c7microe

THE 1992 GENERAL SOCIAL SURVEY - CYCLE 7
TIME USE

Public Use Microdata File Documentation
and
User's Guide

August 1993

Aussi disponible en fran‡ais
The General Social Survey - Cycle 7

Public Use Microdata File Documentation and User's Guide

Table of Contents

⚥1.   INTRODUCTION

This document is designed to enable interested users to access and manipulate the microdata file for the seventh
cycle of the General Social Survey, conducted from January through December, 1992. It contains information
on the objectives, methodology and estimation procedures as well as guidelines for releasing estimates based
on the survey.

This document gives a description on how to correctly use the microdata files. Appendices D, G and I contain
the data dictionaries for the Main File, the Time Use Summary File and the Time Use Episode File,
respectively, which is the major part of this documentation package. The survey questionnaires are contained
in Appendix B, and the variance tables are in Appendix A.

Excluding Appendix B and the two tables at the end of Appendix Q, this package is available in machine
readable form.

2.   OBJECTIVES

Increased pressure, over the last five to ten years, to operate more efficient government funded prog1xmmes,

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 General Social Survey (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 program with a single survey cycle each 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, education, 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 7 was the second cycle to return to previous core content:
time use.   Most of the core content of Cycle 7 repeated Cycle 2, conducted in 1986.

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 for Cycle 7 was
participation in sport and cultural activities sponsored by various government departments and cultural
organizations and Sport Canada.

3.   POPULATION

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

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

The survey employed Random Digit Dialling (RDD), a telephone sampling method. 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 (weighted) to represent persons without telephones.

4.   SURVEY DESIGN

Data for Cycle 7 of the GSS was collected monthly from January to December, 1992. The sample was evenly
distributed over the 12 months to counterbalance seasonal variation in the information gathered.   It was then

divided equally among the seven days of the week.   The sample was selected using the Elimination of
Non-Working Banks technique of Random Digit Dialling (RDD).   A description of this method is provided
in Section 4.2.   Stratification procedures used in the survey design are outlined in Section 4.1, and Section 4.3
discusses sample sizes.

4.1 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 CMA and so did not have a CMA stratum
     - Montreal and Toronto were each separate strata
♀4.2   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 for an area (i.e., to identify all banks with at least
one 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.   The area code prefixes for these areas were identified by
matching the Bell file with a file of all area codes and prefixes.   Area code prefixes from Ontario and Quebec
and not on the Bell file were identified.   All banks within these area code prefixes were generated and added
to the sampling frame. Use of the Waksberg method was not possible for these areas since it requires that an
accurate population estimate be available for the survey area.   Such an estimate was not available for the parts
of Ontario and Quebec not covered by Bell.

A random sample of telephone numbers was generated in each survey month for each stratum (from the
working banks).   An attempt was made to generate the entire sample of telephone numbers on the first day
of interviewing.   Therefore, a prediction of the percentage of numbers dialled that would reach a household
had to be made (this is known as the "hit rate").   The hit rate for January, the first survey month, was
estimated using information from previous RDD surveys.   Hit rates for subsequent months were revised as
required based on January's experience.

For Cycle 7 of the GSS, 46.3% of the numbers dialled reached households.   An attempt was made to conduct
a GSS interview with one randomly selected person from each household.

4.3 Sample Size

The sample consisted of 12,765 households and a GSS Selection Control Form (GSS 7-1) was completed for
each. The GSS 7-1 listed all household members and collected the following basic demographic information:
age, sex, marital status and relation to the household reference person. A person 15 years of age or older
was randomly selected from households which were part of the RDD sample. A GSS 7-2 was then completed
for these selected persons.

The GSS 7-2 collected the following types of information: general questions related to time (Section A); the
time use diary (Section B); a child care diary for respondent's with children less than 15 years of age living
in the household (Section C); information on unpaid help supplied by the respondent to the househsold, as well
as, unpaid help provided by the respondent to persons not living in the household (Section D); perceptions of
time (Section E); educational, cultural and recreational activities of the respondent (Section F); organized sport
(Section G); main activity of the respondent (Section H); main activity of the respondent's partner or spouse,
if applicable (Section J); background socio-economic questions for classification purposes (Section K); and a
final section asking respondents for detailed contact information for follow-up (Section M).

A response was obtained from 9,815 of the selected households, yielding an 77% response rate.

5.   COLLECTION

Two questionnaires were used to conduct the interviews: the Selection Control Form (GSS 7-1) and the main
questionnaire, the GSS 7-2. Respondents were interviewed in the official language of their choice.   The French
and English versions of the main questionnaire were identical with the exception of question K13 "What
language did you first speak in childhood?".   Respondents were not asked if they still understood the language
in which they were being interviewed.

