HEALTH PROMOTION SURVEY, 1990

Microdata User's Guide

Statistics Canada

Table of Contents

1.0 Introduction

The Health Promotion Survey (HPS) was conducted in June, 1990. It was carried out by Statistics Canada for Health and Welfare Canada.

This manual has been produced to facilitate the manipulation of the microdata file of survey results. Any questions about the data set or its use should be directed to:

Client Services
Special Surveys Division
Statistics Canada
Tel: (613) 951-7355 OR 1-888-297-7355
Fax: (613) 951-3012
Email: **ssd@statcan.ca**

IT IS IMPORTANT FOR USERS TO BECOME FAMILIAR WITH THE CONTENTS OF THIS DOCUMENT BEFORE PUBLISHING OR OTHERWISE RELEASING ANY ESTIMATES DERIVED FROM THE MICRODATA FILE OF THE HEALTH PROMOTION SURVEY.

2.0 Background

In 1981, the Health Promotion Directorate of Health and Welfare Canada was given a mandate by the Government of Canada to implement a national health promotion program. One of the elements of that program was the development of a national survey of the prevalence and distribution of a broad range of health practices which influence the health promotion and disease prevention prospects of Canadians.

Planning for "Canada's Health Promotion Survey" began in 1982 and culminated in a national survey conducted by Statistics Canada involving 11,181 adult Canadians in June of 1985. The intention of the 1985 HPS was to establish a national database on health practices and conditions closely related to the health and well-being of our adult population. It established both national and provincial baseline data on the knowledge, attitudes, beliefs, intentions and behaviours of adult Canadians on a wide range of health promotion issues (e.g., fitness, nutrition, safety and use of tobacco, alcohol and drugs).

In order to update the HPS database, a second cycle of the survey was conducted in June 1990. As in 1985, Health and Welfare was the sponsoring department and Statistics Canada was the collection agency.

The HPS was conducted under the authority of the Statistics Act, Revised Statutes of Canada, 1985, Chapter S19. Collection plans are registered under collection registration number STC/HLD-040-03909. The survey was conducted as a voluntary survey with ministerial approval obtained under Section 8 of the Statistics Act.

3.0 Objectives

The objectives of the 1990 HPS were to update and expand the national and provincial baseline data on the knowledge, attitudes, beliefs, intentions and behaviours of adult Canadians on a wide range of health promotion issues.

Among the topics included in the survey are: perceptions of health, blood pressure and cholesterol, alcohol use, exercise, nutrition, dental health, workplace health and safety, environmental health and sexually transmitted diseases.

With the 1990 cycle, comparison with findings from earlier surveys (i.e., the 1978 Canada Health Survey, and the 1985 HPS), will be possible for many health promotion issues. National trends will then be plotted over the years to assess their implications for health promotion and disease prevention programming by the Health Promotion Directorate.

4.0 Concepts/Definitions

Since the HPS data collection was conducted over the telephone, easy to understand terminology was used throughout the questionnaire to avoid long explanations.

However, some basic definitions had to be used in order to standardize the answers of the respondents. The following definitions were either printed on the HPS questionnaire and read to all respondents or given to interviewers to help answers questions from respondents.

High Blood Cholesterol:

If the respondent had been told by a health professional that his blood cholesterol was high then it meant that the respondent had a blood sample taken to measure blood cholesterol level.

Exercise:

Exercise meant vigourous activities such as aerobics, jogging, racquet sports, team sports, dance classes, or brisk walking.

Drink:

A drink meant:
– one bottle of beer or glass of draft;
– one small glass of wine; or
– one shot or mixed drink with hard liquor.

Sexual Intercourse:

Sexual intercourse meant heterosexual and/or homosexual intercourse.

French Canadian:

If the respondent answered French Canadian, Québécois, Acadian or Franco-Ontarian to question R5, then both categories "French" and "Canadian" were marked.

5.0 Survey Design

As for the 1985 survey, this second cycle was again a Random Digit Dialling (RDD) telephone survey, a technique whereby telephone numbers are generated randomly by computer and each number is then dialled.

Based on the experience of the 1985 HPS and other surveys, Statistics Canada did not attempt an RDD telephone survey in the North for this 1990 cycle. A separate data collection will be considered for the Yukon and Northwest Territories, using methodologies more tailored to the conditions in Northern Canada. Health and Welfare is working with the territorial governments on the logistics of doing this survey.

