Catalogue no. 12M0011GPE

1996 General Social Survey, Cycle 11: Social and Community Support

Public Use Microdata File Documentation
and User's guide

Data in many forms . . .

Statistics Canada disseminates data in a variety of forms. In addition to publications, both standard and special tabulations are offered. Data are available on the Internet, compact disc, diskette, computer printouts, microfiche and microfilm, and magnetic tape. Maps and other geographic reference materials are available for some types of data. Direct online access to aggregated information is possible through CANSIM, Statistics Canada's machine-readable database and retrieval system.

How to obtain more information

Inquiries about this publication and related statistics or services should be directed to: Dissemination, General Social Survey, Housing, Family and Social Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone (613) 951-5979) or to the Statistics Canada Regional Reference Centre in:

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(403) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

You can also visit our World Wide Web site: http://www.statcan.ca

Toll-free access is provided for all users who reside outside the local dialling area of any of the Regional Reference Centres.

National enquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Order-only line (Canada and United States)	**1 800 267-6677**

How to order publications

Statistics Canada publications may be purchased from local authorized agents and other community bookstores, the Statistics Canada Regional Reference Centre, or from:

Statistics Canada
Operations and Integration Division
Circulation Management
120 Parkdale Avenue
Ottawa, Ontario
K1A 0T6

Telephone: (613) 951-7277
Fax: (613) 951-1584
Toronto (credit card only): (416) 973-8018
Internet: order@statcan.ca

Standards of service to the public

To maintain quality service to the public, Statistics Canada follows established standards covering statistical products and services, delivery of statistical information, cost-recovered services and services to respondents. To obtain a copy of these service standards, please contact your nearest Statistics Canada Regional Reference Centre.

Statistics Canada
Housing, Family and Social Statistics Division

1996 General Social Survey, Cycle 11: Social and Community Support

Public Use Microdata File Documentation and User's Guide

Published by authority of the Minister responsible for Statistics Canada

January 1998

Price: Canada: $ 75 per issue
United States: US$75 per issue
Other countries: US$75 per Issue

Catalogue no. 12M0011GPE

Ottawa

La version française de cette publication est disponible sur demande (no. 12M0011GPF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing co-operation involving Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Table of Contents

1. Introduction

This package is designed to enable interested users to access and manipulate the microdata file for the eleventh cycle of the General Social Survey (GSS), conducted from February through December, 1996. It contains information on the objectives, methodology and estimation procedures, as well as guidelines for releasing estimates based on the survey.

Appendix A contains the Approximate Variance Tables. A topical index of variables, data dictionary and record layout for the file are available in Appendices B, C and D. The questionnaire is found in a separate document entitled "Cycle 11 Questionnaire Package".

2. Objectives of the General Social Survey

Increased pressure during the past decade, to operate more efficient government funded programmes, has led to a related increase in the information needed for policy formulation, programme development and evaluation. Many of these needs could not be filled through existing data sources or vehicles because of the range or periodicity of the information required. The two primary objectives of the GSS aim at closing these gaps. These objectives are: to gather data on social trends in order to monitor temporal changes in the living conditions and well-being of Canadians; and to provide immediate information on specific social policy issues of current or emerging interest. The GSS is a continuing programme with a single survey cycle every second year.

To meet the stated objectives, the data collected by the GSS are made up of three components: Classification, Core and Focus.

Classification content consists of variables which provide the means of delineating population groups and for use in the analysis of Core and Focus data. Examples of classification variables are age, sex, marital status, language, place of birth, and income.

Core content is designed to obtain information which monitors social trends or measures changes in society related to living conditions or well-being. Cycle 11 marks the first GSS with social support as the core content. Health, the core subject matter in Cycles 1 and 6 was due for repetition in Cycle 11. Social support replaced it, however, as the introduction of the National Population Health Survey in 1994 eliminated the need to collect health data. Social support is not a new topic for the GSS as it received coverage in the first (core=health) and fifth (core=family) cycles.

Focus content is aimed at the second survey objective of GSS. This component obtains information on specific policy issues which are of particular interest to certain federal departments or other user groups. In general, focus content is not expected to be repeated on a periodic basis.

The focus content of Cycle 11 collected information on tobacco use and was sponsored by Health Canada.

3. Definitions, Objectives and Content of Cycle 11

3.1 Definitions

Below are the definitions of terms frequently used throughout this documentation.

Temporary difficult times - A short-term condition lasting or expected to last less than six months including birth of a child, short-term health problems, moving or changing jobs, accident or the death of someone close, etc.

Long-term health or physical limitations - Any condition lasting or expected to last more than six months and which can be either chronic or permanent.

Roster (roster member) - Information was collected about the persons and/or organizations who assisted the respondent in his/her everyday activities due to the respondent's long-term health or physical limitations, or with emotional support or checking up. Information was also collected about the persons and/or organizations whom the respondent assisted due to the receiver's long-term health or physical limitations, or with emotional support or checking up. This data was used to establish the help roster.

Care givers - The roster members (persons/organizations) who assisted the respondent due to the respondent's long-term health or physical limitations or with checking up or emotional support.

Care receivers - The roster members (persons/organizations) whom the respondent assisted due to the receiver's long-term health or physical limitations or with checking up or emotional support.

Help or care- Paid or unpaid assistance provided with everyday activities.

Instrumental activities of daily living - 1) child-care; 2) meal preparation and clean-up; house cleaning; laundry and sewing; house maintenance and outside work; 3) shopping for groceries or other necessities; transportation; banking and bill paying; and 4) personal care.

Informal and formal care - Informal care is defined as the performance of tasks by family and friends, without pay, that helps maintain or enhance people's independence. Formal care is defined as the performance of these same tasks by a paid employee/worker or though a governmental or non-governmental organization.

Dyad - A "care relationship" between two persons, that is between the respondent and the roster member. It involves help related to a long-term health problem or with emotional support or checking up. For example, if person A (the respondent) gives help, as defined above, to person B (the roster member), we can say that we have an A-->B dyad. If B (the roster member) also gives help to A (the respondent), we have a second - and different - dyad, B-->A.

3.2 Objectives and Scope

The objectives and scope of Cycle 11 were: to determine the nature of the help received and provided; to understand the dynamic between an individual's social network and help received and provided; and to identify unmet needs and the reasons for the needs. The 1996 GSS focused on help given or received during temporary difficult times or out of necessity due to long-term health or physical limitations in daily activities either inside or outside the household.

The content of Cycle 11 can be found in a separate document entitled "Cycle 11 Questionnaire Package". The package contains the content, as well as methodology, data collection and sample of the survey. A copy of the questionnaire is in Appendix B of the package and a roadmap of the file is in Appendix D.

4. Survey and Sample Design

Data for Cycle 11 of the GSS were collected monthly from February 1996 to December 1996 inclusive. The sample was evenly distributed over the 11 months to represent the seasonal variation in the information gathered with two exceptions. First, an additional sample was added of approximately 1,250 seniors aged 65 and over (sponsored by the Senior's Directorate of Health Canada) and 700 seniors over-sampled from the province of Quebec (sponsored by the Quebec Bureau of Statistics). These supplemental interviews were drawn from the Labour Force Survey (LFS) rotate-outs[1]. Second, approximately 25% of the regular sample was also drawn from the LFS rotate-outs and was restricted to seniors aged 65 and over, thereby obtaining more reliable estimates from this group.

Most of the sample was selected using the Elimination of Non-Working Banks technique of Random Digit Dialling (RDD). Responses were obtained from 12,756 respondents. The response rate for Cycle 11 was 85.3% A description of these methods is provided in Section 4.3. The target population is discussed in Section 4.1. Stratification used in the survey design is outlined in Section 4.2.

4.1 Target Population

The target population for the GSS was all persons 15 years of age and over residing in Canada, excluding:

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

In the survey, all respondents were contacted by telephone. Households without telephones were therefore excluded; however, persons living in such households represent less than 2% of the target population. Survey estimates have been adjusted (i.e., weighted) to account for persons without telephones. The tacit assumption is that, given the small number of people without telephones, their characteristics are not different enough from those of the rest of the target

[1] The LFS follows a rotating panel sample design, in which households remain in the sample for six consecutive months, and each month a panel is replaced after completing its six month stay in the survey. Rotate-outs refer to the households that have been replaced by new households.

population to have an impact on the estimates. Since no one without a telephone is in the sample, this assumption cannot be verified using GSS data. It has, however, been looked at with data from the Labour Force Survey.

4.2 Stratification

In order to carry out 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 another represented the non-CMA areas.

There were two exceptions to this general rule:

- Prince Edward Island has no CMAs and so did not have a CMA stratum
- Montreal and Toronto were each separate strata.

