Microdata User Guide

Households and the Environment Survey

2007

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction **5**

2.0 Background **7**

3.0 Objectives **9**

4.0 Concepts and Definitions **11**
4.1 Canadian Community Health Survey Concepts and Definitions **11**
4.2 Households and the Environment Survey Concepts and Definitions **11**

5.0 Survey Methodology **17**
5.1 Canadian Community Health Survey Population Coverage **17**
5.2 Canadian Community Health Survey Sample Design **17**
5.3 Sample Size by Province for the Households and the Environment Survey **18**

6.0 Data Collection **19**

7.0 Data Processing **21**
7.1 Data Capture **21**
7.2 Editing **21**
7.3 Coding of Open-ended Questions **22**
7.4 Creation of Derived Variables **22**
7.5 Weighting **23**
7.6 Suppression of Confidential Information **23**

8.0 Data Quality **25**
8.1 Response Rates **25**
8.2 Survey Errors **25**
8.2.1 The Frame **26**
8.2.2 Data Collection **26**
8.2.3 Data Processing **26**
8.2.4 Non-response **26**
8.2.5 Measurement of Sampling Error **27**

9.0 Guidelines for Tabulation, Analysis and Release **29**
9.1 Rounding Guidelines **29**
9.2 Sample Weighting Guidelines for Tabulation **29**
9.3 Definitions of Types of Estimates: Categorical and Quantitative **30**
9.3.1 Categorical Estimates **30**
9.3.2 Quantitative Estimates **30**
9.3.3 Tabulation of Categorical Estimates **31**
9.3.4 Tabulation of Quantitative Estimates **31**
9.4 Guidelines for Statistical Analysis **31**
9.5 Coefficient of Variation Release Guidelines **32**
9.6 Release Cut-off's for the Households and the Environment Survey **34**

10.0 Approximate Sampling Variability Tables 35
10.1 How to Use the Coefficient of Variation Tables for Categorical Estimates **36**
10.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates **37**
10.2 How to Use the Coefficient of Variation Tables to Obtain Confidence Limits **41**
10.2.1 Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits **42**
10.3 How to Use the Coefficient of Variation Tables to Do a T-test **43**
10.3.1 Example of Using the Coefficient of Variation Tables to Do a T-test **43**
10.4 Coefficients of Variation for Quantitative Estimates **43**
10.5 Coefficient of Variation Tables **44**

11.0 Weighting 45
11.1 Weighting Procedures for the Canadian Community Health Survey **45**
11.2 Weighting Procedures for the Households and the Environment Survey **47**

12.0 Questionnaire 49

13.0 Record Layout with Univariate Frequencies 51

1.0 Introduction

The Households and the Environment Survey (HES) was conducted by Statistics Canada from October 2007 to February 2008. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

Any question about the data set or its use should be directed to:

Statistics Canada

Environment Accounts and Statistics Division
Telephone: 613-951-0297
Fax: 613-951-0634
E-mail: environ@statcan.ca

Client Services
Special Surveys Division
Telephone: 613-951-3321 or call toll-free 1-800-461-9050
Fax: 613-951-4527
E-mail: ssd@statcan.ca

2.0 Background

The Households and the Environment Survey (HES) was conducted from October 2007 through February 2008 as a supplement to the Canadian Community Health Survey. The survey was designed to specifically address the needs of its funding source the Canadian Environmental Sustainability Indicators (CESI) project, a joint venture between Statistics Canada, Environment Canada and Health Canada. The CESI project reports annually on air quality, water quality and greenhouse gas (GHG) emissions in Canada using indicators to identify areas of importance to Canadians and monitor progress.

The HES was first conducted in 1991, 1994 and more recently in 2006. The 2007 survey offers an expanded view on household behaviours that relate to the environment but allows for comparisons with the 1994 survey for some indicators and most of the indicators from the 2006 survey.

The target population consisted of households in Canada, excluding households located in the Yukon, Northwest Territories and Nunavut, households located on Indian reserves or Crown lands, and households consisting entirely of full-time members of the Canadian Armed Forces. Institutions and households of certain remote regions were also excluded.

3.0 Objectives

The objective of the Households and the Environment Survey (HES) is to measure the behaviours and practices of households that relate to the environment in terms of their impact on the quality of the air, water and soils as well as contributions to greenhouse gas emissions. Specifically, the following topics were addressed in the 2007 HES:

- Energy use and home heating and cooling
- Consumption and conservation of water
- Water quality concerns of households
- Pesticide and fertilizer use on lawns and gardens
- Recycling, composting and waste disposal practices
- Indoor environment
- Recreational vehicles and gasoline powered equipment
- Transportation decisions, commuting and motor vehicle use
- Impacts of air and water quality on households
- Purchasing decisions

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. The concepts and definitions used in the Canadian Community Health Survey (CCHS) are described in Section 4.1 while those specific to the Households and the Environment Survey (HES) are given in Section 4.2. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaire(s) used.

4.1 Canadian Community Health Survey Concepts and Definitions

Dwelling
A dwelling is defined as any set of living quarters that is structurally separate and has a private entrance outside the building or from a common hall or stairway inside the building.

Types of dwellings

- **Single detached** – A structure with one dwelling only, separated by open space from all other structures except its own garage or shed.
- **Double** – A dwelling joined to only one other dwelling, separated from it by a wall extending from ground to roof.
- **Row or terrace** – A dwelling unit in a row of three or more dwellings, sharing common walls extending from ground to roof in which there are no other dwellings either above or below.
- **Duplex** – Two dwellings situated one above the other, not attached to any other structure and surrounded on all sides by open space.
- **Low-rise apartment** – Dwellings within triplexes, quadruplexes, and apartment buildings of fewer than five stories.
- **High-rise apartment** – Separate dwellings within a residential structure of five or more stories.
- **Mobile homes** – A movable dwelling designed and constructed to be transported (by road) on its own chassis to a site, and placed on a temporary foundation such as block posts or a prepared pad.

Household
A household is defined as any person or a group of persons living in a dwelling. A household may consist of any combination of: one person living alone, one or more families, a group of people who are not related but who share the same dwelling.

4.2 Households and the Environment Survey Concepts and Definitions

Consumption and conservation of water
"Canadians are concerned about how the environment affects their health, thus about the quality of the water they drink".[1] Since public perception, as a determining factor driving public policy, can be as persuasive as empirically-based evidence, it is important to understand how Canadians perceive the quality of their drinking water supply and the behaviours they are exhibiting that may reflect their concerns.

These behaviours are measured in the HES through the purchases of bottled water or the use of water filters, and the reasons for making these purchases – for example, concerns about bacterial contamination. Analysis can be carried out to gain insight into the household characteristics of

[1] Sharing Environmental Decisions: Executive Summary and Recommendations. Final report of the Task Force on a Canadian Information System for the Environment. October 2001, Ottawa.

those that do and do not exhibit behaviours that may indicate uncertainties about the quality of their drinkable water.

