sampler subsamples

Separate subsample data and weight files were created for the activity monitor and indoor air sampler data. These are not true subsamples as respondents were not selected to participate in the measurements. All respondents were given an activity monitor to wear for 7 days and all households were provided with an indoor air sampler to place in their home for 7 days. The activity monitor data and indoor air sampler data are released as subsample files as analysis could only be performed on the records[1] where there was a minimum of 4 valid days of data entries (maximum of 10 days for the indoor air sampler). Due to the high volume of non-response meeting this criteria, a separate weight was created for the valid observations (see Sections 9.11.1 and 9.11.2). Creating a weight for the subsample corrects for potential bias due to differences between respondents who had valid data and those that did not.

6.3.5 Blood and urine subsamples

Subsamples of the survey respondents are selected for different laboratory tests on the blood and urine specimens collected. As some blood lipid tests require that the respondents fast, approximately half of the respondents are asked to fast prior to their MEC appointment and are usually scheduled to attend the MEC in the morning. Some environmental chemicals are also measured on subsamples of survey respondents due to the high cost of performing these tests on the entire sample. All of the subsamples were selected independently; that is, without consideration of who was selected for the other subsamples. This means that a specific respondent can be selected for zero, one, two or three of the subsamples.

The CHMS content summary document (available upon request) indicates which tests have been done on fasted and environmental blood, and environmental urine subsamples.

6.3.5.1 Fasted subsampling

Each sampled dwelling was randomly flagged to indicate whether a respondent should fast prior to the MEC appointment. It required that respondents fast overnight, whereas shorter eating restrictions were imposed on those with non-fasted appointments. Pregnant women, people with diabetes and other special cases were not asked to fast, even if the dwelling was flagged to be fasted. This random allocation reduced the potential for bias, which could occur if respondents were given the option to fast.

During collection, the sampling fractions were adjusted to obtain approximately half of the sample where respondents were selected to fast and were actually fasted prior to the MEC appointment.

6.3.5.2 Environmental blood subsampling

Each respondent aged 12 to 79 who attended the MEC was randomly flagged to indicate if he or she should have a wide range of environmental contaminants measures performed. These measures were the only environmental contaminants on blood done in cycle 2. The original goal was to obtain 250 respondents by sex for the following two age groups: 12 to 19 and 20 to 79, for a total of 1,000 respondents. Prior to Site 11, an extra 500 respondents was added to the target for the 20 to 79 age category, split evenly between males and females, bringing the final total to 1,500. The weighting was adjusted to account for this change in the selection probability for the later sites.

6.3.5.3 Environmental urine subsampling

Each respondent who attended the MEC was randomly flagged to indicate if he or she should have the environmental contaminants measures performed on their urine sample. The goal was to obtain 250 respondents for each sex, for the age groups (3 to 5, 6 to 11, and 12 to 19, and 500 respondents of each sex for the age group 20 to 79), for a total of 2,500 respondents overall.

7 Data collection

7.1 Preparation for collection

7.1.1 The Canadian Health Measures Survey (CHMS) team

The CHMS team is a diverse, well-trained, experienced group of individuals. The group can be subdivided into three sub-teams: field team, mobile examination centre (MEC) team, and head office staff. Each of these three sub-teams was responsible for specific portions of the survey.

7.1.1.1 Field team

The CHMS field team was comprised of Statistics Canada household interviewers and an interviewer manager. The household interviewers were primarily responsible for contacting selected households, conducting the household interview, explaining the MEC portion of the survey to respondents and attempting to secure their participation at the MEC. The CHMS interviewers were supervised by an interviewer manager who was responsible for conducting data quality assurance activities for the household component, overseeing the non-response follow-up and monitoring the household collection rates.

7.1.1.2 Mobile examination centre (MEC) team

The CHMS MEC team consisted of health professionals responsible for various components of the physical measures testing and a site manager who oversaw the day-to-day operation of the MEC. The health measures specialists were responsible for performing the majority of physical measures tests on respondents (e.g., blood pressure, anthropometry, fitness testing, spirometry). The laboratory technologists/phlebotomists conducted the specimen collection (blood and urine), performed the complete blood count (CBC) analysis and processed the biological samples for storage and shipment to the reference labs. In addition to the health professionals, the MEC team also consisted of administrative staff who worked at the reception desk and the appointment booking desk, as well as a site logistics officer who took care of the maintenance of the trailers.

7.1.1.3 Head office staff

The CHMS staff at head office processed data, monitored data collection response rates and data quality, provided human resource support and conducted periodic site visits to ensure staff were following correct protocols. In addition, head office staff prepared and mailed the respondent's report of selected laboratory tests, and provided information about the survey to respondents, the public and the media. Finally, a medical advisor followed-up with respondents about critical or sensitive results.

7.1.2 The mobile examination centre (MEC)

Two sets of trailers were acquired in order to conduct the physical measures and laboratory components of the CHMS. Each MEC was comprised of two trailers, the administrative trailer and the clinic trailer. The trailers had several different rooms including a reception area, restrooms, an administrative office, a fitness testing area, a screening area, an anthropometry area, phlebotomy areas and a laboratory.

Using MECs provided several benefits over a fixed examination centre site (e.g., an examination centre set up in an office building or hospital). They provided a standardized collection environment (equipment set up, room size, etc.) that was designed to meet Statistics Canada security and confidentiality policies as well as the flexibility of locating the MEC near the selected respondents' homes.

7.1.3 Informatics environment

Computer assisted interviewing (CAI) was used to capture the responses for the household, physical measure and laboratory components of the CHMS**.** CAI allowed for custom interviews for every respondent based on their individual characteristics and survey responses. This included:

- Questions that were not applicable to the respondent were skipped automatically.
- Edits to check for inconsistent answers or out-of-range responses were applied automatically and on-screen prompts were shown when an invalid entry was recorded. Immediate feedback was given to the respondent and the interviewer was able to correct any inconsistencies.
- Question text, including reference periods and pronouns, was customised automatically based on factors such as the age and sex of the respondent, the date of the interview and answers to previous questions.

In order to perform computerized data capture within the MEC, a unique data capture architecture had to be developed as there was a requirement for multiple users in different MEC rooms to access a single respondent's case file. This required the development of a complex, fully customized data capture application that used components of the computer assisted telephone interview (CATI) environment.

To reduce data entry errors, increase efficiency of data collection and reduce the need for double entry and data entry verification, the CHMS MEC data capture system was developed to accept direct input from other electronic testing equipment. This included communication (both one and two-way) between the application and the measurement devices (e.g., automated blood pressure cuff, weight scale). In cases where the direct input was not functioning and manual entry was required, the data were entered twice.

In order to support the electronic capture of physical measures data and to support the operational and administrative needs, the MEC was equipped with its own computer server. After each session at the MEC, encrypted data were transmitted from the trailer server via a dedicated out-going phone line to Statistics Canada headquarters. Encryption software was used to ensure the confidentiality of the data during transmissions between the MEC and headquarters.

7.1.4 Questionnaire design

7.1.4.1 Household questionnaire

http://www23.statcan.gc.ca/imdb-bmdi/instrument/5071_Q1_V2-eng.htm

The household questionnaire content was developed with input from stakeholders (Health Canada and the Public Health Agency of Canada) and from external experts who participated as members of various advisory committees. Most of the cycle 2 household questionnaire was identical to the cycle 1 questionnaire.

Focus groups and one-on-one interviews were conducted to look at specific questionnaire content, particularly the content new to cycle 2. As a result of this testing, improvements were made to questionnaire wording and instructions and to the flow of questions.

Some of the main modifications to the household questionnaire between cycle 1 and 2 of the survey included removing the oral health, grooming products use (moved to MEC), fish and shellfish (moved to MEC) and hepatitis A questions. Respondents ages 3 to 5 were also added to the survey. Some questions were not asked to the 3 to 5 year old age group since they were not applicable (e.g., questions related to school or eating disorders), were not standardized for this age group (e.g., Strengths and Difficulties/Health Utility Index) or a parent would not be able to accurately answer through observation (e.g., how refreshing a child's sleep is).

Some other changes from cycle 1 to cycle 2 include the following:

- Questions related to blood pressure/blood cholesterol were administered for respondents 6 to 11 years old for the first time, in order to correspond to the age range for which the blood pressure measure was taken at the MEC.
- Questions related to second hand smoke, cigarettes, etc. and to the indoor air sampler (new to cycle 2) were used to help assess the indoor air lab results.
- A question on vaccination and a question on PAP smear tests were added. These questions helped to correctly interpret the human papillomavirus (HPV) laboratory results and provide an indication of HPV immunization coverage in Canada.
- A question on osteoporosis was added for male and female respondents aged 40 to 79 in order to help monitor the prevalence of the disease in Canada and to provide insight on the serum results of calcium and vitamin D laboratory results.
- Dates for the blood transfusion question changed from "between 1978 and 1985" to "before 1990" to match the time period when the Canadian Red Cross began testing for hepatitis C (starting in 1990).

7.1.4.2 MEC questionnaire

http://www23.statcan.gc.ca/imdb-bmdi/instrument/5071_Q2_V2-eng.htm

Development of the MEC questionnaire (sometimes referred to as the clinic questionnaire) proceeded in much the same way as that of the household questionnaire. Content was determined through a comprehensive consultation process and multiple iterations of the collection application were generated. Each iteration was assessed on flow within the MEC for the respondent and the staff, and the quantity and quality of data collected.

The MEC questionnaire consisted of the indoor air questions (new to cycle 2), fish and shellfish consumption, and grooming product (moved from household) components. The fish and shellfish consumption and grooming products use components were moved to the MEC to decrease the time lapse between asking the questions and analyzing the chemical measurements in the respondents' blood and urine samples. The questionnaire also included a component on medication use as a follow-up to similar questions asked at the household interview as well as a component used to screen respondents for eligibility in physical measures tests (see Appendix 5).

7.1.4.3 Indoor Air Questions

The indoor air block was administered verbally to all respondents who took an indoor air sampler (one per household). Respondents were asked questions about their home environment and housing characteristics (e.g., presence of garage, parking facilities etc.) and certain behaviours and habits which would help determine a possible link to indoor volatile compound (VOC) levels. Their answers will help researchers to better understand the results from the indoor air sampler.

7.1.4.4 Fish and Shellfish Consumption

The fish and shellfish component gathered information on the respondent's consumption of fish and shellfish within the previous month and was used to help us better understand the respondent's blood and urine test results. The fish questions in the component were expanded upon the cycle 1 questions in order to allow for more in-depth analysis of sources regarding possible exposure to biomarkers such as mercury, cadmium, vitamin D and vitamin B12.

7.1.4.5 Grooming Products

The grooming products use component gathered information on respondents' use of chemical based grooming products over the previous 24 hours in order to better understand their blood and urine test results.

7.2 Collection

Cycle 2 data collection took place between August 27, 2009 and November 30, 2011 and included 18 collection sites spread across Canada from St. John's, Newfoundland and Labrador to Richmond, British Columbia (see Appendix 3). While one set of trailers was being used for collection, the second set would be moved to the next location and began the rigorous set of procedures required to prepare for collection.

7.2.1 Collection – Household interview

One to two weeks prior to the start of household interviews at each collection site, introductory material was mailed to the dwellings selected for the survey informing them that they would be contacted to participate in the survey. Interviewers called or drove to each dwelling to book an interview. Once contact was made with the household, the interviewer introduced the survey by outlining the basic steps of the survey and informing the person that participation was voluntary and that any information provided would be kept confidential under the authority of the *Statistics Act*.

Based on the demographic information collected, one or two persons in the household were selected to participate in the survey. A selection algorithm was used to try and reach an equal distribution of people among the different age groups. If two persons were selected in a household, one person was always 3 to 11 years old and the other 12 to 79 years old.

Prior to commencing the interview, the respondent was informed about the survey and was shown a brief four minute introductory video. For respondents between the ages of 3 and 11, an adult/guardian was present during the interview to answer questions with assistance from the child. All respondents aged 12 to 79 years who were able to answer questions on their own were asked to do so.

At the end of the interview the interviewer provided the respondent with an information package, explaining the MEC portion of the survey, information about the tests performed at the MEC and general information about the survey. The interviewer briefly reviewed the material in the information package and answered any questions. At that time, the interviewer informed the respondents that he/she had been assigned to a fasting or non-fasting appointment at the MEC. The fasting appointments required respondents 6 to 79 years of age to fast for 12 hours in preparation for blood lipid tests. See Appendix 4 for a list of the pre-testing guidelines provided to respondents during the household interview.

If the respondent was 14 years or older, the interviewer then asked the respondent to set up an appointment at the MEC. Parents were asked to set up the appointment for their 3 to 13-year-old child. At the end of each day, the interviewers transmitted all completed cases back to Statistics Canada using encryption software to ensure the confidentiality of the data during the transmission.

7.2.2 Collection – MEC

Upon arrival at the MEC, the respondent's information was logged into the database at the reception desk. The MEC staff verified that the respondent's name, sex, date of birth and official language (collected during the household interview) were correct. Adherence to the pre-testing guidelines was verified and documented within the application.

Prior to beginning the physical measures tests, the respondent must have given consent to participate in the MEC portion of the survey. Parents or guardians gave consent on behalf of children aged 3 to 13 while each child provided their assent to participate.

After the consent module, the respondent (or parent/guardian of younger respondents) was asked some screening questions using the application and the Physical Activity Readiness Questionnaire (PAR-Q) (see Appendix 6). Although the PAR-Q was designed for people aged 15 to 69, it was used on all respondents since it was deemed a useful screening tool for all ages. Parents or guardians completed the screening questions on behalf of children aged 6 to 13. Depending on the fasting status, the age of the respondent and the responses to the screening questions, some respondents were excluded from certain measures or laboratory tests (see Appendix 5 – Exclusion Criteria and Appendix 6 – Physical Activity Readiness Questionnaire (PAR-Q)).

Respondents then performed all measures or provided biospecimens for laboratory tests for which they were eligible. Every respondent could, at any time, refuse to participate in a measure or test. The order of the measures and blood tests was set in such a way that the effects of a certain measure (e.g.,

increased blood pressure from mCAFT) did not affect the results of another measure (e.g., resting blood pressure).

After the blood and urine samples were collected from survey respondents, they were processed, analyzed for the CBC and temporarily stored in fridges and freezers within the MEC laboratory. An initial DNA extraction step was carried out at the MEC prior to shipping. Stored samples were sent weekly to reference laboratories in Ottawa, Quebec City and Winnipeg for additional analyses related to general health, diabetes, kidney disease, cardiovascular disease, environmental contaminants and infectious diseases and for storage in the CHMS biobank in Winnipeg.

Prior to leaving the MEC, the respondent received a report of their measurements, and a letter for their health care provider if it was required (e.g., extremely high or low blood pressure). A few months after the visit to the MEC, a report of laboratory tests containing most of the respondent's blood and urine test results as well as spirometry results was sent to them. Respondents 14 years of age or older received their laboratory and spirometry reports, while parents of respondents 3 to 13 years old received their child's reports.

7.2.3 Collection – Home visit

In order to maximize response rates to the CHMS, respondents who were unwilling or unable to attend the MEC, but were willing to have a selected number of measures done in their home were offered the option of a home visit. There were 49 home visits during the entire cycle 2 collection period.

The home visit was conducted by a minimum of two CHMS staff members (most often a health measures specialist (HMS) and a laboratory technologist) using paper questionnaires for recording data. There were no differences in the procedures used to conduct the measurements in the home, however, there were minor differences in the equipment used (e.g., scale, stadiometer) because all home visit equipment needed to be portable. As with the visit to the MEC, the respondent could, at any time, refuse to participate in any measure or laboratory test. The home visit took less time because the indoor air, modified Canadian Aerobic Fitness Test (mCAFT), sit and reach and partial curl-up components were not performed at the home. Once back at the MEC, the HMS entered the data from the home visit into the application and coded it as a home visit. The senior HMS then verified the data entry. A report of measurements was mailed to the respondent within a few days of completion of the home visit and laboratory and spirometry test results were sent to them a few months later.

7.3 Minimizing non-response

7.3.1 Minimizing non-response – Household interview

To ensure the best possible response rate at the household, many practices were used to minimize non-response.

7.3.1.1 Introductory material

Before the start of each collection period, introductory materials were sent to the selected households, explaining the different steps of the survey and emphasizing the importance of the survey by providing examples of how CHMS data would be used.

7.3.1.2 Initiating contact

Interviewers were instructed to make all reasonable attempts to obtain interviews. When the timing of the interviewer's visit was inconvenient, an appointment was made to come back at a more convenient time. If no one was home on first visit, a notice that a CHMS interviewer had stopped by their home was left at the door. Numerous personal visits were made at different times on different days until potential respondents were home and available to do the interview. If interviewers were unable to make contact with anyone at the household, they tried phoning to arrange for a personal visit. In some cases a phone number was not available for the case and the interviewers had to research to find a name and number, using directory listings, neighbours, superintendents, and management/rental offices.

7.3.1.3 Refusal procedures

The interviewer tried to convince the respondent of the importance and potential benefits of participating in the survey. If the individual refused to participate in the survey, a refusal letter was sent and they were then contacted a second time by either another interviewer or the senior interviewer who, again, stressed the importance of the survey and the household's participation in it.

7.3.1.4 Language barriers

Some of the introductory materials (the introductory letter and the CHMS brochure) were available in English, French, and Mandarin. Respondents were interviewed in the official language of their choice (English or French). To remove language as a barrier to conducting interviews, where possible, the CHMS team recruited interviewers with some competencies in a language other than the two official languages. When necessary, cases were transferred to an interviewer or external interpreter with the appropriate language competency so that questions/instructions could be translated for the respondent. If, no one with a certain language competency could be found, it was also acceptable for a household member who was willing and able to translate for the respondent to do so. Note that this was not considered a proxy interview. The household member was simply translating the questions and the respondent’s answers directly to the interviewer, not answering for the respondent.

7.3.1.5 Youth respondents

Interviewers were obliged to obtain verbal permission from parents/guardians to interview youths between the ages of 12 and 17 who were selected for the survey. Several measures were taken to alleviate potential parental concerns and to ensure a completed interview. Interviewers provided the parent or guardian with a copy of the Information and Consent Booklet which contained a section dedicated to parents and guardians. This document explained the purpose of collecting information from youth, listed the subjects to be covered in the survey and explained the need to respect a youth's right to privacy and confidentiality.

When interviewing respondents 12 to 17 years of age, interviewers ensured that the parent was in the home but that the interview took place outside of parents/siblings earshot, unless permission was obtained from the youth for a parent to be present. If the selected youth could not be interviewed in a private setting, the interviewer read the questions out loud with a parent or guardian in the room. The youth then answered the questions directly onto the computer. If privacy and confidentiality could not be respected, the case was coded as a refusal with a permanent note indicating that privacy/confidentiality could not be respected.

If parents asked to know more about the type of questions asked in the survey, interviewers first directed them to the topics listed in the Information and Consent Booklet. If they asked to see the actual questions, interviewers showed them the content section of the Interviewer's Manual. For those parents who requested a copy of the questions, a copy was available through the Data Collection Manager, as well as at Statistics Canada's head office in Ottawa.

7.3.1.6 Proxy interviews

In the CHMS, parents/guardians answered questions about their children aged 3 to 11. This included all household modules that were applicable to children. Children assisted in responding to some questions for which the parents may not have known the answers (e.g., participation in activity during school hours).

In cases where the selected respondent 12 years of age or over was, for reasons of physical or mental limitations, incapable of completing an interview at the household, another knowledgeable member of the household supplied information about the selected respondent. While these proxy respondents were able to provide accurate answers to most of the survey questions, the more sensitive or personal questions were beyond their scope of knowledge. This resulted in some questions from the proxy interview being unanswered. Every effort was taken to keep proxy interviews to a minimum. The variable "PROXY" on the data set indicates whether or not a household interview was completed by proxy.

In cycle 2, 27% of the interviews were proxy, compared to 20% in cycle 1. This increase in the percentage of proxy interviews is primarily due to the addition of three to five year olds to the survey for cycle 2. Of the proxy interviews, 96.6% of the respondents were under 12 years of age and 3.4% were 12 years of age or older.

7.3.2 Minimizing non-response – MEC

Approximately 81.7% of respondents who completed a household interview in cycle 2 agreed to go to the MEC. Some of the main practices used to obtain this high level of participation are described below.

7.3.2.1 Non-response follow-up

MEC staff was responsible for following up with any respondents who did not book an appointment at the end of their household interview and did not call the MEC booking desk to set up an appointment within a few days after their household interview. The staff members followed similar refusal procedures as household interviewers.

7.3.2.2 Flexible MEC hours

Strategies specific to the MEC included the creation of MEC opening hours and appointment times that provided maximum flexibility to the respondent. The MEC staff tried to accommodate as many respondents as possible at each site. In addition, home visits were offered to respondents unwilling or unable to go to the MEC (see Section 7.2.3).

7.3.2.3 Refusal procedures

To minimize the non-response to the CHMS clinic component, the MEC staff were instructed to make all reasonable attempts to convince respondents who participated in the household interview to attend the MEC. The appointment booking desk staff, who had received specific training in handling refusal conversions, followed-up with respondents who refused to participate in the MEC portion of the survey. If they were unsuccessful in booking an appointment, the MEC site manager or senior HMS would call one final time to attempt to book an appointment and at this point they would offer a home visit to the respondent. Respondents who could not be contacted (e.g., no answer at the home phone number) were sent a "No Contact Letter" asking them to phone the MEC to book an appointment.

7.3.2.4 Language barriers

Mobile examination centre (MEC) staff handled language barriers in the same way as household interviewers. CHMS staff, external interpreters or family members with knowledge of the third language were used to help the respondent understand instructions and forms in order to complete the visit at the MEC. The consent forms and screening questions were available in Mandarin since a high percentage of individuals from this group were known to be living within some of the cycle 2 sample collection areas.

7.3.2.5 Youth respondents

As with the household interview, parents/guardians answered all questions about their children aged 3 to 11. Since the age of consent for the MEC portion of the CHMS was 14 years of age, parents/guardians also answered these questions for their 12 and 13 year old youths, though the youths usually assisted. Youths aged 14 and over were responsible for signing their own consent form and answering all questions although parents/guardians were in some cases present during their visit at the MEC and able to assist on difficult questions. To maximize efficiency at the MEC, the selected child or youth usually did the physical measure tests with one CHMS staff member while their selected parent was doing tests with another CHMS staff member.

In cases where the selected respondent 14 years of age or over was, for reasons of physical or mental limitations, incapable of answering questions and completing the consent form at the MEC, the parent/guardian assisted.

7.4 Physical measures protocols - MEC

Major changes to the MEC visit for cycle 2 include the addition of 3 to 5 year old children for many of the measures, the addition of a neck circumference measure (3-19 years of age), and the removal of the oral health component. Modified questions to the fish and shellfish consumption, indoor air (related to the indoor air sampler deployed at the respondent's household) and questions on grooming product were moved from the household questionnaire to the MEC.

In addition, there were numerous protocol changes made for cycle 2, including the adoption of the National Institute for Health (NIH) protocols for waist circumference, revision of the urine collection protocols to accommodate the addition of 3 to 5 year olds, removal of the 3rd set of blood pressure measurements and the 3rd and 4th post mCAFT blood pressure measurements, and the discontinuation of the mCAFT and partial curl-up measures for children ages 6 and 7.

A brief description of the physical measures protocols can be found below, with information on exclusion criteria for physical measures located in Appendix 5.

7.4.1 Anthropometry

The anthropometry component consisted of seven main physical measure tests: standing and sitting height, weight, waist and hip circumference, neck circumference and skin folds. All tests were done on eligible respondents aged 3 to 79 years old except for neck circumference which was done on eligible respondents aged 3 to 19 years old. Respondents having one or more acute or chronic conditions which could affect the results of an anthropometry test result were excluded from participating in that particular test.

7.4.1.1 Standing height

Standing height is an assessment of the maximum vertical size of the respondent. This measure was taken on all the respondents who were able to stand unassisted. Standing height was measured with a fixed stadiometer with a vertical backboard and a moveable headboard using a procedure based on the *Canadian Physical Activity, Fitness and Lifestyle Approach (CPAFLA) 3rd Edition*.[6] A self-reported height was captured for respondents who were not eligible because of an acute or chronic condition (e.g., in a wheelchair) or who refused to have their height measured.

7.4.1.2 Sitting height

Sitting height is an assessment of maximum vertical size when the respondent is sitting. It was measured on respondents who were able to sit unassisted. Sitting height was measured with a fixed stadiometer with a vertical backboard and a moveable headboard. The respondent's sitting height was measured, following the International Society for the Advancement of Kinanthropometry (ISAK) protocol.[7]

7.4.1.3 Weight

The respondent's weight was taken on a Mettler Toledo digital scale, following the CPAFLA protocol.[6]

7.4.1.4 Waist circumference

Waist circumference provides an indication of abdominal fat distribution and is an important indicator of the health risks associated with obesity. Two waist measurements were taken on respondents, one using the World Health Organization (WHO) protocol (as in cycle 1) and the other using the National Institute of Health (NIH) protocol (implemented for cycle 2). The Canadian Society of Exercise Physiology (CSEP) recently adopted the NIH protocol in order to have a more reliable method for measuring waist circumference and standardized measures of waist circumference. The continued use of the WHO protocol from cycle 1 allows for the comparison of data between cycles 1 and 2.