There were slight changes from previous GSS cycles in the allocation of the sample to the various strata. First, with the supplementary sample from the LFS, for some of the CMA strata, no LFS units could be used until May as they were being used by other surveys. Second, Quebec had a buy-in starting in June. Also, the allocation to the various strata of the RDD portion changed slightly because it is based on the total populations of the strata, and these change slightly, relative to each other, every year.

4.3 Elimination of Non-working Banks RDD Design

The Elimination of Non-Working Banks (ENWB) sampling technique is a method of Random Digit Dialling in which an attempt is made to identify all working banks[2] for an area (i.e., to identify all banks containing at least one number that belongs to a household). Thus, all telephone numbers within non-working banks are eliminated from the sampling frame.

For each province, lists of telephone numbers in use were purchased from the telephone companies and lists of working banks were extracted. Each bank was assigned to a stratum within its province.

A special situation existed in Ontario and Quebec because some small areas are serviced by independent telephone companies rather than by Bell Canada. Lists of the phone numbers in use in some of these areas in Quebec were obtained from the Info-Direct company . The area code prefixes for the remaining areas were found by matching the Bell and Info-Direct files with a file of all area codes and prefixes. Area code prefixes from Ontario and Quebec and not on the Bell or Info-Direct files were identified. All banks within these area code prefixes were generated and added to the sampling frame.

In each stratum, a simple random sample without replacement of phone numbers was selected by choosing a simple random sample with replacement of banks from the frame, and then randomly generating the last two digits for each bank to obtain the phone number. The entire monthly

[2] A bank of telephone numbers is a set of 100 numbers with the same first eight digits (i.e., the same Area Code-Prefix-Bank ID). Thus 613-951-9180 and 613-951-9192 are in the same bank, but 613-951-9280 is in a different bank.

sample of telephone numbers was produced before the first day of interviewing for the month. Therefore, a prediction had to be made as to the percentage of numbers dialled that would reach a household (this percentage is known as the "hit rate"). Hit rates from Cycle 10 were used to estimate the hit rates for the Cycle 11 RDD sample, and hit rates from the Cycle 9 LFS sample were used to estimate the hit rate for the Cycle 11 LFS sample (Cycle 10 had no LFS sample).

For Cycle 11 of the GSS, 57.7% of the numbers dialled reached households. The breakdown was a hit rate of 48.1% for the RDD sample and a hit rate of 93.9% for the LFS sample. An attempt was made to conduct a GSS interview with one randomly selected person from each household.

5. Collection

As in the 1994 and 1995 General Social Surveys, data for Cycle 11 were collected using Computer Assisted Telephone Interviewing (CATI) using Computer-Assisted Survey Execution System software (CASES). With CATI, the survey questions appeared on a computer monitor. The interviewer asked the respondent the questions, and entered the responses into the computer as the interview progressed. Built-in edits resulted in fewer processing steps and better quality data. CATI methodology also eliminated the need for paper and pencil questionnaires. Although shown in the questionnaire found in the document “Cycle 11 Questionnaire Package”, skips are built into CATI and do not appear on the screen. As a result, the forms in Appendix A and B of the questionnaire package were produced as reference documents only. In Cycle 11, the CATI system provided the interviewer with two main "components" which can be imagined to represent two paper questionnaires.

QUESTIONNAIRE	AGE GROUP	TITLE
GSS 11-1	All age groups	Survey Control Form
GSS 11-2	Age 15 and over	Social and Community Support Questionnaire

A GSS 11-1 control form was completed for each telephone number generated in the sample. When a private household was contacted, all household members were enumerated and basic demographic information (e.g., age and sex) was collected for everyone. A computer algorithm randomly selected an eligible household member to answer the questionnaire. For most cases, a household member was eligible if aged 15 or older, but for some cases, only household members aged 65 or over were eligible (see Section 4.3 for more information).If the person selected could not be interviewed due to health reasons, a proxy could be interviewed.Data collection involved two possible questionnaires with respondents interviewed in the official language of their choice. A pencil and paper pilot test was conducted in Montreal and Winnipeg in late September and October 1995 using abbreviated questionnaires to determine if inclusion rates were high enough to allow the focus of the survey to be long-term health or physical limitations. The questionnaires, the procedures and the CATI system were then field tested in Montreal and Winnipeg in late November and early December 1995. Data collection began in February 1996 and continued

through December 1996. All interviewing took place using centralized telephone facilities in four of Statistics Canada's regional offices with calls being made from approximately 09:00 until 21:00, Monday to Saturday inclusive. The four regional offices were: Halifax, Montreal, Winnipeg and Vancouver. Interviewers were trained by Statistics Canada staff in telephone interviewing techniques using CATI, survey concepts and procedures in a four day classroom training session. The majority of interviewers had computer and telephone interviewing experience.

It would be too lengthy to include all the survey manuals as part of this documentation package. However, more information can be obtained from the dissemination officer (see Section 11). Shown below is a list of the manuals used in the survey:

Content Manual
CASES Manager's Guide
CATI Interviewer's Manual
GSS Cycle 11 Social and Community Support Questions and Answers

6. Processing

The following is an overview of the processing steps for Cycle 11 of the GSS.

6.1 Data Capture

Using CATI, responses to survey questions were entered directly into computers as the interview progressed. The CATI data capture program allowed a valid range of codes for each question and built-in edits, and automatically followed the flow of the questionnaire. The data were transmitted to Ottawa electronically.

6.2 Edit and Imputation

All survey records were subjected to computer edits throughout the course of the interview. With CATI, built-in edits identified invalid or inconsistent information as the interview progressed. As a result, such problems could be immediately resolved with the respondent.

The system ensured that branchings were correct, that the values were valid, and that each question had a response. For Questionnaire 11-1, the only responses checked were the respondent's age and sex. For Questionnaire 11-2, the CATI system performed checks throughout the interview. In cases where the interviewer was unable to correct the errors detected by the system, he/she was allowed to skip the edit stage and leave the problem for Head Office to solve later.

The Head Office edit system performed the same kind of checks as the CATI system, as well as verifications of greater complexity. For example, where data were missing or incorrect, either non-response codes were assigned to the records or, in some cases, the data were imputed from other responses on the questionnaire. Where data for a particular question were inconsistent with previous responses, the information entered last was usually considered correct. For instance,

if a filter question led to a number of mutually exclusive branches, and if there were data for more than one branch, then the answer to the filter question was considered correct, the data from the branch associated with that answer were retained and the remaining data were assigned the value "N/A". Non-responses were due to one of three possible causes: the question was not asked because the answer to a previous question led to a different branch; the question was not asked because the respondent refused to answer a preceding question; or the respondent refused to answer the question. In the first case, the question was considered N/A, and the corresponding codes – 0, 97, 997, 9997 and so on – were assigned. In the second case, since the respondent refused to answer the filter question, all subsequent questions left unanswered because of that refusal were also coded as refusals (code 9, 99, 999, ...). Refusal codes were also assigned in the third case, since the respondent refused to answer the question.

Non-response was not permitted for items required for weighting, such as age, sex and number of telephone lines. In the case of the age variable, the procedure used to select the respondent ensured that a response would be present. By contrast, values were imputed in the rare cases where the number of residential telephone lines (DVTEL) was missing. DVTEL was assigned a value of one (1) when the respondent failed to provide the information. Values for certain other variables were imputed on the basis of other respondent or household characteristics.

6.3 Coding

Several questions allowing write-in responses had the write-in information coded into either new unique categories, or to a listed category if the write-in information duplicated a listed category. Where possible (e.g., occupation, industry, education, country of birth and religion), the coding followed the standard classification systems as used in the Census of Population.

6.4 Creation of Combined and Derived Variables

A number of variables on the file have been derived by using items found on the GSS 11-1 and 11-2 Questionnaires. Derived variable names generally start with DV and are followed by characters referring to the question number or subject. In some cases, the derived variables are straightforward and involve collapsing of categories. In other cases, several variables have been combined to create a new variable. The data dictionary provides comments indicating the origin of these variables.

6.5 Amount of Detail on Microdata File

In order to guard against disclosure, the amount of detail included on this file is less than is available on the master file retained by Statistics Canada. Variables with extreme values have been capped and information for some variables have been aggregated into broader classes (e.g., occupation, religion, country of birth).

The measures taken to cap, group or collapse data have been indicated in the data dictionary. Variables with a very limited number of observations or refering to small population areas have been excluded from the file.

7. Estimation

When a probability sample is used, as was the case for the GSS, the principle behind estimation is that each person selected in the sample 'represents' (in addition to himself/herself) several other persons not in the sample. For example, in a simple random sample of 2% of the population, each person in the sample represents 50 persons in the population. The number of persons represented by a given person in the sample is usually known as the weight or weighting factor of the sampled person.

