

**National Population Health Survey**

**Health Institutions Component**

**Cycle 1 (1994-1995)**

**Public Use Microdata File Documentation**

**Statistics Canada**



## TABLE OF CONTENTS

1.	Introduction.....	1
2.	Background .....	2
3.	Objectives.....	3
4.	Survey Content.....	4
5.	Sample Design .....	5
	5.1 Design of the Frame .....	5
	5.2 Stratification and Allocation.....	5
	5.3 Sample Selection .....	7
6.	Data Collection .....	8
	6.1 Questionnaire Design and Data Collection Method .....	8
	6.2 Test.....	8
	6.3 Interviewing.....	8
	6.4 Supervision and Control.....	9
	6.5 Non-Response to the NPHS .....	9
7.	Data Processing .....	10
	7.1 Data Capture .....	10
	7.2 Editing.....	10
	7.3 Coding .....	10
	7.4 Creation of Derived Variables .....	10
	7.5 Weighting.....	11
	7.6 Suppression of Confidential Information .....	11
8.	Data Quality.....	12
	8.1 Response Rates .....	12
	8.2 Survey Errors.....	12
	8.3 Imputation .....	14
9.	Guidelines for Tabulation, Analysis and Release .....	15
	9.1 Rounding Guidelines.....	15
	9.2 Sample Weighting Guidelines for Tabulation .....	16
	9.2.1 Definitions of types of estimates: Categorical vs. Quantitative .....	16
	9.2.2 Tabulation of Categorical Estimates.....	17
	9.2.3 Tabulation of Quantitative Estimates.....	18
	9.3 Guidelines for Statistical Analysis .....	18
	9.4 Release Guidelines.....	19
10.	Weighting .....	20
	10.1 Probabilities of selection .....	20
	10.1.1 Probability of selecting an institution .....	20
	10.1.2 Probability of selecting a resident within an institution.....	21
	10.2 Weight Calculations and Adjustments.....	21
	10.2.1 Initial Institutional Weights.....	21
	10.2.2 Institutional Non-response Weight Adjustment.....	21
	10.2.3 Initial Personal Weights.....	21
	10.2.4 Personal Non-response Weight Adjustment.....	22

11. Variance Calculations.....	23
11.1 Running the variance program.....	24

Appendix A: Questionnaire

Appendix B: Record Layout

Appendix C: Data Dictionary

Appendix D: Derived Variables

## **1. Introduction**

The goal of the National Population Health Survey (NPHS) is to collect information related to the health of the Canadian population. To obtain a comprehensive picture of Canadians' health, a special survey was developed for people living in health care institutions—hospitals, nursing homes, and residential facilities for people with disabilities. Data produced from this survey will be both cross-sectional and longitudinal, gathered from a panel of individuals every two years.

In 1994 and 1995, household residents from every province were interviewed for the main component of the NPHS. Data from the household survey were released in the fall of 1995. A separate survey was also conducted in the Yukon and Northwest Territories and the results will be available at a later date.

Statistics Canada conducted the National Population Health Survey of Residents of Health Institutions between January and March 1995. This manual has been produced to facilitate the use of the microdata file of the survey results.

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

Data Access Unit  
(613) 951-1653  
nphs-ensp@statcan.ca

## **2. Background**

In the fall of 1991, the National Health Information Council (NHIC), recommended that an on-going national survey of population health be conducted. This recommendation was based on consideration of the economic and fiscal pressures on the health care system and the commensurate requirement for information to improve the health status of the population in Canada. Existing sources of health data were unable to provide a complete picture of the health status of the population and the myriad of factors having an impact on health.

Beginning in April 1992, Statistics Canada received funding for the development of a National Population Health Survey. The survey was designed to be flexible and to produce valid, reliable, and timely data. Also, it was to be responsive to changing requirements, interests, and policies.

A special survey covering residents of health institutions was undertaken because this population was rarely covered by national surveys and likely had health characteristics different from those of the general population.

### **3. Objectives**

The objectives of the NPHS are to:

- aid in the public policy development by providing measures of the level, trend, and distribution of the population's health status;
- provide data for analytic studies that will assist in understanding the determinants of health;
- collect data on the economic, social, demographic, and environmental correlates of health;
- increase the understanding of the relationship between health status and health care utilization;
- provide information on a panel of people followed over time, to reflect the dynamic process of health and illness and determine the factors affecting institutionalization;
- provide the provinces and territories and other clients with a health survey capacity that will allow supplementation of content or sample;
- allow the possibility of linking survey data to administrative data that is collected routinely, such as vital statistics, environmental measures, community variables, and health services utilization.