5.1 Population Coverage

The target population for the HPS was all persons 15 years of age or older living in Canada with the following two exceptions:

1. residents of the Yukon and the Northwest Territories;
2. full-time residents of institutions.

Because the HPS was conducted using telephone sampling techniques, households (and thus persons living in households) that do not have telephones were obviously excluded from the surveyed population. This accounts for less than 3% of the total population. However, the survey estimates have been weighted to include persons without telephones.

5.2 Stratification

In order to carry out the sampling, each of the ten provinces was divided into strata or geographic areas. Generally, for each province one stratum represented the Census Metropolitan Areas (CMAs) of the province and the other the non-CMA areas. Since Ontario and Saskatchewan are each sampled from two regional offices, more strata were included in the sample design for these areas.

5.3 Sample Selection

The HPS sample was created through RDD using two different methods for generating telephone numbers: the Waksberg method and the Elimination of Non-Working Banks method (ENWB).

5.3.1 The Waksberg method:

The Waksberg method was used in two provinces: Prince Edward Island and the non-Census Metropolitan Area stratum of Québec. The method employs a two-stage sample design which increases the likelihood of contacting households. The following describes what was done for these two strata in the 1990 HPS.

An up-to-date list of all telephone area code and prefix combinations was obtained. To these all possible combinations of the next two digits were added (i.e. all possible "banks" of 100 consecutive numbers within existing area code - prefix combinations were identified). This resulted in a list of all the possible first eight digits of ten digit telephone numbers in each stratum. These eight digit numbers, called "banks", formed the first stage sampling units (i.e. the Primary Sampling Units, PSUs).

Within each stratum, a random selection was made of these eight digit numbers and the final two digits were generated at random. This number, called a Primary number, was then called to determine whether or not it reached a household (i.e. the number was not used by a business, institution, etc.):

- If the number did not reach a household that number was dropped from further consideration.

- If the number reached a household then additional numbers, referred to as Secondary numbers, were generated within the same bank. These secondary numbers were called to determine whether or not they reached a household.

Secondary numbers were generated on a continuing basis until (1) five additional households were reached in each retained bank; or (2) the bank was exhausted (all possible ten digit phone numbers were generated) or; (3) the survey collection period ended.

Primary and secondary numbers were generated throughout the survey period in order to yield a predetermined number of households within each of the two strata sampled using the Waksberg method.

An attempt was made to list all eligible household members and to sample one of these at random.

The principle behind this technique is that, when telephone numbers are generated from clusters within banks which contain at least one residential number, there is a greater chance of contacting other residential numbers. For the 1990 HPS approximately 53.7% of the secondary numbers called reached a household compared to only 14.6% of the primary numbers.

5.3.2 The Elimination of Non-Working Banks (ENWB) Method:

The Elimination Of Non-Working Banks (ENWB) design is a form of RDD in which an attempt is made to identify all working banks for an area (i.e. to identify all banks with at least one household). Thus, all telephone numbers with non-working banks are eliminated from the sample frame. This method was used in all strata except for P.E.I. and the non-CMA stratum in Québec. The following describes what was done for these strata in the 1990 HPS.

A list of all banks that contained at least one residential listing was obtained from the various phone companies across Canada. After assigning each bank to a stratum, a systematic sample of telephone numbers was generated on the first day of interviewing.

Each telephone number was dialled to determine whether or not it reached a household. As with the Waksberg method, for each household reached, an attempt was made to list all eligible household members and to sample one of these at random.

For the 1990 HPS approximately 52.9% of all numbers dialled using the ENWB method reached households.

5.4 Sample Allocation by Province

PROVINCES	EXPECTED SAMPLE	RECORDS ON FINAL FILE
Newfoundland	1000	1088
Prince Edward Island	1000	942
Nova Scotia	1000	1002
New Brunswick	1000	943
Québec	1880	1728
Ontario	2230	2280
Manitoba	1000	1066
Saskatchewan	1000	990
Alberta	2581*	2530*
British Columbia	1250	1223
CANADA	13941	13792

* Sample augmented by the purchase of additional sample by Alberta.

6.0 Data Collection Methodology

6.1 Questionnaire Design

The two main components of the survey were the Control Form and the 1990 HPS Data questionnaire (refer to section 11 for copies of the questionnaires).