7.4.1.5 Hip circumference

Hip circumference is the maximal circumference measured at the hips or buttocks region (whichever is larger). It is used to calculate the waist-to-hip ratio (waist circumference divided by hip circumference) and is a simple method of determining body fat pattern. The protocol for hip circumference was based on the *Canadian Standardized Test of Fitness* (3rd Edition). [8]

7.4.1.6 Neck Circumference

Neck circumference is a new measure added for cycle 2. In adults, neck circumference is well-established as a screening tool for obstructive sleep apnea (OSA), a disease that results in airway obstruction during sleep, and is related to upper body obesity. The eligible age group chosen for the neck circumference measure was 3 to 19 year olds in order to obtain more data on children and youth. Neck circumference was not completed on respondents who were pregnant or had a visible deformity of the neck (e.g. goiter).

7.4.1.7 Skinfolds

Skinfold measurement is used to estimate the amount of body fat and the distribution of subcutaneous adipose tissue. The principle behind this technique is that the amount of subcutaneous fat (fat that lies directly beneath the skin) is proportional to the total amount of body fat. The respondent's skinfolds were measured following the five site method of the CPAFLA protocol. Skinfold measurements were not completed on respondents with a body mass index (BMI) equal to or greater than 30 kg/m^2.

7.4.2 Heart rate and blood pressure

The respondent's resting heart rate and blood pressure (BP) were measured following a new protocol created by the CHMS using information from the report "Hypertension Surveillance in Canada: Minimum Standards for Assessing Blood Pressure in Surveys". [9] This report was published by an expert committee consisting of members of the Canadian Hypertension Society, the Canadian Coalition for High Blood Pressure Prevention and Control and the Heart and Stroke Foundation of Canada.

Heart rate and blood pressure measurements were taken on all eligible respondents aged 6 to 79 years old using an oscillometric blood pressure measurement device. A series of blood pressure and heart rate measurements were taken at one minutes intervals following a five minute rest period. The last five measurements were used to determine the average resting heart rate and blood pressure. For cycle 2, a 3rd blood pressure measurement was removed and a 2nd measurement was only performed when required.

7.4.3 Spirometry

Spirometry was used to assess respondents' lung function. The measurement was taken on all eligible respondents aged 6 to 79 years old, following the 1994 Update of the Standardization of Spirometry article, published by the American Thoracic Society.[10] Reasons for being excluded from doing this test include heart attacks, chest or abdominal surgeries within the last 3 months, eye surgery within the last 6 weeks, tuberculosis, use of certain medications and pregnancy >27 weeks.

7.4.4 Musculoskeletal fitness

The musculoskeletal fitness component provides information on the muscular strength, endurance and flexibility of respondents. The partial curl-up and mCAFT tests were completed on all eligible respondents aged 8 to 69 years old (no longer done on 6 to 7 year olds). The measure was discontinued on this age group due to safety concerns (i.e., mCAFT test steps were too big to step safely) and due to the relatively low level of completed results obtained within this age category for cycle 1 (which could be attributed to the difficulty of this task within this age group).

The grip strength component was done on all eligible respondents aged 8 to 79 years old and the sit and reach test was done on all eligible respondents aged 6 to 69 years old.

Medication use, acute and chronic conditions, pregnancy, and failure to adhere to pre-testing guidelines (see Appendix 4) were reasons why a respondent would not have been eligible for any of the musculoskeletal measures.

7.4.4.1 Grip strength component

The most common method of measuring muscular strength in the field is isometric grip strength. Grip strength refers to the maximum force that can be generated by squeezing a handgrip dynamometer. The protocol for the grip strength component for adults 15 to 69 was based on the *Canadian Physical Activity, Fitness and Lifestyle Approach* (CPAFLA) 3rd Edition[6] and on the 1988 Canada Fitness Survey Longitudinal Study[11] guidelines for 7 to 14 year olds.

7.4.4.2 Sit and reach component

The sit and reach test is a commonly used field test to assess low back and hip-joint flexibility. The protocol for the sit and reach component for adults was based on the CPAFLA (3rd Edition)[6] and on the 1988 Canada Fitness Survey Longitudinal Study[11] guidelines for 7 to 14 year olds.

7.4.4.3 Partial curl-up component

The partial curl-up test evaluates the endurance of the abdominal muscles. The protocol for the partial curl-up component is based on the CPAFLA (3rd Edition).[6]

7.4.4.4 Modified Canadian Aerobic Fitness Test (mCAFT) component

Aerobic fitness is a measure of the combined efficiency of the lungs, heart, vasculature, and exercising muscles. The eligible age group for the mCAFT is 8 to 69 years of age. The protocol for the mCAFT component is based on the CPAFLA 3rd Edition for adults 15 to 69 years of age[6] and on the 1981

Canada Fitness Survey[12] for children aged 8-17 years of age. One of the major changes to the protocol in cycle 2 was to remove the 3rd and 4th post mCAFT blood pressure measurements, as the first two measurements taken at the 2 minute and 3 ½ minute were deemed sufficient.

7.4.5 Activity Monitor

An Actical physical activity monitor was given to all eligible respondents aged 3 to 79 years old, along with an adjustable belt, an XPRESSPOST envelope, and an information sheet. Respondents were asked to wear an activity monitor for 7 days following their visit to the MEC. Activity monitors provide key information on the wearer's level of physical activity, including intensity, timing (day and time), frequency and duration. Respondents in wheelchairs were excluded from this measure.

7.5 Indoor air sampler protocol

Indoor air samplers are small cylindrical devices that were placed in each household in order to establish national baselines for indoor air concentrations of several VOCs. These samplers were distributed to respondents for the first time in cycle 2 in order to measure a number of airborne substances, including benzene, methane and ethanol within their homes for seven days following the MEC visit. At the MEC one indoor air sampler was given per household, along with a pencil, XPRESSPOST envelope and an information sheet. After the 7 day collection period was over, respondents placed their indoor air sampler in a prepaid padded envelope addressed to CASSEN Testing Laboratories, who performed all analyses of the samplers.

7.6 Laboratory measures protocols

A brief description of the laboratory measures protocols used for the CHMS can be found below. Additional supporting documentation is available upon request.

7.6.1 Sample collection

Blood and urine were collected from all eligible respondents at the MEC in order to obtain nationally representative information on a variety of biomarkers (nutrition, chronic and infectious diseases, environmental exposure, etc.).

7.6.1.1 Blood collection

The blood was collected by a phlebotomist using a standardized venipuncture technique. The amount of blood taken from respondents depended on their age:

- 3 to 5 years : 22.0 ml
- 6 to 11 years: 28.5 ml
- 12 to 13 years: 48.8 ml
- 14 to 19 years: 52.8 ml

- 20 to 79 years: 72.8 ml

A deviation from the order of blood collection tubes typically done in a clinical setting was made to accommodate the priority of the test(s) being measured.

7.6.1.2 Urine collection

Respondents were asked to provide urine to conduct all the tests for which they were eligible. The urine was collected using the first catch urine as opposed to the mid-stream urine collected in cycle 1. This protocol was changed to optimize the new infectious disease testing introduced in cycle 2.

7.6.2 Analysis of CBC performed at the mobile examination centre (MEC)

The CBC was analyzed in the MEC laboratory by a technologist and performed for all respondents from whom a sample was collected. Results from any unsuitable samples (e.g., severe lipemia or clot in tube) were not reported.

7.6.3 Processing and storage of the blood and urine samples

It was important to process the specimens as soon as possible because the quality and integrity of the blood and urine specimens would deteriorate over time. Whole blood was centrifuged to separate plasma and serum and to allow for the aliquoting into smaller tubes. The urine was also aliquoted into smaller tubes. These tubes were placed in shipping trays and stored in the MEC laboratory in either the refrigerator or the freezer depending on the test. All specimens were stored as soon as processing was complete to ensure the quality of the samples' viability. A four hour ceiling from the point of collection was placed on the time for blood samples to be processed and stored. For most samples, however, processing and storage was achieved within two hours from the time of collection.

7.6.4 Shipment of the blood and urine samples

The shipping of the blood and urine aliquots was done once a week to the reference laboratory on pre-assigned shipping days. All packages were sent to one of the three reference labs: Health Canada in Ottawa for biomarkers related to chronic disease, general health (chemistry panel) and nutrition, Institut National de Santé Publique du Québec (INSPQ) in Quebec City for environmental biomarkers plus urine creatinine (environmental adjustments and kidney health) and the National Microbiology Laboratories NML in Winnipeg for infectious disease related analyses and the CHMS biobank. Shipments were packaged according to the International Air Transport Association (IATA) and Transport of Dangerous Goods regulations for biological specimens. All shipments were sent by overnight delivery using a courier company certified to handle dangerous goods and were scheduled to arrive at the reference laboratory only on weekdays. A specimen tracking system was also developed so that staff could determine the status of every tube shipped to the reference laboratory. In addition, the temperature of refrigerated shipments was monitored using pre-programmed devices that took the temperature every 15 to 30 minutes during shipping. This provided assurance that the samples received by the reference labs were maintained at an adequate temperature to preserve the sample integrity during shipping.

There were some occurrences where shipments were delayed and the condition of the samples required assessment. If it was determined that the shipment conditions had been inadequate, the affected results were removed to ensure that only the highest quality of data was reported to the CHMS by the reference laboratories.

8 Data processing

8.1 Verification

One of the unique features of the Canadian Health Measures Survey (CHMS) is that three different sets of data are collected for the same respondent: household interview data, physical measures data, and laboratory results data. Each set of data has to be processed on its own, yet they cannot be completely separated from each other because at various points during processing the three sets of data have to be used together.

The processing of the household interview data was performed in a manner similar to that of other health surveys at Statistics Canada. The data are validated first at the record level and then at the individual variable level, followed by detailed top-down editing. During data collection, processing takes place on a daily basis. The household interview responses have to be processed quickly in order for the data to be available at the mobile examination centre (MEC) in time for the respondent's visit to the MEC.

Similarly, the processing of the physical measures data begins with the data being validated first at the record level and then at the individual variable level, followed by detailed top-down editing. Also, because the laboratory tests are determined based on responses received at the MEC, the MEC data are used to generate a file containing a list of the tests for which laboratory results are expected to be received. This laboratory "control" file is used in processing the laboratory results data.

The processing of the laboratory data involves significant file manipulation due to the fact that several different file types are received from the MEC and the four reference laboratories. As with the household and the physical measures data, the laboratory data are validated at the record level and at the individual variable level, and several new variables[1] are then derived. The laboratory data are processed as quickly as possible so that any critical results that have been identified at the reference laboratories and the MEC are immediately available for reporting to respondents.

8.2 Mark-all-that-apply questions

During the initial phase of data processing, mark-all-that-apply questions are expanded, with each response category in the original question becoming a series of separate questions with a yes or no response. In the example below, the respondent selected both 2 and 3 as answers to the original question. The answers to the new questions are based on answers from the original question.

Original question:

CCC_Q96 **What type of hepatitis do you have?**

INTERVIEWER: Read categories to respondent. Mark all that apply.

1 **Hepatitis A**

(2) **Hepatitis B**

3 **Hepatitis C**

Final questions / responses:

CCC_96A **What type of hepatitis do you have? - Hepatitis A**

1 Yes

2 No

Don't Know, Refused

CCC_96B **What type of hepatitis do you have? - Hepatitis B**

1 Yes

2 No

Don't Know, Refused

CCC_96C **What type of hepatitis do you have? - Hepatitis C**

1 Yes

2 No

Don't Know, Refused

8.3 Coding

As in cycle 1, pre-coded answer categories were supplied for all suitable variables, and the interviewers and health measures specialists (HMS) were trained to assign a respondent's answers to the appropriate category. In the event that a respondent's answer could not be easily assigned to an existing category, several questions also allowed the interviewer to enter a long-answer text in the "Other-specify" category. All such questions were closely examined at head office during processing. For some household questions, the long answers were coded into one of the existing listed categories. If this was not possible, the response was coded as 'Other'. For the MEC responses, long answers were reviewed and some responses were coded to existing categories. For the remaining 'Other-specify' answers at both the household and the MEC, some new categories were created when there were sufficient numbers of responses. The remaining responses were coded as 'Other'. For all questions, the "Other-specify" responses will be taken into account when refining the answer categories for future cycles.

Statistics Canada is not permitted to release specific cigarette brands or drug names. As a result, this information, while collected, is not available on the data files. Instead, coded variables representing information about these responses are placed on the data file. The tar values[13] of specific cigarette brands were determined and put on the master file. Codes representing specific drug names were put on the medication file (see Appendix 7 for information on the medical classification systems used).

As in cycle 1, certain data were collected as long answers and had codes assigned. This included medication and other health product names and dosages, cigarette brand names, and job description

information. For medications and cigarette brand names, databases of standard descriptions were available in the computer-assisted interviewing (CAI) applications, and a code could be assigned at the time of collection based on a search of the appropriate database. Any description without a code was extracted during processing and coded manually. Over the course of the full survey, there were 1,960 medications and other health products that were manually assigned a Drug Identification Number (DIN), and 42 cigarette brand names that were manually assigned a brand code. For the assignment of North American Industry Classification System (NAICS) 2002 and National Occupational Classification - Statistics (NOC-S) 2001 codes, all 3,719 records[1] with data in the job description fields were extracted and sent for manual industry and occupation coding.

More information about these classification systems for job descriptions can be found at: http://www.statcan.gc.ca/subjects-sujets/standard-norme/naics-scian/2002/naics-scian-02index-eng.htm and http://www.statcan.gc.ca/subjects-sujets/standard-norme/soc-cnp/2001/noc2001-cnp2001-eng.htm.

8.4 Editing

Most editing of the data was performed at the time of the interview by the CAI application. It was not possible for interviewers/HMS to enter out-of-range values and flow errors were controlled through programmed skip patterns. For example, CAI ensured that questions that did not apply to the respondent were not asked. Edits requiring corrective action were incorporated in the CAI application to deal with inconsistent responses. In addition, warnings not requiring corrective action were also included to identify unusual (i.e., improbable rather than impossible) values as a means of catching potential errors and allowing correction at source.

At head-office, the data underwent a series of processing steps that resulted in some of the data being adjusted. As a final validation step, the CAI edits were re-applied to the processed data. As a result, the final data were complete and contained reserve codes for responses of "less than limit of detection", "not applicable", "don't know", "refusal" and "not stated".

Table 8.1 Reserve code of responses

Reserve Code label	Reserve code
Less than limit of detection	9.5, 99.5 etc.
Not Applicable	6, 96, 99.6 etc.
Don't Know	7, 97, 99.7 etc.
Refusal	8, 98, 99.8 etc.
Not Stated	9, 99, 99.9 etc.

Information on imputation of household income and the treatment of other missing data can be found in Sections 12.1.2 and 12.1.3.

8.5 Creation of derived variables

Derived variables (DV) are created to facilitate data analysis and to minimize the risk of errors. The three most common types of DVs are created by: collapsing data from one variable into groups; combining the data from one or more survey questions for a single respondent; or combining data from more than one respondent. DVs generally have a "D" (derived) or "G" (grouped) in the fourth character of the variable name.

Three other types of DVs can be found on the CHMS data file: converting responses given in various time units (days, weeks, months or a year) into one type of unit such as yearly; calculation of body mass index (BMI); and the creation of the Health Utility Index which is based on responses from a series of questions.

Specifications were received providing details on how to create each derived variable. Cycle 2 of the CHMS has over 500 DVs. Many of these DVs were created within the capture application in order to be able to provide initial results to the respondents at the end of their visit to the MEC, while others were created after the completion of data processing.

All derived variables underwent a validation process after creation in order to ensure that the output provided the requested data. Modifications to the specifications were necessary if, during validation by subject matter experts, the output from the derived variables was determined to be flawed. Complete documentation for all derived variables is available upon request.

8.6 Analytical range

Laboratory data are obtained from blood, urine or air samples. The pre-analytical conditions (e.g., sample volume, exposure time, specimen processing procedures, etc.) and methods used by laboratories define the lower and upper limits between which an analyte can be detected, referred to as the analytical range. The low end of the analytical range is determined by the limit of detection (LOD) of the method, which is the smallest quantity detectable that is statistically different from zero. The upper limit of the analytical range is determined by the highest standard concentration at which precision is no longer met (falling outside the linear range).

Outside the analytical range, the laboratory cannot determine the concentration of the analyte. When laboratory results for blood, urine and air analytes fall below or above the analytical range, the laboratory does not report a value and therefore the data are assigned a reserve code ending in 5 for results less than LOD (e.g.; 99.5) or ending in 7 for results above the upper analytical range (e.g.; 99.7). A full listing of the analytical ranges including LOD values can be found in the content summary document (available upon request). For the purposes of the CHMS data tables, any results less than the LOD are substituted by half the value of the LOD, whereas results above the upper analytical range are considered outliers and are excluded from statistical analyses.

9 Weighting

In order for estimates produced from survey data to be representative of the population covered and not merely of the sample itself, users must incorporate weighting factors (survey weights) into their calculations. A survey weight is assigned to each person included on the final dataset, that is, in the sample of persons who responded to the entire survey. This weight corresponds to the number of people represented by the respondent in the population as a whole.

The survey weight is calculated as the inverse of the probability that the respondent was selected for the survey. As described in chapter 6 (Sample design), the Canadian Health Measures Survey (CHMS) is a multi-stage sample that uses two sampling frames for selecting its sample; an area frame of geographic units (clusters) for constructing and selecting collection sites, and a list frame of the dwellings within each site. The probability of selection for the survey is determined by multiplying the probability of selection at each stage.

In accordance with the weighting strategy, the selection weights for collection sites are multiplied by the selection weights for dwellings (households) and adjusted for non-response. Following the conversion of household weights into person weights, the latter are adjusted for non-response at the interview stage and the mobile examination centre (MEC) stage, and with several other adjustments, this weight becomes the final survey weight. The steps of the weighting process are outlined in Sections 9.1 to 9.9.

9.1 Selection weights for collection sites

The first step is to calculate the selection weight for each collection site. For each site, this weight is calculated as follows within each region[14]:

Selection weight of a collection site within a given region[15] =

$$\frac{\textit{Sum of persons in all sites contained within the region}}{\textit{(Number of persons in the site) x (Number of the sites selected within the region)}}$$

There is no adjustment for non-response at the level of the collection sites, since all sites participated in the survey.

9.2 Selection weights for dwellings

For each collection site selected, a list of all dwellings was obtained from the sample frame which is based on the 2006 Census and updated with other administrative files. These dwellings were stratified into seven groups corresponding to the six age group strata and one other stratum, using household composition as specified in the Section 6.3.2 (Dwelling sampling). The sample of dwellings was allocated among these strata.

For a given dwelling, the selection weight is equal to the inverse of the probability of selection of the dwelling within the stratum to which it belongs.

$$\frac{Count\ of\ all\ dwellings\ in\ the\ strata}{Count\ of\ all\ dwellings\ in\ the\ sample\ from\ the\ strata}$$

This weight is then multiplied by the collection site selection weight.

9.3 Removal of out-of-scope units

Among all the dwellings sampled, a proportion is identified during collection as being outside the scope of the survey. Examples of out-of-scope cases for the CHMS are: demolished dwellings, dwellings under construction, vacant, seasonal or secondary dwellings, institutions, and dwellings in which all household members are under 3 or over 79 years of age or are full-time members of the Canadian Forces. These dwellings are simply removed from the sample, leaving only dwellings in the scope of the survey. The dwellings that remain in the sample retain the same weight as at the previous stage.

9.4 Household non-response

During collection, a proportion of the households sampled inevitably resulted in a non-response. This usually occurs when the household refuses to participate in the survey, provides unusable data or cannot be reached to conduct the interview. The weight of non-respondent households is redistributed to respondents within homogeneous response groups (HRGs). In order to create these HRGs, the score method based on logistic regression is used.[16] First a logistic regression model is created to estimate the response probability, and then these probabilities are used to divide the sample into groups with similar response properties. The logistic regression models are created from the limited amount of information available for all households. This includes data from the frame such as the strata, and geographic location, and paradata about the data collection such as the number of attempts to contact the household and the elapsed time between the first and last contact. An adjustment factor was then calculated within each HRG as follows:

$$\frac{Sum\ of\ weights\ for\ all\ dwellings\ (households)}{Sum\ of\ weights\ for\ all\ respondent\ households}$$

The weight of respondent households is multiplied by this adjustment factor to produce the adjusted household weight. Non-respondent households are eliminated from the weighting process starting at this point.

9.5 Creation of the person weight

Since the final sampling unit for the CHMS is the person, the adjusted household weight up to this point must be converted into a person weight. This is obtained by multiplying the adjusted household weight by the inverse of the probability of selection of the person selected in the household. It should be kept in mind that the probability of a person being selected changes depending on the number of persons in the household and their ages (see Section 6.3.3, Respondent sampling for more details).

9.6 Non-response at the questionnaire level

The CHMS has a three-stage collection process. First, the interviewer obtains the complete list of persons living in the household, then he or she interviews the person(s) selected in the household, and finally, the selected person or persons report to the CHMS MEC.

In some cases, interviewers succeed in completing only the first step, either because they cannot contact the person(s) selected, or because the person or persons selected refuse to be interviewed. Such cases are defined as non-responses at the questionnaire level, and an adjustment factor must be applied to the weights of respondent persons to compensate for this non-response. Just as for non-response at the dwelling (household) level, the adjustment is applied within classes defined by the score method using response probabilities from a logistic regression model. The model is based on the characteristics available for all respondents and non-respondents, which includes all the characteristics collected when the members of the household are listed, such as the number of persons in the household, in addition to geographic information and paradata. An adjustment factor is calculated within each class, as follows:

$$\frac{\textit{Sum of weights for all selected persons}}{\textit{Sum of weights for all selected persons responding to the questionnaire}}$$

Thus, the weight of respondent persons is multiplied by this adjustment factor. Persons not responding to the questionnaires are removed from weighting starting at this point.

9.7 Non-response at the MEC level

Respondents to the questionnaire are then invited to go to the CHMS MEC for physical measurements. In some cases, people refuse to participate or do not keep their appointment at the MEC. Such cases are defined as non-responses at the MEC level, and to compensate for this non-response, an adjustment factor must be applied to the weights of the MEC participants. Just as for non-response at the dwelling (household) and questionnaire levels, the adjustment is applied within classes defined by their probability of attending the MEC. This probability is obtained from a logistic model using the characteristics available for respondents and non-respondents. All the characteristics collected on the questionnaire during the interview (such as income class, whether or not the respondent is employed, general health status, and frequency of smoking), in addition to geographic information and paradata, were made available to create the non-response models. An adjustment factor is calculated within each class as follows:

$$\frac{\textit{Sum of weights for all persons responding to the questionnaire}}{\textit{Sum of weights for all persons participating at the MEC}}$$

The weights of the persons participating at the MEC were accordingly multiplied by this adjustment factor. Persons who did not report to the MEC are removed from the weighting process starting at this point.

9.8 Winsorization

Note that following a series of adjustments applied to the weights, it is possible that some units will have weights that stand out from the other weights to the point of being aberrant. Some respondents may actually represent an abnormally high proportion in their group and therefore strongly influence both the estimates and the variance. To avoid this situation, a respondent weight that contributes aberrantly to the age-sex group is adjusted downward using a method known as "winsorization." In this process, respondent weights that are considered to be outliers are replaced by the highest non-outlier weight for that age and sex group. All of the weights are then adjusted to redistribute the surplus weight (the part of the weight that is higher than the highest non-outlier weight). This is done by multiplying the non-outlier weights by an adjustment factor to create the winsor adjusted weights. The adjustment factor is calculated as:

$$\frac{Sum\ of\ original\ \ final\ \ weights}{Sum\ of\ \ non-outlier\ \ weights}$$

9.9 Calibration

The last step required to obtain the final CHMS weight is calibration. This procedure is applied to ensure that the sum of the final weights corresponds to the estimates of populations defined at the scale of the five Canadian geographic regions[14], for each of the 12 age-sex groups of interest, the six age groups 3 to 5, 6 to 11, 12 to 19, 20 to 39, 40 to 59 and 60 to 79 for each sex. An additional criterion was used to calibrate the 20 to 39 age group to compensate for the fact that persons in this age group living with kids have a greater chance of being selected than those living without kids. In households where there was at least one person aged 3 to 11, a second person aged 12 to 79 was selected for the survey. The second person selected was usually a parent aged 20 to 39. To compensate for any potential bias caused by the selection method the 20 to 39 age group was split into those living with and without children aged 3 to 11.

The population estimates are based on the most recent census counts, as well as on counts of births, deaths, immigration and emigration since then. The calibration was carried out using the mean of the monthly estimates (covering the survey period) for each cross-tabulation of standard regional boundaries and age-sex groups. The population estimate for the 20 to 39 age group in each region was split into those living with and without kids aged 3 to 11 based on the estimated ratio of 20 to 39 year olds living with and without kids from the sampling frame for cycles 1, 2 and 3.

After calibration, the weight adjustment was obtained. The resulting weight is the final CHMS weight found in the data file bearing the variable name WGT_FULL.