#### **4. Survey Content**

The content of the NPHS Health Institutions component was selected according to the following criteria:

- 1) The survey should collect information on the health status of the Canadian population residing in health institutions.
- 2) The data collected should be comparable to that of the household population whenever possible.
- 3) The survey should increase the understanding of conditions relating to institutionalization.
- 4) Information provided should permit the study, over time, of the transitions from households to institutions and vice versa.
- 5) The survey should produce national level data.

Respondents were randomly chosen from selected health care institutions. The questionnaire included components on health status, risk factors, social support, contact with health care providers, and demographic and socio-economic status. For example, health status was measured through questions on self-perception of health, functional ability, chronic conditions, and activity restriction. Behavioural risk factors included smoking and alcohol use. The level of social support was assessed by the frequency of contact with friends and relatives inside and outside the institution. Demographic and socio-economic information included age, sex, education, ethnicity, and personal income. The questionnaire is provided in Appendix A.



## **5. Sample Design**

The target population of the institutional survey included all long-term residents of health institutions from all provinces, excluding the territories. A list frame of facilities with long-term residents was created and stratified by geographic region, type and size of facility. A sample of institutions and a subsequent sample of residents within these institutions were selected.

### **5.1 Design of the Frame**

The sample frame was generated from lists of residential care facilities and hospitals maintained by the Health Statistics Division (HSD) of Statistics Canada. Provincial Ministries of Health verified and updated these lists to ensure their accuracy. The institutions were classified by the dominant type of care provided and only those providing long-term care were retained<sup>1</sup>. From the residential care facilities list, institutions providing long-term care for aged people, emotionally-disturbed children, developmentally- delayed, physically- and psychiatrically-disabled people were retained. Facilities from the hospital list included general hospitals with long-term units, extended/chronic care or rehabilitation facilities, and specialty hospitals with long-term units, such as pediatric and psychiatric hospitals. The number of long-term beds was known for each institution.

The sample population was restricted to those facilities with at least four beds that provided long-term care to residents with health problems. Facilities with fewer than four beds were not included on the HSD lists of residential care facilities and hospitals. Health care institutions on Indian reservations and Canadian Forces Bases or within correctional facilities were removed from the lists for operational reasons.

### **5.2 Stratification and Allocation**

The total sample size was set at 2,600 residents. Assuming a response rate of 85%, this sample size would be sufficient to calculate national estimates with a coefficient of variation (CV) of 10% for variables occurring in a minimum of 10% of the population.

The list of health institutions was initially stratified by geographic region (geographic stratum) and subsequently by the type of institution (characteristic stratum) and number of beds (size stratum).

---

<sup>1</sup> Institutions that exclusively provided short-term care, such as drug rehabilitation centres were excluded because the household component of the NPHS covers short-term institutional residents.

The geographic strata consisted of 5 regions (the Atlantic provinces, Quebec, Ontario, the Prairie provinces, and British Columbia). Within each geographic stratum three characteristic strata were defined:

- Institutions for the Aged - including residential care facilities for the aged and extended/chronic care hospitals.
- Cognitive Institutions - including residential care facilities for emotionally-disturbed children, psychiatrically-disabled and developmentally-delayed people, and psychiatric hospitals.
- Other Rehabilitative Institutions - including rehabilitation, pediatric and other specialty hospitals, general hospitals with long-term units as well as residential care facilities for people with physical disabilities.

Within each of these geographic/characteristic strata, the institutions were grouped into size strata by grouping facilities with a similar number of beds. The number of size strata created depended on the total number of beds in the geographic/characteristic strata. Once the number of size strata was determined, the boundaries for the different size strata were fixed using the  $Cum \sqrt{f(y)}$  rule where  $f(y)$  was the number of beds. The total sample of 2,600 residents was proportionally allocated to each of the size strata based on the number of beds in each stratum. The sample was increased to thirty residents when a size stratum had an initial sample size of less than thirty residents.