Aside from drinking water issues, another important theme involves the water conservation practices of households. Water scarcity is an emerging issue for many Canadians and this concern may be exacerbated by climatic changes. Many regions of Canada have experienced drought or near-drought conditions which has led to regulatory responses by municipal authorities (e.g., water use restrictions) and/or the voluntary adoption of water conservation measures by households.

The HES provides reliable information on household practices such as lawn watering and the use of low-flow showerheads and low-volume toilets, which can be used to determine levels of public participation in water conservation.

Metals and minerals can include any of the following: iron, sulphur, cadmium, zinc, manganese, lead, mercury, arsenic.

Bacteria can include any of the following: E. coli, other coliforms, fecal matter, giardia, cryptosporidium, parasites, protozoa, shigellosis.

Chemicals or other pollutants can include any of the following: chlorine, bromine, pesticides, oil, gasoline, diesel fuel, heating oil, fluoride, nitrate, trichloroethylene (TCE), polycyclic aromatic hydrocarbons (PAH), fire retardants.

Holding tanks are septic tanks that do not have a weeping tile system and which must be pumped out on a regular basis.

A **communal septic system** is a private or public septic system that serves more than one household but is not a part of a municipal sewer system. These are common in places like trailer parks or neighbourhoods where there is not a high enough housing density to warrant full sewage services.

Low flow showerheads are used to regulate the flow of water.

Low volume toilets use a lower volume of water than regular toilets. Usually these toilets use 6 litres as opposed to 12 litres of water per flush.

A rain barrel is a container used to collect and store rainwater. It is usually placed below the downspout of a roof gutter. The collected water is usually used to water the landscape.

A cistern is an artificial reservoir for storing liquids; especially an underground tank for storing rainwater.

Energy use and home heating

The choices regarding what kind of energy a household uses to heat a home, whether the temperature is regulated and if any energy efficient electrical devices are used within the home are all decisions that affect the household's contributions to greenhouse gas (GHG) emissions.

Indirectly, the decisions also indicate to what degree the household has bought into the concept and need for energy conservation. The HES measures not only the types of energy used but those behaviours that indicate whether Canadian households are behaving in a sustainable way regarding energy use.

A **forced air natural gas furnace** is a heating system using a system of ducts and vents to circulate air heated by the combustion of natural gas.

A **forced air oil furnace** is a heating system using a system of ducts and vents to circulate air that has been heated by the combustion of oil.

A **forced air electric furnace** is a heating system using a system of ducts and vents to circulate air that has been heated by electrical current.

A **forced air hot water system** is a heating system using a system of ducts and vents to circulate air that has been heated by hot water.

Hot water radiators are metal structures or pieces of equipment used to heat a room by emitting heat from hot water or steam that circulates through it.

Electric baseboards are heating systems attached to the wall near the floor where elements heat up through use of electrical current. Electric baseboards are usually controlled by individual thermostats on each unit or in some cases one thermostat per room.

Other electric heating is heat produced through electrical current that is delivered through an appliance or other means excluding forced air or baseboards.

A **heating stove** typically uses firewood or wood pellets as a fuel source. A chimney is used for ventilation of smoke and excess heat.

A **central air conditioning system** is part of the home's central heating system and distributes cool air through the home's ductwork as opposed to stand alone air conditioning units that are usually seen in windows and are used to cool a specific part of the home.

Lawn care equipment

Many pieces of lawn care equipment, such as lawnmowers, use internal combustion engines, as do many boats and snowmobiles. Internal combustion engines emit greenhouse gases (GHGs), which are believed to be a main contributor to climate change, and fine particulate matter, which adversely affects air quality. By measuring the extent of the use of these devices, we are establishing a baseline that can be compared to future data to see if the use of these devices are increasing or decreasing.

A **grass trimmer (weed eater)** is a device that trims grass and weeds through the use of a rapidly spinning plastic cord or circular saw.

A **leaf blower** is a device that emits a strong air current that is intended to blow leaves off of one's lawn.

Pesticide and fertilizer use

The usage of fertilizers and pesticides by households is measured in the 2007 HES. By analysing these data with selected household characteristics, policymakers can then use this information to better inform targeted public awareness and information campaigns.

Chemical fertilizers are chemicals given to plants with the intention of promoting growth. They are usually applied either directly to the soil or by spraying.

Herbicides, insecticides, and fungicides are a substance or mixture of substances intended for preventing, destroying, repelling, or mitigating any weeds, insects or fungi, respectively.

Recycling, composting and waste disposal practices
Disposal and recycling of post-consumer waste continues to be an important environmental issue, especially in urban centers. Canadian households generate approximately one third of solid waste by weight. As we move into a "knowledge-based" economy, the amount of "high-tech" waste has increased and represents new waste management problems. In addition, there is a need for better data on household practices regarding composting, participation in product stewardship programs, knowledge of and participation in hazardous waste disposal programs.

Participation in recycling initiatives
Governments are finding it difficult to gauge the level of participation in various waste diversion initiatives concerning certain materials such as used paint, expired medication and used batteries. Scores of such initiatives are underway across Canada and there is a perception that the level of success for these programs is high. As in 1991, 1994, and 2006, the 2007 HES asked questions on access to and use of recycling programs.

Household computer and technological waste
The need for better data on household practices regarding how households deal with their technological waste is an issue that arose through contacts with provincial officials and has been on the agenda at the Federal level.[2] Concerns about the levels of toxic substances (e.g., lead, mercury and dioxins) that can be found in these wastes are heightening as policymakers examine methods to mitigate the impact of these materials on the environment. The HES measures whether households are disposing of technological waste in proper disposal facilities or whether they are including it in the "regular" garbage or are perhaps simply unaware of what to do with these materials and are storing them in basements and garages.

Composting practices
Usage by households of backyard composters and/or curb side organic pick-up containers represents another data gap in waste statistics. There is a high level of interest from all levels of government and NGOs (non-government organizations) regarding household usage of backyard composters and the socio-economic characteristics of users and non-users of backyard composters and centralized composting programs

A **recycling program** can be either:

- A collection system for recyclable materials such as paper, plastics, metals, and glass. This system can be either municipally or privately operated.
- A system whereby residents and/or businesses take their recyclable materials to a central depot or drop-off centre.

Composting involves the separation of kitchen waste (includes food scraps, coffee grinds, eggshells, etc.) and/or yard waste (includes leaves, plants or grass clippings) from the rest of your household garbage. The separated materials can be:

- Put in a compost bin, compost pile or garden
- Picked up by the city, town, municipality or a private company; or
- Taken to a depot or drop off centre

Air and water quality
The quality of the air and the bodies of water (lakes, rivers) used for recreational purposes may influence how Canadians behave in their day to day activities. A failure to alter our behaviour can impact the quality of life and health for all Canadians. The HES measures, for example, how smog advisories or swimming restrictions have influenced people's activities.

Smog is the most visible form of air pollution. It is a brownish-yellow haze caused when heat

[2] Information Technology (IT) and Telecommunication Waste in Canada. EnvirosRIS. Prepared for Environment Canada, National Office of Pollution Prevention. October 2000, Ottawa.

and sunlight react with various pollutants in the air. Smog is a year-round problem but most smog watches and alerts occur from April to September, especially on hot days.