The Control Form was used to select a respondent within the household. The choice of questions for the Control Form had to respect certain constraints associated with Random Digit Dialling surveys.

Although the 1990 questionnaire closely resembled the one used in 1985 several modifications were made:

1) The wording of several questions from the 1985 survey was improved to correct for observed deficiencies. Question I1 -(about the number of breakfasts in the past week) is an example of this kind of question.

2) Some questions asked in 1985 were dropped from the 1990 questionnaire. This was done to avoid duplication with other health surveys and to reflect the new mission statement of the Health Promotion Directorate.

3) New sections were added, such as the section on the prevention of sexually transmitted diseases and dental health section.

In March 1990, prior to the national survey, a pre-test of the two 1990 questionnaires was carried out in two Statistics Canada regional offices, Halifax and Montréal. Approximately 300 respondents were interviewed in each regional Office. The purpose of the pretest was to verify the quality of the collection instrument in both official languages (i.e. interview length, respondent reaction, etc.). Selected respondents from two strata (rural and urban) within each of the two provinces (Nova Scotia and Québec) were interviewed. Based on the pre-test results and interviewer de-briefings some final wording changes were made to the questionnaire used in the full survey.

6.2 Collection Methodology

The Control Form was used each time a different telephone number was dialled by the interviewer. The purpose of this document was to first determine whether or not the number called reached a household, and then if so, to list all household members. One household member 15 years of age or over was then selected at random, using a pre-printed selection grid.

The HPS was then conducted with this selected person by telephone. If this selected person was not available to be interviewed at that time, an attempt was made to determine a convenient time to phone back to complete the interview. Because many of the survey questions were of an attitudinal nature and some were personal, all interviews had to be conducted with the selected respondent only; no proxy reporting was accepted.

6.3 Collection Period

Interviews were conducted from Statistics Canada's eight Regional Offices, from June 1 - 30, 1990.

All interviews were conducted between 8.30 AM to 9.30 PM local time during week days. Interviews were also conducted during daytime on Saturdays.

6.4 Interviewing

The data collection was carried out by experienced interviewers working for Statistics Canada. All questions were administered to respondents in accordance with rigorous interviewer instructions. Their training was particularly oriented towards methods of administering the questionnaire in a neutral manner and to adhere strictly to directives. The sensitive nature of some topics covered in the questionnaire was stressed and they were trained to deal with this fact. The interviewers were provided with a Training Manual and an Interviewer's Manual. The senior interviewers also received a Procedures Manual.

The questionnaire was administered in one of the official languages. If a respondent could not communicate in either language, he/she was not interviewed.

7.0 Data Processing

7.1 Data Capture

The HPS data were captured in June and July of 1990 on minicomputers in the eight Regional Offices of Statistics Canada. The data capture program allowed for a valid range of codes for each question and automatically followed the flow of the questionnaire. Information was then electronically transmitted to Ottawa for the creation of an initial computer file.

7.2 Editing

Initial processing resulted in the formation of 13,960 records. To accommodate most statistical packages, all blank fields were converted to a numeric value.

A customized edit program was designed using the "bottom up" approach to correct for erroneous data flow as a result of either interviewer or data capture error. The "bottom up" method of editing looks at responses within blocks of questions to determine the correct flow. Then, questions which were determined to be "NOT APPLICABLE" for a particular response pattern were given a standard value (i.e. "8", "98", etc,).

Checks were also carried out to compare certain fields. For example, age and year of birth were examined for consistency.

Industry and Occupation were coded using a computerized system developed for the Labour Force Survey.

After all processing the micro data file contains 13,792 records. (Note: 268 records were dropped from the file for several reasons – refer to section 9.2 and 9.3.)

7.3 Derived Variables

After all numerical verification was completed derived variables were created to accommodate user needs. These include items such as the number of household members aged fifteen and older (DVHSIZE), the body mass index (BMI) and the province (DVPROV).

For comparability with other files and to conform with the requirements of the Microdata Documentation Committee, the industry and occupation codes were collapsed into two different coding structures.

7.4 Weighting

The principle behind estimation in a probability sample such as the HPS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example in a simple random sample of 2% of the population, each person represents 50 persons in the population.