For analysis of GSS Cycle 11 data, it is necessary to use the weighting factor (WGHT_FNL), which was placed on the microdata file. As described above, this factor represents the number of persons in the population that the record represents. It refers to the number of times a particular record should contribute to a population estimate. For example, to estimate the number of adults whose mother was born in Canada (G10), the value of WGHT_FNL is summed over all records with this characteristic.

The process of deriving the weighting factor, WGHT_FNL, is described in Section 7.1.

7.1 Weighting

We view each Cycle of the General Social Survey as being composed of a number of independent surveys – one per collection month. Wherever possible, therefore, we weighted each monthly survey independently so that the data collected for each month would contribute equally to the estimates. Where the sample size for a particular month was not large enough, the records for two or more months were grouped at certain stages of the weighting process.

Moreover, Cycle 11 featured a large supplementary sample of telephone numbers belonging to households with at least one member aged 65 or over that had just rotated out of the Labour Force Survey. Thus, the Cycle 11 sample consisted of two parts derived from two independent sources, and it had to be weighted so that the respondents in each part were appropriately represented. Hence, we weighted each part of the sample independently (as if the other part did not exist) and "combined" the two parts at the very end of the process, adjusting their weights by the method described in section III below.

I) RDD Households

1) Basic Weight Calculation

Each household (responding and non-responding) in the sample was assigned a weight equal to the inverse of its probability of selection. This weight was calculated independently for each stratum-month group as follows:

$$\frac{\text{Number of possible telephone numbers within the stratum-month group}}{\text{Number of sampled telephone numbers within the stratum-month group}}$$

(The total number of possible telephone numbers for a stratum is equal to the number of working banks for a stratum times 100).

2) Non-Response Adjustment

Weights for responding households were adjusted to represent non-responding households. This was done independently within each stratum-month group. Records were adjusted by the following factor:

$$\frac{\text{Total of the basic household weights of all sampled households in each stratum-month group}}{\text{Total of the basic household weights of responding households in each stratum-month group}} = \frac{\text{Number of sampled households within the stratum-month group}}{\text{Number of responding households within the stratum-month group}}$$

Non-responding households were then dropped.

3) Multiple Telephone Adjustment

Weights for households in the sample with more than one residential telephone number (i.e., not used for business purposes only) were adjusted downwards to account for the fact that such households had a higher probability of being selected. The weight for each household was divided by the number of residential telephone numbers that serviced the household.

4) Person Weight Calculation

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

5) Adjustment of Person Weight to External Totals

The person weights were adjusted several times using a raking ratio procedure. This procedure ensured that, based on the survey's total sample, estimates produced of the sizes of strata or of province-age-sex groups would match external references. The two sets of groupings used for these adjustments were stratum-month and province-age-sex. The age groupings used were:

15-19, 20-24, 25-29, 30-34, 35-39, 40-44,
45-49, 50-54, 55-59, 60-64, 65-69, 70+.

Sample sizes were too small to allow the province-age-sex adjustments to be made at the survey month level. Also due to small sample sizes, there were cases where two or more adjacent age groups in the same province-sex group or two adjacent months in the same stratum were collapsed before the adjustments were made.

The reference totals for the stratum-month adjustments were one eleventh of the population projections for each month (because Cycle 11 collected information during 11 months). The reference totals for the province-age-sex adjustments were the average of the population projections for each month. At each stage in the adjustment process the weights were adjusted by the factor:

$$\frac{\text{reference total for group}}{\text{sum of person weights for group}}$$

The groupings used for the adjustments alternated between province-age-sex and stratum-month until the weights converged.

It should be noted that persons 15 years and over living in households without telephone service are included in the reference totals even though they were not sampled.

II) LFS Households

For maximum simplicity, we wanted all households of interest (those with one member aged 65 or over) within a given stratum to have the same basic weight. In other words, we expected each of those households, which we will refer to as 65+ households, to have the same probability of being selected for the GSS sample. Since the households were taken from the group just rotated out of the LFS, selection of a 65+ household for inclusion in the GSS sample was a two-stage process: first the household had to be chosen for the LFS; then it had to be one of the LFS households selected for the GSS. To achieve our goal of equal selection probability within strata, we chose households from the LFS with a probability inversely proportional to the probability they had of being included in the LFS. It should be noted that this was impossible in Quebec in June and subsequent months, as the size of the supplementary sample was increased substantially by the purchase of additional units. There were not enough LFS households with at least one person 65 or over in the Montréal stratum, and we were forced to take all of them. As a result, their selection probability was 1 instead of being inversely proportional to their LFS selection probability. We compensated by increasing the number of households selected in the other two strata in Quebec. Once again, because there were too few 65+ households in the LFS, we were unable to select LFS households in those two strata with a probability inversely proportional to their LFS selection probability, and we had to use a uniform selection probability. Hence, the selection probability of 65+ households in the Quebec strata is proportional to their probability of being selected for the LFS, and their basic weight for the GSS is proportional to their LFS weight.

The weights for the supplementary sample were adjusted in the same way as the RDD sample weights. First we adjusted for non-response; then we discarded the non-responses. We multiplied the weight by the number of eligible household members, i.e. by the number of persons aged 65 or over. Note that we did not divide the weight by the number of telephone numbers in the household, since the household could not be included in the LFS sample a second time with another telephone number. We also made sure that the estimated sizes of the month-strata and

the age-sex groups matched the external totals using the iterative quotient method. In this case, of course, the only age groups were 65-69 and 70+.

III) "Combined" Weighting of the Two Samples

For the population under 65 years of age, a given variable can be estimated only with the RDD sample and its weighting. For the population 65 and over, on the other hand, a variable can be estimated either with the RDD sample and its weighting or with the LFS sample and its weighting. One way of "combining" these two estimates is to take the sum of the first estimate multiplied by a factor we will call alpha and the second estimate multiplied by 1 minus alpha. We select the value of alpha so as to minimize the variance of the combined estimator. This occurs when alpha is equal to the variance of the second estimate divided by the sum of the variances of the two estimates. The smaller the variance of the first estimate (relative to the variance of the second estimate), and hence the greater its relative precision, the more important it becomes in the combined estimate. Consequently, we chose a number of variables that we considered the most important in the survey, and for each province separately, we calculated the variance of the two estimates for each variable as well as the value of alpha. Clearly, we had to choose a general value for alpha that did not depend on the variable being estimated: we selected the median of the alphas as the general provincial alpha.

Once this process of combining estimates was complete, respondent weights were adjusted as follows. The weights of respondents under age 65 were left unchanged. The weights of respondents aged 65 or over taken from the RDD sample were multiplied by the appropriate provincial alpha. The weights of respondents from the LFS supplementary sample were multiplied by 1 minus the appropriate provincial alpha.

The result of these adjustments was that the size estimates for the month-strata and age-sex groups no longer matched the external totals exactly. This deficiency was corrected by applying the iterative quotient adjustment to the "combined" weights.

7.2 Weighting Policy

Users are cautioned against releasing unweighted tables or performing any analysis based on unweighted survey results. As was discussed in Section 7.1, there were several weight adjustments performed independently to the records of each province. Sampling rates as well as non-response rates varied significantly from province to province. For example, it is known from other surveys that non-respondents are more likely to be males and more likely to be younger. In addition, there was an important oversample of persons aged 65 and over. As a result, in the responding sample, 2.0% were males between the ages of 15 and 19, while in the overall population, approximately 4.2% are males between 15 and 19. Therefore, it is clear that the sample counts cannot be considered to be representative of the survey target population unless appropriate weights are applied.

Contact was made or attempted with 14,957 households during the survey. Of these, 1,311 (8.8%) were non-responding households. The non-responding households included 743 household refusals, 452 households that could not be reached during the survey period, 115 cases where a response could not be obtained due to language difficulties, illness, or other problems,

and 1 case where the household had already been interviewed. An interview was attempted with a adult randomly selected from the eligible household members of the 13,646 responding households. Usable responses were obtained from 12,756 respondents. The difference consists of 386 person-level refusals, 302 persons that could not be reached during the survey period, 201 cases where the interview could not be completed due to language difficulties, illness, or other problems and here too, 1 case where the person had already been interviewed. A response rate of 85.3% was obtained when it is assumed that all of the households for which there was no response were "in scope" (i.e., had at least one eligible member).

7.3 Types of Estimates

Two types of 'simple' estimates are possible from the results of the General Social Survey. These are qualitative estimates (estimates of counts or proportions of people possessing certain characteristics) and quantitative estimates involving quantities or averages. More complex estimation and analyses are covered in Section 7.4.