An **air quality advisory** is an advisory that is issued for smoke, smog or poor air quality. Alerts are based on the air quality index.

Transportation decisions

Emissions from automobiles are a significant component of greenhouse gas emissions in Canada, and thus of climate change. Policies that seek to educate the public about the impact of their commuting choices as well as policies around public transportation issues require reliable information about the behaviours and motivations of Canadians regarding their commuting choices.

Ethanol blended fuel is gasoline blended with ethanol. Ethanol is an alcohol produced from the starch portion of corn. This gasoline is usually marketed as more “environmentally friendly” than other types of gasoline.

5.0 Survey Methodology

The Households and the Environment Survey (HES) was administered from October 2007 to February 2008 to a sub-sample of the dwellings that were part of the Canadian Community Health Survey (CCHS) Cycle 4.1 between January 1st and June 30th, 2007. Therefore its sample design is closely tied to that of the CCHS. The CCHS design is briefly described in Sections 5.1 and 5.2.[3]

5.1 Canadian Community Health Survey Population Coverage

The CCHS data is collected from people aged 12 years and over living in private dwellings within the 10 provinces and three territories. Specifically excluded from the survey's coverage are residents of Indian Reserves and Crown land, full-time members of the Canadian Armed Forces, inmates of institutions and residents of isolated areas. The CCHS represents approximately 98% of the Canadian population aged 12 years and over.

5.2 Canadian Community Health Survey Sample Design

To provide reliable estimates to the 121 Health Regions (HR), and given the budget allocated to the CCHS (Cycle 4.1), a sample of 130,000 respondents was desired. The sample allocation strategy consisting of three steps, gave relatively equal importance to the HRs and the provinces. In the first two steps, the sample was allocated among the provinces according to their respective populations and the number of HRs they contain. In the third step, each province's sample was allocated among its HRs proportionally to the square root of the estimated population in each HR.

The CCHS used three sampling frames to select the sample of households: 50% of the sample of households came from an area frame, 49% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialling (RDD) sampling frame. For most of the health regions, 50% of the sample was selected from the area frame and 50% from the list frame of telephone numbers.

The CCHS used the area frame designed for the Canadian Labour Force Survey (LFS) as its primary frame. The sampling plan of the LFS is a multistage stratified cluster design in which the dwelling is the final sampling unit. In the first stage, homogeneous strata were formed and independent samples of clusters were drawn from each stratum. In the second stage, dwelling lists were prepared for each cluster and dwellings, or households, were selected from the lists.

For the purpose of the plan, each province is divided into three types of regions: major urban centres, cities and rural regions. Geographic or socio-economic strata are created within each major urban centre. Within the strata, dwellings are regrouped to create clusters. Some urban centres have separate strata for apartments or for census enumeration areas (EA) in which the average household income is high. In each stratum, six clusters or residential buildings (sometimes 12 or 18 apartments) are chosen by a random sampling method with a probability proportional to size (PPS), the size of which corresponds to the number of households. The number six was used throughout the sample design to allow a one-sixth rotation of the sample every month for the LFS.

The other cities and rural regions of each province are stratified first on a geographical basis, then according to socio-economic characteristics. In the majority of strata, six clusters (usually census EAs) are selected using the PPS method. Where there is low population density, a three-step plan is used whereby two or three primary sampling units (PSU), which normally correspond

[3] For a detailed description of the CCHS Cycle 4.1 sample design see the Public Use Microdata File guide, Catalogue no. 82M0013GPE.

to groups of EAs, are selected and divided into clusters, six of which are sampled. The final sample is obtained using a systematic sampling of dwellings.

5.3 Sample Size by Province for the Households and the Environment Survey

The following table shows the number of dwellings that were selected for the 2007 HES. This table excludes dwellings which were non-respondents to the CCHS.

Province	Sample Size - Number of dwellings
Newfoundland and Labrador	870
Prince Edward Island	609
Nova Scotia	1,106
New Brunswick	1,227
Quebec	6,198
Ontario	10,173
Manitoba	1,752
Saskatchewan	1,586
Alberta	2,642
British Columbia	3,794
Canada	**29,957**

6.0 Data Collection

An introductory letter was mailed to respondents approximately one week before data collection began. Collection for the Households and the Environment Survey (HES) was carried out from October 2007 to the end of February 2008 using a computer-assisted telephone interviewing (CATI) system.

The CATI system has a number of generic modules which can be quickly adapted to most types of surveys. A front-end module contains a set of standard response codes for dealing with all possible call outcomes, as well as the associated scripts to be read by the interviewers. A standard approach set up for introducing the agency, the name and purpose of the survey, the survey sponsors, how the survey results will be used, and the duration of the interview was used. We explained to respondents how they were selected for the survey, that their participation in the survey is voluntary, and that their information will remain strictly confidential. Help screens were provided to the interviewers to assist them in answering questions that are commonly asked by respondents.

The CATI application ensured that only valid question responses were entered and that all the correct flows were followed. Edits were built into the application to check the consistency of responses, identify and correct outliers, and to control who gets asked specific questions. This meant that the data was already quite "clean" at the end of the collection process.

The survey manager met with senior staff responsible for collection to discuss issues and questions before the start of the training session. A description of the background and objectives as well as a detailed description of the concepts and definitions particular to the 2007 HES was provided for interviewers in their Interviewer Manual. A glossary of terms and a set of questions and answers were also included.

Interviewers were trained on the survey content through a classroom training session. In addition, the interviewers completed a series of mock interviews to become familiar with the survey, its concepts, definitions and the CATI application itself. Question and answer documentation was provided to the interviewers to assist them in answering questions that are commonly asked by respondents.

The data collection was conducted by specialized staff at Statistics Canada offices in Edmonton, Sturgeon Falls, Toronto, Halifax, Winnipeg and Sherbrooke. The workload and interviewing staff within each office was managed by a project manager. The automated scheduler used by the CATI system ensured that cases were assigned randomly to interviewers and that cases were called at different times of the day and different days of the week to maximize the probability of contact. There were a maximum of 25 call attempts per case; once the maximum was reached, the case was reviewed by a senior interviewer who determined if additional calls would be made.

The average interview time was estimated to be 20 minutes. However, the length of the interviews varied depending on the circumstances of the responding households. For example, the average interview time was slightly higher for respondents residing in a single family household as opposed to respondents living in an apartment.

The team of interviewers was under the supervision of senior interviewers responsible for ensuring that everyone was familiar with the concepts and procedures of the survey. Periodical monitoring of interviewers and the review of completed documents was done in accordance with collection protocol.

7.0 Data Processing

The main output of the Households and the Environment Survey (HES) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

The microdata file contains data from the following sections:

- HH Household demographics
- DC Dwelling characteristics
- EH Energy use and home heating
- WA Water
- FP Fertilizer and pesticide use
- RC Recycling
- CP Composting
- IE Indoor environment
- GP Recreational vehicles/Gasoline powered equipment
- TD Transportation decisions
- MV Motor vehicles
- TW Transportation to work
- AQ Air quality
- PD Purchasing decisions
- HD Income

7.1 Data Capture

Responses to survey questions are captured directly by the interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission. The response data are encrypted to ensure confidentiality and sent via modem to the appropriate Statistics Canada Regional Office. From there they are transmitted over a secure line to Ottawa for further processing.