<b>Table 1</b>			
<b>Sample Size by Characteristic Stratum</b>			
Aged	Cognitive	Other Rehabilitative	Total
1 931	411	282	2 624

### **5.3 Sample Selection**

The number of institutions selected from a size stratum depended on the amount of sample allocated to the stratum and the size of the institutions within the stratum. In strata comprised of larger institutions, a larger sample of residents was selected from each institution. This reduced the total number of institutions visited. Once the number of institutions to be selected from each size stratum was determined, a systematic sample of institutions was taken from the stratum list with the probability of selection proportional to size (PPS). Size was determined by the number of long-term beds. It was possible that the listing indicated a head office for several smaller institutions. In this case, a listing of all of the institutions under this head office was obtained and two were selected—the largest (in terms of beds) and another randomly selected using PPS sampling. A list of long-term residents was compiled for each selected institution and a systematic sample of residents was chosen from this list just prior to interviewing.

## **6. Data Collection**

### **6.1 Questionnaire Design and Data Collection Method**

The NPHS Institutional component questions were designed to be conducted by personal interview using paper and pencil. Telephone interviews were acceptable when a proxy respondent could not be contacted in person.

The administrator of the institution or a contact within the institution determined which of the selected residents required a proxy interview because of illness or incapacity. The proxy respondent could be a relative, a staff member, or a volunteer at the institution. Proxy respondents completed 57% of the interviews (40% were relatives of the resident). A staff member from the institution provided information on each selected resident's use of medications and contact with health professionals.

### **6.2 Test**

A test was conducted in the fall of 1994, before the main survey was implemented in the field. The test was carried out by experienced interviewers. The main objective of the test were to observe respondent reaction to the survey, to obtain an estimate of interview time, and to evaluate the ability of achieving a high response rate.

### **6.3 Interviewing**

Collection took place between January 30 and March 31, 1995. The interviews were conducted by Statistics Canada interviewers.

Prior to collection, all institutions were contacted by telephone by senior interviewers to arrange a meeting between an interviewer and the administrator of the institution. During this liaison visit the interviewer administered a short questionnaire on the policies of the institution and made the sample selection. The residents requiring proxy interviews were determined at this time. The name and telephone number of the next-of-kin were obtained in these cases. The next-of-kin was then phoned and given the option to complete the interview primarily themselves or have it completed by a knowledgeable institutional staff member.

Most interviews were conducted in person. The total interview took an average of 40 minutes for non-proxy and 30 minutes for proxy respondents.

#### **6.4 Supervision and Control**

All interviewers are under the supervision of a staff of senior interviewers. The seniors are responsible for ensuring that interviewers are familiar with the concepts and procedures of the survey. They periodically monitor their interviewers and review their completed documents. The senior interviewers are, in turn, under the supervision of program managers, located in each of the Statistics Canada regional offices.

#### **6.5 Non-Response to the NPHS**

Interviewers were instructed to make all reasonable attempts to obtain interviews with selected residents. Refusals at the institutional level were followed-up by senior interviewers, project managers or by other interviewers to try to convince the institution to participate in the survey.

## **7. Data Processing**

### **7.1 Data Capture**

The programme written for the data capture of the questionnaire prevented out-of-range values from being entered. Editing for correct flow was performed after data capture.

### **7.2 Editing**

After completing an interview, the interviewer reviewed the questionnaire to ensure the skip patterns were correctly followed. Further editing was done at the Regional Offices to check for completeness, legibility and consistency of entries on the questionnaire. This allowed for immediate follow-up.

Head office edits included the verification of the demographic variables and response codes prior to data capture. After data capture, top-down editing was performed on all records to check the skip patterns. With the exception of the Health Utility Index (HUI), no imputation was performed (see Section 8.3).

### **7.3 Coding**

Information from write-in responses was coded to new unique categories. Conditions or health problems causing activity restrictions were coded based on the International Classification of Diseases, Version 9 (ICD-9) or according to the Musculoskeletal Impairment Supplementary Coding Scheme developed for the Canadian Health and Disability Survey. Drugs and medications were coded using the Canadian Anatomical Therapeutic Chemical (ATC) Classification System.

### **7.4 Creation of Derived Variables**

To facilitate data analysis, a number of variables on the file have been derived using responses to the NPHS questionnaire for residents of health institutions. Derived variable names generally start with "DV" and are followed by characters referring to the question number or subject. Details of how these variables were created can be found in Appendix D.