The weighting phase is a step which calculates, for each record, what this number is and places it on the microdata file for each record. This weight must be used to derive estimates from the microdata file. For example, if the number of persons who have partially completed elementary school is to be estimated, it is done by selecting the records referring to persons with that characteristic and summing the weights of those records.

Details of the method used to calculate these sampling weights are presented in Section 12.

8.0
Sampling Error

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

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 also use.

The basis for measuring the potential size of sampling errors is the standard error (se) of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey such as this, 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 ($\bar{X}$) and is expressed as a percentage of the estimate:

$$cv_{\bar{x}} = se_{\bar{x}} / \bar{X}$$

For example, suppose that, based upon the H.P.S. results, one estimates that 25% of all adults (persons aged 15 and older) say that they are in "excellent health" and that this estimate is found to have a standard error of 0.012. Then the coefficient of variation of the estimate is calculated as:

$$cv_{\bar{x}} = se_{\bar{x}} / \bar{X} = 0.012 / 0.25 = 0.048 = 4.8\%$$

Note: Refer to section 10.3 for sampling variability guidelines.

Before discussing how these measures can be obtained it is useful to describe the two main types of point estimates of population characteristics which can be generated from the microdata files for the HPS.

(1) Categorical Estimates

Categorical estimates are estimates of the number, proportion or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of persons aged 15-24 who are in "excellent health" or the proportion of Nova Scotia's population that consists of females in "poor health" are examples of such estimates.

In this context, an estimate of the number of persons possessing a certain characteristic is referred to as an estimate of an aggregate.

(2) Quantitative Estimates

Quantitative estimates are estimates of totals or of means, and other measures of central tendency 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 total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total.

An example of a quantitative estimate in this survey is the mean number of sexual partners in the past twelve months of Canadians aged 15 and above.

8.1 Coefficient of Variation for a Categorical Estimate

In order to supply cv's 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 "look-up" tables, referred to as Approximate Sampling Variability Tables, has been produced and included as Section 15.

The cv's in these tables 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, has been determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the look-up tables which would then apply to the entire set of characteristics. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis.

The following rules should enable the user to determine the approximate cv's 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 estimates.

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

The cv for an aggregate 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 cv.

Rule 2: Estimates of Proportions or Percentages Possessing a Characteristic

The cv 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 subset of the total population. (Note that in the tables the cv's 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 given age group), 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 cv.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The cv for a difference between two estimates is given by the standard error of the difference divided by the difference. The standard error (se) of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately.

With $\hat{d} = \hat{X}_1 - \hat{X}_2$

$$se_{\hat{d}} = \sqrt{(\hat{X}_1 * \alpha_1)^2 + (\hat{X}_2 * \alpha_2)^2}$$

$$cv_{\hat{d}} = se_{\hat{d}} / \hat{d}$$

where: $\hat{X}_1$ and $\hat{X}_2$ are estimates
α_1 and α_2 are the cv's of $\hat{X}_1$ and $\hat{X}_2$ respectively

This formula is accurate for the difference between separate and uncorrelated estimates 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, to the case where the denominator is the number of males and the numerator is the number of males with a given characteristic.

In the case where the numerator is not a subset of the denominator, the cv of the ratio of two estimates ($\hat{R} = \hat{X}_1 / \hat{X}_2$) is equal to the standard error of the ratio divided by the ratio ($cv_R = se_R / \hat{R}$). The standard error of the ratio is approximately equal to the square root of the sum of squares of each cv considered separately multiplied by the ratio. This means that the cv of the ratio is approximately equal to the square root of the sum of squares of each cv considered separately:

With $\hat{R} = \hat{X}_1 / \hat{X}_2$

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

$$cv_R = se_R / \hat{R}$$

$$cv_R = (\hat{R} * \sqrt{\alpha_1^2 + \alpha_2^2}) / \hat{R}$$

$$cv_R = \sqrt{\alpha_1^2 + \alpha_2^2}$$

where: $\hat{X}_1$ and $\hat{X}_2$ are estimates

α_1 and α_2 are the cv's of $\hat{X}_1$ and $\hat{X}_2$ respectively

This 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 cv's for the two ratios are first determined using Rule 4, and then the cv of their difference is found using Rule 3.

8.2 Examples

The following two examples, using the HPS data, are included to assist users in applying the Approximate Sampling Variability Tables. Refer to section 13 for the variable names and codes.