Some editing is done directly at the time of the interview. Where the information entered is out of range (too large or small) of expected values, or inconsistent with the previous entries, the interviewer is prompted, through message screens on the computer, to modify the information. However, for some questions, interviewers have the option of bypassing the edits and of skipping questions if the respondent does not know the answer or refuses to answer. Therefore, the response data are subjected to further edits once they arrive in head office.

7.2 Editing

The first stage of survey processing undertaken at head office was the replacement of any "out-of-range" values on the data file with blanks. This process was designed to make further editing easier.

The first type of error treated was errors in questionnaire flow, where questions that did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

The second type of error treated involved a lack of information in questions that should have been answered. For this type of error, a non-response or "not-stated" code was assigned to the item.

This was followed by a series of edits to ensure consistency in the responses for a household.

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. These questions required coding for inclusion on the HES data file.

The second type of coding performed was for questions which allow for numeric values to be entered. These numeric values were first reviewed for outliers and then grouped into ranges. An example of a question which allows for numeric values would be total household income from all sources.

7.4 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. This was done by using one variable or a combination of variables. The following is a list of the derived variables for the HES.

HES Derived Variables

WAD04	Indication if any treatment is applied to the drinking water in the household
WAD11	Indication if any maintenance is done to the septic system
WADREDUC	Indication if any of the noted devices are used by the household to conserve or reduce consumption of water
EHD09	During the winter season, at what temperature is the dwelling usually kept when you are there and awake? (Celsius)
EHD10	During the winter season, at what temperature is the dwelling usually kept when you are asleep? (Celsius)
EHD11	During the summer season, at what temperature is the dwelling usually kept when you are there and awake? (Celsius)
EHD12	During the summer season, at what temperature is the dwelling usually kept when you are asleep? (Celsius)
EHD13	During the summer season, at what temperature is the dwelling usually kept when you are not at home? (Celsius)
GPD03	Quantity of fuel used
GPDEQUIP	Usage of snow blower, lawn mower, weed eater, or leaf blower
RCDPROGR	Usage of any recycling program (paper, plastics, glass or metal cans and containers)
CMA	Census Metropolitan Area (CMA) - 2006 Census Code
HHEDUCLV	Highest level of education ever completed by any member of the household
HHAG0005	Number of persons aged 0 to 5 in the household
HHAG0612	Number of persons aged 6 to 12 in the household
HHAG1315	Number of persons aged 13 to 15 in the household
HHAG1617	Number of persons aged 16 to 17 in the household
HHAG1819	Number of persons aged 18 to 19 in the household
HHAG2024	Number of persons aged 20 to 24 in the household
HHAG2534	Number of persons aged 25 to 34 in the household
HHAG3544	Number of persons aged 35 to 44 in the household

HHAG4554	Number of persons aged 45 to 54 in the household
HHAG5564	Number of persons aged 55 to 64 in the household
HHAG65PL	Number of persons aged 65 and over in the household
HHSIZE	Number of people in the household
HHTYPE	Type of household, based on age composition
HDD02	Household income
MV1D06	Total kilometres 1st vehicle driven
MV2D06	Total kilometres 2nd vehicle driven
MV3D06	Total kilometres 3rd vehicle driven
MV4D06	Total kilometres 4th vehicle driven
TWDWORK	Total number of people who worked (self-employed or employee) outside the home in the last 12 months

7.5 Weighting

The principle behind estimation in a probability sample such as the HES is that each unit in the sample "represents", besides itself, several other units not in the sample. For example, in a simple random 2% sample of the population, each unit in the sample represents 50 units in the population.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example if the number of households who treat their drinking water is to be estimated, it is done by selecting the records referring to those households in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 11.0.

7.6 Suppression of Confidential Information

It should be noted that the "Public Use" Microdata Files (PUMF) may differ from the survey "master" files held by Statistics Canada. These differences usually are the result of actions taken to protect the anonymity of individual survey respondents. The most common actions are the suppression of file variables, grouping values into wider categories, and coding specific values into the "not stated" category. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 9.0 of this document.

The 2007 HES Public Use Microdata File will not be available until the fall of 2009.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Canadian Community Health Survey (CCHS) and to the Households and the Environment Survey (HES).

Province	CCHS Selected Households	CCHS Response Rate (%) *	HES Selected Households	HES Responding Households	HES Response Rate (%)**
Newfoundland and Labrador	1,167	89.1	870	639	73.4
Prince Edward Island	839	86.4	609	431	70.8
Nova Scotia	1,481	86.7	1,106	802	72.5
New Brunswick	1,634	85.6	1,227	831	67.7
Quebec	7,749	84.0	6,198	4,597	74.2
Ontario	14,411	83.5	10,173	7,271	71.5
Manitoba	2,165	88.3	1,752	1,313	74.9
Saskatchewan	2,237	89.3	1,586	1,181	74.5
Alberta	3,709	85.4	2,642	1,900	71.9
British Columbia	5,192	82.1	3,794	2,725	71.8
Canada	**40,584**	**84.6**	**29,957**	**21,690**	**72.4**

* The CCHS response rate is the number of CCHS responding households as a percentage of the number of CCHS selected households for the January 1st to June 30th, 2007 period.

** The HES response rate is the number of HES responding households as a percentage of the number of HES selected households.

8.2 Survey Errors

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

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

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort were taken to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures include the use of highly

skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized, and coding and edit quality checks to verify the processing logic.

8.2.1 The Frame

Because the 2007 HES was a supplement to the CCHS Cycle 4.1, the frame used was the CCHS frame. The CCHS frame was based on both the Labour Force Survey (LFS) area frame and a telephone frame including a random digit dialling component. The CCHS survey coverage was very good (98% of all households in Canada). It is unlikely that the 2% exclusion introduces any significant bias into the survey data.

It is important to note that the HES interview took place between 3 and 14 months after the CCHS Cycle 4.1 interview.

8.2.2 Data Collection

Interviewer training consisted of reading the HES Procedures Manual and Interviewer's Manual, practicing with the HES training cases on the computer and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. Interviewers collected the HES information after the CCHS information was collected. The interviews were conducted between October 2007 and February 2008.

8.2.3 Data Processing

Data processing of the HES was done in a number of steps including verification, coding, editing, estimation, etc. At each step a "picture" of the output files is taken and an easy verification can be made comparing files at the current and previous step. This greatly improved the data processing stage.

8.2.4 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households that responded to the survey to compensate for those that did not respond.

In most cases, item non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information. Values were not imputed when these were missing. They were coded to "not-stated".

Partial non-response occurs when the interview is started but not completed for various reasons. In the case of the HES, less than 1% of interviews were started but not completed and the missed questions were treated as multiple item non-response and coded to "not-stated".