9.10 Bootstrap weights

The CHMS uses a complex sampling design to select the sample and there are no simple formulas that can be used to calculate the variance of the survey estimates. Instead, a re-sampling approach known as the bootstrap method is used to approximate the sample variance. The bootstrap method involves creating subsamples of the full sample by randomly selecting « n-1 » collection sites with replacement among the « n » collection sites in each region.[17] An adjusted weight is then calculated for each

respondent in the selected subsample. This is repeated 500 times to create the bootstrap weights. To calculate the variance of a point estimate (such as the mean), the estimate for each of the 500 replicates is calculated using the bootstrap weight. The variability among the 500 estimates gives the variance estimate. The bootstrap weights are provided on the final CHMS weight file with the names BSW1-BSW500. Refer to Section 11.5 on the use of the weight variable and Section 12.2.4 on available software for more information.

9.11 Weighting for selected subsamples

9.11.1 Weighting for activity monitor data

A separate weight was created for the analysis of the activity monitor data, even though all respondents were asked to wear the activity monitor for one week following the MEC appointment. For this reason the activity monitor data is not a true subsample. An analysis could only be performed for respondents who had at least 4 days of valid data on the activity monitor. To compensate for any bias due to a difference in the respondents who had at least 4 valid days of data and those that did not a separate weight was created for the activity monitor data.

The weighting steps described in Sections 9.1 to 9.7 were carried out and then an additional adjustment was made to account for respondents who did not have at least 4 valid days of data. Cases where there were not at least 4 valid days of data were treated as non-respondents to the activity monitor. The weights of the persons with four or more valid days were multiplied by an adjustment factor to account for non-response. Similar to the non-response adjustment at the collection site, household and MEC levels, the adjustment is applied within homogeneous response classes. The classes are created based on the probability of response from a logistic model using the characteristics available for all respondents and non-respondents. All of the characteristics collected during the interview and at the MEC, in addition to geographic information and paradata, were available to create the non-response models. The "score" method is then used to define the classes. The adjustment factor within each class is calculated as follows:

$$\frac{\textit{Sum of weights for all persons participating at the MEC}}{\textit{Sum of weights for all persons with four or more days of valid entries}}$$

The weight of the persons who have valid entries for four or more days were multiplied by this adjustment factor. Persons who had less than four days of valid entry were removed from the weighting process. The final three steps, winsorization, calibration and bootstrap weights (see Sections 9.8 to 9.10) were then applied to the respondents. After calibration, the weight adjustment was obtained. The resulting weights can be found on the activity monitor file with the name wgt_acmo and the corresponding bootstrap weights are labelled as BSW1-BSW500.

9.11.2 Weighting for indoor air sampler data

Similar to the activity monitor weighting, weights are created for the indoor air sampler (IAS) data to compensate for any bias between the respondents with valid measures and those without. A valid measure for the indoor air sampler was an exposure time of 4 to 10 days. The indoor air sampler is not

a true subsample as all households were supposed to receive an indoor air sampler to place in their home. The indoor air sampler data is representative of the household and it is the same for two respondents in the same household. Two sets of weights were created to allow the data to be analysed at both the household level and the person level. The household level weights allow for the creation of baseline measure of contaminants in Canadian households. The person level weights allow the measures from the indoor air sampler to be related to the characteristics of the persons selected for the survey.

9.11.2.1 Household-level weights

To create the household-level weights the first four steps described in Sections 9.1 to 9.4 remain the same. The person design weights (i.e. the inverse of the probability of selecting the chosen respondent from the household) are not considered for the IAS household weights. Instead the initial household weight is adjusted to account for non-response at the questionnaire and MEC levels with an additional adjustment for households that did not have valid IAS data. The non-response adjustments are again done using the score method with logistic regression. The difference is that the models consider whether or not at least one person responded for the household rather than looking at individual persons if two people were selected from the same household. However, in households where two people were selected, one person was chosen to represent the household for the non-response models because whether or not a person chooses to respond to each level of the survey is in part dependent on the characteristics of the individual selected. The variables[1] used in the non-response model at each level are a combination of the household level and person level characteristics available for all respondents and non-respondents.

The calibration of the final weights is done at the household level so that the weighted estimate reflects all dwellings in Canada that are not an institution or a collective dwelling. The sum of the final household weights corresponds to the estimates of the number of dwellings in each of the five Canadian geographic regions[14] for households of size 1, 2 and 3 or more persons. The household counts are based on the most recent census counts available, with updates to account for births, deaths, immigration and emigration. The values used for calibration were the mean of the monthly estimates (covering the survey period) for each cross-tabulation of household size and geographic region minus an estimate of the number of households with only persons aged over 79 as they are not part of the CHMS target population. The proportion of households where all occupants are over the age of 79 was estimated from the dwelling sampling frame of CHMS cycles 1, 2 and 3.

The final indoor air sampler household weights were obtained after calibration. The weights are labelled as wgt_iash and can be found on the indoor air sampler household file. The bootstrap weights for variance estimation were created as described in Section 9.10 and are labelled BSW1-BSW500.

9.11.2.2 Person-level weights

The weighting steps described in Sections 9.1 to 9.7 were carried out and then an additional adjustment was made to account for respondents who did not have 4 to 10 valid days of data. Households that did not satisfy this condition were treated as non-respondents to the indoor air sampler. An adjustment was calculated at the household level to adjust the weight of persons with valid IAS data to account for non-response. Similar to the other non-response adjustments, the adjustment is applied within homogeneous response classes. The classes are created based on the probability of response from a logistic model using the characteristics available for all respondent and non-respondent households. As response to the

IAS is at the household level, household level variables available from the household questionnaire, in addition to geographic information and paradata, were available to create the non-response models. The respondents and non-respondents are divided into equal size classes based on their predicted probability of response from the logistic model. The adjustment factor within each class is calculated as follows:

$$\frac{\textit{Sum of weights} \text{ for all households with at least one person participating at the MEC}}{\textit{Sum of weights} \text{ for all households with 4 to 10 days of valid data}}$$

The weight of the persons whose household has 4 to 10 valid days of data were multiplied by this adjustment factor. Persons whose household did not meet this criteria were removed from the weighting process. The final three steps, winsorization, calibration and bootstrap weights (see Sections, 9.8 to 9.10) were then applied to the respondents. After calibration, the weight adjustment was obtained. The resulting weights can be found on the activity monitor file with the name wgt_iasp and the coordinating bootstrap weights are labelled as BSW1-BSW500.

9.11.3 Weighting for blood and urine data

Three subsamples were selected during the sample collection: the fasted subsample, the environmental blood subsample and the environmental urine subsample (see Section, 6.3.5). For each of these subsamples a sample weight was created that accounts for this extra step in the sample selection process. As the selection for each subsample was done independently, a separate weight is calculated for each of these subsamples.

9.11.3.1 Weighting for fasted subsample

The fasted subsample was selected when the sample of dwellings were selected, and thus occurred prior to completion of the household questionnaire. To create the fasted subsample weights, the steps described in Sections 9.1 to 9.4 remain the same and then at the person weight creation step (see Section 9.5) the subsample flags that were assigned to the dwellings were attributed to the selected person(s). Before adjusting for non-response at the questionnaire level (see Section 9.6), the person weight of those selected for the fasted subsample was adjusted to incorporate the subsample sampling weight. An adjustment factor was derived by collection site and stratum as follows:

$$\frac{\textit{Sum of weights of selected persons}}{\textit{Sum of weights of persons selected for the fasted sub-sample}}$$

The weights of the persons selected for the subsample were accordingly multiplied by this adjustment factor. Persons who were not selected for the fasted subsample were removed from the weighting process.

An additional step was required to adjust for persons who were selected for the subsample but who did not fast or did not provide blood. Such cases were defined as non-respondents to the fasted subsample and to compensate for this non-response an adjustment factor was applied to the weights of the persons with a valid measure. Just as for non-response at the collection site, dwelling (household) and MEC

levels, the adjustment is applied within homogeneous response classes. The classes are defined using the score method with logistic regression based on the characteristics available for all respondents and non-respondents. All the characteristics collected on the questionnaire during the interview and measures taken at the MEC, in addition to geographic information and paradata, were available for creating the homogeneous response classes. An adjustment factor was calculated within each class as follows for the fasted subsample:

$$\frac{\textit{Sum of weights for all persons selected to fast and participating at the MEC}}{\textit{Sum of weights for all persons selected who were fasted and had a valid measure}}$$

The weight of the persons who were selected for the subsample who had a valid measure were accordingly multiplied by this adjustment factor. Persons who did not have a valid measure were removed from the weighting process.

The final three steps, winsorization, calibration and generating bootstrap weights (see Section, 9.8 to 9.10) were then applied to the fasted subsample. After calibration the weight adjustment was obtained. The resulting weight is the final CHMS weight found on the subsample data file and it is called wgt_fast. The bootstrap weights are labelled BSW1-BSW500.

9.11.3.2 Weighting for environmental blood and urine subsamples

Respondents were selected for the blood and urine subsamples when they attended the MEC appointment, so the non-response adjustments that are applied to the full sample weights remain the same for the roster, household questionnaire and MEC levels (steps 9.1 to 9.7). The selection for these subsamples was done independently so two separate weights were created, however the same procedure was used for both subsamples. Two additional adjustments are applied to the weights from step 9.7 to adjust for respondents not selected for the subsample and to account for non-response to the subsample. First the weight of the respondents not selected for the subsample is redistributed to the weights of the selected respondents using the following adjustment factor within each combination of site, age group and sex:

$$\frac{\textit{Sum of weights of all persons participating at the MEC}}{\textit{Sum of weights of persons participating at the MEC and selected for the sub – sample}}$$

The weights of the persons selected for the subsample were accordingly multiplied by this adjustment factor. Persons who were not selected for the subsample were removed from the weighting process.

The weights were then adjusted to account for non-response to the subsample, which occurred when a respondent did not provide blood or urine or a valid measure could not be obtained on at least one of the laboratory tests. The adjustment is applied within homogeneous non-response classes, similar to the other non-response adjustments at the collection site, household and MEC levels. The classes are defined using the score method with logistic regression based on the characteristics available for all respondents and non-respondents. All the characteristics collected on the questionnaire during the interview and measures taken at the MEC, in addition to geographic information and paradata, were available for creating the homogeneous response classes. An adjustment factor was calculated within each class as follows for the environmental subsamples:

$$\frac{\textit{Sum of weights for all persons selected for the sub-sample and participating at the MEC}}{\textit{Sum of weights for all persons selected for the sub-sample who had a valid measure}}$$

This adjustment factor is multiplied by the weight of all the respondents selected for the subsample who had a valid measure. Persons who did not have a valid measure were removed from the weighting process.

The weights are then winsorized and calibrated and the bootstrap weights are generated following the steps described in sections 9.8 to 9.10. The resulting weight is the final subsample weights and it is labelled wgt_eb for the environmental blood subsample and wgt_eu for the environmental urine subsample. The weights are available on the subsample data file and the bootstrap weights, labelled BSW1-BSW500, are available on the weight file for each subsample.

10. Data quality

10.1 Response rates

One of the most important ways to ensure that the survey data collected are truly representative of the Canadian population is to maximize participation of selected respondents in all parts of the survey. Response rates are very useful data quality indicators that help to measure success in achieving this goal. The following section describes how response rates are derived for the sample and subsamples.

10.1.1 Household and MEC response rates

In all, 8,520 of the dwellings selected in the Canadian Health Measures Survey (CHMS) were within the scope of the survey.[18] Of these dwellings, 6,465 agreed to provide information on the composition of the household, for a household response rate of 75.9%. From these respondent households, 8,652 persons were selected (one or two persons per household) to participate in the survey, of whom 7,830 agreed to respond to the questionnaire, for a response rate of 90.5%. Of these persons, 6,395 then reported to the CHMS mobile examination centre (MEC) for physical measurements, for a response rate of 81.7%. At the Canadian level, a combined response rate of 55.5% was observed for cycle 2 of the CHMS. It is important to note that the combined response rate is not obtained by multiplying the response rates at the person and household levels (or questionnaire level and the MEC level), since two persons were selected in some households. Appendix 8A shows the combined response rates and the relevant information for calculating them for the given age groups and gender.

Below is a description of how the different components of the equation must be used to calculate combined response rates correctly.

Response rate at the household level

$$HR = \frac{R}{S} = \frac{\textit{\# of respondent households}}{\textit{\# of households within the scope of the survey}}$$

Response rate at the person level among households where <u>one person</u> was selected (questionnaire)

$$PQ1 = \frac{Q1}{PS1} = \frac{\textit{\# of respondents to the questionnaire among households where one person was selected}}{\textit{\# of persons selected among households where one person was selected}}$$

Response rate at the person level among households where <u>two persons</u> were selected (questionnaire)

$$PQ2 = \frac{Q2}{PS2} = \frac{\text{\# of respondents to the questionnaire among households where two persons were selected}}{\text{\# of persons selected among households where two persons were selected}}$$

Response rate at the person level among households where <u>one person</u> was selected (MEC)

$$PC1 = \frac{C1}{Q1} = \frac{\text{\# of participants at the MEC among households where one person was selected}}{\text{\# of respondents to the questionnaire among households where one person was selected}}$$

Response rate at the person level among households where <u>two persons</u> were selected (MEC)

$$PC2 = \frac{C2}{Q2} = \frac{\text{\# of participants at the MEC among households where two persons were selected}}{\text{\# of respondents to the questionnaire among households where two persons were selected}}$$

Ratio for households where <u>one person</u> was selected

$$RR1 = \frac{R1}{R} = \frac{\text{\# of respondent households among those where one person was selected}}{\text{\# of respondent households}}$$

Note: The "*# of respondent households among those where one person was selected*" is equal to the "number of persons selected among households where one person was selected" in Appendix 8A.

Ratio for households where <u>two persons</u> were selected (area frame)

$$RR2 = \frac{R2}{R} = \frac{\text{\# of respondent households among those where two persons were selected}}{\text{\# of respondent households}}$$

Note: The "*# of respondent households among those where two persons were selected*" is obtained by dividing by 2 "the number of persons selected among household where two persons were selected" in Appendix 8A.

Once all the above components have been calculated, a user can calculate the combined response rate using the following formula:

Combined response rate

$$COMBRR = HR * [(RR1 * PQ1 * PC1) + (RR2 * PQ2 * PC2)]$$

The household response rate (HR) is determined by the total number of respondent households (6,465) and the households in scope of the survey (8,520). It cannot be calculated separately for each age group and sex. For this reason, the household response rate is not included in Appendix 8A.

$$HR = \frac{R}{S} = \frac{\textit{\# of respondent households}}{\textit{\# of households within the scope of the survey}}$$

$$HR = \frac{6{,}465}{8{,}520} \qquad 0.759$$

$$= \mathbf{75.9\%}$$

Below you will find an example of calculating the combined response rate for Canada using the information provided in Appendix 8A.

$$PPQ1 = \frac{Q1}{PS1} = \frac{3{,}828}{4{,}278} = 0.895$$

$$PPC1 = \frac{C1}{Q1} = \frac{3{,}057}{3{,}828} = 0.799$$

$$PPQ2 = \frac{Q2}{PS2} = \frac{4{,}002}{4{,}374} = 0.915$$

$$PPC2 = \frac{C2}{Q2} = \frac{3{,}338}{4{,}002} = 0.834$$

$$PRR1 = \frac{R1}{R} = \frac{4{,}278}{6{,}465} = 0.662$$

$$PRR2 = \frac{R2}{R} = \frac{4{,}374 \div 2}{6{,}465} = 0.338$$

Combined response rate =

$0.759 * [(0.662 * 0.895 * 0.799) + (0.338 * 0.915 * 0.834)]$

$= 0.555$
$= \mathbf{55.5\%}$

10.1.2 Activity monitor response rates

Of the 6,395 participants who reported to the CHMS mobile examination centre for physical measurements, 6,375 participants were offered an activity monitor. Of these persons, 4,948 returned the activity monitor with at least four days of valid entries. The combined response rate for the activity monitor was 42.4%. It is important to note that the combined response rate is not obtained by multiplying the response rates by the person and household scales, since two persons were selected in some households. Appendix 8B shows the combined response rates and the relevant information for calculating them for the given age groups by gender.

Below is a description of how the different components of the equation must be manipulated to calculate combined response rates correctly. Two additional response rates are required to derive the activity monitor combined response rates:

Response rate at the person level among households where <u>one person</u> was selected (activity monitor)

$$AM1 = \frac{V1}{OF1} = \frac{\textit{\# of persons who returned the activity monitor with at least four days of valid entries among MEC participants where one person was selected}}{\textit{\# of participants at the MEC who were offered an activity monitor among households where one person was selected}}$$

Response rate at the person level among households where <u>two persons</u> were selected (activity monitor)

$$AM2 = \frac{V2}{OF2} = \frac{\textit{\# of persons who returned the activity monitor with at least four days of valid entries among MEC participants where two persons were selected}}{\textit{\# of participants at the MEC who were offered an activity monitor among households where two persons were selected}}$$

Activity monitor combined response rate

$$AMCBRR = HR * [(RR1 * PQ1 * PC1 * AM1) + (RR2 * PQ2 * PC2 * AM2)]$$

where HR, RR1, PQ1, PC1, RR2, PQ2 and PC2 are described in section 10.1.1.

Below is an example of calculating the activity monitor combined response rate for Canada using the information provided in Appendices 8A and 8B and the calculations in section 10.1.1.

$$AM1 = \frac{V1}{OF1} = \frac{2{,}254}{3{,}042} = 0.741$$

$$AM2 = \frac{V2}{} = \frac{2{,}694}{} = 0.808$$

OF2	3,333

Activity monitor combined response rate =

0.759 * [(0.662 * 0.895 * 0.799 * 0.741) + (0.338 * 0.915 * 0.834 * 0.808)]

= 0.424

= **53.6%**

10.1.3 Indoor air sampler response rates

Since only one indoor air sampler (IAS) was provided per household even if sometimes two respondents were selected in the household for the survey, the response rate for the IAS is given at the household level. There were 6,465 household where at least one person was selected to participate in the survey. Of these households, 5,874 households had a least one person complete the household questionnaire. Of those households, 4,782 households had at least one person who reported to the CHMS mobile examination centre for physical measurements and 4,709 of these persons were offered an IAS to place in their home. Of those households, 3,857 returned the IAS and had a valid exposure time between 4 and 10 days. The combined response rate for the IAS is 46.0%. Appendix 8C shows the combined response rates and the relevant information for calculating them for the household of size 1, 2 and 3 or more persons.

Since the indoor air sampler response rates are produced at the household level, they are calculated as the product of the individual response rates. Contrary to the person level response rates there is no adjustment for households with one or two persons selected. The following equations are used to calculate the household level response rates:

Response rate at the household level for the questionnaire

$$PQ = \frac{Q}{R} = \frac{\textit{\# of households where at least one person completed the questionnaire}}{\textit{\# respondent households}}$$

Response rate at the household level for the MEC

$$PC = \frac{C}{Q} = \frac{\textit{\# of households where at least one person participated at the MEC}}{\textit{\# of households where at least one person completed the questionnaire}}$$

Response rate at the household level for the indoor air sampler

$$IAS = \frac{Val}{Off} = \frac{\text{\# of households that returned the indoor air sampler with four to ten days of valid entries}}{\text{\# of households with at least one participant at the MEC that were offered an indoor air sampler}}$$

Indoor air sampler combined response rate

IASCRR = HR * PQ * PC * IAS

where HR, is described in section 10.1.1.

10.1.4 Blood draw and urine response rates

Although the file with the laboratory data includes all MEC participants, some laboratory analysis could not be performed because the respondents did not or could not provide blood or urine. Of the 6,395 participants who reported to the CHMS MEC for physical measurements, 6,128 participants provided blood and 6,327 provided urine. The combined response rate for blood draw was 53.6% whereas the combined response rate for urine was 55.0%. It is important to note that the combined response rate is not obtained by multiplying the response rates at the person and household levels, since two persons were selected in some households. Appendix 8D shows the combined response rates and the relevant information for calculating them for the given age groups by gender.

Below is a description of how the different components of the equation must be manipulated to calculate combined response rates correctly. Four additional response rates are required to derive the blood draw and the urine combined response rates:

Response rate at the person level among households where <u>one person</u> was selected (blood draw)

$$BC1 = \frac{B1}{C1} = \frac{\text{\# of persons who provided blood among households where one person was selected}}{\text{\# of participants at the MEC among households where one person was selected}}$$

Response rate at the person level among households where <u>two persons</u> were selected (blood draw)

$$BC2 = \frac{B2}{C2} = \frac{\text{\# of persons who provided blood among households where two persons were selected}}{\text{\# of participants at the MEC among households where two persons were selected}}$$

Response rate at the person level among households where <u>one person</u> was selected (urine)

$$UC1 = \frac{U1}{C1} = \frac{\textit{\# of persons who provided urine among households where one person was selected}}{\textit{\# of participants at the MEC among households where one person was selected}}$$

Response rate at the person level among households where <u>two persons</u> were selected (urine)

$$UC2 = \frac{U2}{C2} = \frac{\textit{\# of persons who provided urine among households where two persons were selected}}{\textit{\# of participants at the MEC among households where two persons were selected}}$$

Blood draw combined response rate

$$BCOMBRR = HR * [(RR1 * PQ1 * PC1 * BC1) + (RR2 * PQ2 * PC2 * BC2)]$$

Urine combined response rate

$$UCOMBRR = HR * [(RR1 * PQ1 * PC1 * UC1) + (RR2 * PQ2 * PC2 * UC2)]$$

where HR, RR1, PQ1, PC1, RR2, PQ2 and PC2 are described in section 10.1.1.

10.1.5 Response rates for blood and urine subsamples

As described in section 6.3.5, subsamples of the survey respondents were selected for three subsamples: fasting, environmental contaminants in blood (i.e., perfluorinated compounds) and environmental contaminants in urine. A combined response rate for each of the subsamples can be obtained by modifying the response rate calculations shown in section 10.1.1. It is important to note that the combined response rate is not obtained by multiplying the response rates at the person and household level, since two persons were selected in some households. Below is a description of how the different components of the equation must be used to calculate the response rate correctly.

Response rate at the person level among households where <u>one person</u> was selected for the subsample (questionnaire)

$$SSQ1 = \frac{SQ1}{SPS1} = \frac{\textit{\# of respondents to the questionnaire among households where one person was selected for the subsample}}{\textit{\# of persons selected for the subsample among households where one person was selected for the subsample}}$$

Response rate at the person level among households where <u>two persons</u> were selected for the subsample (questionnaire)

$$SSQ2 = \frac{SQ2}{SPS2} = \frac{\text{\# of respondents to the questionnaire among households where two persons were selected for the subsample}}{\text{\# of persons selected for the subsample among households where two persons were selected for the subsample}}$$

Response rate at the person level among households where <u>one person</u> was selected for the subsample (MEC)

$$SSC1 = \frac{SC1}{SQ1} = \frac{\text{\# of participants at the MEC among households where one person was selected for the subsample}}{\text{\# of respondents to the questionnaire among households where one person was selected for the subsample}}$$

Response rate at the person level among households where <u>two persons</u> were selected for the subsample (MEC)

$$SSC2 = \frac{SC2}{SQ2} = \frac{\text{\# of participants at the MEC among households where two persons were selected for the subsample}}{\text{\# of respondents to the questionnaire among households where two persons were selected for the subsample}}$$

Response rate at the person scale among households where <u>one person</u> was selected for the subsample and had a valid measure (i.e. provided blood or urine and/or fasted)

$$SSB1 = \frac{SB1}{SC1} = \frac{\text{\# of respondents with a valid measure for the subsample among households where one person was selected for the subsample}}{\text{\# of participants at the MEC among households where one person was selected for the subsample}}$$

Response rate at the person scale among households where <u>two persons</u> were selected for the subsample and had a valid measure (i.e. provided blood or urine and/or fasted)

$$SSB2 = \frac{SB2}{SC2} = \frac{\text{\# of respondents with a valid measure for the subsample among households where two persons were selected for the subsample}}{\text{\# of participants at the clinic among households where two persons were selected for the subsample}}$$

Subsample combined response rate

$$SSCOMBRR = HR^* [(RR1^*SSQ1^*SSC1^*SSB1) + (RR2^*SSQ2^* SSC2^* SSB2)]$$

where HR, RR1 and RR2 are described in section 10.1.1.

10.1.5.1 Fasted subsample response rates

From the 6,465 respondent households, 4,380 persons were selected (one or two persons per household) to participate in the fasted subsample, of whom 3,944 responded to the questionnaire, for a response rate of 90.0%. Of these persons, 3,224 then reported to the CHMS MEC for physical measurements, for a response rate of 81.7%. Of these persons, 2,793 had a valid measure (i.e were actually fasted and provided blood). At the Canadian scale, a combined response rate of 48.4% was observed, using the formula given in section 10.1.5. It is important to note that the combined response rate is not obtained by multiplying the response rates at the person and household level, since two persons were selected in some households. Appendix 8E shows the fasted subsample combined response rates and the relevant information for calculating them for the given age groups by gender.

Fasted subsample combined response rate

$$FSCOMBRR = HR^* [(RR1^*SSQ1^*SSC1^*SSB1) + (RR2^*SSQ2^* SSC2^* SSB2)]$$

where HR, RR1 and RR2 are described in section 10.1.1.