Example 1

A user can estimate from the microdata file that 200,260 adults in Manitoba reported that they were "breakfast skippers" (meaning that they had "just coffee, tea or nothing at all for breakfast" for the 7 days before they responded to the HPS; I1A = 7). How does the user determine the coefficient of variation of this estimated total?

1) Refer to the table for Manitoba.

2) The estimated aggregate, 200,260, 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, 200,000.

3) The cv for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 6.1%.

4) So the approximate cv of the estimated total is 6.1%.

Example 2

Suppose that the user then estimates that of the 200,260 "breakfast skippers" in Manitoba 40.6% (or 81,208) of these had a body mass index (BMI) of 20.0 to 24.9. How does the user determine the cv of this estimated percentage?

1) Refer to the table for Manitoba.

2) Because the estimated percentage of 40.6% is based on a subset of the total population (i.e., "breakfast skippers"), it is necessary to use both the percentage (40.6%) and the numerator portion of the percentage (81,208) in determining the cv.

3) The numerator, 81,208, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closet to it, namely 80,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, namely, 40.0%.

4) The figure at the intersection of the row and column used, namely, 8.6% is the cv to be used.

5) So the approximate cv of the estimated percentage is 8.6%.

8.3 Coefficients of Variation for Quantitative Estimates

Most of the major variables of interest in the H.P.S. are categorical in nature. For quantitative estimates, special tables would have to be produced upon request to determine their sampling error.

As a rule, however, if the total number of persons (weighted) on which the quantitative estimate is based is itself not releasable, then the quantitative estimate is not releasable. This implies that for any tabulations involving quantitative estimates, tables giving the estimated number of persons in each cell should be produced.

8.4 Confidence Intervals

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 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\bar{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\{\bar{X}-k, \bar{X}+k\}$ 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 the cv of the estimate $\bar{X}$ from the appropriate table, and then using the following formula to produce a confidence interval "CI":

$$CI_{\bar{x}} = \{\bar{X} - (t * \bar{X} * \alpha), \bar{X} + (t * \bar{X} * \alpha)\}$$

where: α is the determined coefficient of variation of $\bar{X}$

t = 1.0 if a 68% confidence interval is desired
t = 1.6 if a 90% " " "
t = 2.0 if a 95% " " "
t = 3.0 if a 99% " " "

Example

A 95% confidence interval for the estimated proportion of adults in Manitoba who are "breakfast skippers" with a Body Mass Index between 20.0 and 24.9 (from Example 2 in section 8.2) would be calculated as follows:

$\bar{X}$ = 40.6% (or expressed as a proportion = 0.406)

t = 2

$cv_{\bar{x}}$ = 8.6% (0.086 expressed as a proportion)

$CI_{\bar{x}} = \{0.406 - (2 * 0.406 * 0.086), 0.406 + (2 * 0.406 * 0.086)\}$

$CI_{\bar{x}} = \{0.406 - 0.0698, 0.406 + 0.0698\}$

$CI_{\bar{x}} = \{0.3362, 0.4758\}$

With 95% confidence it can be said that between 33.6% and 47.6% of adults in Manitoba who are breakfast skippers have a Body Mass Index between 20.0 and 24.9.

9.0 Non-sampling Error

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 generally 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 has been made to reduce non-sampling errors in the HPS. Quality assurance measures have been 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 HPS procedures and questionnaires, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors are minimized and coding and edit quality checks to verify the processing logic. Despite these efforts non-sampling error is bound to have some impact on HPS estimates. The following section outlines the most likely sources of this error and its probable impact on the survey estimates.

9.1 Total Non-response

Total non-response can be a major source of non-sampling error in many surveys depending on the degree to which respondents and non-respondents differ with respect to characteristics of interest. In the HPS, total non-response occurred because the selected individual could not be contacted or the selected individual refused to participate in the survey. Total non-response is handled by adjusting the sampling weight of responding individuals to compensate for missing individuals.

9.2 Partial Non-response

Partial non-response in the HPS occurred if the respondent refused to answer a question or did not understand a question.

For HPS, a set of questions that must be answered by respondents were chosen. If one of these questions was not answered then the record was dropped from the file. 156 records were dropped from the file. As a result, it is unlikely that partial non-response contributed substantially to non-sampling error.