8.2.5 Measurement of Sampling Error

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

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results.

However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose that, based upon the 2006 HES results, one estimates that 34.9% of households had a lawn and used chemical fertilizers in 2005 and this estimate is found to have a standard error of 0.0051. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0051}{0.349}\right) X\ 100\,\% = 1.46\,\%$$

There is more information on the calculation of coefficients of variation in Chapter 10.0.

9.0 Guidelines for Tabulation, Analysis and Release

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

The microdata file should be used for analysis and estimation of the majority of the content of the survey where the unit of measurement was the household.

9.1 Rounding Guidelines

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

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

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

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

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

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

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Households and the Environment Survey (HES) was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users must apply the proper survey weight.

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

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

9.3 Definitions of Types of Estimates: Categorical and Quantitative

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

9.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of households that have access to a glass recycling program or the proportion of households that primarily drink bottled water are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: In the last 12 months, did you/your household have access to a municipal, provincial, or private household hazardous waste depot?
R: Yes / No

Q: During the last 12 months, what type of water did you/your household primarily use for drinking? Was it...?
R: Tap water / Bottled water, including purchased water in a water cooler, tank or other dispenser / Both / Other

9.3.2 Quantitative Estimates

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

An example of a quantitative estimate is the average number of motor vehicles owned or leased by households. The numerator is an estimate of the total number of motor vehicles owned or leased by households, and its denominator is the number of households with a vehicle.

Examples of Quantitative Questions:

Q: How many motor vehicles were owned or leased for personal use in the last 12 months?
R: |_|_| motor vehicles

Q: How many kilometres or miles was this vehicle driven in the past 12 months? Exclude the use of this vehicle in the conduct of a business or self-employment.
R: |_|_|_|_|_|_| kilometres or miles

9.3.3 Tabulation of Categorical Estimates

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

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

9.3.4 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of motor vehicles owned or leased by households, multiply the value reported in TD_Q02 (number of motor vehicles owned/leased) by the final weight for the record, then sum this value over all records with TD_Q01 = 1 (households that own or lease a motor vehicle for personal use).

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of vehicles owned or leased by households,

a) estimate the total number of motor vehicles owned or leased ($\hat{X}$) as described above,
b) estimate the number of households that owned or leased a motor vehicle ($\hat{Y}$) by summing the final weights of all records with TD_Q01 = 1, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

9.4 Guidelines for Statistical Analysis

The HES is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

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

simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

For example, suppose that analysis of all Quebec households is required. The steps to rescale the weights are as follows:

1) select all households from the file in PROV = 24, Quebec;

2) calculate the AVERAGE weight for these records by summing the original household weights from the microdata file for these records and then dividing by the number of households in PROV = 24;

3) for each of these respondents, calculate a RESCALED weight equal to the original household weight divided by the AVERAGE weight;

4) perform the analysis for these households using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the HES, users should first determine the quality level of the estimate. The quality levels are *acceptable, marginal* and *unacceptable.* Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 8.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 8.0 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to rounded weighted estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality Level Guidelines

Quality Level of Estimate	Guidelines
1) Acceptable	Estimates have a sample size of 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%. Estimates should be flagged with the letter E (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3) Unacceptable	Estimates have a sample size of less than 30, or very high coefficients of variation in excess of 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 some similar identifier) and the following warning should accompany the estimates: “Please be warned that these estimates [flagged with the letter F] do not meet Statistics Canada’s quality standards. Conclusions based on these data will be unreliable, and most likely invalid.”

9.6 Release Cut-off's for the Households and the Environment Survey

The following tables provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 10.0.

For example, the table shows that the quality of a weighted estimate of 8,000 households possessing a given characteristic in Newfoundland and Labrador is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explains the correct procedure to be used for ratios.

Table of Release Cut-offs – Household File

Province and Region	Acceptable CV 0.0% to 16.5%	Marginal CV 16.6% to 33.3%	Unacceptable CV > 33.3%
Newfoundland and Labrador	21,800 & over	5,800 to < 21,800	under 5,800
Prince Edward Island	7,300 & over	2,000 to < 7,300	under 2,000
Nova Scotia	32,300 & over	8,500 to < 32,300	under 8,500
New Brunswick	27,800 & over	7,300 to < 27,800	under 7,300
Quebec	96,200 & over	24,200 to < 96,200	under 24,200
Ontario	71,000 & over	17,600 to < 71,000	under 17,600
Manitoba	35,200 & over	9,200 to < 35,200	under 9,200
Saskatchewan	23,300 & over	6,000 to < 23,300	under 6,000
Alberta	64,300 & over	16,400 to < 64,300	under 16,400
British Columbia	60,700 & over	15,300 to < 60,700	under 15,300
Atlantic Provinces	27,900 & over	7,000 to < 27,900	under 7,000
Prairie Provinces	51,900 & over	13,000 to < 51,900	under 13,000
Canada	**68,100 & over**	**16,800 to < 68,100**	**under 16,800**

10.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation (CV) which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These CV tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value (usually the 75th percentile) to be used in the CV tables which would then apply to the entire set of characteristics.

The table below shows the conservative value of the design effects as well as sample sizes (number of responding households) and population counts by province and region, which were used to produce the Approximate Sampling Variability Tables for the Households and the Environment Survey (HES) household file.

Household File

Province and Region	Design Effect	Sample Size	Population
Newfoundland and Labrador	2.12	639	200,833
Prince Edward Island	1.80	431	54,895
Nova Scotia	2.02	802	382,550
New Brunswick	2.29	831	303,461
Quebec	3.76	4,597	3,303,877
Ontario	2.98	7,271	4,794,437
Manitoba	3.01	1,313	455,002
Saskatchewan	2.08	1,181	384,354
Alberta	2.66	1,900	1,316,021
British Columbia	2.69	2,725	1,736,920
Atlantic Provinces	2.25	2,703	941,739
Prairie Provinces	2.96	4,394	2,155,377
Canada	**3.13**	**21,690**	**12,932,350**

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. Since the approximate CV is conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance a *marginal* estimate could become *acceptable* based on the exact CV calculation.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate is most likely unacceptable and Statistics Canada recommends not to release such an estimate, regardless of the value of the coefficient of variation.

10.1 How to Use the Coefficient of Variation Tables for Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Approximate Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

Rule 1: Estimates of Numbers of Households Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the Approximate Sampling Variability Table for the appropriate geographic area, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: Estimates of Proportions or Percentages of Households Possessing a Characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, in the 2006 HES, the proportion of households that had a lawn (WA_Q16) and used chemical fertilizers in 2005 (FP_Q01) is more reliable than the estimated number of households that had a lawn and used chemical fertilizers in 2005. (Note that in the tables the coefficients of variation decline in value reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular province) reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} \sqrt{\left(\hat{X}_1 \alpha_1\right)^2 + \left(\hat{X}_2 \alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example in the 2006 HES, to the case where the denominator is the number of households that had a lawn and the numerator is the number of households that had a lawn and used chemical fertilizers in 2005.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of households that own their dwelling, had a lawn and used chemical fertilizers in 2005 as compared to the number of households that rent, had a lawn and used chemical fertilizers in 2005 the standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is:

$$\sigma_{\hat{R}} = \hat{R}\sqrt{\alpha_1^{\ 2} + \alpha_2^{\ 2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of Differences of Ratios

In this case, Rules 3 and 4 are combined. The CVs for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

10.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates

The following examples based on the HES 2006 are included to assist users in applying the foregoing rules. Please note that the data for these examples are different than the results obtained from the current survey and are only to be used as a guide.