7.3.1 Qualitative Estimates

It should be kept in mind that the target population for the GSS was non-institutionalized persons 15 years of age or over, living in the ten provinces. Qualitative estimates are estimates of the number or proportion of this target population possessing certain characteristics. The number of adults who were given emotional support (SC90) is an example of this kind of estimate. The simplest estimate to produce is that of the number of people in the target population who would have responded that they received such support if a census had been conducted. This estimate is simply the sum of the weights for those respondents with a value of '1' for question 90: 8,816,219. This estimate does not however adjust for non-response in any way. If we make the assumption that those who either refused to answer the question or who responded 'don't know' have the same distribution as those who responded, then an adjusted estimate can be made. To do this, the proportion of the target population with this characteristic is estimated by ignoring the respondents with 'Not stated' or 'Don't know' answer to question 90 and calculating the ratio of the total of the weights of those respondents who answered 'Yes' or 'No'. This proportion is then multiplied by the size of the target population to produce the final estimate (it should be noted that this adjustment does not have to be done, but it can be if needed) :

$$8{,}845{,}529 = 23{,}604{,}792 \times \frac{8{,}816{,}219}{23{,}526{,}577}$$

The difference in this example is small because only .3% of respondents didn't answer this question. When the proportion of responses that are 'don't know' or 'refused' is higher the differences between the two estimates will be larger.

7.3.2 Quantitative Estimates

Some variables on the 1996 General Social Survey microdata file are quantitative in nature (e.g., number of hours per week a person usually works - H9). From these variables, it is possible to obtain such estimates as the average number of hours per week a person usually works. These

estimates are of the following ratio form:

Estimate (average) = X / Y

The numerator (X) is a quantitative estimate of the total of the variable of interest (for example, number of hours per week person usually works) for a given sub-population (for example, persons who were employed in the past 12 months). X would be calculated by multiplying the person weight WGHT_FNL by the variable of interest when it is known, i.e. not equal to '997' or '999', and summing this product over all records which are in the subpopulation. The denominator (Y) is the qualitative estimate of the number of participants within that subpopulation (those who worked in the past 12 months and for whom the weekly hours of work was known). Y would be calculated by summing the person weight, WGHT_FNL, over all records for persons who reported working in the past 12 months and who were not unable or unwilling to give the number of hours per week they usually worked. The two estimates X and Y are derived independently and then divided to provide the quantitative estimate. The average weekly number of hours persons who worked in the past 12 months worked per week when they were working is then estimated to be:

$$38.51 = \frac{585{,}812{,}610}{15{,}210{,}279}$$

7.4 Guidelines for Analysis

As is detailed in Section 4 of this document, the respondents from the GSS do not form a simple random sample of the target population. Instead, the survey had a complex design, with stratification and multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used.

The GSS used a stratified design with significant differences in sampling fractions between strata. Thus, some areas are over-represented in the sample (relative to their populations) while some other areas are relatively under-represented. This means that the unweighted sample is not representative of the target population.

The survey weights must be used when producing estimates or performing analyses in order to account for this over- and under-representation. While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures often differs from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, 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, or any subset of the data, are rescaled so that the average weight is one (1), then the variances produced by the

standard packages will be more reasonable; they still will not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. This rescaling can be accomplished by dividing each weight by the overall average weight before the analysis is conducted.

For example, if an analysis of the “help respondents gave to people with long-term health problems” is desired, then the following steps are required:

- Calculate the Average Weight equal to the average of WGHT_FNL for all of the records (The total of WGHT_FNL for all 12,756 respondents is 23,604,792, so the average weight is 1,850.49 = 23,604,792 / 12,756)

- For each of these respondents calculate a "working" weight equal to WGHT_FNL / 1,850.49

- Perform the analysis for the “help respondents gave to people with long-term health problems” using the "working" weight.

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

7.5 Methods of Estimation and Interpretation of Estimates

Select: Records with FLAG_1st=1

Person Weight: WGHT_FNL

The basic sampling weight assigned to each sampled individual has been adjusted to reflect the age and sex composition of the various provincial populations as projected by the Labour Force Survey for each month of 1996.

$$\sum_{1}^{12{,}756^{3}} \text{WGHT\_FNL} = 23{,}604{,}792$$

= an estimate of the number of persons 15 years of age and older in the population.

Examples and Interpretation:

(I) 14% (3.2 million) of adult Canadians gave help to someone with a long-term health problem or physical limitation (HGBR_FG=1).

[3] The number of responding households (with one randomly chosen respondent per household).

(ii) 19% (605,000) of female (DVSEX=2) adult Canadians aged 45 to 64 (08<=DVAGE<=11) gave informal help to someone with a long-term health problem or physical limitation (RFR_F_IN=1).

(iii) 12% (1.5 million) of adult Canadians who had a job outside the home (DVH16COL=1) also gave informal help to someone with a long- term health problem or physical limitation (RFR_F_IN=1).
.

(iv) 56% (1.6 million) of adult Canadians who gave informal help to someone with a long-term health problem or physical limitation (RFR_F_IN=1) did not feel burdened by their caregiving duties (F2K=1).

7.6 Using the Household Weight

Select: Records with FLAG_1st=1.

A household weight could be calculated for this file. It is simply the ratio:

$$WGHTHLD = \frac{WGHT\_FNL}{DVELLCOL}$$

Where DVELLCOL is the number of eligible households members 15+ (capped at 5 or more members on the microdata file for confidentiality reasons).

$$\sum_{1}^{12,756} WGHTHLD = 12,156,292$$

= an estimate of the number of households that include someone in the target population

The household weight should be used with variables that represent characteristics of the household rather than of the respondent. Examples of such variables on the microdata file are: DVHHSCOL (household size collapsed) and DVHHINC (household income). The household weight should be used when one wants to estimate the number of households rather than the number of people. For example, to estimate the number of households where a person received help because of their long-term health or physical limitations one would sum the household weights for those cases with DVHR_LTH=1:

7.4% (893,649) of Canadian households have a member receiving help because of their long-term health or physical limitations.[4]

[4] While 7.4% of households had a member receiving help because of their long-term health or physical limitations, 5.9% (1,397,801) of Canadians received help because of their long-term health or physical limitations. This can be interpreted as the proportion of households where a member received help being greater than the proportion of disabled people there was

8. Release Guidelines And Data Reliability

It is important for users to become familiar with the contents of this section before publishing or otherwise releasing any estimates derived from the General Social Survey microdata file.

This section of the documentation provides guidelines to be followed by users. With the aid of these guidelines, users of the microdata should be able to produce figures consistent with those produced by Statistics Canada and that respect the established guidelines for rounding and release. The guidelines can be broken into four broad sections: Minimum Sample Sizes for Estimates, SamplingVariability Policy, Sampling Variability Estimation and Rounding Policy.

8.1 Minimum Sample Size For Estimates

Users should determine the number of respondents on the microdata file which contribute to the calculation of a given estimate. When the number of contributors (i.e., respondents) to the weighted estimate is less then 15 the weighted estimate should not be released regardless of the value of the Approximate Coefficient of Variation.

8.2 Sampling Variability Guidelines

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

Errors that 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 into the CATI system and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

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

A major source of non-sampling errors in surveys is the effect of non-response on the survey

in the population.

results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, a language problem prevented the interview from taking place, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households who responded to the survey to compensate for those who did not respond.

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

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error.

Although the exact sampling error of the estimate, as defined above, cannot be measured from sample results alone, it is possible to estimate a statistical measure of sampling error, the standard error, from the sample data. Using the standard error, confidence intervals for estimates (ignoring the effects of non-sampling error) may be obtained under the assumption that the estimates are normally distributed about the true population value. 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 it is virtually certain that the differences would be less than three standard errors.

Type of Estimate	Coefficient of Variation	Policy Statement
1. Unqualified	0.0 to 16.5%	Estimates can be considered for general unrestricted release.
2. Qualified	16.6 to 33.3%	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. Not for release	33.4% or over	Estimates should not be released in any form under any circumstances. In statistical tables, such estimates should be deleted.

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

Because of the large range in size of estimates that can be produced from a survey, the standard

error is usually expressed relative to the estimate to which it pertains. The resulting measure, known as the coefficient of variation of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate. Before releasing and/or publishing any estimates from the microdata file, users should determine whether the estimate is releasable based on the guidelines shown in the above box.

8.3 Estimates of Variance

Variance estimation is described separately for qualitative and quantitative estimates.

8.3.1 Sampling Variability for Qualitative Estimates

Derivation of sampling variabilities for each of the estimates which could be generated from the survey would be an extremely costly procedure, and for most users, an unnecessary one. Consequently, approximate measures of sampling variability, in the form of tables, have been developed for use and are included in Appendix A "Approximate Sampling Variability Tables".

Variance tables for estimates using WGHT_FNL are provided at the Canada and province levels, as well as for the Atlantic and Prairie Regions.