9.3 Response Rates

For the HPS 35,077 phone numbers were called and 17,674 of these were determined to belong to households. Of these households, 2,288 (12.9%) were non-responding households because, either they refused to respond or could not respond to the survey. Included here, as well, are households that could not be reached during the entire survey collection period.

For the 15,386 responding households where an interview was attempted, 511 selected persons refused to complete the survey (one person was randomly selected per responding household). In addition, there were 168 records which were dropped during Head Office Processing either because of partial non-response or because the respondent was less than 15 years old.

If it is assumed that all the non-responding and dropped households were all "in scope" (i.e., had at least one member 15 years old or older), then the overall survey response rate was 78.0%.

9.4 Coverage

As mentioned in section 5.1, less than 3% of the total population did not have telephones. This part of the population may have unique characteristics which will not be reflected in the survey estimates. Users should be cautioned about this downward bias on estimates of populations at risk since risk is often, but not always, correlated with non-telephone ownership.

9.5 Underestimation

Due to the sensitive nature of some questions, such as the sexual health questions, more respondents refused to answer such questions compared to non-personal questions (10% refusals for age of first sexual intercourse compared to 2% for effect of environmental pollution on health). This part of the population may represent unique characteristics which have not been accounted for in the estimates.

10.0 Publication and Release Guidelines

IT IS IMPORTANT FOR USERS TO BECOME FAMILIAR WITH THE CONTENTS OF THIS SECTION BEFORE PUBLISHING OR OTHERWISE RELEASING ANY ESTIMATES DERIVED FROM THE MICRODATA FILE OF THE HEALTH PROMOTION SURVEY.

This section of the documentation outlines the guidelines to be adhered to by users publishing or otherwise releasing any data derived from the survey microdata file. 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. This section consists basically of four sub-sections – the rounding guidelines, the sample weighting guidelines and the sampling variability guidelines and guidelines for statistical analysis.

10.1 Rounding Guidelines

In order that estimates for publication or other release derived from these microdata files will 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 thousand 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 1000, if the last three digits are between 000 and 499, they are changed to 000 and the preceding digit (the thousands digit) is left unchanged. If the last digits are between 500 and 999 they are changed to 000 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 1000 units using normal rounding.

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

d) Sums and differences of aggregates or ratios are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 1000 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 that actually exists.

10.2 Sample Weighting Guidelines for Tabulation

The sample design used for the HPS was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the sampling weights placed on the individual microdata tape records. Otherwise, the estimates derived from the microdata tapes 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, because of their treatment of the weight field, may not allow the generation of estimates that exactly match those available from Statistics Canada.

10.3 Sampling Variability Guidelines for the Release of Estimates

Users should first determine the number of respondents on the micro data file who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate (the coefficient of variation (cv) is the percent standard deviation).

Before releasing and/or publishing any estimate from the microdata tape the user should determine the approximate coefficient of variation for each estimate and follow the guidelines below. Section 8 contains a detailed description on how to obtain an approximate coefficient of variation for each estimate.

Type of Estimate	**Coefficient of Variation (in %)**	**Release Guideline**
1. Unqualified	0.0 to 16.5%	Estimates can be considered for general unrestricted release. No special notation is required.
2. Qualified	16.6 to 25.0%	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning users of the high sampling variability associated with the estimates.
3. Restricted	25.1 to 33.3%	Estimates can be considered for general unrestricted release only when sampling variabilities are obtained using an exact variance calculation procedure. Otherwise, the estimate should be deleted. When sampling variabilities are obtained using exact variance calculation procedures, the estimates should be accompanied by a warning of high sampling variability associated with the estimates.
4. Not for release	33.4% or over	Estimates should not be released in any form under any circumstances. In such statistical tables, such estimates should be deleted.

Note: The sampling variability guidelines should be applied to rounded estimates.

10.4 Guidelines for Statistical Analysis

The HPS 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.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures 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 almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, estimation of rates and proportions and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful. If the weights on the data file are rescaled so that the average weight is one (1), then the variances produced by the standard packages will be more reasonable; while they still will not take into account the stratification and clustering of the sample's design, they will take into account the unequal probabilities of selection. The rescaling can be accomplished by dividing each weight by the overall average weight before the analysis is conducted.