## **7.5 Weighting**

Estimation in a probability sample (such as the NPHS) is based on the principle that each person in the sample "represents," besides himself or herself, several others who are not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 people in the population. In the terminology used here, it can be said that each person has a weight of 50.

The weighting phase calculates the associated weight for each person. This weight appears on the microdata file and must be used to derive meaningful estimates from the survey. For example, the number of individuals who smoke daily (see question SMOK-Q62 in section 9.2.1) is estimated by selecting the records referring to those individuals in the sample having that characteristic. The weights entered on those records are then summed.

Details of the method used to calculate these weights are presented in Chapter 10.

## **7.6 Suppression of Confidential Information**

'Public Use' microdata files differ in many important respects from the survey 'master' files held by Statistics Canada. These differences result from action taken to protect the anonymity of individual survey respondents. 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 Section 9 of this document.

## **8. Data Quality**

### **8.1 Response Rates**

Two separate response rates can be calculated for the Institutional component of the NPHS. The institutional response rate identifies the percentage of in-scope institutions that agreed to allow the survey to be conducted in their facility. The residents could not be interviewed without agreement from the institution. The institutional response rate is calculated by:

$$\begin{aligned} & \frac{\text{Number of selected institutions which agreed to participate}}{\text{Total number of in - scope selected institutions}} \times 100 \\ &= \frac{214}{224} \times 100 \\ &= 95.5\% \end{aligned}$$

The individual response rate identifies the percentage of selected residents from the responding institutions with whom an interview was conducted. It is calculated by:

$$\begin{aligned} & \frac{\text{Number of residents with fully or partially completed interviews}}{\text{Total number of selected residents within the responding institutions}} \times 100 \\ &= \frac{2287}{2444} \times 100 \\ &= 93.6\% \end{aligned}$$

Note: Multiplying the two rates together does not give a meaningful result because different numbers of residents were selected within each institution.

### **8.2 Survey Errors**

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



conditions is called the sampling error of the estimate.

Errors 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 numerous 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 spent to reduce non-sampling errors in the survey. Quality assurance measures were carried out at each step of the data collection and processing cycle to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and the questionnaire, and observation of interviewers to detect the misunderstanding of instructions.

Non-response to the survey is a major source of non-sampling error. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total and partial non-response to the institutional component of the NPHS was small. Partial non-response occurred when the respondent refused to answer a question or could not recall the requested information. Total non-response occurred because the interviewer was unable to contact the proxy-respondent or because of refusal to participate in the survey either at the institutional or individual level. Total non-response was handled by adjusting the weight of the residents who responded to the survey to compensate for those who did not respond.

It is unavoidable 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 that Statistics Canada commonly uses. Users producing estimates from this microdata file are also urged to use these measures.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results. 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 (which is equal to the square root of the variance of the estimate) by the estimate itself. It is expressed as a percentage of the estimate.

For example, suppose that, based on the survey results, one estimates that 12% of residents of health institutions are daily cigarette smokers. This estimate yields a standard error of .0096. The coefficient of variation of the estimate is then calculated as:

$$\left(\frac{.0096}{.12}\right) \times 100\% = 8.28\%$$

For more information on the calculation of CVs for this survey, see Section 11.

### 8.3 Imputation

Imputation was used to derive scores for one variable in the Institutional component of the NPHS. The variable IMPHST denotes the resident's Health Status Index (HST) score. This measure of overall health assesses vision, hearing, speech, mobility, dexterity, emotions, memory, cognition and pain. Overall HST scores ranging from zero to one are calculated based on responses to a series of health status questions<sup>2</sup>. A complete HST score could not be calculated if one or more of the components were not answered. At least one component of the HST was missing for 20.5% of respondents. A form of hot deck imputation was used to impute values for the missing components so that an overall HST could be computed for these cases.

Records where the respondent had answered all components of the HST were used as donors. Twelve donor groups were created. Within each of the three characteristic strata (aged, cognitive, other rehabilitative), the donors were classified into one of four age/sex categories (male/female, aged 0-64/65+). The recipients were classified in the same manner. For each recipient, a partial HST (PHST) value was generated using the components answered. Within the associated donor group, a PHST value was generated for each donor using the same components. The individual from the donor group that had the closest PHST value to the recipient's PHST became the imputation donor for that individual. The missing components in the recipient's record were imputed from the donor's record and an imputed HST value was computed for the recipient. If several potential donors had the same PHST value, then one was selected at random to be the donor. Records having at least one imputed component are identified by IMPFLAG=1.