10.2 Errors in surveys

A survey yields estimates based on the information collected from a sample of persons. Somewhat different estimates may have been obtained if a complete census had been conducted using the same questionnaire, the same interviewers, the same measurement experts, the same supervisors, the same processing methods, etc. as used for the survey. The difference between the estimates based on the sample and those resulting from a complete enumeration conducted under similar conditions is called the sampling error of the estimates.

In addition, errors that are not related to sampling may be made at almost any stage of a survey. Interviewers may have misunderstood the instructions, respondents may have made errors when completing the questionnaire, responses may have been incorrectly captured, measurements may have been made incorrectly, and errors may have crept in when the data were processed and totalled. These are all examples of non-sampling errors.

10.2.1 Non-sampling errors

Over a great number of observations, random errors will have little effect on the estimates drawn from the survey. However, errors that occur systematically will contribute to biases in the estimates from the survey. Much time and effort was devoted to reducing non-sampling errors in the survey. Quality assurance measures were applied at each stage of the data collection and processing cycle to control the quality of the data. Further details on the quality assurance procedures for each stage of the survey are provided in the Quality assurance and control Section 10.3.

The effect of non-response on survey results is a major source of non-sampling error in surveys. The scope of non-response varies from partial non-response (where the respondent does not respond to one or more questions) to total non-response. In cycle 2 of the CHMS, there is little partial non-response, since once the questionnaire began, respondents tended to complete it. There was total non-response when the person selected to participate in the survey refused to do so or could not be contacted by the interviewer. Cases of total non-response were taken into account during weighting by correcting the weights of persons who responded to the survey in order to compensate for those who did not respond. See chapter 9 for more information on how the survey weights were adjusted to account for non-response.

10.2.2 Sampling errors

Since the estimates from a sample survey inevitably include sampling errors, good statistical methods require researchers to provide users with some indication of the scope of this error. Measuring the possible scope of sampling errors is based on the standard error of the estimates drawn from the survey results. For a survey with a complex design, such as the CHMS, the standard error is calculated from the bootstrap replicates (see Section, 9.10 on the creation of the bootstrap weights and Chapter 12 on guidelines for tabulation for more information). To get a better indication of the size of the standard error, it is often more useful to express the standard error in terms of the estimate being measured. The resulting measure, called the coefficient of variation (CV), is obtained by dividing the standard error of the estimate by the estimate itself, and it is expressed as a percentage of the estimate.

For example, assume that a person estimates that 20% of Canadians aged 12 to 79 smoke regularly and this estimate has a standard error of 0.005. The CV of this estimate is then calculated as follows:

$$(0.005/0.20) \times 100\% = 2.5\%$$

Statistics Canada often uses the CV results for data analysis, and it strongly advises users producing estimates based on the data files from cycle 2 of the CHMS to do the same. Table 10.1 provides the Statistics Canada guidelines for releasing estimates based on their CV.

Table 10.1 Sampling variability guidelines

Type of Estimate	CV (in %)	Guidelines
Acceptable	$0.0 \leq CV \leq 16.6$	Estimates can be considered for general unrestricted release. Requires no special notation.
Marginal	$16.6 < CV \leq 33.3$	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 E (or in some other similar fashion).
Unacceptable	$CV > 33.3$	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 F (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."

10.3 Quality assurance and control

There are many problems that can introduce errors in a direct measures survey. These errors can significantly affect the integrity of survey results. To ensure the success of the CHMS in meeting its objectives, quality assurance (QA) and quality control (QC) measures were implemented in all processes including those described below and previously in chapter 7 (Data Collection).

QA anticipates problems and therefore consists of those activities that take place before data collection or in improving and refining data collection. QC responds to observed problems and thus consists of those activities that take place during and after data collection. The goal of QA and QC is to ensure the reliability and validity of the data and to reduce systematic bias to the lowest possible level.

10.3.1 Training of household interviewers and Mobile Examination Centre (MEC) staff

10.3.1.1 Initial training

Training of all staff emphasized the goals and objectives of the survey, survey methodology and quality control guidelines. Training also included questionnaire/application content and functionality, standardization of survey procedures, data transmission, refusal conversion techniques and administrative procedures.

Training involved both formal classroom training and mandatory reading of procedures and training manuals. The core of the position specific training involved hands-on practice with instructors. Experts from various fields related to the survey measures (e.g., blood pressure) and biospecimen collection/processing protocols provided seminar sessions to the appropriate staff and participated in aspects of the hands-on training.

Household interviewers:
A week of training was provided to household interviewers. Household interviewers took part in mock interviews to familiarize themselves with the household questionnaire, to simulate difficult situations and to practice potential non-response situations. They also discussed techniques for dealing with sensitive questions.

Retraining was conducted with household interviewers anytime a change was made with the application or when clarification was needed on the household questionnaire. Formal and informal training was on-going in order to ensure that proper protocols were followed.

MEC staff:
A month of training was provided to MEC staff prior to collection. The administrative staff received specific training on refusal conversion techniques and telephone skills.

The health measures specialists were provided additional training on calibration and maintenance of equipment, health and occupational safety guidelines (including both respondent and staff safety), emergency procedures and media awareness. They also received specific training on *Canadian Physical Activity, Fitness and Lifestyle Approach* (CPAFLA)[6] protocols, blood pressure and heart rate measurement, spirometry and on how to accommodate respondents with disabilities.

The laboratory technologists received supplementary training on blood and urine collection, processing, storage and running of laboratory tests, as well as, re-enforcement training on laboratory protocols.

The site logistics officer received specific training on how to set-up the trailers for the beginning of collection at each site and how to prepare the trailers for their move to the next collection site once collection was finished. As well, the site logistics officers received training on information technology maintenance and troubleshooting.

10.3.1.2 Dress rehearsal

Before the start of the actual survey, a one month dress rehearsal was done in Ottawa using Statistics Canada, Health Canada and Public Health Agency of Canada employees and their relatives who volunteered. The purpose of the dress rehearsal was to allow both the household interview staff and the MEC staff to practice their skills before beginning collection. The MEC staff had the opportunity to set-up, run and prepare the MEC trailers for transportation to another location in the same format as was to be done during the actual survey. This included booking volunteers into the MEC using the same schedule that the staff was going to use when operating at a site. The dress rehearsal also allowed the MEC staff the opportunity to refine the flow through the MEC and work on other processes that needed testing (e.g., shipping). The dress rehearsal allowed for verification of the accuracy of the documentation (e.g., procedures manuals), as well as, the household interview and MEC staff's understanding of the procedures, processes, and the flows. It also allowed for training/re-training issues to be identified prior to going into the field.

10.3.1.3 Ongoing training – Dry run day

Prior to the start of collection at each site, one day was set aside for community volunteers to participate in a visit to the MEC. These days were referred to as dry runs. The purpose of the dry runs was to ensure that all the equipment was functioning correctly. It also provided the MEC staff with the opportunity to practice their skills before the beginning of collection sites, as well as perform on-going training.

10.3.1.4 Annual retraining

Half way through cycle 2, the household interviewers took part in a debriefing and retraining session. During this session, items that were discussed included refusal avoidance and conversion techniques, techniques on dealing with non-response and how to handle sensitive questions.

An eight day annual retraining at head office was attended by MEC staff during the middle of cycle 2. These sessions were similar to the initial training and were performed in collaboration with experts in different fields related to the survey but focused on elements that specifically needed retraining.

10.3.2 Household component

10.3.2.1 Monitoring – Household interview

Monitoring of the work performed by the household interviewers was an integral part of the household collection processes. Observations were completed on all interviewers at the beginning of the collection cycle and, thereafter, two to three scheduled observations were done at each site as well as on an ad hoc basis. Debriefing sessions were held in four of the eighteen sites where factors affecting data quality were discussed.

In addition to monitoring the work of the interviewers, staff from head office performed interview observations to monitor the functionality of the household interviewing system, the respondents' understanding of the household survey content and the usefulness of the communications tools. Observers provided feedback on these items to the content development and communications teams at head office, and problems were addressed as required.

10.3.2.2 Household questionnaire response rates

Monitoring the household collection response rates was conducted throughout the cycle by staff at head office preparing collection progress reports. Staff monitored the reasons for non-response by age and sex, the number of contact attempts, the distribution of contact attempts by time of day, the refusal rates and refusal conversions attempts and the distribution of the fasting/non-fasting flag to ensure that the target number of respondents per age and sex group would be achieved.

10.3.2.3 Validation of questionnaire responses

At the end of each site, notes and remarks made by interviewers within a respondent's case file were reviewed and adjustments to the data were made when required. In addition, the frequency of answer categories within a question was determined for "other-specify", "don't know", and "refusal". Questions with a high rate of response for these categories were monitored according to expected rates from other Statistics Canada health surveys and investigated if the rate was higher than expected. These data were also monitored between CHMS collection sites to identify site-specific problems.

Validation of questionnaire responses was also performed on questions that are included in other Statistics Canada health survey with similar content, namely cycles 3.1 and 4.1 of the Canadian Community Health Survey.

10.3.3 Mobile examination centre (MEC) component

The following section will provide information concerning the quality assurance and quality control procedures that were put in place and are specific to collection of the physical measures and laboratory data during the visit to the MEC.

10.3.3.1 Equipment selection

The quality of the equipment used for collection was essential to ensure data accuracy and validity. In selecting the appropriate equipment, a combination of consulting, researching, testing, and evaluation was employed. This was done considering industry standards and in conjunction with partners (Health Canada, Public Health Agency of Canada), experts from other physical measures surveys (e.g., National Health and Nutrition Examination Survey (NHANES) in the United States) and CHMS advisory committees.

When determining the MEC laboratory equipment needs, many considerations were taken into account such as the size of the equipment (due to space constraints), the cost of the equipment, the accuracy and precision of the testing equipment, as well as the comparability to other international surveys. The reliability of the instrumentation, including frequency of breakdowns, repairs and maintenance, were also examined. Other items considered when selecting the equipment used for data collection include the infrastructure needs (e.g., use of water, energy consumption, waste disposal), the ease of operation and maintenance, training courses included, the availability and timeliness of service throughout the country, the laboratory biosafety guidelines and the test throughput.

10.3.3.2 Protocols and procedures

To ensure consistency between MEC staff on all measurement techniques, several procedures manuals containing detailed protocols for each measure were developed. These protocols were developed in consultation with, and reviewed by experts in each measurement field (when appropriate), ensuring the highest quality and least biased data collection. For standardization purposes, these protocols were covered thoroughly during staff training and scripted within the data capture application. Staff were required to review these protocols periodically during collection so as to keep themselves up to date. When changes to protocols were made all staff members were informed, provided with the updated protocols and re-training was provided if necessary. All changes to protocols were documented (date of update, reason for updating, process that was followed).

The Equipment Verification, Calibration and Maintenance Manual was also developed to ensure that calibration and maintenance of all testing equipment during collection was performed to meet or exceed the established standards. These standards were established by the equipment manufacturers and the experts that were consulted for the MEC laboratory standard operating procedures (SOP). This included pre-analytical functions (mixing, aliquoting), testing protocols (e.g., complete blood count (CBC)), non-testing procedures (e.g., specimen storage and shipping), quality control procedures and equipment use, calibration and maintenance.

All CHMS reference laboratories also followed standard operating procedures that were developed for every test and technique performed in these laboratories. These provided uniform assay protocols that laboratory staff used to ensure similar results and consistent performance.

10.3.3.3 Mobile examination centre (MEC) environment

All efforts were made to ensure measurements were carried out under controlled and standardized conditions and according to specified procedures. Due to the fact that certain measures and equipment were highly sensitive to changes in room temperature (e.g., spirometry, blood pressure, laboratory equipment), every effort was made to keep the MEC at a comfortable and constant room temperature (21°C ± 2°C). The environmental conditions of the MEC testing rooms (temperature, humidity and barometric pressure) were recorded at a minimum once per shift as well as anytime the temperature went outside of the ± 2°C range. In addition, careful monitoring of the conditions within the MEC laboratory was undertaken to ensure that the collection, analysis, storage and shipment of samples were performed under the appropriate conditions.

10.3.3.4 Adherence to pre-testing guidelines

At the beginning of the visit to the MEC, adherence to the pre-testing guidelines (see Appendix 4) was verified and documented within the data capture application and adherence rates were assessed at head office. The purpose of these guidelines was to ensure testing standardization by minimizing the potential that external factors would affect the results of certain tests. Pre-testing standardization was done to enhance the quality of the data collected.

10.3.3.5 Equipment monitoring

Regular verification, calibration and maintenance of all the equipment used for data collection during the CHMS was performed to ensure data accuracy and validity. This testing was performed to meet or exceed the standards established by the equipment manufacturers and experts in the field.

10.3.3.6 Data entry verification

All paper forms such as the respondent verification sheet, the consent form, the Physical Activity Readiness Questionnaire (PAR-Q), as well as all the worksheets from the home visits or MEC visits done on paper, were manually entered and subsequently verified by a manager to ensure data entry accuracy.

10.3.3.7 Spirometry data review

All spirometry tests were reviewed by an external reviewer via a custom application that was developed for the CHMS. The application was designed to identify unacceptable efforts based on American Thoracic Society testing[19] criteria. The reviewer's role was to validate the effort acceptability information set by the application, and assign reliability and quality information to assist users with data analysis.

All tests were reviewed and assigned a test reliability factor (SPM_RELI; reliable or not reliable) to identify whether the test is recommended for inclusion in analysis. All acceptable trials were reviewed to determine whether the test results for FVC and FEV1 represented a maximal effort by the respondent

(quality) and met the reproducibility criteria of 150ml. The FVC quality factor (SPM_QFVC) and FEV1 quality factor (SPM_QV1) indicate whether a reproducible FVC and/or FEV1 were obtained.

Table 10.2 Quality factors for FVC and FEV1 spirometry results

Quality factor	Description	Comments
A	Excellent quality and reproducibility	Use in analysis
B	Good quality and reproducibility	Use in analysis. Example: A common case is for children and adolescents who do not exhale for the required length of time (3 or 6 seconds). If these respondent‘s trials show a 1 second plateau and the reviewer judges that the curves represent a maximum volume (FVC) then a code B is assigned.
C	Questionable quality and reproducibility	Use in analysis with caution. Example: A respondent has only 2 acceptable curves which are not reproducible and the other curves are unacceptable due to early termination and large extrapolated volumes. However, the curves with the large extrapolated volumes do confirm the FVC reproducibility. Example: A respondent with chronic obstructive pulmonary disease (COPD) and a low FEV1/FVC ratio (e.g., <0.45). Because of COPD, the respondent cannot provide a repeatable FVC with even 8 manoeuvres. So, the grade should be at least a C or better as the lack of an acceptable curve due to end of test failures is not a sufficient reason to exclude or grade the subject‘s results with a D or F.
D	Highly questionable reproducibility and quality	Do not use in analysis
F	Unacceptable test results	Do not use in analysis

10.3.3.8 Data validation

Data validation was performed to ensure that the CHMS data were consistent with other similar data sources including other Statistics Canada surveys, as well as international surveys. Data validation was done to compare various physical measures and the laboratory measures and to compare self-reported data from the household interview to directly measured data from the MEC visit (e.g., height, weight), by site and overall. The physical measures data were also compared against data sources that contained directly measured data, such as the Canadian Community Health Survey cycle 2.2 and the National Health and Nutrition Examination Survey (NHANES) in the United States.

All CBC data received from the MEC were monitored by comparing the mean values and standard deviations to a normal reference range in order to identify any increased outliers and trends in the QC findings and respondent results. The laboratory measures analyzed by the reference laboratories were also examined and compared to other surveys such as NHANES.

Prior to the release of cycle 2 data, external data validation was performed by making information available to deemed Statistics Canada employees at Health Canada, and the Public Health Agency of Canada; who informed Statistics Canada of any concerns or anomalies related to data quality.

10.3.3.9 Activity Monitor Data Review

A data review process was developed to clean and process the activity monitor data. The review process was broken down into four distinct steps: Step 1 involved the downloading and saving of respondent data, ensuring that no malfunctioning monitors were returned to the field, and following up with respondents who hadn't returned the monitor. Step 2 involved stacking of all respondent data into a single file, and dropping non-valid/bad data (such as initialization errors, or spurious data); Step 3 involved accepting only those with at least 1 day of viable wear-time (at least 5 hours of data for 3 to 5 year olds and at least 10 hours of data for 6 to 79 year olds), and calculating the activity intensity per minute; and Step 4 involved applying reserve codes to missing data and generating the derived variables.[1] See Section 12.2.2 and Appendix 9 for more information on the quality of the activity monitor data, and on research being done.

10.3.3.10 Indoor Air Sampler (IAS)

Duplicates were done once a week to evaluate the precision of the analytical method of the reference laboratory. Each time, a pair of samplers was deployed at the MEC for seven days. A total of 82 pairs of duplicates were done during cycle 2. The duplicates were blind QC since the testing laboratory did not know that they were QC samples. See the last paragraph of Section 10.3.3.16 for information on the blanks used for the indoor air samplers.

10.3.3.11 Replicate Testing

Replicate testing was also done regularly at the MEC on the anthropometry component. Replicate samples were collected during dry-run days at the beginning of each site for a variety of laboratory tests, including the complete blood count, and tests analyzed at the Health Canada reference laboratories. Approximately 8 to 10 dry-run replicate samples were performed by splitting the blood and urine samples from the participants into two distinct specimen tubes with different identification numbers (IDs). As the corresponding split sample IDs were unknown to the technician/technologist testing the samples, these "blind QC samples" were meant to monitor the precision of the assay and poor performance was inferred if the coefficient of variation obtained from the replicate samples was greater than a pre-set criteria. All replicate samples were sent to the reference laboratories along with other respondent samples and were analyzed following the same procedures as respondent samples. Data from the dry-run replicate samples were transmitted as usual along with the other respondent results. The lab section at head office analysed them and all results outside of the acceptable limits were followed up on with the testing laboratory.

10.3.3.12 Mobile examination centre (MEC) laboratory

Procedures were put into place to allow for quick detection of errors related to the MEC laboratory CBC analysis. These procedures included internal and external QC monitoring, and allowed the CHMS to ensure accurate results and data quality for laboratory measures. Aside from weekly reviews, regular

comparisons were made with QC results obtained by peer users of the same haematology analyzer employed by the CHMS. In addition, external QC material from the College of American Pathologists (CAP) and Quality Management Program – Laboratory Service (QMP-LS) were tested according to their respective schedules. The results of these blind QC samples provided an evaluation of the testing accuracy.

10.3.3.13 Proficiency Testing

All CHMS reference laboratories were responsible for having their own quality control programs in place. However, the CHMS also sent reference quality control materials as a form of proficiency testing for the reference laboratories. The CHMS used BioRad control samples with known concentration for all tests included in the CHMS for which reference QC samples were available. The use of these materials allowed the CHMS to monitor the accuracy of the analytical testing being performed at the reference labs. The CHMS mobile examination centre laboratory sent these BioRad samples to the reference laboratories on a weekly basis during each collection site along with the regular shipments. The test samples were aliquotted into the same type of shipping tubes used for respondent samples and were labelled with unique sample identification numbers so as to blind the laboratories to the process. Testing results were sent back to the CHMS head office and were compared to the reference ranges provided by BioRad. The test results were assessed by head office laboratory staff and feedback was provided to the reference laboratories for review and remedial action, if necessary.

10.3.3.14 Processing and storage of the blood and urine samples

All blood specimens collected at the MEC were centrifuged in the MEC's laboratory within two hours of collection to preserve the quality and integrity of the specimens. The specimens were stored at the appropriate temperature (e.g., 2-8 ^{o}C fridge or -20 ^{o}C or colder freezer) and the fridge and freezer temperatures were monitored via readings during each staff's shift and by an alarm system at all times. All samples were processed and stored within four hours of collection. Urine samples were refrigerated immediately upon collection and were subsequently processed and stored at the appropriate temperature as soon as possible.

10.3.3.15 Shipping

As described in the Laboratory Measures Protocol (see Section 7.6), shipment temperature monitoring ensured that only samples whose integrity remained intact were used for analysis.

10.3.3.16 Blanks

In order to ensure that urine and blood samples were not being contaminated by the MEC laboratory environment and processes, at the beginning of every site the MEC lab sent field blanks, in triplicate, to the laboratory testing the respondent samples. A blank solution that was tested to be free of any environmental contaminants was used to mimic the same processes carried out with respondent samples for collection and processing of environmental measures on blood and urine. Deionized water was used as the blank solution for the baseline testing of urine and blood environmental measures with the exception of mercury for which a nitric acid and gold solution was used.

The field blank results were compared to an acceptable upper limit that was set at three times the limit of detection of each analyte or to an acceptable upper limit that was not more than 10% of the 95th percentile of the population. Results above these limits were reviewed by head office staff and deviations were investigated, in consultation with the analytical lab and other experts and, if necessary, corrective action was taken.

For the indoor air samplers, two types of blanks (cleaning and field) were done at the MEC at each site. The cleaning blanks were used to evaluate the amount of contamination from the cleaning process at the testing lab (canisters were cleaned for re-use) and during shipments. The MEC staff took unopened canisters that were received and sent them to the testing lab (without any manipulation to the canister) once a week. The field blanks were used to evaluate the amount of contamination while the canisters were being prepared for the respondents at the MEC, from the cleaning process and during shipments. Field blanks were prepared by the MEC staff (staff changed the caps on the canisters as is done for all respondent samplers) and sent the canisters to the testing lab twice a week. All blanks were sent as blind samples, with no indication given to the testing lab that they were quality controls rather than samplers from respondent households. A total of 173 field blanks and 75 cleaning blanks were completed for indoor air sampling during cycle 2. Results from the field and cleaning blanks are available upon request.

10.3.4 Head office

10.3.4.1 Correcting for bias

The CHMS experienced several levels of non-response. First, the selected dwelling may refuse to provide the household composition. Second, the person(s) selected amongst the household members may refuse to answer the questionnaire. Third, the person may refuse to participate in the MEC component. Finally, the person may refuse to provide blood and/or urine for the laboratory tests for current or future analysis.

At each level of non-response characteristics available for respondents and non-respondents are used in a logistic regression model to identify variables which explain most non-response. The variables which were highly correlated with response or non-response and which were used in the logistic regression models included collection site, age group, sex, household size, education, income, travel distance to the MEC, and the number of days and weeks to contact the dwelling and complete the household questionnaire. Based on the results of each regression, homogeneous response groups are created. The non-response adjustments are applied within these groups to adjust the survey weights (see chapter 9, Weighting). Using the survey weights to create estimates will minimize non-response bias due to differences in the survey variables between respondents and non-respondents.

Several studies were done on the cycle 1 data to test for bias in the respondents to the MEC, because it was believed that less healthy people were less likely to go to the MEC, and for bias caused by the oversampling of respondents aged 20 to 39 who were living with children aged 6 to 11.[20] The results showed that MEC respondents, with the adjustment for MEC non-response, are similar to the household questionnaire respondents. Although the cycle 1 data did not show a significant bias due to the 2-person selection strategy which favoured the selection of a 20 to 39 year old living with a child aged 6 to 11, it was anticipated that a greater bias would be observed in cycle 2 with the addition of the 3 to 5 age group. To compensate for any potential bias, the survey weights and bootstrap weights for

cycle 2 were created using a post-stratification by age group and sex, with an additional adjustment for 20 to 39 year olds living with and without children aged 3 to 11.

Information on imputation of household income and the treatment of other missing data can be found in Sections 12.1.2 and 12.1.3.

10.3.4.2 Indoor air data

Blank corrections were applied when more than 50% of the field blanks for the indoor air samplers had values that were greater than the limit of detection (LOD).[29] Sample data were then adjusted by subtracting the median value of the field blanks. Field blanks with values below the LOD were replaced with a value of ½ the LOD for calculation of the median.[29] Blank corrections were applied to the measured amount, and the corresponding concentration was recalculated. When the blank correction resulted in a measured amount changing to a value less than the LOD the measured amount was coded as <LOD with a reserve code ending in 5. Also see Section 8.6 on analytical range and Section 10.3.3.16 on blanks.

10.3.4.3 Insulin data

After a few sites of data collection for cycle 2, the insulin measures were put on hold after a problem was discovered with the reagents of the Immulite instrument. This was the instrument that the CHMS had used in cycle 1 and had started to use for cycle 2 to measure insulin levels in serum samples. Despite the problem with the reagents, insulin samples were still collected at every site and shipped frozen to the analytical lab for analysis of vitamin B12. They were then frozen again for future analysis once a new instrument and new reagents became available. It was decided that the survey would use the newly purchased cycle 3 Centaur instrument instead for the analysis of the cycle 2 insulin samples (see Section 12.2.3.5 for information on data comparability between cycles 1 and 2). The samples from early cycle 2 collection sites were frozen for a period of over two years, well beyond the generally accepted period of two to six months for which serum insulin is expected to remain stable after being frozen.[30] Consequently, the CHMS undertook a study to investigate whether the long term freezing and storage of the cycle 2 samples impacted the quality of the insulin data. The study (available upon request), concluded that the cycle 2 insulin data seem to be of good quality and were not affected by the freezing period.

11. File Usage

11.1 Description of data files

The Canadian Health Measures Survey (CHMS) dissemination files only contain data from respondents who attended the mobile examination centre (MEC) and agreed to share their responses with Health Canada (HC) and the Public Health Agency of Canada (PHAC). For cycle 2 of the CHMS, after the completion of processing, and with the removal of non-share records[1], 6,395 records remain on the main master file.