It should be noted that all coefficients of variation in these tables are approximate and therefore unofficial. Estimates of actual variance for specific variables may be purchased from Statistics Canada. Use of actual variance estimates may allow users to release otherwise unreleasable estimates, i.e., estimates with coefficients of variation in the "Not for Release" range (see the policy regarding the release of the survey estimates on preceding pages). Statistics Canada is currently investigating the feasibility of releasing to GSS microdata file users a feature that would allow them to calculate an exact variance for any variable included on the microdata file. The variance calculation would be done using the bootstrap method. A large number of additional weights, known as bootstrap weights, would be provided for each case. Suppose an exact variance estimate was required for a certain variable. By using the first bootstrap weight for each case, a first bootstrap estimate would be calculated in the exact same way that the estimate for the original variable was found, i.e., the first bootstrap weights would replace the original weights used. The same thing would be done for the second bootstrap weights to produce a second bootstrap estimate, and so on, until all the bootstrap weights were used. The bootstrap estimates calculated in this way constitute a large set of numbers. The variance of this set can be used to calculate an estimate of the variance of the original variable. Please contact Statistics Canada for more information on the bootstrap method for the calculation of exact variances by microdata file users.

The Approximate Variance tables have been produced using the coefficient of variation formula based on a simple random sample and the straightforward expansion estimator. Since estimates for the General Social Survey were based on a complex sample design and the complicated raking ratio estimator alluded to earlier, a factor called the Design Effect was introduced into the variance formula. The Design Effect for an estimate is the actual variance for the estimate (taking into account the design and estimator that were used) divided by the variance that would result if the estimate had been derived from a simple random sample and a simple expansion estimator. The Design Effect used to produce the Approximate Variance Tables has been determined by first calculating Design Effects for a wide range of characteristics and then choosing among these a

conservative value which will not give a false impression of high precision. These Design Effects are specified in the table below.

GENERAL SOCIAL SURVEY CYCLE 11 DESIGN EFFECTS	
Geographic Area	*15+ Population*
Canada	2.04
Newfoundland	1.49
Nova Scotia	1.61
P.E.I.	1.32
New Brunswick	1.54
Atlantic Region	1.66
Quebec	2.20
Ontario	1.61
Manitoba	1.67
Saskatchewan	1.66
Alberta	1.58
Prairie Region	1.65
British Columbia	1.53

8.3.2 Sampling Variability For Quantitative Estimates

Approximate variances for quantitative variables cannot be as conveniently summarized. As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding qualitative estimate (i.e., the number of persons contributing to the quantitative estimate). If the corresponding qualitative estimate is not releasable, then the quantitative total will in general not be releasable.

8.4 Rounding

In order that estimates produced from the General Social Survey microdata file correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates. It is improper to release unrounded estimates, as they imply greater precision than actually exists.

8.4.1 Rounding Guidelines

(1) Estimates of totals in the main body of a statistical table should be rounded to the nearest thousand using the normal rounding technique (see definition in Section 8.4.2).

(2) 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 thousand units using normal rounding.

(3) Averages, proportions, rates and percentages are to be computed from unrounded components and then are to be rounded themselves to one decimal using normal rounding.

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

(5) In instances in which, due to technical or other limitations, a different rounding technique is used, which results in estimates being released which differ from the corresponding estimates produced by Statistics Canada, users are encouraged to note the reason for such differences in the released document.

8.4.2 Normal Rounding

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, the number 8499 rounded to thousands would be 8 and the number 8500 rounded to thousands would be 9.

9. Structure Of File

Normally, the GSS microdata file is structured in such a way that there is one record per respondent and this record includes variables that represent characteristics of the respondent and the household. For Cycle 11, however, this is not the most efficient file structure because the amount of data collected for a respondent depends on the number of roster members (i.e., people the respondent has given help to and/or received help from).

To optimize analysis options for users it was decided to allow the possibility of multiple records per respondent. The number of basic records per respondent is determined by the number of roster members a respondent reported. A roster member is one who provides help to the respondent or receives help from the respondent. Each basic record contains personal level information (screening sections, sections C, F, G, H, I, J), and space for one roster member (sections A-Demographic Information on Non-Household Roster Members, B- Help Received by Respondent, and E- Help Given by Respondent). Typically, only section B or E will be completed unless there were exchanges of help in both directions, for example, exchanges of emotional support.

Each record has a unique record identification (REC_ID) and each respondent is identified by a unique CASE_ID number that is repeated (as part of the person level information) on each of the basic records needed to contain all the roster members for that respondent. There are 27,567 records (unique REC_IDs) on the analysis file, but only 12,756 different respondents (unique CASE_IDs).

Because of this file structure, certain variables or flags must be selected to ensure the population of

interest has been chosen for analysis. "Select" will be used throughout the documentation to indicate this computer operation which creates a sub-file containing only those records that match on the SELECT criteria. For example, to create a working file of all respondents, SELECT all records with FLAG_1ST=1.

Some of the most commonly used flags found on the file, when to use them, and how to use them are described in the following section. In order to understand Section 9.1 and 9.2, one should be familiar with the definitions in Section 3.

9.1 Respondent Level Flags

9.1.1 FLAG_1st

Used when the population of interest is respondents.

FLAG_1ST has two possible values - '1' indicating the first occurrence of the respondent level information - '0' indicating that the record is not the first occurrence of the respondent level information.

FLAG_1ST= 1 must be selected for analysis involving respondent level data in order to avoid multiple counting of respondent level information. The respondent level information is contained on ALL records (FLAG_1ST=1 or FLAG_1ST=0). This information is provided so that tables may be produced for the care giver/care receiver data broken down by respondent level information.

NOTE: FLAG_1ST should not be selected when giver/receiver variables (Sections A, B or E) are the variables to be analysed.

12,756 records have FLAG_1ST=1
14,811 records have FLAG_1ST=0

9.1.2 Type of Care

Used to select the type of care of interest - informal, formal or a mix of formal and informal. This flag summarizes the type of care across all the giver/receiver records for a respondent. Thus, if a respondent gave help to 10 individuals and in 9 of those cases the help was provided on an informal basis while in the 10th the help was provided on a formal basis, this respondent level variable would have a value of '3' indicating a mix of formal and informal help across all the records for that CASE_ID.

RFR_F_IN [5] [6]

[5] This variable is defined as help received from respondent for one or more of the following activities: childcare; meal preparation; house cleaning; household maintenance; grocery shopping; transportation; banking and bill paying; or personal care.

6 For a breakdown of formal and informal care received from respondent including checking up and emotional support (childcare, meal preparation, house cleaning, household maintenance, grocery shopping, transportation, banking and bill paying, personal care, checking up or emotional support) use RFR_FIAL.

RFR_F_IN (received help from respondent - formal and informal) has five possible values - '1' indicating only informal help received from respondent - '2' indicating only formal help received from respondent - '3' indicating a mix of formal and informal help received from respondent - '0' indicating no one received help from respondent - '9' indicating did not state if help was received from respondent.

If analysis involves differentiating between informal and formal help received from respondent then FLAG_1ST=1 and RFR_F_IN=1 must be selected for analysis of informal care (or '2' for formal care or '3' for a mix).

1,480 records have RFR_F_IN=1
159 records have RFR_F_IN=2
42 records have RFR_F_IN=3
11,057 records have RFR_F_IN=0
18 records have RFR_F_IN=9

GTR_F_IN [7] [8]

GTR_F_IN (help given to respondent - formal and informal) has five possible values - '1' indicating only informal help given to respondent - '2' indicating only formal help given to respondent - '3' indicating a mix of formal and informal help given to respondent - '0' indicating no one gave help to respondent - '9' indicating did not state if help was given to respondent.

If analysis involves differentiating between informal and formal help given to respondent then FLAG_1ST=1 and GTR_F_IN=1 for analysis of informal care (or '2' for formal care or '3' for a mix).

829 records have GTR_F_IN=1
485 records have GTR_F_IN=2
451 records have GTR_F_IN=3
10,946 records have GTR_F_IN=0
45 records have GTR_F_IN=9

9.1.3 Reason for giving/receiving help[9]

Detailed information was collected on help given by the respondent and help received by the

[7] This variable is defined as help given to respondent for one or more of the following activities: childcare; meal preparation; house cleaning; household maintenance; grocery shopping; transportation; banking and bill paying; or personal care.

[8] For a breakdown of formal and informal care given to respondent including checking up and emotional support (childcare, meal preparation, house cleaning, household maintenance, grocery shopping, transportation, banking and bill paying, personal care, checking up and emotional support) use GTR_FIAL.

[9] Help with one of more of the following activities: childcare; meal preparation; house cleaning; household maintenance; grocery shopping; transportation; banking and bill paying; or personal care.

respondent for two different reasons - due to long-term health or physical limitations or due to temporary difficult times. Flags or variables have been created that can be selected for analysis of the help given or help received due to long-term health or physical limitations or due to temporary difficult times.