---

<sup>2</sup> For more information on the calculation of the HST see Appendix D: Derived Variables.

## **9. Guidelines for Tabulation, Analysis and Release**

The following guidelines should be followed when tabulating, analyzing, publishing or otherwise releasing any data derived from the survey microdata files. Figures produced using these guidelines should be consistent with estimates produced by Statistics Canada.

### **9.1 Rounding Guidelines**

The following guidelines should be followed when rounding estimates derived from the microdata files:

- 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 subtotals 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 should be computed from unrounded components (i.e. numerators and/or denominators) and then rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.
- d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.
- e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).
- f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

## **9.2 Sample Weighting Guidelines for Tabulation**

The sample design used for the NPHS Institutional component was not self-weighting. The sampling weights are not identical for all individuals in the sample. When producing simple estimates, including the production of ordinary statistical tables, users must apply the sampling weight.

If proper weights are not used, the estimates derived from the microdata files cannot be considered 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.

### **9.2.1 Definitions of types of estimates: Categorical vs. Quantitative**

Two main types of point estimates of population characteristics can be generated from the microdata file for the NPHS Institutional component.

#### Categorical Estimates:

Categorical estimates (also referred to as estimates of an aggregate) are estimates of the number or percentage of the surveyed population possessing certain characteristics or falling into a defined category. The number of individuals who smoke daily is an example of such an estimate.

#### Example of Categorical Question:

SMOK-Q62 At the present time do you (does . . .) smoke cigarettes daily, occasionally or not at all?

- Daily
- Occasionally
- Not at all

### 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  $\hat{Y} / \hat{X}$  where  $\hat{Y}$  is an estimate of surveyed population quantity total and  $\hat{X}$  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 cigarettes smoked per day by individuals who smoke daily. The numerator is an estimate of the total number of cigarettes smoked per day by individuals who smoke daily. Its denominator is an estimate of the number of individuals who smoke daily.

### Example of Quantitative Question:

SMOK-Q64: How many cigarettes do you (does ...) smoke each day now?

|\_|\_| Number of Cigarettes

## **9.2.2 Tabulation of Categorical Estimates**

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

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

### 9.2.3 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 weight. This quantity is then summed over all records of interest. For example, to obtain an estimate of the total number of cigarettes smoked each day by individuals who smoke daily, multiply the value reported in question SMOK-Q64 by the weight for the record, then sum this value over all records with a response of 'daily' to SMOK-Q62.

To obtain a weighted average of the form  $\hat{Y} / \hat{X}$ , the numerator ( $\hat{Y}$ ) is calculated as for a quantitative estimate and the denominator ( $\hat{X}$ ) is calculated as for a categorical estimate. For example, to estimate the average number of cigarettes smoked per day by individuals who smoke daily:

- a) estimate the total number of cigarettes smoked per day by individuals who smoke daily as described above,
- b) estimate the number of daily smokers by summing the weights of all records with a response of 'daily' to SMOK-Q62, then
- c) divide estimate (a) by estimate (b).

### 9.3 Guidelines for Statistical Analysis

The NPHS Institutional component has a complex design with stratification, multiple stages of selection, and unequal probabilities of respondent selection. 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.

Many analysis procedures found in statistical packages allow weights to be used. However, the meaning or definition of the weight in these procedures is not appropriate in a sample survey framework. Typically the estimates produced by the packages are correct, but the calculated variances are almost meaningless.

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

## 9.4 Release Guidelines

The number of sampled residents contributing to the calculation of the estimate should be determined before releasing and/or publishing any estimate from the microdata file. If this number is less than 30, the weighted estimate should not be released regardless of this estimate's CV. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines below.

### Sampling Variability Guidelines

Type of Estimate	CV (in %)	Guidelines
1. Unqualified	0.0 - 16.5	Estimates can be considered for general unrestricted release. Requires no special notation.
2. Qualified	16.6 - 33.3	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter E (or in some other similar fashion).
3. Not for Release	33.4 or greater	Estimates cannot be released in any form under any circumstances. In statistical tables, such estimates should be deleted and replaced by dashes (--).