There will be 12 data files released for CHMS cycle 2, with the releases occurring in three waves. A master file will be released with each wave, with data accumulating as time goes on. Only the most recent master file, with the most complete and up-to-date data should be used. In other words, on November 21, 2012 (wave 2 release date), the September 20, 2012 (wave 1) master file should no longer be used and on April 17, 2013 (wave 3 release date), the November 21, 2012 (wave 2) master file should no longer be used. The April 17, 2013 (wave 3) master file is expected to be the final master file for cycle 2 CHMS, and should be used from that date on.

See Section 12.2.3.6 for information on the combined cycle 1,2 file.

Wave	Files Released	Data File Name	Corresponding Bootstrap Weights File	Date of Release
1	• Master file Contains household and physical measures data, except for activity monitor and indoor air data.	HC.txt	wgt_full.txt	Sept 20, 2012
	• Medication file Contains data on medication use collected at both the household and the MEC.	MED.txt	wgt_full.txt	
	• Postal code file Contains postal code information for each respondent in the survey (special restrictions are in place for the use of this file).	PC.txt	*	
	• Climate and air quality file Ancillary data on climate and air quality for each MEC site across Canada (data comes from national and provincial weather organizations such as Environment Canada).	ENV.txt	*	
2	• Master file Contains activity monitor and non-environmental laboratory data that have been added to the contents	HCL1.txt	wgt_full.txt	Nov 21, 2012

	of the wave 1 master file. • Activity monitor subsample file Contains activity monitor data for people who wore their activity monitor for at least 10 hours per day (5 hours for 3 to 5 year olds) for 4 or more days. • Indoor air subsample file – household level Contains indoor air data for each household that deployed their indoor air sampler in their home for between 4 and 10 days. • Indoor air subsample file – person level Contains indoor air data for each person that deployed their indoor air sampler in their home for between 4 and 10 days. • Fasted subsample file Contains data on insulin, low density lipoproteins and triglycerides for those respondents who fasted.	AM.txt IASH.txt IASP.txt FASTED.txt	wgt_acmo.txt wgt_iash.txt wgt_iasp.txt wgt_fast.txt	
3	• Final master file Contains environmental laboratory data that have been added to the contents of the wave 2 master file. • Environmental blood subsample file Contains data on nine different perfluorinated compounds measured in respondents' blood. • Environmental urine subsample file Contains data on 63 different compounds measured in respondents' urine.	HCL2.txt EB.txt EU.txt	wgt_full.txt wgt_eb.txt wgt_eu.txt	Apr 17, 2013

* The postal code and climate and air quality files are always used in conjunction with another file and its corresponding bootstrap weights file. In many cases, it will be the master file's bootstrap weights file (wgt_full.txt) that is used with these files.

Most of the data files listed in the table above contain a weight variable. The bootstrap weight files contain more detailed information (i.e. 500 bootstrap weight variables for variance calculation purposes). More information on bootstrap weights and weighting can be found in Chapter 9. Note that SAS versions of the text files are also available.

The survey includes respondents 3 to 79 years of age. However, some of the measures or tests were only done for one sex, on fasting respondents or on a random subgroup of ages. Refer to CHMS Content summary for cycles 1, 2 and 3 (available upon request) for more information.
For information regarding the actual questions asked of the respondents, users can consult the household and MEC questionnaires at the following link:
http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getInstrumentLink&SurvItem_Id=129548&Query_Id=62444&Query=instance&lang=en&db=imdb&adm=8&dis=2

11.1.1 Master file (wave 1)

This data file contains the responses collected during the household interview and the MEC visit, except for activity monitor and indoor air sampler data, which will be added in wave 2. It also contains some derived variables (DVs) created by the collection applications and other DVs created after data processing. Two separate DV documents are available that describe how these DVs are created/calculated: household/MEC DV document and activity monitor DV document.

Each of the 6,395 records represents one respondent who went to the MEC and participated in at least some of the measures and then agreed to share their household and MEC results with HC and PHAC.

The data contained in this file covers the following topics:

- Alcohol use
- Anthropometry
- Cardiovascular health and fitness
- Chronic conditions
- Environmental exposure
- Family medical history
- General health
- Infectious markers
- Lung health
- Musculoskeletal fitness
- Nutrition
- Physical activity
- Pregnancy/Birth
- Sexual health
- Sleep
- Smoking
- Socio-demographic characteristics
- Sun exposure

11.1.2 Medication file

In order to make the medication data collected during the CHMS household interview and MEC visit easier to work with, a separate database containing all medication responses was created. This database contains information for all 6,395 respondents and therefore uses the same weights as the master files.

During the household interview, respondents were asked to report any prescription medication, over-the-counter medication and health products or herbal remedies that they took during the previous month. The interviewer was able to capture information for up to 15 prescription medications, 15 over-the-counter medications, and 15 health products or herbal remedies, for a total of 45 possible entries.

During the MEC visit, it was confirmed if each medication, over-the-counter product and health product or herbal remedy reported at the household was still being taken. Respondents then had an opportunity to add up to five new prescription medications, five new over-the-counter products, and five new health products or herbal remedies that they started to take after the household interview had been completed.

Using this information, a series of derived variables was created to indicate where the products were reported, how many total products were reported, and when the products were last taken by product type (prescription medications, over-the-counter products or health products or herbal remedies). A separate document describing how these DVs were created/calculated is available. In addition, both Anatomical Therapeutic Chemical (ATC) and American Hospital Formulary Service (AHFS) codes were assigned for each reported medication and a second AHFS code was added when applicable. See Appendix 7 for more information on the ATC and AHFS classification systems.

11.1.2.1 Naming convention for medication variables

As per the general rules for naming CHMS variables, the medication database variable names are limited to 8 characters and each character of the variable name contains information about the data contained in the variable.

Positions 1-3: Identifies the abbreviated name of the variable
- "MED" or "MHR": collected or derived variables from the household and/or MEC interviews
- "ATC": Anatomical Therapeutic Chemical (ATC) classification variables
- "AHF": American Hospital Formulary Service (AHFS) classification variables

Position 4: Identifies whether the variable was collected during the household or MEC interview ("_") or was later derived ("D").

Position 5: Identifies the medication category:
- "1": prescription medication
- "2": over-the-counter medication
- "3": health product or herbal remedy

Positions 6-7: Identifies the list number of the medication:
- "01" to "15": Medications collected during the household interview which were then confirmed during the MEC visit
- "31" to 35": Additional medications collected during the MEC visit

Position 8: Identifies one of two parts of a related concept:
- "A": e.g., AHF_101A is the first AHFS code derived from the drug identification number (DIN)
- "B": e.g., AHF_101B is the second AHFS code derived from the drug identification number (DIN)

11.1.2.2 Description of key medication variables

The key variables are described below. An "X" is used in position 5 of the variable name as the same logic applies for each medication category (prescription medication, over-the-counter medication, and health products or herbal remedy).

The variable MEDDX00A indicates whether or not the respondent took any medication within a given category in the past month. The variable MEDDX00B is the sum of the medications within a given category taken by the respondent during the past month. Both MEDDX00A and MEDDX00B are derived from information collected during both the household interview and the MEC visit. For more information, refer to the derived variables documentation.

It should be noted that although a maximum of 20 medications were captured for each medication category[21], the value for MEDDX00B represents the actual number of medications being taken by the respondent within a given category, and therefore may exceed a value of 20. Due to the limitations of the data capture application, details are only available for the first 20 medications reported. For example, if a respondent indicated taking 17 different prescription medications in the last month at the time of the household interview then MEDD100B will equal 17, although details would only have been captured for the first 15 of the 17 prescription medications.

Six additional variables are associated with each of the medications reported within a given category at the household and five additional variables are associated with each new medication reported at the MEC.

MED_X01B to MED_X15B	This information was collected at the household and indicates the time the respondent last took each medication from the time of the household interview. Note: This variable is suggested for use during analysis if attempting to relate medication use to one or more other variables from the household interview.
MEDDX01A to MEDDX15A and MEDDX31A to MEDDX35A	This variable is derived from the information collected at the MEC during the confirmation of medication-use reported during the household interview. It provides information about whether the respondent was still taking the medication, had stopped taking the medication since the household interview, or had never taken the medication.
MHR_X01B to MHR_X15B and MHRDX31B to MHRDX35B	These variables are collected at the MEC and indicate the last time the respondent took each medication from the time of the MEC interview. Note: This variable is suggested for use during analysis if attempting to relate medication use to one or more other variables from the MEC visit.
ATC_X01A to ATC_X15A and ATC_X31A to ATC_X35A	The ATC code derived from the drug identification number (DIN) collected during the household and/or MEC interview.
AHF_X01A to	The first AHFS code derived from the DIN collected during the

AHF_X15A and AHF_X31A to AHF_X35A	household and/or MEC interview. Note: If the DIN was not available for a given medication, the AHF code will appear as "Not Stated" in the database.
AHF_X01B to AHF_X15B and AHF_X31B to AHF_X35B	The second AHFS code derived from the DIN collected during the household and/or MEC interview (if applicable). Note: If the DIN was not available for a given medication, the AHF code will appear as "Not Stated" in the database.

11.1.3 Postal code file

This file contains the postal code (DHH_DPC) for all 6,395 persons who participated in the survey, along with their CLINICID (see Section 11.2). The file is only available in Statistics Canada Research Data Centers (RDCs). Users will need to link the postal code file to another data file, such as the master file, as well as to the corresponding bootstrap weights file in order to produce estimates for the total population (see Section 11.4). Researchers must specifically state the need for the file in their RDC application (as it is separate from the CHMS master file) and must clearly explain what analysis will be done using postal codes.

Users need to be aware of the restrictions on the use of this file. The file can only be used for deriving variables that are to be created and analyzed at the national level. It cannot be used to try to produce small area estimates as the survey is designed only for national level estimates. The file must be used at the person level with the appropriate weights. It cannot be used to compare members of the same household or for producing household estimates since household weights are not available.

11.1.4 Climate and air quality file

One of the modules included in the CHMS MEC visit is a lung function test called spirometry. As lung function can be affected by the atmosphere, adjustments based on the weather and air quality at the time of the test may be required in order to analyze the data. To address these needs, a file listing the hourly climate and air quality data for the collection period at each of the 18 MEC collection sites has been created. The data have been obtained from Environment Canada's National Climate Data and Information Archive (http://www.climate.weatheroffice.gc.ca) and from the National Air Pollution Surveillance Network (NAPS) (http://www.ec.gc.ca/rnspa-naps/), a federal/provincial collaborative network through which provincial NAPS partners collect their data and send them to Environment Canada.

The following indicators are on the climate and air quality file:

Air quality: Ozone, nitrogen dioxide, particulate matter (2.5 microns)

Climate: Temperature, barometric pressure, precipitation, relative humidity, humidex, wind speed, wind direction and wind chill

This file may be linked to the master file by using the variables SITE, V2_YEAR, V2_MONTH, V2_DAY, and V2_HOUR, which are present on both files. Upon merging the two files, the climate

and air quality information measured on the hour the respondent visited the MEC will be appended to the master file, meaning that there will be climate and air quality information available for all 6,395 respondents. Additional supporting documentation is available upon request.

11.1.5 Master file (wave 2)

In addition to the data in the wave 1 master file, this file contains data from the activity monitor respondents wore and from the non-environmental laboratory blood and urine tests (6,395 records). There are over 52 added variables (not including DVs), related to topics such as cardiovascular health, chemistry panel, diabetes, general health, infection markers, kidney health and nutritional status.

11.1.6 Activity monitor subsample file

This file contains data for the 4,948 respondents who wore their activity monitor for at least 10 hours per day (5 hours for 3 to 5 year olds) for 4 or more days. The file also contains a weight specific to this sub-population. The vast majority of variables on this file are also located on the master file. The variables are denoted by the prefix "AMM" on the master file and "AMS" on the subsample file. See Section 12.2.3.2 and Appendix 9 for more information on using this file.

11.1.7 Indoor air subsample file – household level

This file contains data for the 3,857 households that deployed the indoor air sampler in their home for between 4 and 10 days. The file also contains a weight specific to this sub-population. The variable names in this file are identical to the variable names found in the indoor air person level file.

11.1.8 Indoor air subsample file– person level

This file contains data for the 5,191 persons that deployed their indoor air sampler in their home for between 4 and 10 days. The file also contains a weight specific to this sub-population. The variable names in this file are identical to the variable names found in the indoor air household level file.

11.1.9 Fasted subsample file

This file contains data for a subset of the CHMS respondents, the 2,793 respondents who had fasted for a minimum of 10 hours. The file contains data for three variables: insulin, low density lipoproteins (LDL) and triglycerides. The file also contains a weight specific to this sub-population.

11.1.10 Final master file (wave 3)

This file contains all of the data from the wave 2 master file, as well as the laboratory variables related to environmental contaminants and chemical substances for which all respondents aged 3 to 79 were tested (6,395 records). These variables contain laboratory data on cotinine and 29 of the 35 metals that were part of CHMS cycle 2.

11.1.11 Environmental blood subsample file

The environmental blood subsample file contains data on the following nine different perfluorinated compounds:

perfluorobutane sulfonate (PFBS)
perfluorodecanoic acid (PFDA)
perfluorohexane sulfonate (PFHxS)
perfluorohexanoic acid (PFHxA)
perfluoro-n-butyric acid (PFBA)
perfluorononanoic acid (PFNA)
perfluorooctane sulfonate (PFOS)
perfluorooctanoic acid (PFOA)
perfluoroundecanoic acid (PFUDA)

The tests were done on a random subsample of the population aged 12 to 79 (target of 1,500 respondents) and the file contains a weight specific to this sub-population.

11.1.12 Environmental urine subsample file

The environmental urine subsample file contains data on the following 63 different compounds:

bisphenol A
triclocarbon
triclosan
benzene metabolites (3)
carbamate insecticides (2)
chlorophenols (5)
metals and trace elements (5 speciated arsenic and fluoride)
organophosphate insecticides (6)
phenoxy herbicide 2,4-D
phthalate metabolites (12)
polyaromatic hydrocarbons (17)
pyrethroids (metabolites) (5)
triazine herbicides (3)

The tests were done on a random subsample of the population aged 3 to 79 (target of 2,500 respondents) and the file contains a weight specific to this sub-population.

11.2 Key variables for linking data files

As a result of the large number of files disseminated for cycle 2, it may be necessary to be able to link these files to have access to a larger pool of information. In particular, users may want to link the subsample files and medication file to the master file in order to obtain contextual questionnaire and

physical measures information for their analysis. Users will also need to link the appropriate bootstrap weights file to each data file in order to produce estimates for the total population (see Section 12.1).

In order to facilitate the linking of two or more files, a variable that uniquely identifies each respondent on each file is required. For the CHMS, the variable to be used is called CLINICID. With this variable, data users are able to join the data from any subsample file to the master file for a particular respondent. The only data file that does not use CLINICID as the linking variable is the climate and air quality file (see Section 11.1.4).

11.3 Key variables and definitions

Other variables which may be particularly useful for data users are listed below:

CLC_AGE	Respondent's age at the time of the mobile examination centre visit
CLC_SEX	Gender of Respondent
DHH_MS	Marital Status of Respondent
EDUDR04	Highest level of education - respondent, 4 levels
EDUDR10	Highest level of education - respondent, 10 levels
PROXY	Proxy or non-proxy interview
WGT_FULL	Weight for the full sample file and the medication file
WGT_ACMO	Weight for the Activity Monitor subsample file
WGT_IASH	Weight for the indoor air household level subsample file
WGT_IASP	Weight for the indoor air person level subsample file
WGT_FAST	Weight for the fasted subsample file
WGT_EB	Weight for the environmental blood (perfluorinated compounds) subsample file
WGT_EU	Weight for the environmental urine subsample file

The naming convention used for the weighting variables listed above is the same as the naming convention for the corresponding bootstrap weight files, with the only difference being the suffix in the file name. For example, ".txt" can be removed from the bootstrap file "WGT_FULL" to obtain the name of the weighting variable on the master file.

11.4 Use of age and sex variables

Age and sex are collected twice during the survey process for the CHMS, first through the household interview, then again during the MEC component. Because the two appointments could be several days apart, it is possible that a respondent's age could be different on the household and MEC files. Each application uses the age and sex information that was collected as the reference for coverage and skip patterns for sections within that application. For that reason, it is important to use the appropriate age and sex variables[1] for the data being analyzed. Incorrect usage of age and sex variables could lead to errors in analyses, due to respondents whose ages changed between the two appointments. For example, a respondent who is 11 at the time of the household interview and 12 at the time of the MEC visit will be in different age groups depending on the data being analyzed (i.e. 6 to 11 for household analysis and 12 to 19 for MEC analysis).

For household variables, DHH_AGE and DHH_SEX should be used; for MEC variables, CLC_AGE and CLC_SEX should be used. If variables from both components are analyzed, use the age and sex variables for the module containing the most important variable of interest.

11.5 Use of weight variables

The CHMS is a sample survey, which means that the respondents "represent" many other Canadians not included in the survey. For example, a 1% sample would mean that each CHMS respondent represented 100 Canadians. In order that the results of the survey are representative of the population, survey weights have been created. These survey weights, when applied to the survey results, enable data users to create estimates for the entire population.

Each respondent record on the master file (and many of the other files) has a unique survey weight attached to it. In order to produce estimates for a particular characteristic, the data user must sum the weights for each respondent with that characteristic. The total created by that calculation would produce an estimate of that characteristic in the total population. There are various software packages available that will use survey weights to produce estimates (see Section 12.2.4 for more information).

Because of the small sample size for cycle 2 of the CHMS the results should only be used to produce national estimates. Due to the different number of respondents contained within the various files output for cycle 2 each file produced will contain a different weight variable. If subsample files are linked to the master file then the weight stored on the subsample file should be used for creating the estimates.

11.6 Variable naming convention

The CHMS naming convention for cycle 2 of the survey follows the same pattern as that used by CHMS cycle 1 and many other Statistics Canada surveys. The variable name is constructed in a way that allows the data users to easily identify the originating section of the survey, the type of variable and the survey question that collected the data. The variable names have also been created in such a way as to identify similar data between different cycles of the CHMS.

Each variable name must adhere to a mandatory requirement which restricts variable names to a maximum of 8 characters for ease of use by analytical software products. As a result, each character of the variable name contains information about the data contained in the variable.

Generally speaking:

Positions 1 - 3: Section name
Position 4: "_" or Variable type
Positions 5 - 8: Question reference (question number or acronym/short form to represent the concept of the variable)

For example: The variable from question 101 of the household Chronic Conditions section, CCC_101:

Positions 1 - 3: CCC chronic conditions section

Position 4: _ underscore (_= collected data)
Positions 5 - 7: 101 question number

AND

SPM_QFVC:

Positions 1 - 3: SPM spirometry section of the mobile examination centre questionnaire
Position 4: _ underscore (_= collected data)
Positions 5 - 8: QFVC quality code for the FVC results in spirometry

11.6.1 Position 4: Place Holder or Variable Type

_ Collected variable: A variable that appeared directly on the questionnaire

C Coded variable: A variable coded from one or more collected variables (e.g., SIC, NAICS, ATC, Tar)

D Derived variable (DV): A variable calculated from one or more collected or coded variables

F Flag variable: A variable calculated from one or more collected variables (like a DV) but usually calculated by the data collection computer application for later use during the interview or in the case of the income imputation flag (INCFIMP), derived in head office after imputation

G Grouped variable: Collected, coded, or derived variables collapsed into groups (e.g., age groups)

L Limit of detection (LOD) variable for indoor air laboratory data

11.6.2 Positions 5-8: Question reference

In general, the fifth to seventh positions follow the variable numbering used on the questionnaire. The letter "Q" used to represent the word "question" is removed, and all question numbers are presented in a two- digit format. For example, question Q01A in the questionnaire becomes simply 01A, and question Q15 becomes simply 15. In other cases, an acronym or short form is used to represent the concept of the variable and an eighth character is sometimes required.

For questions which have more than one response option, the final position in the variable naming sequence is represented by a letter. For this type of question, new variables were created to differentiate between a “yes” or “no” answer for each response option. For example, if Q2 had multiple response options, the new questions would be named Q2A for option 1, Q2B for option 2, Q2C for option 3, etc. If only options 2 and 3 were selected, then Q2A = No (“2” on file), Q2B = Yes (“1” on file), Q2C = Yes (“1” on file) and Q2D = No (“2 on file).

To help reconcile the variable names found on the master from those on the household and MEC questionnaires, users can consult the questionnaires and the data dictionaries, and other documentation available upon request.

11.7 Access to data files

Access to CHMS data is provided through the RDCs. These research data centres require researchers to submit a research project that will use respondent data from the released files to Statistics Canada. These projects are accepted based on a set of specific rules. When the project is accepted, the researcher is designated as a "deemed employee" of Statistics Canada for the duration of the research, and given access to the data from designated Statistics Canada sites. For more information, please consult the Statistics Canada webpage: http://www.statcan.gc.ca/rdc-cdr/index-eng.htm.

Another means of access to the data files is to provide Statistics Canada specifications for tabulations. On a fee for service basis, these tables are programmed and run against the data files by employees of Statistics Canada. This service allows users who do not possess knowledge of tabulation software products or who do not have access to the RDCs, to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, contact Health Statistics Division Client Services at 613-951-1746 or by e-mail at hd-ds@statcan.gc.ca.

Finally, HC and PHAC government employees have access to CHMS data files through a share agreement. At the MEC, respondents provide their consent to “share the information collected during the survey with its partners” who “will keep the information confidential, and use it for statistical purposes only”. Within HC/PHAC, employees who require CHMS information in order to do their work apply for access and are granted approval upon filling out and signing an acknowledgment document and providing a purpose of research. To arrange access, HC employees should contact the DAIS team at DAIS@hc-sc.gc.ca while PHAC employees should contact the Data Coordination and Access Program at DCAP-PCAD@phac-aspc.gc.ca.

12. Guidelines for tabulation, analysis and release

This chapter provides guidelines to be applied by data users in tabulating, analyzing, publishing or otherwise releasing any data derived from the survey files. In addition, two data user presentations[22] are available upon request.

12.1 Guidelines for tabulation

The sample design used for this survey is not self-weighted. In other words, the sampling weight is not the same for all the persons included in the sample. Even to produce simple estimates, including ordinary statistical tables, the user must employ the appropriate sampling weight. Otherwise, the estimates calculated on the basis of the master file cannot be considered representative of the population observed, and they will not correspond to those of Statistics Canada. For further information on the creation of the survey weights (the sampling weights that have been adjusted for non-response), refer to chapter 9. Information on the use of the weight variable is found in Section 11.5 and information on the key variables used for linking data files is found in Section 11.3.

Users should also keep in mind that because of the treatment reserved for weights, some software packages do not yield estimates that exactly match those of Statistics Canada.

12.1.1 Tabulation of categorical and quantitative estimates

There are two main types of point estimates of population characteristics that can be generated from the data files: categorical estimates and quantitative estimates. A brief explanation of each estimate type is given prior to describing how the survey data can be tabulated.

12.1.1.1 Categorical estimates

A categorical estimate is the estimate of the number or percentage of the surveyed population that possess a certain characteristic or fall into some defined category. For example, the proportion of respondents with diabetes and the number of persons with each type of diabetes are both categorical estimates.

Examples of categorical questions:

Do you have diabetes? (CCC_Q51)
1 Yes
2 No

If Yes, were you diagnosed with: (CCC_Q52)
1 … insulin dependent diabetes (Type 1)?
2 … non-insulin dependent diabetes (Type 2)?
3 … gestational diabetes ?

Estimates of the number of people with a certain characteristic can be obtained from the data file by summing the final weights of all records[1] possessing the characteristic of interest. A proportion is calculated as $\hat{X}/\hat{Y}$ using the following steps:

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 ($\hat{Y}$)
c) Dividing the numerator estimate by the denominator estimate

For example, to obtain the proportion of respondents 20-79 with Type 2 diabetes, the numerator is obtained by summing the weights (WGT_FULL) for all respondents aged 20 to 79 who answered non-insulin dependent diabetes (Type 2), answer 2 to CCC_Q52. The denominator is the sum of the weights for all 20-79 year old respondents to the survey.

12.1.1.2 Quantitative estimates

A quantitative, or continuous, estimate is an estimate of the total or of the mean, median or other measure of central tendency of quantities based on some or all members of the surveyed population. For example, the average age of first diagnosis with type 2 diabetes is a quantitative estimate.

Example of a quantitative question:

If CCC_Q52 = 1 or 2 then:
How old were you when this was first diagnosed? (CCC_Q53)
|_|_|_| Age in years

For CHMS data, two different estimates of the mean are commonly generated for continuous variables: the arithmetic mean and the geometric mean. The arithmetic mean is the simple average of the data and is best suited for variables that are evenly distributed around the mean. The arithmetic mean is calculated by:

a) Multiplying the value of the variable of interest by the final weight and summing this quantity over all records of interest to obtain the numerator ($\hat{X}$)
b) Summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$)
c) Dividing the numerator estimate by the denominator estimate.