HGBR_FG

HGBR_FG (help given by respondent for long-term health problems) has three possible values - '1' indicating the respondent gave help to someone with a long-term health problem or physical limitation - '0' indicating the respondent did not give help to someone with a long-term health problem or physical limitation - '9' indicating the respondent did not state if they gave help to someone with a long-term health problem or physical limitation

FLAG_1ST=1 and HGBR_FG=1 can be selected for analysis of respondents as care givers due to receivers long-term health problems.

11,031 records have HGBR_FG=1
1,699 records have HGBR_FG=0
26 records have HGBR_FG=9

NOTE: The variable DVHG_LTH can be used instead of HGBR_FG as they both measure "help given by respondent" and have the same values.

DVHG_LTH

DVHG_LTH (help given by respondent for long-term health problems) has three possible values - '1' indicating the respondent gave help to someone with a long-term health problem or physical limitation - '0' indicating the respondent did not give help to someone with a long-term health problem or physical limitation - '9' indicating the respondent did not state if they gave help to someone with a long-term health problem or physical limitation

FLAG_1ST=1 and DVHG_LTH=1 can be selected for analysis of respondents as care givers due to receivers long-term health problems.

11,031 records have DVHG_LTH =1
1,699 records have DVHG_LTH =0
26 records have DVHG_LTH =9

NOTE: The variable HGBR_FG can be used instead of DVHG_LTH as they both measure "help given by respondent" and have the same values.

DVHR_LTH

DVHR_LTH (help received by respondent for long-term health problems) has four possible values - '1' indicating the respondent received help because of a long-term health problem or physical limitation - '2' indicating the respondent did not receive help because of a long-term health problem or physical limitation - '8' or '9' respondent did not know or did not state or if they received help because of a long-term health problem or physical limitation.

FLAG_1ST=1 and DVHR_LTH=1 can be selected for analysis of respondents as care receivers

due to their long-term health problems.

1,810 records have DVHR_LTH=1
10,937 records have DVHR_LTH=2
1 record has DVHR_LTH=8
8 records have DVHR_LTH=9

DVHG_TDT
DVHG_TDT (help given by respondent because of temporary difficult times) has three possible values - '1' indicating the respondent gave help because of someone else's temporary difficult time - '2' indicating the respondent did not give help because of someone else's temporary difficult time - '9' respondent did not state if they gave help because of someone else's temporary difficult time

FLAG_1ST=1 and DVHG_TDT=1 can be selected for analysis of respondents as care givers due to receiver's temporary difficult times.

904 records have DVHG_TDT=1
11,827 records have DVHG_TDT=2
25 records have DVHG_TDT=9

DVHR_TDT
DVHR_TDT (help received by respondent because of temporary difficult times) has four possible values - '1' indicating the respondent received help because of temporary difficult times - '2' indicating the respondent did not receive help because of temporary difficult times - '8' or '9' respondent did not know or did not state if they received help because of a temporary difficult time

FLAG_1ST=1 and DVHR_TDT=1 can be selected for analysis of respondents as care receivers due to their temporary difficult times.

652 records have DVHR_TDT=1
12,088 records have DVHR_TDT=2
2 records have DVHR_TDT=8
14 records have DVHR_TDT=9

9.1.3 Type of activity

Used to select the particular activity for which help was received by respondent.

HC_FGALL
HC_FGALL (house cleaning) has four possible values - '1' indicating the respondent received help with house cleaning because of the respondent's long-term health or physical limitations- '0' indicating that the respondent did not receive help with house cleaning - '8' or '9' indicating the respondent did not know or did not state if they received help with house cleaning

(This variable can also be interpreted as at least one caregiver (to respondent) on the

respondent's roster gave the respondent help with house cleaning, B21=1, i.e, at least one roster member had B21=1.)

FLAG_1ST=1 and HC_FGALL=1 can be selected for analysis of respondents who received help with house cleaning.

1,143 records have HC_FGALL=1
11,579 records have HC_FGALL=0
2 records have HC_FGALL=8
32 records have HC_FGALL=9

A similar variable exists for each activity CC_FGALL (received help with childcare), MP_FGALL (received help with meal preparation), HM_FGALL (received help with house maintenance), GS_FGALL (received help with grocery shopping), TR_FGALL (received help with transportation), BB_FGALL (received help with banking or bill paying) and PC_FGALL (received help with personal care) and can be used the same way as HC_FGALL.

9.2 Roster Level Flags

Certain variables represent characteristics of the respondent, while others represent characteristics of the roster members. By NOT selecting FLAG_1ST=1 a user has access to all the records on the file for information on the care givers/care receivers (i.e., roster members). This can also be referred to as the 'roster level' data.

9.2.1 Roster member received help from respondent by group of activities

Selected when the population of interest is the roster members who received help because of a long-term health or physical limitation or with checking up or emotional support (or respondents as care givers). Used to select the group of activities for which roster members received help from respondent.

R2_MP_FG
R2_MP_FG (roster member received help with meal preparation and/or house cleaning and/or house maintenance from respondent) has two possible values - '1' indicating roster member received help with meal preparation and/or house cleaning and/or house maintenance from respondent and '0' - indicating roster member did not receive help with meal preparation and/or house cleaning and/or house maintenance from respondent

R2_MP_FG=1 can be selected for analysis of help received by roster members with meal preparation and/or house cleaning and/or house maintenance in order to obtain each record, as more than one roster member can receive help from respondent.

1,022 records have R2_MP_FG=1
26,545 records have R2_MP_FG=0

A similar variable exists for each group of activities R1_CC_FG (roster member received help from respondent with childcare), R3_GS_FG (roster member received help from respondent with

grocery shopping and/or transportation and/or banking and bill paying), R4_PC_FG (roster member received help from respondent with personal care), R5_CU_FG (roster member received help from respondent by checking up) and R6_ES_FG (roster member received help from respondent by providing emotional support) and can be used the same way as R2_MP_FG.

9.2.2 Roster member gave help to respondent by group of activities

Selected when the population of interest is the roster members who gave help to respondent because of a long-term health or physical limitation or with checking up or emotional support (or respondents as care receivers). Used to select the group of activities for which roster members gave help to respondents.

G2_MP_FG (roster member gave help with meal preparation and/or house cleaning and/or house maintenance to respondent) has two possible values - '1' indicating roster member gave help with meal preparation and/or house cleaning and/or house maintenance to respondent and '0'- indicating roster member did not give help with meal preparation and/or house cleaning and/or house maintenance to respondent.

G2_MP_FG=1 must be selected for analysis involving help given by roster member with meal preparation and/or house cleaning and/or house maintenance in order to obtain each record, as more than one person can give help to respondent.

2,090 records have G2_MP_FG=1
25,477 records have G2_MP_FG=0

A similar variable exists for each group of activities G1_CC_FG (roster member gave help to respondent with childcare), G3_GS_FG (roster member gave help to respondent with grocery shopping and/or transportation and/or banking and bill paying), G4_PC_FG (roster member gave help to respondent with personal care), G5_CU_FG (roster member gave help to respondent by checking up) and G6_ES_FG (roster member gave help to respondent by providing emotional support) and can be used the same way as G2_MP_FG.

9.3 Examples from the Record Layout of the File

Record Layout

CASE_ID	REC_ID	FLAG_1ST	GIVRECID	G2_MP_FG	R2_MP_FG	DVA1COL	Number of records for this CASE_ID (respondent)
12020191	00257	1	01	0	1	18	1
12020194	00258	1	01	1	0	04	6
12020194	00259	0	02	1	0	04	-
12020194	00260	0	03	0	0	03	-

Record Layout							
12020194	00261	0	04	0	0	24	-
12020194	00262	0	05	0	0	24	-
12020194	00263	0	06	0	0	24	-
12020196	00264	1	01	0	0	05	1
12020198	00265	1	01	0	0	24	1
12020247	00331	1	97	0	0	97	1

The above table shows 6 variables from several case_ids from the file.

Example 1:
CASE_ID 12020191 is an example of a respondent with 1 roster member. The roster member did not give help with meal preparation and/or house cleaning and/or house maintenance to respondent (G2_MP_FG=0). The roster member received help with meal preparation and/or house cleaning and/or house maintenance from respondent (R2_MP_FG=1). The roster member was the respondent's sister-in-law (DVA1COL=18)[10].

Example 2:
CASE_ID 12020194 is an example of a respondent with 6 roster members. Two roster members gave help with meal preparation and/or house cleaning and/or house maintenance to respondent (G2_MP_FG=1). Four roster members did not give help with meal preparation and/or house cleaning and/or house maintenance to respondent (G2_MP_FG=0). None of the 6 roster members received help with meal preparation and/or house cleaning and/or house maintenance from respondent (R2_MP_FG=0). Two roster members were the respondent's daughters (DVA1COL=04), 3 were the respondent's close friends (DVA1COL=24), and 1 was the respondent's son (DVA1COL=03) .