Example 1: Estimates of Numbers of Households Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 8,375,473 households had a lawn in 2005. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA.

2) The estimated aggregate 8,375,473 does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 8,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 0.7%.

4) So the approximate coefficient of variation of the estimate is 0.7 %. The finding that 8,375,473 (to be rounded according to the rounding guidelines in Section 9.1) households had a lawn in 2005 is publishable with no qualifications.

Households and the Environment Survey, 2006 - Household File

Approximate Sampling Variability Tables - Canada

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE										
	0.1%	1.0%	...	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	107.6	107.1		99.3	96.3	93.3	90.1	86.8	83.4	76.1	59.0	34.1
2	76.1	75.8		70.2	68.1	65.9	63.7	61.4	59.0	53.8	41.7	24.1
3	62.1	61.9		57.3	55.6	53.8	52.0	50.1	48.2	44.0	34.1	19.7
4	53.8	53.6		49.6	48.2	46.6	45.0	43.4	41.7	38.1	29.5	17.0
5	48.1	47.9		44.4	43.1	41.7	40.3	38.8	37.3	34.1	26.4	15.2
6	43.9	43.7		40.5	39.3	38.1	36.8	35.4	34.1	31.1	24.1	13.9
7	40.7	40.5		37.5	36.4	35.2	34.1	32.8	31.5	28.8	22.3	12.9
8	38.1	37.9		35.1	34.1	33.0	31.9	30.7	29.5	26.9	20.9	12.0
9	35.9	35.7		33.1	32.1	31.1	30.0	28.9	27.8	25.4	19.7	11.4
10	34.0	33.9		31.4	30.5	29.5	28.5	27.5	26.4	24.1	18.7	10.8
11	32.5	32.3		29.9	29.0	28.1	27.2	26.2	25.1	23.0	17.8	10.3
12	31.1	30.9		28.7	27.8	26.9	26.0	25.1	24.1	22.0	17.0	9.8
13	*****	29.7		27.5	26.7	25.9	25.0	24.1	23.1	21.1	16.4	9.4
14	*****	28.6		26.5	25.7	24.9	24.1	23.2	22.3	20.4	15.8	9.1
15	*****	27.7		25.6	24.9	24.1	23.3	22.4	21.5	19.7	15.2	8.8
16	*****	26.8		24.8	24.1	23.3	22.5	21.7	20.9	19.0	14.7	8.5
125	*****	9.6		8.9	8.6	8.3	8.1	7.8	7.5	6.8	5.3	3.0
150	*****	*****		8.1	7.9	7.6	7.4	7.1	6.8	6.2	4.8	2.8
200	*****	*****		7.0	6.8	6.6	6.4	6.1	5.9	5.4	4.2	2.4
250	*****	*****		6.3	6.1	5.9	5.7	5.5	5.3	4.8	3.7	2.2
300	*****	*****		5.7	5.6	5.4	5.2	5.0	4.8	4.4	3.4	2.0
350	*****	*****		5.3	5.1	5.0	4.8	4.6	4.5	4.1	3.2	1.8
400	*****	*****		5.0	4.8	4.7	4.5	4.3	4.2	3.8	2.9	1.7
450	*****	*****		4.7	4.5	4.4	4.2	4.1	3.9	3.6	2.8	1.6
500	*****	*****		4.4	4.3	4.2	4.0	3.9	3.7	3.4	2.6	1.5
750	*****	*****		3.6	3.5	3.4	3.3	3.2	3.0	2.8	2.2	1.2
1,000	*****	*****		3.1	3.0	2.9	2.8	2.7	2.6	2.4	1.9	1.1
1,500	*****	*****		2.6	2.5	2.4	2.3	2.2	2.2	2.0	1.5	0.9
2,000	*****	*****		*****	2.2	2.1	2.0	1.9	1.9	1.7	1.3	0.8
3,000	*****	*****		*****	*****	1.7	1.6	1.6	1.5	1.4	1.1	0.6
4,000	*****	*****		*****	*****	*****	*****	1.4	1.3	1.2	0.9	0.5
5,000	*****	*****		*****	*****	*****	*****	*****	1.2	1.1	0.8	0.5
6,000	*****	*****		*****	*****	*****	*****	*****	*****	1.0	0.8	0.4
7,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	0.7	0.4
8,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	0.7	0.4
9,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	*****	0.4
10,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	*****	0.3

Note: For correct usage of these tables, please refer to the microdata documentation.

Example 2: Estimates of Proportions or Percentages of Households Possessing a Characteristic

Suppose that the user estimates that 2,920,763 / 8,375,473 = 34.9% of households had a lawn and used chemical fertilizers in 2005. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., households that had a lawn in 2005), it is necessary to use both the percentage 34.9% and the numerator portion of the percentage 2,920,763 in determining the coefficient of variation.

3) The numerator, 2,920,763, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 3,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 35.0%.

4) The figure at the intersection of the row and column used, namely 1.6% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 1.6%. The finding that 34.9% of households had a lawn and used chemical fertilizers in 2005 can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 2,724,073 / 7,331,151 = 37.2% of households that own their dwelling, had a lawn and used chemical fertilizers in 2005 while 196,690 / 1,044,323 = 18.8% of households that rent, had a lawn and used chemical fertilizers in 2005. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for households that own their dwelling, had a lawn and used chemical fertilizers in 2005 as 1.6%, and the CV of the estimate for households that rent as 6.8%.

2) Using Rule 3, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1 (households that own their dwelling) $\hat{X}_2$ is estimate 2 (households that rent) and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d} = 0.372 - 0.188 = 0.184$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(0.372)(0.016)\right]^2 + \left[(0.188)(0.068)\right]^2} \\ &= \sqrt{(0.000035) + (0.000163)} \\ &= 0.014\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.014 / 0.184 = 0.076$

4) So the approximate coefficient of variation of the difference between the estimates is 7.6%. The difference between the estimates is considered acceptable and this estimate can be released with no qualifications.

Example 4: Estimates of Ratios

Suppose that the user estimates that 2,724,073 households that own their dwelling, had a lawn and used chemical fertilizers in 2005, while 196,690 households that rent, had a lawn and used chemical fertilizers in 2005. The user is interested in comparing the estimate of owners versus renters in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of households that own their dwelling, had a lawn and used chemical fertilizers in 2005. The denominator of the estimate ($\hat{X}_2$) is the number of households that rent, had a lawn and used chemical fertilizers in 2005.

2) Refer to the coefficient of variation table for CANADA.