11.0
Questionnaires

12.0 Weighting Procedures

12.1 Estimation

When a probability sample is used, as was the case for the HPS, the principle behind estimation is that each person selected in the sample represents (besides himself/herself) several other persons not in the sample. For example, in a simple random sample of 1% of the population, each person in the sample represents 100 persons in the population.

For the HPS microdata file an overall statistical weight (called "WEIGHT") was placed on each record to represent the number of sampled persons that the record represents. This weighting factor refers to the number of times a particular record should contribute to a population estimate. For example, to estimate the number of persons who describe their lives as being "Very Stressful" the value of WEIGHT is summed over all records with question A2 having a code of 1. The HPS weighting process is described below in Section 12.2.

12.2 Weighting of Health Promotion Survey

Because the HPS employed two different sampling techniques (as discussed in Section 5), two slightly different weighting procedures were employed. The preliminary weighting procedures for each type of HPS design are described separately, depending on the sampling technique used. The final adjustments made to all records are described in a third section.

12.2.1 Preliminary Weighting Procedure for Waksberg Design

The Waksberg method was used in Prince Edward Island stratum and the non-Census Metropolitan Area stratum of Québec.

A self-weighting sample design is one for which the weights for all units in the sample are the same. For a two-stage sample design, this happens if the first stage units (i.e., the Primary Sampling Units) are selected using proportional to size sampling and a fixed number of units are selected within each selected Primary Sampling Unit with equal probability.

The following outlines the steps that were used in weighting the HPS records using the Waksberg technique.

1) Basic Weight

In the first stage of weighting all households that were selected into the sample within a given stratum were assigned an identical weight.

2) Non-Response Adjustment

Weights for responding households were adjusted to represent non-responding households. Within each working bank of telephone numbers selected in the sample, the HPS sampling scheme required that six households be contacted. In some cases, one or more of these six households refused to participate in the survey. Weights of responding households were adjusted to compensate for non-responding households by multiplying the basic weight of responding households within a bank by the following ratio:

$$\frac{6}{\text{(No. of responding households within the bank)}}$$

3) Multiple Telephone Adjustment

Weights for households with more than one private telephone number were adjusted downwards to account for the fact that such households have a higher probability of being selected (i.e. the weight for each household was divided by the number of distinct telephone numbers that serviced the household).

4) Person Weight Calculation

A person weight was then calculated for each person who responded to the survey by multiplying the household weight for that person by the number of persons in the household who were eligible to be selected for the survey (i.e., the number of household members 15 years old or older).

12.2.2 Preliminary Weighting Procedure for E.N.W.B. Design

As was the case for the Waksberg design, when the Elimination of Non-Working Banks (E.N.W.B.) design is used, each household within a stratum has an equal probability of selection. This probability is equal to:

$$\frac{\text{No. of telephone numbers sampled within the stratum}}{\text{Total number of possible telephone numbers within the stratum}}$$

Note that the total number of possible telephone numbers for a stratum is equal to the number of working banks for a stratum times 100.

The following steps outline the weighting procedure that was used for E.N.W.B. records.

1) Basic Weight

Each household (responding and non-responding) was assigned a weight equal to the inverse of its probability of selection:

$$\frac{\text{Total number of possible telephone numbers within the stratum}}{\text{No. of telephone numbers sampled within the stratum}}$$

2) Non-Response Adjustment

Weights for responding households were adjusted to represent non-responding households. This was done independently within each area code prefix. Records were adjusted by the following factor:

$$\frac{\text{Sum of the household weights of all households within the area code prefix}}{\text{Sum of the household weights of responding households with the area code prefix}}$$

Non-responding households were then dropped from further weighting procedures.

3) Multiple Telephone Adjustment

Weights for households with more than one private telephone number were adjusted downwards to account for the fact that such households have a higher probability of being selected. The weight for each household was divided by the number of distinct telephone numbers that serviced the household.

4) Person Weight Calculation

A person weight was then calculated for each person who responded to the survey by multiplying the household weight for that person by the number of persons in the household who were eligible to be selected for the survey (i.e., the number of household members 15 years old or older).

12.2.3 Combined Weighting Procedure for Both Designs

After the preliminary weighting procedures for both the Waksberg and the E.N.W.B. design were completed the separate files were combined for the remaining "combined" weighting procedure.