The CV is defined as the standard error (equal to the square root of the variance of the estimate), multiplied by 100 and divided by the estimate. See section 11 for more details on calculating the variance.

## 10. Weighting

The weights given to responding members of the institutional survey were based on the probability of selecting the individual as well as any adjustments for non-response within the stratum or institution.

Notation:

$M_h$  - number of beds in stratum  $h$  (from list of hospitals and residential care facilities)

$M_{h,i}$  - number of beds in stratum  $h$ , institution  $i$  (from list of hospitals and residential care facilities)

$n_h$  - number of institutions to be selected from size stratum  $h$

$L_{h,i}$  - actual number of long-term residents in stratum  $h$ , institution  $i$  (obtained during initial visit)

$r_{h,i}$  - number of residents allocated to be selected from stratum  $h$ , institution  $i$

### 10.1 Probabilities of selection

#### 10.1.1 Probability of selecting an institution

Institutions were selected from the frame with probability proportional to the number of beds. Therefore, the probability of selecting an institution  $i$  in most cases was:

When a head office was selected (see Section 5.3 for more details) the probability was:

$$n_h \times \frac{M_{h,i}}{M_h}$$

where  $P_{h,i,j}$  was the probability that an institution  $j$  under the authority of head office  $i$  is

$$n_h \times \frac{M_{h,i}}{M_h} \times P_{h,i,j}$$

selected. For the largest institution under  $i$ ,  $P_{h,i,j}=1$ .

For the other  $j$ 's where  $\mathcal{N}$  consists of all of the institutions under head office  $i$ , excluding the largest one.

$$P_{h,i,j} = \frac{M_{h,i,j}}{\sum_{j \in i'} M_{h,i,j}}$$



### **10.1.2 Probability of selecting a resident within an institution**

Once an institution had been selected, each resident had an equal probability of selection:

$$\frac{r_{h,i}}{L_{h,i}} \text{ if } L_{h,i} \geq r_{h,i}$$

or  $1$  if  $L_{h,i} < r_{h,i}$

## **10.2 Weight Calculations and Adjustments**

At this point, initial weights can be calculated at both the institutional and residential level. However, there may be non-response at both levels. Adjustments have to be made at both levels to account for those units that do not respond.

### **10.2.1 Initial Institutional Weights**

Institutional weights correspond to the number of institutions represented by the sampled institution. The **initial institutional weight** is equal to the inverse of the probability of selecting the institution.

### **10.2.2 Institutional Non-response Weight Adjustment**

If interviewing did not take place at a selected in-scope institution, then an adjustment is made to the other institutions within the same size stratum to account for the non-responding institution. This adjustment is equivalent to:

$$\frac{\text{number of responding and non - responding institutions}}{\text{number of responding institutions}}$$

Multiplying this weight adjustment by the initial institutional weight gives the **final institutional weight**.

### **10.2.3 Initial Personal Weights**

An **initial personal weight** can be calculated as the final institutional weight multiplied by the inverse of the probability of selecting a resident within the institution.

#### **10.2.4 Personal Non-response Weight Adjustment**

In some cases certain selected residents did not respond to the questionnaire. An additional adjustment has to be made to the weights to compensate for the non-respondents. This adjustment is similar to the institutional non-response adjustment.

$$\frac{\textit{number of responding and non - responding selected residents}}{\textit{number of responding selected residents}}$$

This adjustment is made at the institutional level.

Multiplying the initial personal weight by the personal non-response weight adjustment gives the **final personal weight** that appears on the file. Since this survey focussed on individuals and not institutions themselves, the final institutional weight does not appear on the file.

## 11. Variance Calculations

The institutional component of the NPHS uses a well-known, simple variance formula to compute the variances and the CVs of estimates. It assumes that institutions are selected with unequal probabilities and with replacement. In reality, the institutions were selected without replacement, that is, once selected, an institution could not be chosen a second time. This was done for operational reasons rather than for variance improvement so the impact of assuming sampling with replacement should be negligible.