3) The numerator of this ratio estimate is 2,724,073. The figure closest to it is 3,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 1.6%.

4) The denominator of this ratio estimate is 196,690. The figure closest to it is 200,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 6.8%.

5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

$$\alpha_{\hat{R}} = \sqrt{{\alpha_1}^2 + {\alpha_2}^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. That is:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.016)^2 + (0.068)^2} \\ &= \sqrt{0.000256 + 0.004624} \\ &= 0.0698\end{aligned}$$

6) The obtained ratio of households that own their dwelling versus households that rent, had a lawn and used chemical fertilizers in 2005 is 2,724,073 / 196,690 which is 13.8 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 7.0% which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of households that own their dwelling had a lawn and used chemical fertilizers in 2005, to households that rent, had a lawn and used chemical fertilizers in 2005 is 11.0 for Saskatchewan, while it is 13.8 for Manitoba. The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First calculate the approximate coefficient of variation for the Saskatchewan ratio ($\hat{R}_1$) and the Manitoba ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Saskatchewan ratio is 17.3% and 24.5% for Manitoba.

2) Using Rule 3, the standard error of a difference ($\hat{d} = \hat{R}_1 - \hat{R}_2$) is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{R}_1\alpha_1\right)^2 + \left(\hat{R}_2\alpha_2\right)^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 11.0 – 13.8 = -2.8 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(11.0)(0.173)\right]^2 + \left[(13.8)(0.245)\right]^2} \\ &= \sqrt{(3.6214) + (11.4312)} \\ &= 3.880\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 3.880 / (-2.8) = -1.386.

4) So the approximate coefficient of variation of the difference between the estimates is 138.6%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter F (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

10.2 How to Use the Coefficient of Variation Tables to Obtain Confidence Limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

> If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.
>
> Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the difference would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.
>
> Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\left(\hat{X} - k,\ \hat{X} + k\right)$ where k is

determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval ($CI_{\hat{x}}$):

$$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}},\ \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

where $\alpha_{\hat{x}}$ is the determined coefficient of variation of $\hat{X}$, and

$t = 1$ if a 68% confidence interval is desired;
$t = 1.6$ if a 90% confidence interval is desired;
$t = 2$ if a 95% confidence interval is desired;
$t = 2.6$ if a 99% confidence interval is desired.

Note: Release guidelines which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

10.2.1 Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of households, that had a lawn and used chemical fertilizers in 2005. (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 34.9% (or expressed as a proportion 0.349)

t = 2

$\alpha_{\hat{x}}$ = 1.6% (0.016 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.349 – (2) (0.349) (0.016), 0.349 + (2) (0.349) (0.016)}

$CI_{\hat{x}}$ = {0.349 – 0.011, 0.349 + 0.011}

$CI_{\hat{x}}$ = {0.338, 0.360}

With 95% confidence it can be said that between 33.8% and 36.0% of households that had a lawn used chemical fertilizers in 2005.

10.3 How to Use the Coefficient of Variation Tables to Do a T-test

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

10.3.1 Example of Using the Coefficient of Variation Tables to Do a T-test.

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of households that own their dwelling, had a lawn and used chemical fertilizers in 2005 and the proportion of households that rent, had a lawn and used chemical fertilizers in 2005. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.014. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.372 - 0.188}{0.014} = \frac{0.184}{0.014} = 13.1$$

Since t = 13.1 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the HES are primarily categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the total number of months a year a household composts would be greater than the coefficient of variation of the corresponding proportion of households that compost their kitchen waste. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

10.5 Coefficient of Variation Tables

Refer to HES2007_CVTabsE.pdf for the coefficient of variation tables.

11.0 Weighting

Since the Households and the Environment Survey (HES) used a sub-sample of the Canadian Community Health Survey (CCHS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the CCHS. The CCHS weighting procedure is briefly described below.

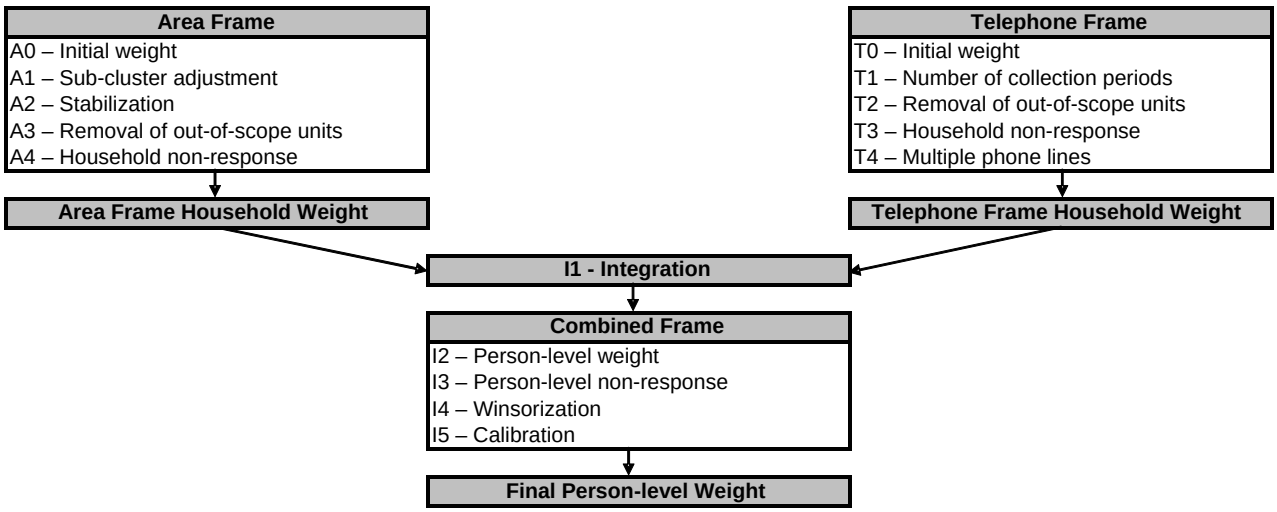
11.1 Weighting Procedures for the Canadian Community Health Survey

The CCHS has recourse to three sampling frames for its sample selection: an area frame acting as the primary frame and two frames formed of telephone numbers complementing the area frame. Since only minor differences differentiate the two telephone frames in terms of weighting, they are treated as one and referred to as being part of the telephone frame.

The weighting strategy was developed by treating both the area and telephone frames independently. Household-level weights resulting from these two frames are then combined into a single set of household weights through a step called "*integration*". After applying person-level weights and some further adjustments, this integrated weight becomes the final weight. Note that depending on the need, one or two frames were used for the selection of the sample within a given health region (HR). The weighting strategy deals with this aspect at the integration step.

Note: The CCHS household subweight (after the **I1 - Integration** step) corresponds to the initial HES weight.

Diagram A: CCHS Weighting Strategy Overview



The CCHS household subweight (required for the HES weighting) is available once steps A0 to A4 (for the area frame), T0 to T4 (for the telephone frame) and I1 (integration) are completed. , Each CCHS weighting step is described below.