5) Adjustment for External Stratum Totals

An adjustment was made to the person weights on records within each stratum in order to make population estimates consistent with Census projected population counts for persons 15 and older. This was done by multiplying the person weight for each record within the stratum by the following ratio:

$$\frac{\text{Census population projection for persons 15 and older for the stratum}}{\text{Sum of the person weights of HPS Waksberg Records within the stratum}}$$

6) Adjustment for Province - Sex - Age Group Totals

The next weighting step was to ratio adjust the weights of all records within a province to agree with Census projected age-sex distributions. Census projected population counts were obtained for the reference date of June 1990.

The following age groups were used for both males and females:
15-19, 20-24, 25-29, 30-34, 35-39, 40-44
45-49, 50-54, 55-59, 60-64, 65-69, 70 and older

For each of the resulting classifications (10 provinces X 12 age groups X 2 sexes) the person weights for records within the classification were adjusted by multiplying by the following ratio:

$$\frac{\text{Projected census population for Population - Sex - Age group}}{\text{Sum of the person weights of records in the Province - Sex - Age group}}$$

Some collapsing of age groups was required prior to applying this ratio to ensure minimum size requirements were met.

It should be noted that persons living in households without telephone service are included in these projections even though such persons were not sampled.

7) Raking Ratio Adjustment

The weights of each respondent were adjusted several times using a raking ratio procedure. This procedure ensured that estimates produced for a stratum and for each Province - Sex - Age Group would agree.

This adjustment was made by repeating steps 5) and 6) of the weighting procedures, using the weights obtained from the previous step, until the two sets of estimates were both correct. The final statistical weight became the variable "WEIGHT".

12.3 Weighting Policy

Users are cautioned against releasing unweighted tables or performing any analysis based on unweighted survey results since there were several weight adjustments performed independently on records within each province. As well, sampling and response rates varied significantly from province to province and from age group to age group.

The HPS was designed so that estimates could be reliably produced at the provincial level. Due to the difference in the population of many of the strata this resulted in a large difference between the lowest and highest average weights for the survey. For example, P.E.I. respondents had an average weight of about 105 compared to an average weight of above 3,000 for respondents from Quebec.

Also, it is known that non-respondents are more likely to be males and more likely to be younger. In the HPS sample males aged 15-19 represented 3.6% of the raw sample but 4.6% of the weighted population. On the other hand females, 70 years old or older, represented 4.0% of the raw sample but only 2.8% of the population.

Clearly, sample counts cannot be considered to be representative of the survey target population unless appropriate weights are applied.

13.0
Record Layout

14.0
Notes on Record Layout

Notes on Interpretation of the Record Layout for the Health Promotion Survey (1990)

These notes are intended to provide additional information for fields that may not be clearly defined on the record layout. The notes should be read in conjunction with the record layout.

14.1 Acronym

The acronyms used on the record layout refer to the question numbers as they appear on the questionnaire. Exceptions are the derived variables and the multiple choice questions. An example of a multiple choice question is C1 which is shown on the layout as following:

– C1CAT01: C1 is the question number, CAT refers to the category within the question, 01 is the category number.

Note: Questions using this format (C1, C5, C8, I2, N4, Q2, Q4, R4 and R5) did not have their answer categories read to the respondent but were marked by interviewers when given as responses.

14.2 Coding

Throughout the questionnaire standard codes were used.

Code 8, 98, 998:

This code means that the respondent did not have to answer to this question, the question was not applicable.

Since section L applies to female respondents only, this code was used for all the questions within this section when the respondent is a male. The reverse logic was applied to section M which applies to male respondents only.

Code 9, 99, 999:

This code means that the respondent did not answer the question, however an answer should have been given.

For multiple choice questions, if a respondent did not answer to all categories, then all categories will be coded to 9. If one category was answered, then the answered category will be coded to 1 and the others to 2.

14.3 Values (Codes)

The values shown on the record layout do not correspond to the values on the questionnaire in all cases. Also, values shown in the "notes" portion of the record layout refer to the values on the record layout. For example, the note for question E3a states this question is not asked if E1=1 or E2=2. If one were to look at the questionnaire, they would see the corresponding value for E1 is valid but the value for E2 is a 4 and not a 2.

15.0 Approximate Sampling Variability Tables (C.V. Tables)