Example 3:
CASE_IDs 12020196 and 12020198 are both examples of respondents with 1 roster member. The roster members did not receive help (G2_MP_FG=0) or give help (R2_MP_FG=0) with meal preparation and/or house cleaning and/or house maintenance to respondent, however, they gave or received help with tasks not reported in this table. One respondent's roster member was his/her father (DVA1COL=05) and the other respondent's roster member was his/her close friend (DVA1COL=24).

Example 4:
CASE_ID 12020247 is an example of a respondent with no roster members. Therefore, all roster information is not applicable (G2_MP_FG=0, R2_MP_FG=0, DVA1COL=97 and GIVRECID=97.)

[10] Refer to variable DVA1COL in the data dictionary for value labels for this variable.

10. What can you do with the file?

There are three main types of analyses that can be done with this file: 1) respondent level analysis; 2) respondent level analysis that involves looking across the roster members and then creating a respondent level variable; and 3) roster level analysis. Sections 10.1 and 10.2 provide "caregiving" (in other words, respondents as caregivers)examples of these three types of analyses.

Dyads

As well as collecting respondent level information, this survey collected information on "dyads". A dyad is defined as a "care relationship" between two persons, that is between the respondent and the roster member. It involves help due to a long-term health or physical limitation or with emotional support or checking up. For example, if person A (the respondent) gives help, as defined above, to person B (the roster member), we have an A-->B dyad. If B (the roster member) also gives help to A (the respondent), we have a second - and different - dyad, B-->A. It should be stressed that the number of activities, as well as the time spent, are not implicitly considered in a dyad; a dyad only indicates that there is a care relationship between two persons.

When put in context of the file structure (Section 9.0), this means that, most often, each record represents a dyad. There are, however, two exceptions: 1) when a respondent did not give or receive any help because of a long-term health problem or with emotional support or checking up, therefore, there are no dyads for that record and 2) when the respondent gave help and received help from the same person, therefore, there are two dyads for the same record.

If one wants to analyse dyad data, a note is in order. Counting the weight once, there are over 48,000,000 exchanges of help. In other words, each Canadian gives help to or receives help from an average of 2.0 people (48,605,738 dyads / 23,604,792 Canadians) [1].

[1] When exchanges are in both directions, that is when the roster member both gives help to the respondent and receives help from the respondent, the weight should be counted twice.

10.1 Respondent Level Analysis

To examine the total number of caregivers in Canada, one should NOT take into account each dyad, as a person might be counted more than once since a caregiver could provide care to more than one person. For example, if a caregiver helped her mother, father and aunt it was counted as one caregiver but three "care relationships" or dyads.

Example 1:
If one wants to determine the average age of caregivers, again one should not take into account

each dyad, rather, one should identify the caregivers and then compute the average age. To do this one would select FLAG_1ST=1 (because the population of interest is respondents) and HGBR_FG=1 (respondents as care givers due to receivers long-term health problems) and then compute the average age using the respondent level variable DVAGE.

Example 2:
If one wants to know the percentage of caregivers giving help to their parent(s), again the respondent level, rather than the dyad level approach, should be taken. However, this time one would have to look across all of the respondent's roster members to see if each one of the dyads had involved the respondent giving care to his or her parents. A new respondent level variable could then be created that would indicate if the respondent helped his or her parents (there was a dyad involving help to parents) or if the respondent did not help (there was no dyad involving help to parents).

Once the new respondent level variable has been created, one would select FLAG_1ST=1 (because the population of interest is respondents) and HGBR_FG=1 (respondents as care givers due to receivers long-term health problems). The new "caregiving to parents" variable would be divided by the total number of caregivers.

10.2 Dyad Level Analysis

To examine the total number of caregiving relationships or dyads in Canada, one should take into account each dyad, as one now wants to accumulate the person weight more than once if the respondent provided help to more than one person. For example, if a caregiver helped her mother, father and aunt one would want to count the three caregiving relationships not the one caregiver.

Example 1:
If one wants to know the percentage of caregiving relationships or dyads that involve a parent as receiver, then the dyad level, rather than the respondent level approach, should be taken.

To do this one would select R1_CC_FG=1 or R2_MP_FG=1 or R3_GS_FG=1 or R4_PC_FG=1 or R5_CU_FG=1 or R6_ES_FG=1 (because the population of interest is the roster members who received help from respondents). The result is then obtained by dividing the number of dyads where the relationship involves a father or a mother (DVA1COL=05 or DVA1COL=06) by the total number of caregiving dyads (when respondent is care provider).

5,929,484 / 27,591,817=21%

11. Additional Information

Additional information about this survey can be obtained from the individuals listed below. Data from the survey are available through published reports, special request tabulations, and this microdata file. The microdata file is available from the Housing, Family and Social Statistics Division of Statistics Canada at a cost of **$1500.00.** Tabulations can be obtained at a cost that will reflect the resources required to produce the tabulation.

Sample Selection Procedures, Weighting and Estimation
Dave Paton
Household Survey Methods Division
(613) 951-1467

or
Paul Matthews
Household Survey Methods Division
(613) 951-1480

Subject Matter, Data Collection and Data Processing
Dissemination Officer
Housing, Family and Social Statistics Division
(613) 951-5979

APPENDIX A

Approximate Variance Tables

APPROXIMATE VARIANCE TABLES

By using the Approximate Variance Tables and the following rules, users should be able to determine approximate coefficients of variation for aggregates (totals), percentages, ratios, differences between totals and differences between ratios.

This publication gives one table or set of tables for each of the following types of estimate produced:
1) Estimates for the total adult population 15 and over
2) Estimates for the population aged 65 and over
3) Estimates for dyad data, with the respondent as a giver
4) Estimates for dyad data, with the respondent as a receiver

Although the same weight variable is used in all cases (WGHT_FNL), its values sometimes differ considerably from one subpopulation to the other, for an estimate of the same value.

In the first case (total adult population), tables are shown for Canada, the provinces as well as for the Atlantic region (Newfoundland, Prince Edward Island, Nova Scotia and New Brunswick) and the Prairie region (Manitoba, Saskatchewan and Alberta). In the three others, only Canada tables are shown in this guide, but provincial and regional tables are available upon request. Furthermore, in order to help users, a summary table is shown on the next page, defining the cutoff values to determine if the estimates can be released "Unqualified", "Qualified", or if they should not be released at all. For each type of estimate, the "Unqualified" column gives the number which corresponds to the 16.6% coefficient of variation, and the "Qualified" column gives the value for the 33.3% coefficient of variation. For example, for the population aged 65 and over, in the Atlantic region, if the estimate is 12,000 or more, it can be released without qualification. If it is between 3,000 and 11,000, it can be released, but should be qualified, while no estimates less than 3,000 should be released.

As noted in Section 8.2, estimates having a coefficient of variation (cv) of more than 33.3% are not releasable. In addition, as mentioned in 8.1, each estimate should be derived from at least 15 respondents in order to be released, regardless of the approximate coefficient of variation.

Users should ensure that the Approximate Variance Table or cutoff table used corresponds with the type of estimate being considered (Chapter 7 and Appendix A contain information on the correct production and use of these types of estimates).

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

The coefficient of variation (cv) depends only on the size of the estimated aggregate itself. In the Approximate Variance Table, locate the estimated aggregate in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks across to the first figure encountered. This figure is the estimated coefficient of variation.

GENERAL SOCIAL SURVEY CYCLE 11
MINIMUM RELEASEABLE POPULATION 65+ AND DYAD ESTIMATES (000s)

	POPULATION 65+		*DYADS-RESPONDENT AS GIVER*		*DYADS-RESPONDENT AS RECIPIENT*	
Geographic Area	*UNQUALIFIED*	*QUALIFIED*	*UNQUALIFIED*	*QUALIFIED*	*UNQUALIFIED*	*QUALIFIED*
Canada	40	10	200	45	150	40
Atlantic Region	12	3	75	19	70	18
Newfoundland	9	3	50	13	50	14
Prince Edward Isl	5	2	24	7	35	9
Nova Scotia	13	4	85	23	80	20
New Brunswick	12	3	60	15	60	16
Quebec	35	9	200	45	200	50
Ontario	55	13	250	60	250	60
Prairie Region	19	5	125	30	125	30
Manitoba	16	4	80	20	70	18
Saskatchewan	14	4	70	18	75	20
Alberta	22	6	125	30	150	40
British Columbia	35	9	150	40	125	35

*

Example 1:

A user estimates that in Canada 232,202 females aged 15 years and over describe their state of health as poor compared to other people their age (question J24=5). How does the user determine the approximate coefficient of variation for this estimate?