For example, to obtain the average age when respondents were first diagnosed with Type 2 diabetes, first compute the numerator ($\hat{X}$) by summing the product between the age given in CCC_Q53 and the survey weight (WGT_FULL) for all respondents who answered 2 to question CCC_Q52 (have Type 2 diabetes). Next, sum the value of WGT_FULL for all respondents who answered 2 to question CCC_Q52 to obtain the denominator ($\hat{Y}$). Divide $\hat{X}$ by $\hat{Y}$ to obtain the estimate of the average age of first diagnosis with Type 2 diabetes.

The geometric mean is also a measure of central tendency; however it is more robust to the presence of extreme values then the arithmetic mean. Thus, the geometric mean provides a better indication of central tendency for data that is highly skewed, meaning the data is unevenly spread towards higher or lower values. The geometric mean is typically used for environmental chemicals. It is calculated by:

a) Computing the natural log transform of the value of the variables of interest and multiply it by the final weight. The numerator ($\hat{X}$) is then created by summing all of these values.
b) Summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$)

c) Dividing the numerator estimate by the denominator estimate and then calculating the exponential value.

To determine if the arithmetic mean or the geometric mean is a more appropriate measure for a given variable it is best to plot the data with a histogram. If the data has a relatively even distribution of high and low values than the arithmetic mean will be a good indicator of central tendency. In this case the value of the mean and the median will be very similar. If the data show an uneven spread towards high or low values, then a geometric mean would be more appropriate for the data. In this case the mean and median would be quite different.

12.1.2 Imputation of household income

Largely due to the sensitive nature of reporting income, only 71% of the CHMS respondents reported their total household income (INC_21) in cycle 2. Part of the non-response was also due to a problem with the questionnaire application, in that the income questions were never asked for 835 of the respondents that belonged to households with two selected respondents.

Non-respondents to total household income were asked a series of questions (INC_22 to INC_28) in order to determine a household income range. In cycle 2, 42% of total household income non-respondents provided a response or partial response for the range. Therefore, there is some form of household income information for 83% of respondents. In addition to the low response rate, a study conducted in 2009 using data from the Canadian Community Health Survey (CCHS), suggested that the income non-respondents had different health characteristics than the respondents.[22] Because of this difference in health profiles and the low response rate, analyses that are based exclusively on income respondents may be biased. It was therefore decided to impute total household for the CHMS cycle 2 release.

To impute the household income an auxiliary variable, the modelled household income, is first created. A regression model is used to predict the personal income of each member of all responding CHMS households. Variables used in the model include age group, sex, education level, main source of income, CHMS collection site, marital status, household size and its composition (number of persons in each age group) and whether the dwelling is owned or rented. It is worthwhile to note that no health variables are used in the model in order to prevent the creation of artificial relationships between income and certain health variables. The personal income is then summed for each household to create the modelled household income. This variable is then used to impute the household income using nearest neighbour imputation. The modeled household income defined above is used as a distance measure to determine which pair of respondent-non-respondent records is the “nearest” within imputation classes. The data from the respondent, or donor, is then copied to the non-respondent or recipient. For respondents who provided an income range, a nearest neighbour is selected within the same income range and household size. For respondents who did not provide any income range, the donor record is selected within the same collection site and household size.

With the implementation of the imputation process, there is no non-response to total household income (variable INC_21) starting in cycle 2. An imputation flag INCFIMP4 is included on the file to identify which records have an imputed value for total household income. In addition, derived income variables such as INCDHH, INCDDIA2, INCDDIA4, INCDDIA5, etc., will be based on the imputed household income value.

It should be noted that total personal income (variable INC_31) is not imputed. Users should take the appropriate steps when analyzing this variable.

For users interested in conducting analysis focusing specifically on household income, it is highly recommended to rely on other income data sources rather than using the CHMS. Those interested in analysing household income as it relates to health should use derived income variables such as income quintiles[23] rather than the actual income value directly. The imputation process performs well at predicting income quintiles but there is variability in predicting the exact income value.

12.1.3 Other missing data and values below the limit of detection

Missing data are the result of non-response to some or all questions on the survey. Instances of total non-response, where there is no data for a selected respondent, are adjusted for in the calculation of the survey weights. Thus by using the survey weights to create estimates, the bias that can be introduced by total non-response is reduced. The CHMS data files contain only respondents to the survey; however some variables may be missing for some respondents. There are three main options for analysing a variable with some missing data.

a) Keep the records with missing values in the analysis and report the results for the missing category separately. In this case careful interpretation of the missing category is necessary.
b) Remove records with missing values from the analysis. This is valid if the missing values are random and do not represent any non-response bias. To test this assumption compare the respondents with valid data to the respondents with missing data to look for any differences in some key variables. If there are differences, then it is an indication of non-response bias due to eliminating the missing values from the analysis.
c) Impute the missing values to a new value prior to the analysis.

Besides income imputation (see Section 12.1.2), the most common form of imputation for CHMS data is the replacement of values of the laboratory variables that were below the limit of detection (LOD). The values below the limit of detection are coded as 95, 995, 99.5, etc. depending on the units of the measure. These values are often replaced by the LOD divided by two, but other methods can be used. The LOD values for each laboratory variable can be found in the content summary document by contacting Statistics Canada's National Contact Centre (toll-free 1-800-263-1136; 613-951-8116; infostats@statcan.gc.ca).

If the percentage of values below LOD is greater than 40% then measures of central tendency, such as the geometric mean, are not calculated. This is because the mean is rendered somewhat meaningless when a large proportion of the records have the same value. In this case, percentiles provide a more useful tool of summarising the data.

12.2 Guidelines for statistical analysis

12.2.1 Precise variances or coefficients of variation

Calculation of a precise variance or coefficient of variation is not an easy matter, since there is no simple mathematical formula that can take into account all aspects of the CHMS sample design and the selection probabilities. It is therefore necessary to turn to other methods to estimate these measures of precision, such as re-sampling methods. Among these, the bootstrap method is the one recommended for analysing CHMS data (see Section 9.10 Bootstrap weights). 500 bootstrap replicates have been created for the analysis of the full sample data and for each of the subsamples. The bootstrap replicates are available on each weight file, labelled as BSW1-BSW500. There are several different software packages available that will use the bootstrap replicates to create an estimate of the variance. This is done by calculating the value of the desired estimate for each of the bootstrap replicates and then measuring the variability between the bootstrap estimates.

12.2.2 Some recommendations for doing analysis with data from cycle 2 of the CHMS

The CHMS was designed to provide national baseline prevalence estimates for a variety of health indicators. Because of cost considerations, 18 collection sites from 5 regional strata were chosen; 2 sites from the Atlantic region, 4 sites from the Quebec region, 6 sites from the Ontario region, 3 sites from the Prairies region and 3 sites from the British Columbia region. Then a sample of individuals of all ages was selected from each chosen site. This design should yield approximately unbiased national prevalence estimates that would have CV's of approximately 16.5% for a prevalence of 10% for each of 5 age groups (6-11, 12-19, 20-39, 40-59, and 60-79) by sex and for 3-5 year olds of both sexes combined.

While the small number of sampled collection sites can produce baseline prevalence estimates that meet the above criteria, it has the drawback of leaving at most 13 "degrees of freedom"[24] for variance estimation. Limited degrees of freedom have several consequences for analysis and inference[25]; in particular:

- The variability of variance estimates of estimated quantities needs to be taken into account when doing analyses,
- Estimated covariance matrices of vectors of estimates (such as the vector of estimated coefficients of a model) could be singular or close to singular, thus possibly not invertible[26],
- It may not be possible to calculate some test statistics,
- The usual asymptotic distributions of many test statistics may not hold when there is only a small number of primary sampling units (PSUs), in this case the collection sites in the sample, even when the total sample size is large.

Because of the possible consequences of having a small number of PSUs, a researcher is advised to consider the following recommendations when analyzing CHMS data:

- Produce only national estimates since regional estimates will have even fewer than 13 degrees of freedom because fewer PSUs would be involved in variance estimation. This still allows the

researcher to do analyses of many subpopulations such as a single age group or one sex, since all ages and both sexes are in the sample from each collection site.

- Avoid fitting models with a large number of coefficients. Limit the number of parameters in a regression model to 12 as one degree of freedom is used to estimate the intercept. A continuous explanatory variable is equivalent to one parameter in the model. The number of parameters used by a categorical variable is the number of categories minus 1. For example, sex has two categories therefore it uses 1 parameter in a regression model.
- Use analytical methods that are less impacted by the limited degrees of freedom, or are conservative. In particular, when testing a hypothesis involving a vector of quantities, avoid Wald statistics and their modifications since they have been found to be unstable. Better choices would be Satterthwaite-adjusted statistics or conservative Bonferroni tests.
- As the number of degrees of freedom is at most 13 and may be less for some analyses, consider using a smaller value of alpha to determine if a test is significant. For example, a test could be considered significant if the p-value is less than 0.01 rather than the usual cut-off of 0.05.
- Since bootstrap weights are used for variance estimation with the CHMS, it will likely be necessary to specify the degrees of freedom in the software being used because the default degrees of freedom is likely to be incorrect.

12.2.3 Data comparability over time

12.2.3.1 Normative scales

To help give some context to the raw data that was collected at the MEC, normative scales used by analysts for specific types of data have been added to the master file. The norms used for cycle 2 data collection were agreed upon at the start of collection, and respondents were informed of how they measured against those norms in the final report delivered to them. The norms used throughout cycle 2 collection do not reflect changes that may have been made since the start of collection in 2009. In most cases cycle 1 norms were used again for cycle 2 in order to maintain as much consistency as possible. As the raw data is also included on the file, users are free to use norms other than those included on the file.

12.2.3.2 Activity monitor data for 3 to 5 year olds

Researchers analyzing activity monitor (accelerometer) data for 3 to 5 year olds are advised to take note of an important methodological difference between cycles 2 and 3. In cycle 2, all activity monitor data were collected in 60 second epochs and this may not be the optimal approach for 3 to 5 year old children. Data are being collected in shorter epochs (15 seconds) in cycle 3 (for 3 to 5 year olds only) to align with current research. See Appendix 9 for more information on research being conducted to examine the impact of epoch length on physical activity derived variables as well as assessing adherence to the new preschool physical activity guidelines using accelerometer data. Reliability and quality factors to be used when analyzing spirometry data can be found in Section 10.3.3.7.

12.2.3.3 Phthalate data

CHMS data on phthalate metabolites have been adjusted to correct a bias introduced by the inaccuracy of the standards used to calibrate the analytical instruments and to allow the comparison of data from one CHMS cycle to the next. The adjustment factors are available upon request.

12.2.3.4 Glucose data

The glucose matrix was changed from plasma in cycle 1 to serum in cycle 2. A Vacutainer specific to the glucose analysis (grey top) was used to prepare the plasma in cycle 1. In cycle 2, glucose was analysed in serum as part of the chemistry panel because of the limited amount of blood that could be collected from the respondents. A comparability study was undertaken to confirm that the glucose results in both matrices (serum and plasma) when using the CHMS procedures (e.g., Vacutainer types used and their time of processing) were not significantly different. More information about this comparability study is available upon request.

12.2.3.5 Insulin data

Insulin data was measured using the Advia's Centaur XP instrument in cycle 2, a change from the instrument that was used in cycle 1 (Siemens Diagnostics HealthCare's Immulite 2000). Caution should be used when comparing the insulin data between cycles due to significant bias between different methods, the use of different antibodies, and changes in calibration for each method over time. However, a crossover study (available upon request) of more than 600 specimens was performed between the two methods and the following equation can be used to produce new point estimates for cycle 1: 0.90 * cycle 1 non-adjusted result +7.61. This equation is obtained from the Deming regression, a method that considers that both instruments have variation in the measures. Also see Section 10.3.4.3 for information on the quality of the insulin data.

12.2.3.6 Combining data from cycles 1 and 2

Researchers combining data from cycles 1 and 2 should use the official Statistics Canada combined cycle 1, 2 weight files, and accompanying documentation. This will ensure that the data and methods used in analysing the files are as consistent as possible and that analysis done on the combined cycle 1 and 2 data is representative of the population. Most of the official 1, 2 weight files and accompanying documentation are expected to be released on February 27, 2013. Some additional information related to the environmental lab variables will be released on April 17, 2013, when the cycle 2 environmental variables will be released.

12.2.4 Software packages available

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures can differ from what is appropriate in a sample survey framework. As a result, the procedure may produce an estimate that is correct but a variance estimate that is almost meaningless. A software package with procedures for the analysis of survey data should be used instead. A suitable software package for analysing CHMS data should allow the use of sampling weights, bootstrap weights to estimate the variance and specification of the number of degrees of freedom to use for confidence intervals and significant tests. Examples of available software

packages are SUDAAN 10.0.1, SAS 9.2/9.3, STATA 11.2, WESVAR 5.1 and BootVar 3.2. Table 12.1 gives a comparison of the estimates available in each software package. For more detailed comparison and examples of the code for each program refer to Serré (2012).[27] The estimates produced by these programs may yield slightly different estimates due to differences in the formulas for the standard error.

Table 12.1 Comparison of the procedures available in various software packages

Test	Estimate	Software package				
		SUDAAN 10.0.1	SAS 9.2/ 9.3	STATA 11.2	WESVAR 5.1	BOOTVAR 3.2
Means and Total	Point Estimate	Yes	Yes	Yes	Yes	Yes
	Standard Error	Yes	Yes	Yes	Yes	Yes
	Confidence Interval	Yes	Yes	Yes	Yes	Yes
	Coefficient of Variation	can be calculated	Yes	Yes	Yes	Yes
Proportions	Point Estimate	Yes	Yes	Yes	Yes	Yes
	Standard Error	Yes	Yes	Yes	Yes	Yes
	Confidence Interval	Yes	Yes	Yes	Yes	Yes
	Coefficient of Variation	can be calculated	can calculate	Yes	Yes	Yes
Percentiles	Point Estimate	Yes	In 9.3 only	Yes	Yes	Yes
	Standard Error	Yes	Yes	No	Yes	Yes
	Confidence Interval	Yes	Yes	No	Yes	Yes
	Coefficient of Variation	can be calculated	can calculate	No	Yes	Yes
Linear Regression	coefficients and standard errors	Yes	Yes	Yes	Yes	Yes
	Satterthwaite adjusted statistics	Yes	No	No	No	No
Logistic Regression	coefficients and standard errors	Yes	Yes	Yes	Yes	Yes

12.3 Guidelines for releasing data

12.3.1 Sample size and coefficient of variation

Before releasing or publishing any estimates from the data files users must first determine the number of sampled respondents having the characteristic of interest to ensure that enough observations are available to calculate a quality estimate. For the calculation of a point estimate it is recommended to have at least 10 observations in the numerator and 20 in the denominator. For example, to report the number of Canadian adult females aged 20 to 39 who have high blood pressure there should be at least 10 respondents who have high blood pressure out of at least 20 respondents who are females aged 20 to 39. For the creation of a data table it is recommended that there are at least 10 observations per cell.

As mentioned in Section 10.2.2, an indicator of the scope of the sampling error as measured by the coefficient of variation (CV) should also be produced for each estimate. Table 10.1 gives the Statistics Canada guidelines for releasing estimates based on their CV. Users are strongly advised to adhere to these guidelines.

12.3.2 Rounding guidelines

In order to ensure that estimates for publication or other releases derived from the data 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 proceeding digit is increased 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;

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 that 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.

13. References and end notes

1. The terms "record" and "variables" are used throughout this chapter. A record corresponds to a line on the file and represents a respondent. A "variable" corresponds to a column on the file and represents either a question, a measure, a weight or a bootstrap weight.

2. Giroux, S., F. Labrecque and A. Quigley. 2012. *Sampling documentation for cycle 2 of the Canadian Health Measures Survey*. Statistics Canada internal document.

3. MacIsaac, K. 2011. *The effects of cluster sampling in the Canadian Health Measures Survey*. Statistics Canada internal document.

4. For the first 8 collection sites, the 3 to 5 and 6 to 11 strata were combined. They were split for sites 9 to 18 to improve the sample selection and to ensure the target sample size for each age group was obtained.

5. Stratum defined according to composition of the household in the 2006 Census, updated with administrative information available at the time of sample selection.

6. Canadian Society for Exercise Physiology. 2003. *Canadian Physical Activity, Fitness and Lifestyle Approach (CPAFLA).* 3rd edition. Ottawa.

7. Ross, W. D. & M.J. Marfell-Jones. 1991.*"Kinanthropometry" in Physiological Testing of the High-Performance Athlete.* 2nd edition. J. D. MacDougall, H. A. Wenger, & H. J. Green, eds., Human Kinetics Books, Champaign, Illinois, p. 223-308.

8. Fitness Canada. 1986. *Canadian Standard Test of Fitness (CSTF) Operations Manual.* 3rd edition. Ottawa.

9. Campbell, Norm R.C., Joffre Michel R., McKayDonald W. 2005. "Hypertension Surveillance in Canada: Minimum Standards for Assessing Blood Pressure in Surveys." *Canadian Journal of Public Health* Vol 96, number 3. p. 217 - 220.

10. Miller M.R., J. Hankinson, V. Brusasco *et al*. 2005. "Standardization of spirometry." *Eur Respir*. Vol. 26 p. 319-338.

11. Canadian Fitness and Lifestyle Research Institute (1988) Canada Fitness Survey Longitudinal Study: Reference Booklet. Canadian Fitness and Lifestyle Research Institute Ottawa: ON.

12. Stephens, T. & C.L. Craig. 1985. *Fitness and activity measurement in the 1981 Canada Fitness Survey*. In: Drury, T. (ed.). 1989. *Assessing physical fitness and activity patterns in general population surveys*. Department of Health and Human Services, Public Health Service, Center for Disease Control, National Center for Health Statistics. Hyattsville, MD. pub. No. (PHS) 89-1253.

13. The tar value was provided by Health Canada and was determined as follows:

"Five conditioned cigarettes are smoked per port, using an automated 20-port constant volume smoking machine equipped with a CO analyzer, onto a conditioned, pre-weighed glass fiber filter disc (pad). The gas phase is collected in a Vapour Phase (VP) collection bag and then introduced into a Non-Dispersive Infra-Red analyzer (NDIR) and the % CO determined. The pad is then re-weighed and the difference is the Total Particulate Matter (TPM). The pad is extracted with Isopropanol (IPA) containing the internal standards, and the extract analyzed for nicotine and water by gas chromatography. The tar value is determined by subtracting the water and nicotine from the TPM"[28]

14. The five regions are: British Columbia (including Whitehorse, Yukon), the Prairies (Alberta, Manitoba, Saskatchewan, and Yellowknife, Northwest Territories), Ontario, Quebec and the Atlantic provinces (Newfoundland and Labrador, Prince Edward Island, Nova Scotia and New Brunswick).

15. The number of sites selected in each region is provided in Table 5.1 in Section 5.3.1 (Sampling of collection sites).

16. Haziza, D. and J.-F. Beaumont. 2007. On the construction of imputation classes in surveys. *International Statistical Review.* 75(1): 25-43.

17. Rao JNK, Wu CFJ, Yue K. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1992; 18(2): 209-17

18. Among the dwellings initially selected, some are not within the scope of the survey. These include, for example, vacant or demolished dwellings, non-residential dwellings or dwellings in which all household members are under 3 or over 79 years of age or are full-time members of the Canadian Forces. These dwellings are identified during collection; otherwise they would have been excluded during the selection process. They are not included when calculating response rates.

19. American Thoracic Society. 1995. "Standardization of Spirometry 1994 Update." *Am J Respir Crit Care Med.* Vol. 152, p. 1107-1136.

20. Dion, S.M. And S. Giroux. 2012. *Cycle 1 of the Canadian Health Measures Survey: Bias Study.* Statistics Canada internal report.

21. During the household interview, respondents were asked to name all prescription, over the counter and herbal remedies they were taking, to a maximum of 15 per category. At the MEC interview, the respondent was then asked if they were still taking the medication(s) they listed during the household interview, as well as the names of any new ones they started taking since the household interview. A maximum of five additional medications could be added into each category (prescription, over the counter and herbal). In the case where respondent indicated more than 5 new medications for a given category, only the first five of those medications were chosen and coded.

22. Presentations include the following:
Using data from cycle 1 of the Canadian Health Measures Survey (CHMS): Part 1
Using data from cycle 1 of the Canadian Health Measures Survey (CHMS): Part 2

23. Sarafin, C. 2009. *CCHS Health Indicators by Income Quintiles*. Statistics Canada internal document

24. "Degrees of freedom" is being used as a generic term to reflect the amount of information used to estimate variances and covariances. An approximation often used for the value of "degrees of freedom" is # of PSUs - # of strata. For cycle 2 of the CHMS, the 18 collection sites are the PSUs and there are 5 regions (strata) resulting in 13 degrees of freedom (18-5). This is an approximate estimate of the degrees of freedom and provides only the maximum value.

25. In particular confidence intervals and tests of hypotheses.

26. Invertible covariance matrices are needed to perform Wald tests for tests on vectors of parameters.

27. Serré, L. 2012. *Comparison of software programs for analysing the CHMS.* Statistics Canada internal document.

28. Health Canada. 1999. *Determination of "Tar", Nicotine and Carbon Monoxide in Mainstream Tobacco Smoke*. Ottawa

29. Héroux ME, Clark N, Van Ryswyk K, Mallick R, Gilbert NL, Harrison I, Rispler K, Wang D, Anastassopoulos A, Guay M, MacNeill M, Wheeler AJ. Predictors of indoor air concentrations in smoking and non-smoking residences. *Int J Environ Res Public Health*. 2010; 7:3080-3099.

30. Young D.S. 2007. *Effects of preanalytical variables on clinical laboratory test*. 3rd ed.

Appendix 1 - Acronyms and Abbreviations

AHFS	American Hospital Formulary Service
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
BP	blood pressure
CAI	Computer assisted interviewing
CAP	College of American Pathologists
CATI	Computer assisted telephone interview
CBC	complete blood count
CCHS	Canadian Community Health Survey
CHMS	Canadian Health Measures Survey
CMA	Census Metropolitan Area
CPAFLA	Canadian Physical Activity, Fitness and Lifestyle Approach
CSEP	Canadian Society of Exercise Physiology
CV	Coefficient of variation
DIN	drug Identification Number
DV	Derived variables
HC	Health Canada
HMS	Health measures specialist
HR	household response rate
HRGs	homogeneous response groups
HPV	Human Papillomavirus
IAS	Indoor air sampler
IATA	International Air Transport Association
INSPQ	Institut National de Santé Publique du Québec
ISAK	International Society for the Advancement of Kinanthropometry
mCAFT	Modified Aerobic Fitness Test
MEC	Mobile Examination Clinic
NAICS	North American Industry Classification System
LFS	Labour Force Survey
LOD	Limit of detection
NIH	National Institute for Health
NHANES	National Health and Nutrition Examination Survey
NML	National Microbiology Laboratories
NOC-S	National Occupational Classification – Statistics
OSA	obstructive sleep apnea
PAR-Q	Physical Activity Readiness Questionnaire
PFBS	perfluorobutane sulfonate
PFBA	perfluoro-n-butyric acid
PFDA	perfluorodecanoic acid
PFHxA	perfluorohexanoic acid
PFNA	perfluorononanoic acid
PFHxS	perfluorohexane acid
PFOA	perfluorooctanoic acid
PFOS	perfluorooctane sulfonate
PFUDA	perfluoroundecanoic acid

PHAC	Public Health Agency of Canada
PSU	primary sampling units
QA	Quality assurance
QC	Quality control
QMP-LS	Quality Management Program - Laboratory Service
VOC	volatile organic compound
RDC	Research Data Centers
SOP	Standard operating procedures
TTY	Teletypewriter (telecommunication device for the hearing impaired)
WHO	World Health Organization

Appendix 2 - List of available Canadian Health Measures Survey (CHMS) documents

CHMS Content summary for cycles 1, 2 and 3

- The content summary document is divided into separate tables which list all of the content topics in the survey by age group of respondent. There are tables on the household questionnaire, mobile examination centre (MEC) physical measures, MEC questionnaire, laboratory blood and urine tests, laboratory indoor air sample tests and laboratory tap water sample tests. The laboratory tables also provide information on analytical range and conversion factors for cycles 1 and 2.

CHMS Data User Guide – Cycle 2

- There are three different versions of the User Guide, corresponding to the three main data release dates. Version 1 is released on September 20, 2012 (household and physical measures data), Version 2 on November 21, 2012 (activity monitor, indoor air and non-environmental laboratory data) and Version 3 on April 17, 2013 (environmental laboratory data). Background and contextual information related to a particular data file is included in the earliest possible version of the User Guide, even though it may correspond to a data file that has not yet been released.
- The Table of Contents can be accessed using the following link: http://www23.statcan.gc.ca/imdb-bmdi/document/5071_D2_T1_V2-eng.htm

CHMS Derived Variables (DVs) documentation – Cycle 2

- There are separate DV documents for the following different types of DVs: household and mobile examination centre (MEC), medication, activity monitor, non-environmental laboratory measures and environmental laboratory measures.

CHMS Data Dictionaries – Cycle 2

- There are separate data dictionaries for the following different data files: master, medication, activity monitor subsample, indoor air subsample – household level, indoor air subsample – person level, fasted blood measures subsample, environment blood measures subsample and environment urine measures subsample.