A variance computation program written in SAS is provided as part of this microdata package. This program can be used to calculate variances for means and totals. The formulas used for calculating the variances for a total Y or a ratio R=Y/X are:

$$V(\hat{Y}_{total}) = \sum_{h=1}^N \frac{\sum_{i=1}^{n_h} (\hat{Y}_{h,i} - \hat{Y}_h)^2}{n_h (n_{subh} - 1)} \quad V(\hat{R}) = \frac{1}{\hat{X}^2} \sum_{h=1}^N \frac{\sum_{i=1}^{n_h} (\hat{Y}_{h,i} - \hat{R}_h \hat{X}_{h,i})^2}{n_h (n_h - 1)}$$

where  $\hat{Y}_h$  is the stratum h estimate for a response variable Y based on all of the respondents in stratum h

$\hat{Y}_{h,i}$  is the stratum h estimate for a response variable Y based on all of the respondents in stratum h, institution i

N is the number of strata

$n_h$  is the number of sampled institutions in stratum h

$\hat{X}_h$  is the stratum h estimate for the ratio denominator variable X based on all of the respondents in stratum h

$\hat{X}_{h,i}$  is the stratum h estimate of the ratio denominator variable X based on all of the respondents in stratum h, institution i

$\hat{X}$  is the overall estimate of the ratio denominator variable X

$\hat{R}_h$  is the ratio of  $\hat{Y}_h / \hat{X}_h$

## **11.1 Running the variance program**

The SAS program included with the microdata file allows the user to calculate means and totals with minimal work. The following outlines the steps of the program:

STEP 1: In the data step under 'STEP 1' the user identifies the variables for which he/she wants totals and/or ratios. For totals, a 0/1 variable is assigned to each characteristic of interest. Likewise, for ratios, a 0/1 variable is defined for both the numerator and denominator of the ratio. In the example, three totals are being computed and the 0/1 variables are defined as tot1-tot3. In the ratio example, the variables are num1-num4 and denom1-denom4 where num1/denom1 identifies a ratio requiring variance estimates. Maintain the naming convention tot1-tot $n$ , num1-num $m$  and denom1-denom $m$  as the program automatically uses these names. In the keep statement at the end of this step, change the tot $n$ , num $m$  and denom $m$  variables to indicate the number of totals or ratios computed.

Quantitative estimates are calculated in a similar manner. The only difference is that the 0/1 variable is replaced by a quantity variable, where the quantity represents the value of the characteristic for the respondent.

STEP 2: In the proc format statement, the user can assign descriptive names to replace the totals and ratio names generated by the program. "Totfrmt" defines the names for totals and "ratfrmt," the names for ratios. For example, in totfrmt, 2='Popn 65+' since tot2 is calculating the estimate and variance of the total population aged 65 and over.

STEP 3: In this step, the user simply has to change the array references so that the correct number of totals, numerators and denominators are shown. For example, if the user has defined two totals and three ratios in STEP 1 then the array statements would then read:

```
array totarray{*}tot1-tot2;  
array numarray{*}num1-num3;  
array denarray{*}denom1-denom3;
```

The program can then be run like any other SAS program.

Note: The program calls a different subroutine when calculating totals or ratios. If only totals or only ratios are calculated it is not necessary to run both subroutines. At the end of the program there are two lines:

```
%totals;  
%rates;
```

To save time the unnecessary subroutine can be "commented out" by surrounding the appropriate statement with `/*` and `*/`. For example, if only totals are being calculated then:

```
%totals;  
/* %rates; */
```

Only the subroutine associated with variances for totals will be called.

In this example, it is not necessary to define the `num1-numm` and `denom1-denomm` variables in STEP 1 since no rates are calculated. Likewise, if the `%totals` line is "commented out," the `tot1-totn` variables are not required. Simply remove the reference to these variables from the `keep` statement in STEP 1.

Running the program using the variables in the example produces output similar to this:

Estimates, Variances and CVs of Totals					1
OBS	DESCRIPTION	TOTAL	VARIANCE	CV	
1	Total Popn	227841.79	22325494.69	2.07	
2	Popn 65+	186496.38	22765493.02	2.56	
3	Women 80+	99343.07	15047102.4	3.9	
Estimates, Variances and CVs of Rates					2
OBS	DESCRIPTION	RATE	VARIANCE	CV	
1	Close Staff Members 65+	0.39964	0.00039088	4.95	
2	English 65+	0.56058	0.00031905	3.19	
3	French 65+	0.15752	0.00012785	7.18	
4	English/French 65+	0.10677	0.00009991	9.36	