Refer to the approximate variance table for Canada level estimates. The estimated aggregate does not appear in the left-most column (the "Numerator of Percentage" column), so it is necessary to use the closest figure, namely 250,000. The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry for that row, in this case 12.2%. This cv falls within the range of cv's for "Unqualified" estimates (i.e. 0.0% -16.5%, see section 8.2) allowing the estimate to be released without restriction.

Rule 2: <u>Estimates of Percentages or Proportions Possessing a Characteristic</u>

The coefficient of variation of an estimated percentage or proportion depends on both the size of the percentage or proportion and the size of the total upon which the percentage is based. Estimated percentages or proportions are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, at least when the denominator is not the total population. (Note that in the tables the cv's decline in value reading from left to right).

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

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

Example 2 (a):

A user estimates that in Canada 1.94% of females aged 15 years and over describe their state of health as poor compared to others their age (question J24=5). This is the expression of the estimate obtained in Example 1 as a percentage of all females aged 15 years and over in Canada. How does the user determine the approximate coefficient of variation for this estimate?

Refer to the approximate variance table for Canada level estimates. Because the estimate is a percentage which is based on a subset of the population covered by the table, it is necessary to use both the percentage (1.94%) and the numerator portion of the percentage (232,202) to determine the approximate coefficient of variation. Since the numerator does not appear in the left-most column (the 'Numerator of Percentage' column), it is necessary to use the figure closest to it, namely 250,000. Similarly, the percentage estimate does not appear among the column headings, so it is necessary to use the figure closest to it, namely 2.0%. The figure at the intersection of the row and column selected, namely 12.2%, is the coefficient of variation. This cv falls within

the range of cv's for 'Unqualified' estimates (i.e. 0.0% - 16.5) allowing the estimate to be released without restriction.

Example 2 (b):

A user estimates that in Canada, among women aged 15 and over who consider their state of health as poor compared to others their age, 19.3% are usually free of pain or discomfort (DVPAS_FC = 1). How does the user determine the approximate coefficient of variation for this estimate?

Refer to the approximate variance table for Canada level estimates. Because the estimate is a percentage which is based on a subset of the population covered by the table, it is necessary to use both the percentage (19.3%) and the numerator portion of the percentage (44,723) to determine the approximate coefficient of variation. Since the numerator does not appear in the left-most column (the 'Numerator of Percentage' column), it is necessary to use the figure closest to it, namely 45,000. Similarly, the percentage estimate does not appear among the column headings, so it is necessary to use the figure closest to it, namely 20%. The figure at the intersection of the row and column selected, namely 25.9%, is the coefficient of variation. This cv falls within the range of cv's for 'Qualified' estimates (i.e. 16.6% - 33.3%) allowing the estimate to be released with qualification.

Rule 3: <u>Ratios</u>

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. In the case where the numerator is not a subset of the denominator, the coefficient of variation of the ratio of two estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately. That is, the standard deviation of a ratio

$$R = X / Y$$

is

$$sd(R) = (R) * (cv(X)^2 + cv(Y)^2)^{½}$$

The coefficient of variation of R is approximately:

$$cv(R) = sd(R) / R$$
$$= (cv(X)^2 + cv(Y)^2)^{½}$$

This formula will tend to overstate the error if X and Y are positively correlated and understate the error if X and Y are negatively correlated.

Example 3:

A user estimates that in Canada among females aged 15 years and over, 232,202 describe their state of health as poor compared to other people their age (question J24=5) and 3,083,401 describe their state of health as excellent as compared to others their age. The user is interested in the ratio of females

describing their health as excellent versus those describing their health as poor. How does the user

determine the approximate coefficient of variation for this ratio estimate?

The numerator of the ratio estimate is 3,083,401 (X). Using Rule 1 (refer to Example 1), the coefficient of variation for this estimate is determined to be 3.3% cv(X). The denominator of the ratio estimate is 232,202 (Y). Again using Rule 1, the coefficient of variation is determined to be 12.2% cv(Y). Using Rule 3, the coefficient of variation of the ratio estimate is

$$cv = (0.033^2 + 0.122^2)^{½}$$

$$= 0.126$$

Therefore at the Canada level, the ratio of females who describe their health as excellent versus females who describe their health as poor is 3,083,401/232,202 or 13.28 to 1. The coefficient of variation of this estimate is 12.6%, and so the estimate can be released without restriction.

Rule 4: <u>Differences Between Totals or Percentages</u>

The standard deviation of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard deviation considered separately. That is, the standard deviation of a difference:

$$d = X - Y$$

is

$$sd(d) = ((X * cv(X))^2 + (Y * cv(Y))^2)^{½}$$

The coefficient of variation of d is approximately:

$$cv(d) = sd(d) / d$$

This formula is accurate for the difference between separate and uncorrelated characteristics but is only approximate otherwise.

Example 4:

A user estimates that in Canada, among those 15 years and over, 1.94% (X) of females describe their state of health as poor compared to others their age and 2.35% (Y - an estimated 272,700) of males describe their state of health as poor compared to other people their age. The user is interested in the difference between these two estimates. How does the user determine the approximate coefficient of variation for the estimate of the difference?

From Rule 2 (refer to example 2(a)), the coefficient of variation for the female estimate is 12.2%. The coefficient of variation for the male estimate is also 12.2%.

The difference between the estimates is 0.41%. Using Rule 4, standard deviation of the difference between the estimates is

$$sd = ((0.0194 \times 0.122)^2 + (0.0235 \times 0.122)^2)^{1/2}$$
$$= 0.0037$$

and the coefficient of variation is

$$cv = \frac{0.0037}{0.0041}$$
$$= 0.9024$$

Therefore the coefficient of the difference between the estimates is 90.24% and the estimate should not be released.

Rule 5: Difference of Ratios

In this case, Rules 3 and 4 are combined. The cv's of the two ratios are first determined using Rule 3, and the cv of their difference is found using Rule 4.

Confidence Limits

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

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the 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 are generally expressed as two numbers: one below the estimate and one above the estimate, that is, more explicitly, as $(\hat{Y}-k, \hat{Y}+k)$ where k is determined from the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate, and then using the following formula to convert to a confidence interval (CI):

$$CI_Y = \{\hat{Y} - (t)(\hat{Y})(\alpha\hat{Y}), \hat{Y} + (t)(\hat{Y})(\alpha\hat{Y})\}$$

where $\alpha\hat{Y}$ is the determined coefficient of variation of $\hat{Y}$ and

t = 1 if a 68% confidence interval is desired
t = 1.6 if a 90% confidence interval is desired
t = 2 if a 95% confidence interval is desired
t = 3 if a 99% confidence interval is desired

Example 5(a):

An estimated 504,902 persons described their state of health as poor (question J24=5) as compared to other people their age. This estimate has an approximate coefficient of variation of 8.5% (obtained from the 500,000 row, left-most column, of the Canada approximate variance table). The 95% confidence interval for this estimate is thus:

CI = {504,902 - (2)(504,902)(0.085), 504,902 + (2)(504,902)(0.085)}
= {504,902 -85,833, 504,902 + 85,833}
= {419,069, 590,735}

With 95% confidence it can be said that between 419,069 and 590,735 persons aged 15 years and over in Canada, describe their state of health as poor, compared to other people their age.

Example 5(b):

An estimated 1.94% of females aged 15 years and over in Canada described their state of health as poor when compared to other people their age or .0194 expressed as a proportion. From Example 2a this estimate has an approximate coefficient of variation of 12.2%. A 95% confidence interval for this estimate (expressed as a proportion) is

CI = {.0194 - (2 x .0194 x .122, .0194 + (2 x .0194 x .122)}
= {0.0147, 0.0241}

With 95% confidence it can be said that between 1.47% and 2.41% of females aged 15 years and over in Canada, describe their state of health as poor, compared to other people their age.

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

T-test

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

Let X_1 and X_2 be sample estimates for 2 characteristics of interest. Let the standard error of the difference $X_1 - X_2$ be σ_d

$$\text{If } t = \frac{X_1 - X_2}{\sigma_d} \text{ is between -2 and 2,}$$

then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 5% level (Note:at the 1% level, values of -3 and +3 should be used, etc.).

Example 6:

A user wishes to test at the 5% level of significance the hypothesis that at the Canada level there is no difference between percentage estimates of males and females who describe their state of health as poor, as compared to other people their age. From Example 4 the estimate of the standard deviation of the difference between the estimates is 0.0041.

$$\text{Hence } t = \frac{0.0194 - 0.0235}{0.0037}$$
$$= -1.11$$

Since t = -1.11 is between -2 and 2, there is no evidence to reject the hypothesis at the 5% significance level.