Supporting documentation for the climate and air quality file – Cycle 2

CHMS sampling documentation – Cycle 2

Presentations on using CHMS data – Cycles 1 and 2

For more information or to obtain copies of the documents in the list above, please contact Statistics Canada's National Contact Centre (toll-free 1-800-263-1136; 613-951-8116; infostats@statcan.gc.ca).

Appendix 3 - List of Collection Sites for Cycle 2

The CHMS collection sites for cycle 2 are as follows:

- Central and East Ottawa, Ontario
- Oakville, Ontario
- South of Brantford, Ontario
- Laval, Québec
- South Montérégie, Quebec
- St. John's, Newfoundland and Labrador
- Winnipeg, Manitoba
- Richmond, British Columbia
- Edmonton, Alberta
- Central and East Kootenay, British Columbia
- Calgary, Alberta
- Southwest Toronto, Ontario
- Kingston, Ontario
- Colchester and Pictou counties, Nova Scotia
- East Toronto, Ontario
- Gaspésie, Quebec
- North Shore Montréal, Quebec
- Coquitlam, British Columbia

Note: As the CHMS was designed to produce national estimates only, it is not recommended to do analysis at lower geographic levels as it could result in either extreme sampling variability or unstable estimates of the sampling variability.

For more information, please contact Statistics Canada's National Contact Centre (1-800-263-1136; 613-951-8116; infostats@statcan.gc.ca).

Aussi disponible en français.

Appendix 4 - Pre-testing Guidelines

Pre-testing Guidelines for fasting appointments

GUIDELINES TO FOLLOW BEFORE YOUR APPOINTMENT (does not apply to children aged 3-5)

- During the **2 days prior** to your clinic appointment:
  - Do not donate blood (blood tests are permitted).
- During the **12 hours prior** to your clinic appointment:
  - Do not eat or drink anything other than water (no food, candies, gum, cough lozenges, flavoured water, coffee, etc.);
  - Do not drink any alcoholic beverages.
- During the **2 hours prior** to your clinic appointment:
  - Do not smoke or use other tobacco and nicotine products;
  - Do not go to the bathroom, as you will be asked to **provide a urine sample upon your arrival**.
- On **the day** of your clinic appointment:
  - Take your medications as usual;
  - Do not exercise (from midnight);
  - Do not wear scented products.

- **Medication use:**
  - **Take your medications as usual** on the day of your clinic appointment.
  - Please bring with you **all medications** (prescription or over the counter), **herbal remedies** or **supplements** that you began taking since, or did not disclose during, the household interview.
  - If you have a breathing condition (e.g., asthma), bring **your inhaler or medication**.

WHAT TO BRING TO YOUR APPOINTMENT

- **All medications** (prescription or over-the-counter), **herbal remedies** or **supplements** that you did not mention during the household interview.
- Clothes and footwear **appropriate for exercise** (no jeans, flip flops or high heels).
- **Provincial health insurance card.**
- Name, address and phone number of **two contact persons.**
- If you have a breathing condition (e.g., asthma), bring **your inhaler or medication**.

Pre-testing Guidelines for non-fasting appointments

GUIDELINES TO FOLLOW (does not apply to children aged 3-5)

- During **the 2 days prior** to your clinic appointment:
  - Do not donate blood (blood tests are permitted).
- During the **6 hours prior** to your clinic appointment
  - Do not drink any alcoholic beverages.
- During the **2 hours prior** to your clinic appointment:
  - Do not eat a heavy meal or consume caffeinated products (e.g., chocolate, coffee, tea, pop or energy drinks) — a light snack is permitted;
  - Do not smoke or use other tobacco and nicotine products;
  - Do not go to the bathroom, as you will be asked to **provide a urine sample upon your arrival**.
- On **the day** of your clinic appointment:
  - Take your medications as usual;
  - Do not exercise (from midnight);
  - Do not wear scented products.

WHAT TO BRING TO YOUR APPOINTMENT

- **All medications** (prescription or over-the-counter), **herbal remedies** or **supplements** that you did not mention during the household interview.
- Clothes and footwear **appropriate for exercise** (no jeans, flip flops or high heels).
- **Provincial health insurance card.**
- Name, address and phone number of **two contact persons.**
- If you have a breathing condition (e.g., asthma), bring **your inhaler or medication**.

Appendix 5 - Exclusion Criteria

The exclusion criteria for the physical measures were separated into automatic application exclusions and staff decision exclusions. These can be defined as the following;

Automatic application exclusion criteria include exclusions that are made automatically by the application based on previously asked questions in order to prevent the respondent from completing certain physical components (e.g., a respondent who is pregnant is automatically screened out of the waist circumference measurement).

Staff decision exclusion criteria is based on the staff assessing if the respondent's condition will negatively impact the data quality of the physical measure or could compromise the safety of the respondent (e.g., the respondent has difficulty breathing at rest therefore, the staff decision would be to exclude the respondent from the mCAFT test).

Measure	**Exclusion Criteria**	
	Automatic application exclusions	**Staff Decision**
Blood Pressure	• Respondent < 6 years of age	<u>Test Screen Out</u> • Blood pressure cuff too small or too large to fit arm • Acute or chronic conditions on both arms (e.g,rashes, gauze dressings, casts, edema, paralysis, tubes, open sores or wounds, withered arms or a-v shunts) • Double mastectomy • Double arm amputee <u>Right Arm Exclusion</u> • Acute or chronic condition on the right arm (e.g., rashes, gauze dressing, cast, edema, paralysis, tubes, open sore or wound, withered arm or a-v shunt) • Right mastectomy • Right arm amputation
Standing Height	None	• Acute or chronic condition preventing respondent from standing upright unassisted (e.g., cast on leg)
Sitting Height	None	• Acute or chronic condition hindering respondent from sitting upright unassisted (e.g., full leg cast)
Weight	None	• Fibreglass/plaster cast which cannot be removed
Waist Circumference	• Pregnancy	• Acute or chronic condition (e.g., unable to correctly landmark a wheelchair bound respondent, a colostomy bag which interferes with taking an accurate measurement)
Hip Circumference	• Pregnancy	• Respondent is wheelchair bound • Acute or chronic condition preventing respondent from standing unassisted
Neck Circumference	• Respondent > 19 years of age • Pregnancy	• Visible deformity of the neck (e.g., goiter) • Presence of a tracheotomy or other appliance that cannot be removed • Acute or chronic condition preventing the respondent from holding their head in the correct

Measure	Exclusion Criteria	
	Automatic application exclusions	**Staff Decision**
		position
Skinfolds	• BMI ≥ 30 kg/m^2 • No BMI was calculated • Pregnancy	Test Screen Out • Quadriplegic Site-specific exclusions : • Site specific paralysis • Missing right arm or leg • Acute or chronic condition (e.g., varicose veins, skin condition, unable to landmark a wheelchair bound respondent)
Phlebotomy	• Chemotherapy within the past 4 weeks • Haemophilia	• Acute or chronic conditions on both arms (e.g., rashes, gauze dressings, casts, edema, paralysis, tubes, open sores or wounds, withered arms or missing limbs, damaged, sclerosed or occluded veins, allergies to cleansing reagents, burned or scarred tissue, shunt or IV)
Urine	• Wheelchair bound and has a catheter	• Important language barrier preventing proper instruction for collection • Mental/physical disability preventing providing a sample
Spirometry	• Respondent < 6 years of age • Pregnancy (> 27 weeks) • Heart attack within the last 3 months • Major surgery on chest or abdomen within the last 3 months • Recent eye surgery (≤ 6 weeks)	• Acute respiratory condition (e.g., cold, bronchitis, flu, etc.) • Respondent with a stoma • Difficulty breathing at rest • Respondent taking medication for tuberculosis • Important language barrier • Persistent cough • Acute condition or chronic condition (e.g., persistent cough) • Any other reason as assessed by the HMS (e.g, cleft pallet)
Activity Monitor	• None	• Respondent is wheelchair bound • Acute or chronic condition
Indoor Air Sampler	• Home visit • Second person from a two person household to complete the clinic visit	• None
Grip Strength	• Respondent < 6 years of age • Positive response(s) to PAR-Q questions 6 and condition is "aneurysm"	Test Screen Out • Positive response(s) to PAR-Q questions 5 or 7 (depending upon probing). • Drank alcohol within the last 6 hours (depending upon probing) • Double arm/hand amputee • Acute or chronic condition preventing respondent from gripping dynamometer with both hands (e.g., severe burns, neurological condition) • Any other reason as assessed by the HMS

Measure	Exclusion Criteria	
	Automatic application exclusions	**Staff Decision**
		Hand-specific exclusions : • Positive response(s) to PAR-Q questions 5 or 7 (depending upon probing). • Right/left arm/hand amputee • Acute or chronic condition preventing respondent from gripping dynamometer with one hand (e.g., severe burn, neurological condition)
mCAFT	• Home visit • Age < 8 or > 69 years • Pregnancy (> 12 weeks) • Positive response(s) to PAR-Q questions 1, 2, 3 • Positive response to PAR-Q question 4 and condition is "During or after exercise" or "At random" • Taking quick relief type medication for breathing condition (e.g.,. asthma) and did not bring medication with them • Resting BP > 144/94 mmHg (age>14) • Resting HR > 99 bpm (age>14) • Blood donation in past 24 hours	• Positive response(s) to PAR-Q questions 4, 5, 7 (depending upon probing). • Heart rate or blood pressure altering medications • Paralysis • Difficulty breathing at rest • Use of supplemental oxygen by respondent • Large meal within 2 hour of appointment start time (depending upon probing) • Drank alcohol within the last 6 hours (depending upon probing) • Acute or chronic condition (e.g., lower limb injury, neurological condition, ill or feverish, persistent cough) • Mentally or physically impaired • Any other reason as assessed by the HMS
Sit and Reach	• Home visit • Age < 6 or > 69 years • Pregnancy (> 12 weeks)	• Positive response(s) to PAR-Q questions 5 and 7 (depending upon probing). • Lower body paralysis • Drank alcohol within the last 6 hours (depending upon probing) • Acute or chronic condition (e.g., upper/lower limb cast, herniated disc, colostomy bag etc.) • Mentally or physically impaired • Any other reason as assessed by the HMS
Partial Curl-Ups	• Home visit • Age < 8 or > 69 years • Pregnancy (> 12 weeks) • Positive response(s) to PAR-Q questions 1, 2, 3 • Positive response(s) to PAR-Q questions 6 and condition is "aneurysm" • Resting BP > 144/94 mmHg* (age>14) • Resting HR > 99 bpm* (age>14)	• Positive response(s) to PAR-Q questions 5 and 7 (depending upon probing). • Lower body paralysis • Drank alcohol within the last 6 hours (depending upon probing) • Acute or chronic condition (e.g., upper/lower limb cast, herniated disc, colostomy bag, ill or feverish, persistent cough) • Mentally or physically impaired • Any other reason as assessed by the HMS

Appendix 6 – Physical Activity Readiness Questionnaire (PAR-Q)

Physical Activity Readiness Questionnaire - PAR-Q (revised 2002)

PAR-Q & YOU

(A Questionnaire for People Aged 15 to 69)

Regular physical activity is fun and healthy, and increasingly more people are starting to become more active every day. Being more active is very safe for most people. However, some people should check with their doctor before they start becoming much more physically active.

If you are planning to become much more physically active than you are now, start by answering the seven questions in the box below. If you are between the ages of 15 and 69, the PAR-Q will tell you if you should check with your doctor before you start. If you are over 69 years of age, and you are not used to being very active, check with your doctor.

Common sense is your best guide when you answer these questions. Please read the questions carefully and answer each one honestly: check YES or NO.

YES	NO		
☐	☐	1.	**Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?**
☐	☐	2.	**Do you feel pain in your chest when you do physical activity?**
☐	☐	3.	**In the past month, have you had chest pain when you were not doing physical activity?**
☐	☐	4.	**Do you lose your balance because of dizziness or do you ever lose consciousness?**
☐	☐	5.	**Do you have a bone or joint problem (for example, back, knee or hip) that could be made worse by a change in your physical activity?**
☐	☐	6.	**Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?**
☐	☐	7.	**Do you know of any other reason why you should not do physical activity?**

If you answered

YES to one or more questions

Talk with your doctor by phone or in person BEFORE you start becoming much more physically active or BEFORE you have a fitness appraisal. Tell your doctor about the PAR-Q and which questions you answered YES.

- You may be able to do any activity you want — as long as you start slowly and build up gradually. Or, you may need to restrict your activities to those which are safe for you. Talk with your doctor about the kinds of activities you wish to participate in and follow his/her advice.
- Find out which community programs are safe and helpful for you.

NO to all questions

If you answered NO honestly to all PAR-Q questions, you can be reasonably sure that you can:

- start becoming much more physically active – begin slowly and build up gradually. This is the safest and easiest way to go.
- take part in a fitness appraisal – this is an excellent way to determine your basic fitness so that you can plan the best way for you to live actively. It is also highly recommended that you have your blood pressure evaluated. If your reading is over 144/94, talk with your doctor before you start becoming much more physically active.

DELAY BECOMING MUCH MORE ACTIVE:

- if you are not feeling well because of a temporary illness such as a cold or a fever – wait until you feel better; or
- if you are or may be pregnant – talk to your doctor before you start becoming more active.

PLEASE NOTE: If your health changes so that you then answer YES to any of the above questions, tell your fitness or health professional. Ask whether you should change your physical activity plan.

Informed Use of the PAR-Q: The Canadian Society for Exercise Physiology, Health Canada, and their agents assume no liability for persons who undertake physical activity, and if in doubt after completing this questionnaire, consult your doctor prior to physical activity.

No changes permitted. You are encouraged to photocopy the PAR-Q but only if you use the entire form.

NOTE: If the PAR-Q is being given to a person before he or she participates in a physical activity program or a fitness appraisal, this section may be used for legal or administrative purposes.

"I have read, understood and completed this questionnaire. Any questions I had were answered to my full satisfaction."

NAME _____

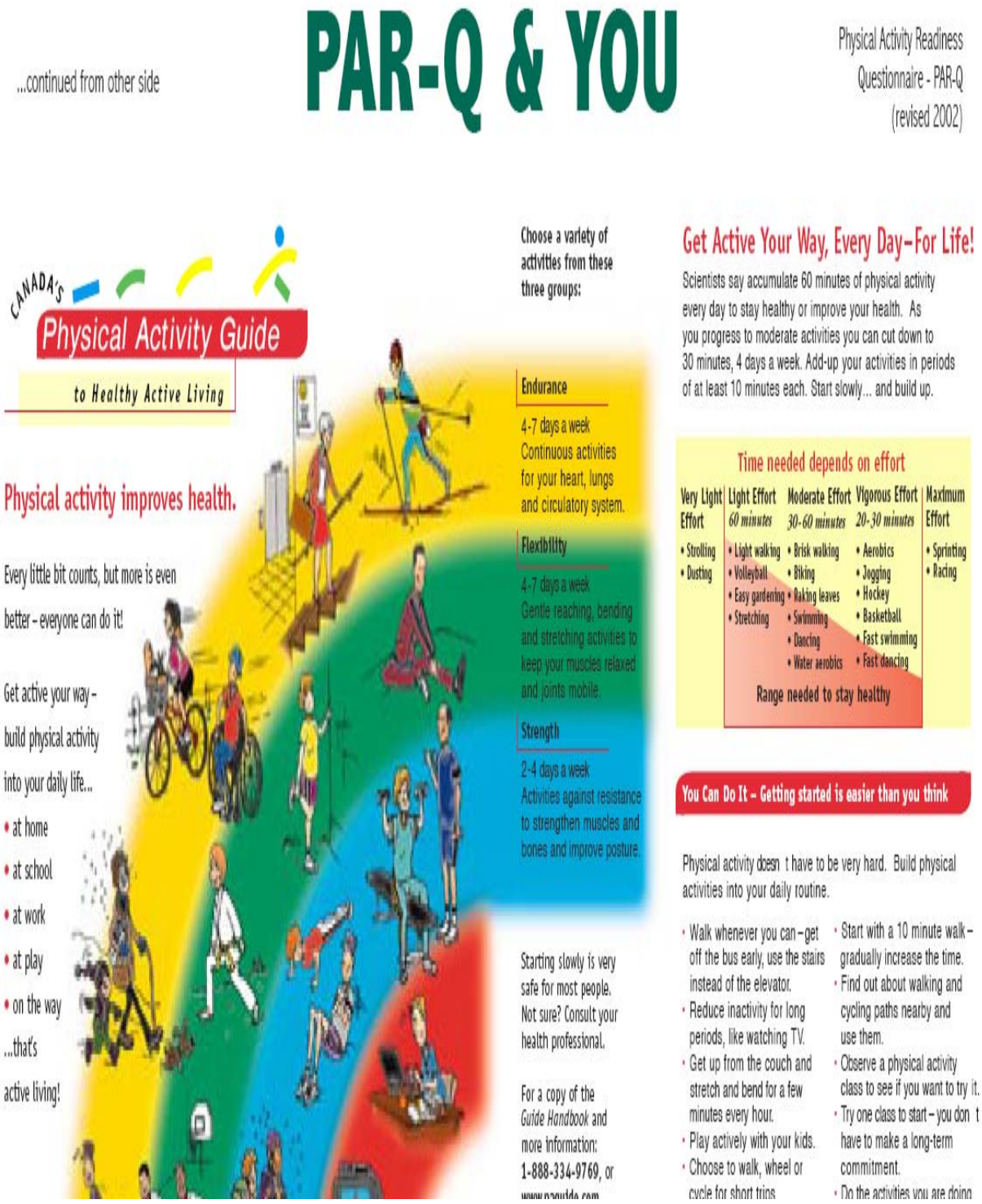
SIGNATURE _____ DATE _____

SIGNATURE OF PARENT or GUARDIAN (for participants under the age of majority) _____ WITNESS _____

Note: This physical activity clearance is valid for a maximum of 12 months from the date it is completed and becomes invalid if your condition changes so that you would answer YES to any of the seven questions.

Supported by: Health Canada Santé Canada

continued on other side...



...continued from other side

PAR-Q & YOU

Physical Activity Readiness Questionnaire - PAR-Q (revised 2002)

CANADA'S Physical Activity Guide to Healthy Active Living

Physical activity improves health.

Every little bit counts, but more is even better - everyone can do it!

Get active your way - build physical activity into your daily life...

- at home
- at school
- at work
- at play
- on the way

...that's active living!

Choose a variety of activities from these three groups:

Endurance

4-7 days a week
Continuous activities for your heart, lungs and circulatory system.

Flexibility

4-7 days a week
Gentle reaching, bending and stretching activities to keep your muscles relaxed and joints mobile.

Strength

2-4 days a week
Activities against resistance to strengthen muscles and bones and improve posture.

Starting slowly is very safe for most people. Not sure? Consult your health professional.

For a copy of the *Guide Handbook* and more information: 1-888-334-9769, or www.paguide.com

Get Active Your Way, Every Day-For Life!

Scientists say accumulate 60 minutes of physical activity every day to stay healthy or improve your health. As you progress to moderate activities you can cut down to 30 minutes, 4 days a week. Add-up your activities in periods of at least 10 minutes each. Start slowly... and build up.

Time needed depends on effort

Very Light Effort	Light Effort *60 minutes*	Moderate Effort *30-60 minutes*	Vigorous Effort *20-30 minutes*	Maximum Effort
• Strolling • Dusting	• Light walking • Volleyball • Easy gardening • Stretching	• Brisk walking • Biking • Raking leaves • Swimming • Dancing • Water aerobics	• Aerobics • Jogging • Hockey • Basketball • Fast swimming • Fast dancing	• Sprinting • Racing

Range needed to stay healthy

You Can Do It - Getting started is easier than you think

Physical activity doesn t have to be very hard. Build physical activities into your daily routine.

- Walk whenever you can - get off the bus early, use the stairs instead of the elevator.
- Reduce inactivity for long periods, like watching TV.
- Get up from the couch and stretch and bend for a few minutes every hour.
- Play actively with your kids.
- Choose to walk, wheel or cycle for short trips
- Start with a 10 minute walk - gradually increase the time.
- Find out about walking and cycling paths nearby and use them.
- Observe a physical activity class to see if you want to try it.
- Try one class to start - you don t have to make a long-term commitment.
- Do the activities you are doing

Appendix 7 - Medication Classification Systems

- **Anatomical Therapeutic Chemical (ATC) classification system:** The Anatomical Therapeutic Chemical (ATC) classification system was developed by the World Health Organization (WHO). It classifies pharmaceutical products according to the organ or system on which they act and their chemical, pharmacological and therapeutic properties. An ATC code was assigned to each medication using the Drug Identification Number (DIN) and Health Canada's ATC coding system. When a medication had more than one indication, the ATC code was decided based on the main indication of the medication. The main indication was determined by Health Canada reviewers using the Product Monograph (a factual document on a drug product that describes the properties, claims, indications, and conditions of use of the drug product). In cases where it was still not clear from the Product Monograph which ATC code to assign, Health Canada contacted the ATC group at the WHO Collaborating Centre for Drug Statistics Methodology in Norway for clarification. A medication was not assigned an ATC code if the DIN was missing or if the DIN did not exist in the Health Canada drug database (for example experimental drugs). In these cases the ATC code appears as "not stated" on the file. The classification system is only available in English. For more information on Anatomical Therapeutical Chemical System please refer to: http://www.whocc.no/atc_ddd_index/.

- **American Hospital Formulary Service (AHFS) classification system:**
  The American Hospital Formulary Service (AHFS) classification system is published by the American Society of Health-System Pharmacists to describe the mode of action of pharmaceutical products, including vitamin and mineral supplements. The classification system is only available in English. For more information on the American Hospital Formulary Service (AHFS) classification system please refer to: http://www.ahfsdruginformation.com/.

Appendix 8 - Response Rates

Appendix 8a - CHMS Cycle 2 Response Rates by age group and sex

			Households where one person was selected					Households where two persons were selected				
Sex	**Age Group**	**# respondent households**	**# persons selected (PS1)**	**# respondents to the questionnaire (Q1)**	**HHLD questionnaire response rate (PQ1)**	**# participants at the MEC (C1)**	**MEC response rate (PC1)**	**# persons selected (PS2)**	**# respondents to the questionnaire (Q2)**	**HHLD questionnaire response rate (PQ2)**	**# participants at the MEC (C2)**	**MEC response rate (PC2)**
Both Sexes	3 to 5	409	2	2	100.0	2	100.0	814	754	92.6	599	79.4
	6 to 11	682	1	1	100.0	1	100.0	1361	1272	93.5	1074	84.4
	12 to 19	1134	884	784	88.7	664	84.7	499	462	92.6	396	85.7
	20 to 39	1390	882	752	85.3	578	76.9	1015	902	88.9	751	83.3
	40 to 59	1357	1037	927	89.4	746	80.5	639	573	89.7	486	84.8
	60 to 79	1495	1472	1362	92.5	1066	78.3	46	39	84.8	32	82.1
	3 to 79	**6465**	**4278**	**3828**	**89.5**	**3057**	**79.9**	**4374**	**4002**	**91.5**	**3338**	**83.4**
	6 to 79	**6056**	**4276**	**3826**	**89.5**	**3055**	**79.8**	**3560**	**3248**	**91.2**	**2739**	**84.3**
Males	6 to 11	344	1	1	100.0	1	100.0	685	635	92.7	533	83.9
	12 to 19	600	471	412	87.5	343	83.3	258	241	93.4	207	85.9
	20 to 39	638	450	378	84.0	290	76.7	376	325	86.4	266	81.8
	40 to 59	690	510	453	88.8	359	79.2	359	308	85.8	260	84.4
	60 to 79	670	661	612	92.6	498	81.4	18	15	83.3	11	73.3
	6 to 79	**2941**	**2093**	**1856**	**88.7**	**1491**	**80.3**	**1696**	**1524**	**89.9**	**1277**	**83.8**
Females	6 to 11	676						676	637	94.2	541	84.9
	12 to 19	534	413	372	90.1	321	86.3	241	221	91.7	189	85.5
	20 to 39	752	432	374	86.6	288	77.0	639	577	90.3	485	84.1
	40 to 59	667	527	474	89.9	387	81.6	280	265	94.6	226	85.3
	60 to 79	825	811	750	92.5	568	75.7	28	24	85.7	21	87.5
	6 to 79	**3115**	**2183**	**1970**	**90.2**	**1564**	**79.4**	**1864**	**1724**	**92.5**	**1462**	**84.8**

Appendix 8a - CHMS Cycle 2 Response Rates by age group and sex, continued

				Response rate not adjusted for 1 or 2 persons selected				
Sex	**Age Group**	**# respondent households**	**Overall Combined Response Rate ***	**# persons selected**	**# respondents to the questionnaire**	**HHLD questionnaire response rate**	**# participants at the MEC**	**MEC response rate**
	3 to 5	409	55.9	816	756	92.6	601	79.5
Both Sexes	6 to 11	682	59.9	1362	1273	93.5	1075	84.4
	12 to 19	1134	57.7	1383	1246	90.1	1060	85.1
	20 to 39	1390	52.1	1897	1654	87.2	1329	80.4
	40 to 59	1357	55.3	1676	1500	89.5	1232	82.1
	60 to 79	1495	54.9	1518	1401	92.3	1098	78.4
	3 to 79	**6465**	**55.5**	**8652**	**7830**	**90.5**	**6395**	**81.7**
	6 to 79	**6056**	**55.4**	**7836**	**7074**	**90.3**	**5794**	**81.9**
Males	6 to 11	344	59.1	686	636	92.7	534	84.0
	12 to 19	600	56.5	729	653	89.6	550	84.2
	20 to 39	638	50.3	826	703	85.1	556	79.1
	40 to 59	690	53.8	869	761	87.6	619	81.3
	60 to 79	670	57.0	679	627	92.3	509	81.2
	6 to 79	**2941**	**54.9**	**3789**	**3380**	**89.2**	**2768**	**81.9**
Females	6 to 11	676	60.7	676	637	94.2	541	84.9
	12 to 19	534	59.1	654	593	90.7	510	86.0
	20 to 39	752	53.6	1071	951	88.8	773	81.3
	40 to 59	667	56.9	807	739	91.6	613	82.9
	60 to 79	825	53.2	839	774	92.3	589	76.1
	6 to 79	**3115**	**55.9**	**4047**	**3694**	**91.3**	**3026**	**81.9**

* The number of households within the scope of the survey was 8520. This number cannot be broken down by age group or sex because no information is known about the households that did not respond. The overall household response rate is 75.9% (6465 / 8520). This rate is used to calculate all of the combined response rates by age group and sex.