Weighting of the area frame sample

A0 – Initial weight

The weighting on the area frame sample begins with a weight provided by the Labour Force Survey (LFS). The initial weight A0 is based on the LFS design since the CCHS sample is derived from the LFS. The LFS design consists of a sample of dwellings within the selected clusters of the LFS strata.

A1 – Sub-cluster adjustment
In clusters that experience significant growth, a sub-sampling methodology is used to ensure that the workload of the interviewers is kept at a reasonable level. This can consist of sub-sampling from the selected dwellings, dividing the cluster into sub-clusters, or reclassifying the cluster as a stratum and creating new clusters within the stratum. In all these cases, a sub-sample adjustment is calculated and applied to the CCHS weight. This adjustment is applied to weight A0 to produce weight A1.

A2 – Stabilization
In some Health Regions (HR), the increase of the sample size results in a significantly larger sample than necessary. Stabilization is used to bring the sample size back down to the desired level. The stabilization process consists of randomly sub-sampling dwellings at the HR level from the dwellings originally obtained within each cluster. An adjustment factor representing the effect of this stabilization is calculated in order to adjust the probability of selection appropriately. This factor, multiplied by weight A1, produces weight A2.

A3 – Removal of out-of-scope units
Among all dwellings sampled, a certain proportion is identified during collection as being out-of-scope. Dwellings that are demolished or under construction, vacant, seasonal or secondary, and institutions are examples of out-of-scope cases for the CCHS. These dwellings and their associated weight are simply removed from the sample. This leaves a sample that consists of, and is representative of, in-scope dwellings. These remaining dwellings maintain the same weight as in the previous step, which is now called A3.

A4 – Household non-response
During collection, a certain proportion of sampled households inevitably result in non-response. Weights of the non-responding households are redistributed to responding households within response homogeneity groups (RHG). In order to create the response groups, a scoring method based on logistic regression models is used to determine the propensity to respond and these response probabilities are used to divide the sample into groups with similar response properties. The information available for non-respondents is limited so the regression model uses characteristics such as the collection period and geographic information, as well as paradata, which includes the number of contact attempts, the time/day of attempt, and whether the household was called on a weekend or weekday. An adjustment factor is calculated within each class as follows:

$$\frac{\textit{Sum of weight A3 for all households}}{\textit{Sum of weight A3 for all responding households}}$$

Weight A3 is multiplied by this factor to produce weight A4 for the responding households. Non-responding households are dropped from the process at this point.

Weighting of the telephone frame sample

T0 – Initial weight
The initial weight T0 for the units on the telephone frames is defined as the inverse of the probability of selection and is computed separately for the random digit dialling (RDD) and list frame samples since the method of selection differs between these two frames.

T1 – Number of collection periods
On the area frame, the entire sample is selected at the beginning of the year. This is in contrast to the telephone frame, where samples are drawn every two months. Each of these samples comes with an initial weight that allows each sample to be representative at the HR level. To ensure that the total sample represents the population only once, an adjustment factor is applied to reduce the weights of each two-month sample. The adjustment factor

applied to each two-month sample is equal to the the inverse of the number of samples being combined (i.e. the number of collection periods). Following this adjustment, the entire list frame sample corresponds to the average over the entire combined collection period. The initial weights are multiplied by this adjustment factor to produce weight T1.

T2 – Removal of out-of-scope numbers
Telephone numbers associated with businesses, institutions or other out-of-scope dwellings, as well as numbers not in service or any other non-working numbers are all examples of out-of-scope cases for the telephone frame. Similar to the methods used on the area frame, these cases are simply removed from the process, leaving only in-scope dwellings in the sample. These in-scope dwellings keep the same weight as in the previous step, now called weight T2.

T3 – Household non-response
The adjustment applied here to compensate for the effect of household non-response is identical to the one applied for the area frame (adjustment A4) although the paradata used does differ because of the differences in collection applications for personal and telephone interviews. The adjustment factor calculated within each class was obtained as follows:

$$\frac{\textit{Sum of weight } T2 \textit{ for all households}}{\textit{Sum of weight } T2 \textit{ for all responding households}}$$

The weight T2 of responding households is multiplied by this factor to produce the weight T3. Non-responding households are removed from the process at this point.

T4 – Multiple phone lines
Some households can possess more than one residential telephone line. This has an impact on the weighting as these households have a higher probability of being selected so the weights for these households need to be adjusted for the number of residential telephone lines within the household. The adjustment factor represents the inverse of the number of lines in the household. The weight T4 is obtained by multiplying this factor by the weight T3.

11.2 Weighting Procedures for the Households and the Environment Survey

The principles behind the calculation of the weights for the HES are identical to those for the CCHS. However, further adjustments are made to the CCHS household subweight in order to derive a final weight for the records on the HES microdata file.

Diagram B: HES Weighting Strategy Overview

Weighting steps
H0 - CCHS subweight
H1 - HES initial weight
H2 - HES non-response
H3 - Calibration

H0 – CCHS subweight
The CCHS subweight is obtained once the I1 step of the CCHS weighting process is completed. The CCHS I1 step consists of integrating the weights for households common to the area and telephone frames into a single weight by applying a method of integration. The integration factor can be calculated as follows:

$$\alpha = n_A/(n_A + n_T)$$

where n_A and n_T represent the area and telephone frames sample sizes respectively. The weight of the area frame units is multiplied by this factor α, while the weight of the telephone frame units is multiplied by $1-\alpha$. The product between the factor derived here and the final household weight calculated earlier (A4 or T4, depending on which frame the unit belongs to), gives the integrated household weight I1, also called the CCHS subweight.

H1 – HES initial weight
The HES sample is a random sub-sample of the CCHS respondents. The probability of being selected in the HES sample is first calculated. For each household selected for the HES, an adjustment factor is defined as the inverse of its probability of selection. This factor, multiplied by weight H0, produces weight H1.

H2 – HES non-response
The weights of the non-responding HES households are redistributed to responding households within response homogeneity groups (RHG). In order to create the RHGs, a scoring method based on logistic regression models is used to determine the propensity to respond and these response probabilities are used to divide the sample into groups with similar response properties. An adjustment factor is calculated within each class as follows:

$$\frac{Sum\ of\ weight\ H1\ for\ all\ households}{Sum\ of\ weight\ H1\ for\ all\ responding\ households}$$

The weight H1 of responding households is multiplied by this factor to produce the weight H2. Non-responding households are removed from the process at this point.

H3 – Calibration
The last step necessary to obtain the final HES weight, H3, is calibration. Calibration is done to ensure that the sum of the final weights corresponds to the Census projections defined at the province and household size (one person, two persons or three persons and more) levels. The weight H3, produced at this step, is the final weight, WTHM, on the Master microdata file.

12.0 Questionnaire

The Households and the Environment Survey (HES) questionnaire was used to collect the information for the supplementary survey. The file HES2007_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

See HES2007_Master_CdBk.doc for the record layout with univariate counts

The TW file contains the questions relating to travel to work asked to every household member over the age of 16. See HES2007_TW_Master_CdBk.doc for the record layout with univariate counts