Appendix 8b - CHMS Cycle 2 Activity Monitor Response Rates by age group and sex

		Households where one person was selected			Households where two persons were selected				Response rate not adjusted for 1 or 2 persons selected		
Sex	**Age Group**	**# participants at the MEC offered an activity monitor (OF1)**	**# persons who returned monitor with at least 4 days of valid entries (V1)**	**Activity monitor response rate (AM1)**	**# participants at the MEC offered an activity monitor (OF2)**	**# persons who returned monitor with at least 4 days of valid entries (V2)**	**Activity monitor response rate (AM2)**	**Activity monitor combined response rate (AMCBRR)**	**# MEC respondents offered an activity monitor**	**# persons who returned monitor with at least 4 days of valid entries**	**Activity monitor response rate**
Both Sexes	3 to 5	2	0	0.0	597	459	76.9	42.7	599	459	76.6
	6 to 11	1	1	100.0	1072	893	83.3	49.9	1073	894	83.3
	12 to 19	662	429	64.8	395	302	76.5	38.9	1057	731	69.2
	20 to 39	578	391	67.6	751	606	80.7	37.9	1329	997	75.0
	40 to 59	744	617	82.9	486	405	83.3	45.9	1230	1022	83.1
	60 to 79	1055	816	77.3	32	29	90.6	42.6	1087	845	77.7
	3 to 79	**3042**	**2254**	**74.1**	**3333**	**2694**	**80.8**	**42.4**	**6375**	**4948**	**77.6**
	6 to 79	**3040**	**2254**	**74.1**	**2736**	**2235**	**81.7**	**42.4**	**5776**	**4489**	**77.7**
Males	6 to 11	1	1	100.0	532	438	82.3	48.7	533	439	82.4
	12 to 19	342	212	62.0	207	160	77.3	37.0	549	372	67.8
	20 to 39	290	184	63.4	266	219	82.3	34.9	556	403	72.5
	40 to 59	358	284	79.3	260	205	78.8	42.6	618	489	79.1
	60 to 79	492	380	77.2	11	11	100.0	44.2	503	391	77.7
	6 to 79	**1483**	**1061**	**71.5**	**1276**	**1033**	**81.0**	**40.9**	**2759**	**2094**	**75.9**
Females	6 to 11				540	455	84.3	51.2	540	455	84.3
	12 to 19	320	217	67.8	188	142	75.5	41.1	508	359	70.7
	20 to 39	288	207	71.9	485	387	79.8	40.4	773	594	76.8
	40 to 59	386	333	86.3	226	200	88.5	49.4	612	533	87.1
	60 to 79	563	436	77.4	21	18	85.7	41.3	584	454	77.7
	6 to 79	**1557**	**1193**	**76.6**	**1460**	**1202**	**82.3**	**43.9**	**3017**	**2395**	**79.4**

Appendix 8c - CHMS Cycle 2 Indoor Air Sampler (IAS) Response Rates

Household Size	# respondent households (R)	# HHLD where at least 1 person completed the questionnaire (Q)	HHLD questionnaire response rate (PQ)	# HHLD with at least 1 participant at the MEC (C)	MEC response rate (PC)	# HHLD offered an IAS (Off)	# HHLD that returned IAS with valid data (Val)	IAS response rate (IAS)	Overall Combined Response Rate (IASCRR)
1 person	1013	978	96.5	753	77.0	736	594	80.7	45.5
2 persons	1775	1594	89.8	1273	79.9	1257	1048	83.4	45.4
3 or more persons	3677	3302	89.8	2756	83.5	2716	2215	81.6	46.4
Total	**6465**	**5874**	**90.9**	**4782**	**81.4**	**4709**	**3857**	**81.9**	**46.0**

Appendix 8d: CHMS Cycle 2 Blood Draw and Urine Response Rates by age group and sex

		Households where one person was selected					Households where two persons were selected						
Sex	**Age Group**	**# participants at the MEC (C1)**	**# persons who provided blood (B1)**	**Blood draw response rate (BC1)**	**# persons who provided urine (U1)**	**Urine response rate (UC1)**	**# participants at the MEC (C2)**	**# persons who provided blood (B2)**	**Blood draw response rate (BC2)**	**# persons who provided urine (U2)**	**Urine response rate (UC2)**	**Blood draw combined response rate**	**Urine combined response rate**
Both Sexes	3 to 5	2	2	100.0	2	100.0	599	516	86.1	571	95.3	48.2	53.3
	6 to 11	1	1	100.0	1	100.0	1074	973	90.6	1063	99.0	54.3	59.3
	12 to 19	664	637	95.9	659	99.2	396	371	93.7	387	97.7	55.1	57.1
	20 to 39	578	573	99.1	577	99.8	751	743	98.9	748	99.6	51.6	51.9
	40 to 59	746	744	99.7	743	99.6	486	483	99.4	486	100.0	55.1	55.2
	60 to 79	1066	1054	98.9	1058	99.2	32	31	96.9	32	100.0	54.3	54.5
	3 to 79	**3057**	**3011**	**98.5**	**3040**	**99.4**	**3338**	**3117**	**93.4**	**3287**	**98.5**	**53.6**	**55.0**
	6 to 79	**3055**	**3009**	**98.5**	**3038**	**99.4**	**2739**	**2601**	**95.0**	**2716**	**99.2**	**54.0**	**55.1**
Males	6 to 11	1	1	100.0	1	100.0	533	493	92.5	531	99.6	54.7	58.9
	12 to 19	343	330	96.2	340	99.1	207	199	96.1	203	98.1	54.3	55.8
	20 to 39	290	288	99.3	289	99.7	266	265	99.6	264	99.2	50.0	50.1
	40 to 59	359	359	100.0	356	99.2	260	260	100.0	260	100.0	53.8	53.5
	60 to 79	498	497	99.8	496	99.6	11	11	100.0	11	100.0	56.9	56.8
	6 to 79	**1491**	**1475**	**98.9**	**1482**	**99.4**	**1277**	**1228**	**96.2**	**1269**	**99.4**	**53.9**	**54.6**
Females	6 to 11						541	480	88.7	532	98.3	53.9	59.7
	12 to 19	321	307	95.6	319	99.4	189	172	91.0	184	97.4	55.9	58.5
	20 to 39	288	285	99.0	288	100.0	485	478	98.6	484	99.8	52.9	53.5
	40 to 59	387	385	99.5	387	100.0	226	223	98.7	226	100.0	56.5	56.9
	60 to 79	568	557	98.1	562	98.9	21	20	95.2	21	100.0	52.2	52.7
	6 to 79	**1564**	**1534**	**98.1**	**1556**	**99.5**	**1462**	**1373**	**93.9**	**1447**	**99.0**	**54.1**	**55.5**

Appendix 8d: CHMS Cycle 2 Blood Draw and Urine Response Rates by age group and sex, continued

		Response rate not adjusted for 1 or 2 persons selected				
Sex	**Age Group**	**# participants at the MEC**	**# persons who provided blood**	**Blood draw response rate**	**# persons who provided urine**	**Urine response rate**
Both Sexes	3 to 5	601	518	86.2	573	95.3
	6 to 11	1075	974	90.6	1064	99.0
	12 to 19	1060	1008	95.1	1046	98.7
	20 to 39	1329	1316	99.0	1325	99.7
	40 to 59	1232	1227	99.6	1229	99.8
	60 to 79	1098	1085	98.8	1090	99.3
	3 to 79	**6395**	**6128**	**95.8**	**6327**	**98.9**
	6 to 79	**5794**	**5610**	**96.8**	**5754**	**99.3**
Males	6 to 11	534	494	92.5	532	99.6
	12 to 19	550	529	96.2	543	98.7
	20 to 39	556	553	99.5	553	99.5
	40 to 59	619	619	100.0	616	99.5
	60 to 79	509	508	99.8	507	99.6
	6 to 79	**2768**	**2703**	**97.7**	**2751**	**99.4**
Females	6 to 11	541	480	88.7	532	98.3
	12 to 19	510	479	93.9	503	98.6
	20 to 39	773	763	98.7	772	99.9
	40 to 59	613	608	99.2	613	100.0
	60 to 79	589	577	98.0	583	99.0
	6 to 79	**3026**	**2907**	**96.1**	**3003**	**99.2**

Appendix 8e - CHMS Cycle 2 Fasted Sub-Sample Response Rates by age group and sex

		Households where one person was selected							Households where two persons were selected							
Sex	**Age Group**	**# persons selected for fasted sub-sample (SPS1)**	**# respond-ents to the question-naire (SQ1)**	**HHLD quesion-naire response rate (SSQ1)**	**# partici-pants at the MEC (SC1)**	**MEC response rate (SSC1)**	**# respond-ents who actually fasted (SB1)**	**Fasted response rate (SSB1)**	**# persons selected for fasted sub-sample (SPS2)**	**# respond-ents to the question-naire (SQ2)**	**HHLD quesion-naire response rate (SSQ2)**	**# partici-pants at the MEC (SC2)**	**MEC response rate (SSC2)**	**# respond-ents who actually fasted (SB2)**	**Fasted response rate (SSB2)**	**Overall Combined Response Rate**
Both Sexes	3 to 5 *								1	1	100.0	1	100.0	0	0.0	0.0
	6 to 11								742	696	93.8	585	84.1	466	79.7	47.7
	12 to 19	493	435	88.2	366	84.1	316	86.3	288	265	92.0	229	86.4	184	80.3	48.6
	20 to 39	487	414	85.0	330	79.7	296	89.7	568	505	88.9	414	82.0	354	85.5	46.6
	40 to 59	586	524	89.4	423	80.7	389	92.0	351	311	88.6	263	84.6	238	90.5	50.6
	60 to 79	841	773	91.9	597	77.2	537	89.9	23	20	87.0	16	80.0	13	81.3	48.4
	3 to 79	**2407**	**2146**	**89.2**	**1716**	**80.0**	**1538**	**89.6**	**1973**	**1798**	**91.1**	**1508**	**83.9**	**1255**	**83.2**	**48.4**
	6 to 79	**2407**	**2146**	**89.2**	**1716**	**80.0**	**1538**	**89.6**	**1972**	**1797**	**91.1**	**1507**	**83.9**	**1255**	**83.3**	**48.4**
Males	6 to 11								381	356	93.4	299	84.0	240	80.3	47.8
	12 to 19	260	223	85.8	184	82.5	161	87.5	145	135	93.1	117	86.7	94	80.3	47.5
	20 to 39	251	207	82.5	163	78.7	144	88.3	203	174	85.7	140	80.5	118	84.3	43.7
	40 to 59	295	260	88.1	208	80.0	188	90.4	198	164	82.8	137	83.5	126	92.0	48.3
	60 to 79	384	353	91.9	285	80.7	260	91.2	10	9	90.0	5	55.6	5	100.0	51.2
	6 to 79	**1190**	**1043**	**87.6**	**840**	**80.5**	**753**	**89.6**	**937**	**838**	**89.4**	**698**	**83.3**	**583**	**83.5**	**47.8**
Females	6 to 11								361	340	94.2	286	84.1	226	79.0	47.5
	12 to 19	233	212	91.0	182	85.8	155	85.2	143	130	90.9	112	86.2	90	80.4	49.8
	20 to 39	236	207	87.7	167	80.7	152	91.0	365	331	90.7	274	82.8	236	86.1	49.0
	40 to 59	291	264	90.7	215	81.4	201	93.5	153	147	96.1	126	85.7	112	88.9	53.1
	60 to 79	457	420	91.9	312	74.3	277	88.8	13	11	84.6	11	100.0	8	72.7	46.0
	6 to 79	**1217**	**1103**	**90.6**	**876**	**79.4**	**785**	**89.6**	**1035**	**959**	**92.7**	**809**	**84.4**	**672**	**83.1**	**49.0**

***** 3 to 5 year olds were excluded from the fasted sub-sample, however a 3-5 year old could have been selected for the sub-sample if their age was mistakenly listed as at least 6 at the time of the household interview

Appendix 8e - CHMS Cycle 2 Fasted Sub-Sample Response Rates by age group and sex, continued

		Response rate not adjusted for 1 or 2 persons selected						
Sex	**Age Group**	**# persons selected for fasted sub-sample**	**# respondents to the questionnaire**	**HHLD questionnaire response rate**	**# participants at the MEC**	**MEC response rate**	**# respondents who actually fasted**	**Fasted response rate**
Both Sexes	3 to 5 *	1	1	100.0	1	100.0	0	0.0
	6 to 11	742	696	93.8	585	84.1	466	79.7
	12 to 19	781	700	89.6	595	85.0	500	84.0
	20 to 39	1055	919	87.1	744	81.0	650	87.4
	40 to 59	937	835	89.1	686	82.2	627	91.4
	60 to 79	864	793	91.8	613	77.3	550	89.7
	3 to 79	**4380**	**3944**	**90.0**	**3224**	**81.7**	**2793**	**86.6**
	6 to 79	**4379**	**3943**	**90.0**	**3223**	**81.7**	**2793**	**86.7**
Males	6 to 11	381	356	93.4	299	84.0	240	80.3
	12 to 19	405	358	88.4	301	84.1	255	84.7
	20 to 39	454	381	83.9	303	79.5	262	86.5
	40 to 59	493	424	86.0	345	81.4	314	91.0
	60 to 79	394	362	91.9	290	80.1	265	91.4
	6 to 79	**2127**	**1881**	**88.4**	**1538**	**81.8**	**1336**	**86.9**
Females	6 to 11	361	340	94.2	286	84.1	226	79.0
	12 to 19	376	342	91.0	294	86.0	245	83.3
	20 to 39	601	538	89.5	441	82.0	388	88.0
	40 to 59	444	411	92.6	341	83.0	313	91.8
	60 to 79	470	431	91.7	323	74.9	285	88.2
	6 to 79	**2252**	**2062**	**91.6**	**1685**	**81.7**	**1457**	**86.5**

***** 3 to 5 year olds were excluded from the fasted sub-sample, however a 3-5 year old could have been selected for the sub-sample if their age was mistakenly listed as at least 6 at the time of the household interview

Appendix 9 - Activity Monitor Research

In cycle 2 of the CHMS, activity monitor data was collected in 60 second epochs for all respondents. Emerging research has demonstrated that epochs shorter than 60 seconds may more accurately capture the sporadic and intermittent physical activity that is typical in preschool aged children. This led the CHMS to collect activity monitor data in 15 second epochs for 3 to 5 year olds in cycle 3. Work is underway to investigate how 15 and 60 second epoch data relate to each other in this age group to facilitate between-cycle comparisons for the survey. This information will be made public via a methodological publication (projected completion mid- 2013). This work is important because the current intensity cut-points available for the Actical for this age group were developed specifically for use on data collected in 15 second epochs. For cycle 2 of the CHMS, the 15 second cut-points were multiplied by 4 to work with 60 second epoch data; however, researchers have expressed concern with this approach (Pfeiffer, McIver et al., 2006; Kim, Beets et al., 2012). When the cycle 3 data are published, the cross-over methodological work will help analysts interpret the impact of the methodological change from 60 to 15 second epochs on the physical activity outcomes.

Further research is also needed to understand how to assess the new preschool physical activity guidelines using accelerometer data. The new guidelines state that "preschoolers should accumulate at least 180 minutes of physical activity at any intensity spread throughout the day" (Tremblay, Leblanc et al., 2012). Intensity cut-points have been published for moderate and vigorous physical activity in this age group for the Actical accelerometer; however, more work is needed to understand what the bottom cut-off should be (i.e. the transition between sedentary and light). Another gap is the lack of a daily step count target for this age group. Based on CHMS data from cycle 1, a daily step target of 12,000 steps was published in 2012 for children and youth aged 6 to 19 (Colley, Janssen et al., MSSE 2012). Similar analyses need to be done for 3 to 5 year old children to find an appropriate daily step target.

References (Activity Monitor Research) **:**

Colley RC, Janssen I, Tremblay MS. Daily step count target to measures adherence to physical activity guidelines in children. Med Sci Sports Exerc 2012; 44(5): 977-982.

Kim Y, Beets MW, Pate RR, Blair SN. The effect of reintegrating Actigraph accelerometer counts in preschool children: Comparison using different epoch lengths. J Sci Med Sport 2012. Jun 30. [Epub ahead of print]

Pfeiffer KA, McIver KL, Dowda M, Almeida MJ, Pate RR. Validation and calibration of the Actical accelerometer in preschool children. Med Sci Sports Exerc. 2006; 38(1):152-157.

Tremblay MS, Leblanc AG, Carson V, Choquette L, Connor Gorber S, Dillman C et al. Canadian physical activity guidelines for the early years (aged 0-4 years). Appl Physiol Nutr Metab 2012; 37(2):345-356.

Appendix 10 – Changes to Wave 1 Variables

The following changes have been made between the wave 1 and wave 2 master files.
Only the changes to the WGT_FULL variable also apply to the medication file. With the exception of WGT_FULL, which appears first in the list below, all other changes to variables are listed below in the order in which they appear on the master file. Items #23 to #26, at the end of the list, describe changes that apply to variables from several survey modules.

As a result of these changes, the wave 2 files (master and medication) should be used instead of the wave 1 versions.

1. WGT_FULL: The full-sample weight variable has been adjusted as a result of corrections that were applied to the age of several respondents. (When the age reported during the household interview did not match the age confirmed at the clinic, taking the respondent's birthday into account, the age confirmed at the clinic has been accepted as the true age of the respondent.) The corresponding bootstrap weight variables (BSW1-BSW500) on the WGT_FULL file have also been adjusted.

2. GEN_13, GEN_14, GEN_18, GEN_19, and GENDMHI: Respondents less than 12 years of age (1688 cases) were incorrectly assigned a code of "Not stated". These cases have all been recoded to "Not applicable" since children less than 12 years of age were not eligible for these questions.

3. SLP_14 and SLP_15: Respondents less than 6 years of age (612 cases) were incorrectly assigned a code of "Not stated". These cases have all been recoded to "Not applicable" since children less than 6 years of age were not eligible for these questions.

4. HWT_4: Respondents less than 12 years of age (1688 cases) were incorrectly assigned a code of "Not stated". These cases have all been recoded to "Not applicable" since children less than 12 years of age were not eligible for this question.

5. WSD_23: There were 129 cases where the data were incorrectly coded to "Not applicable", a code which should not exist for this question (all respondents are eligible). The data have been recovered for 126 cases, and recoded to "Not stated" for the other 3 cases.

6. WSD_24E and WSD_24F: There were 3261 cases where the data were incorrectly coded to "Not applicable". The data have been recovered for these cases.

7. WSD_24H: There were 3 cases where this question was incorrectly coded to "No". The data have been recovered for these cases.

8. CPA_13 and CPA_14: Respondents less than 6 years of age (612 cases) were incorrectly assigned a code of "Not stated". These cases have been recoded to "Not applicable", since children less than 6 years of age were not eligible for these questions.

9. HEP_13B: Respondents whose age was equal to the year of interview minus 1990 were not eligible for this question; however, they were incorrectly assigned a code of "Not stated" (39 cases). These cases have all been recoded to "Not applicable".

10. HPV_13: Respondents with HPV_12 = 2 were not asked question HPV_13; however, these cases were incorrectly assigned a non-response code of "Not stated" (1239 cases). These cases have all been recoded to "Not applicable".

11. SMK_52 and SMK_53: There were 304 cases where data for daily smokers was incorrectly coded to "Not applicable". The data have been recovered for 301 of these cases, and coded to "Not stated" for the other 3 cases.

12. ETS_12BA to ETS_12BD, ETS_12C, ETS_12D, ETS_12E and ETS_12F: These are household-level questions that were designed to be asked only to the youngest respondent in households where 2 respondents had been selected. There were 135 cases where the data were not properly copied from the record for the youngest respondent to the record for the other respondent. The data for these cases have now been copied.

13. HSC_12B: This is a household-level question that was designed to be asked only to the youngest respondent in households where 2 respondents had been selected. There were 1,653 cases where the data was not properly copied from the record for the youngest respondent to the record for the other respondent. The data for these cases have now been copied.

14. EDU_01 and EDU_02: Education questions were not asked for respondents under the age of 15. These cases were automatically coded as EDU_01 = 1 (currently attending a school, college or university) and EDU_02 = 1 (full-time student). This approach was also used in cycle 1; however, for cycle 1 the minimum respondent age was 6 as opposed to 3 for cycle 2. These questions have been recoded to "Not applicable" for respondents less than 6 years of age (612 respondents).

15. INC_21: An inconsistency was discovered between some of the imputed values and self-reported household income range information reported in INC_22 to INC_28 when only partial range information had been provided by the respondent. As a result 87 cases were re-imputed, resulting in a modified imputed household income for 26 cases. See Section 12.1.2 of the User Guide for more information on imputation of household income.

16. INCDHH, INCDDIA2, INCDDIA4, INCDDIA5: These variables have been rederived based on the new value of INC_21 (affected 22, 7, 17 and 17 cases respectively).

17. INCFIMP4: The imputation flag INCFIMP has been replaced with a new flag, INCFIMP4. Additional categories have been added to provide more detailed information regarding the method used to impute INC_21.

18. INC_32, INC_33, INC_34, INC_35, INC_36, INC_37 and INC_38: There were 5 cases where these variables were incorrectly assigned a non-response code of "Not applicable". These cases have all been recoded to "Not stated".

19. INCDPER: As a result of the incorrect code for INC_32 to INC_38, there were 5 cases where INCDPER was incorrectly assigned a result of "Zero income". These cases have been recoded to "Not stated".

20. PHC_37: There were 343 cases where the data were incorrectly coded to a non-response ("Not stated" or "Not applicable") code. The data have been recovered or recoded for all cases.

21. IAS_11, IAS_12, IAS_21, IAS_22, IAQD04 and IAQ_11to IAQ_21: These are questions that were designed to be asked to only one respondent per household. The data for these questions were not properly copied to the second respondent's record in two-respondent households. The data for these cases have now been copied.

22. BPM_160, BPM_200 to BPMD263: Data for the second blood pressure result was missing for 73 cases. The data have been recovered for these cases.

23. There were 3 cases in two-respondent households where only the selected respondent aged 12 or older completed the household interview. As a result, household-level questions, which were designed to be asked only to the youngest selected respondent in a household, were never asked. These variables were incorrectly assigned a code of "Not applicable" for the respondent aged 12 or older. These cases have all been recoded to "Not stated". This affects the following variables: SLT_13, ETS_11, ETS_12, ETS_12BA, ETS_12BB, ETS_12BC, ETS_12BD, ETS_12C, ETS_12D, ETS_12E, ETS_12F, WSD_23, WSD_24A to WSD_24H, all FMH variables, HSC_11, HSC_12B to HSC_44.

24. In households where two respondents are selected for the survey, one is a child under the age of 12, and the other 12 years of age or older. There was one case where, following confirmation of the age at the clinic, both respondents selected in the household were found to be 12 years of age or older. As a result, some of the household-level data that had only been asked to the youngest respondent were incorrectly wiped out on both records. These data have been recovered. This affects the following variables: WSD_23, WSD_24 A to H, SLT_13, ETS_11 and HSC_11 to HSC_41.

25. There was one case where, following confirmation of the age at the clinic, both respondents selected in the household were found to be under the age of 12. This resulted in a number of issues that have now been corrected. On the record where the age was changed, the following corrections have been made: PROXY was recoded to 1; data was recovered for HSC_12B; GEN_13, GEN_14, GENDMHI, GEN_16 to GEN_19, HWT_4 and ETS_12BA to ETS_13 have been recoded to "Not applicable". On both records, data have been recovered for: SLT_13, WSD_23 to WSD_24H, all questions in FMH, ETS_11, ETS_12, HSC_11, HSC_13A to G and HSC_31 to HSC_44.

26. There was 1 case where a child was recorded as being 6 years old during the household interview, but later confirmed to be 5 years old at the clinic. Several variables on this record had already been modified on the first version of the master file; however the variables SLP_14, CPA_13 and CPA_14, which are not applicable to a 5-year old, have now also been recoded to "Not applicable".