Questionnaires and procedures were field tested in July 1991 in Halifax and Montreal.   Data collection began
the third week of January 1992 and continued through the third week of December 1992.   The sample was
evenly distributed over the 12 months. All interviewing took place using centralized telephone facilities in five
of Statistics Canada's regional offices with calls being made from approximately 9 a.m. until 9:30 p.m.,
Monday to Saturday inclusive.   The five regional offices were: Halifax, Montreal, Sturgeon Falls, Winnipeg
and Vancouver.   Interviewers were trained by Statistics Canada staff in telephone interviewing techniques,
survey concepts and procedures in a two day classroom training session.   The majority of interviewers had
previous telephone interviewing experience.

It would be too lengthy to include all the survey manuals as part of this documentation package, however, they
can be purchased (see Chapter 10). Shown below is a list of the manuals used in the survey:

GSS 7-3: Procedures Manual
GSS 7-4: Interviewer's Manual
GSS 7-6: Interviewer's Exercise Book
GSS 7-7: Senior Interviewer's Exercise Book.
GSS 7-8: Interviewer's Training Guide

6.   PROCESSING

The following is an overview of the processing steps for Cycle 7 of the GSS.

6.1 Data Capture

Data from the survey questionnaires were entered directly into mini-computers at Statistics Canada's regional
offices (ROs) and subsequently transmitted to Head Office in Ottawa. The data capture program allowed for
a valid range of codes for each question and automatically followed the flow of the questionnaire.

6.2 Edit and Imputation

All survey records were subjected to an exhaustive computer edit to identify and correct invalid or inconsistent
information on the questionnaires. For the second time, a batch edit system was implemented for use in the
Regional Offices. The system mainly edited the GSS 7-2 for possible flow errors, values out of range and
missing values. Edits on the GSS 7-1 were limited to a few edits for the respondent's age and sex. In the event
the interviewer was unable to correctly resolve the detected errors, it was possible for the interviewer to bypass
the edit and forward the data to head office for resolution.

Head office edits performed the same checks as the batch edit system as well as more detailed edits. Records
with missing or incorrect information were assigned non-response codes or corrected from other information
from the respondent's questionnaire. In most cases editing was 'bottom-up', meaning that specific related
information following a question with a branching pattern was employed to ensure the branching was correct.
For example, question D5 'Do you pay anyone, on a regular basis, to help out with cleaning your house?' was
edited in relation to question D6 'How often do you use this service?' Correlation edits were also conducted,
for example, question K11 of the Time Use Questionnaire was 'In what year did you first immigrate to
Canada?'. This question was edited in relation to the respondent's age as derived from question K12 'What
is your date of birth?'. These edits ensured that the information was consistent and complete among questions.

Due to the nature of the survey, imputation was not appropriate for most items and thus 'not stated' codes were
usually assigned for missing data. In some cases, the answer was not known but could be obtained
deterministically by the questions which followed or from information from other

areas of the survey.

Non-response was not permitted for those items required for weighting. Values were imputed in the rare cases
where any of the following were missing: age, sex, and number of residential telephone lines. The imputation
was based on a detailed examination of the questionnaire and the consideration of any useful data such as age
and sex of other household members, and interviewer's comments.

DVTEL (number of residential telephone lines) was derived from questions K4 to K9 of the Time Use
Questionnaire (GSS 7-2). When the questionnaire did not contain adequate information to derive DVTEL,
it was assigned a value of one (1).

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, language, country of birth for the respondent, as well as, the respondent's mother
and father, and religion), the coding followed either the standard classification systems as used in the Census
of Population. The coding of the daily activities was done in the Regional offices within 24 hours of data
collection by the senior interviewers.

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 7-1 and GSS 7-2
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 dictionaries provide comments indicating the origin of these variables. The coding of the daily activities
was done in the Regional offices by the senior interviewers within 24 hours of data collection.

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, industry, country
of birth).

The measures taken to cap or group data have been indicated in the data dictionaries.
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.

Three microdata files were created for the General Social Survey based on information from the Time Use
Questionnaire (i.e. the GSS 7-2): the Main File which contains information from 9,815 respondents who
answered questions on unpaid help, cultural activities and organized sport, the Time Use Summary File which
contains information from 8,996 respondents who answered the time use questions and the Time Use Episode
File which contains information describing detailed time use activities for the 8,996 respondents on the Time
Use Summary File as well as the activities of those who refused to complete a full diary. The 8,996
respondents who answered time use questions are a subset of the 9,815 respondents who answered the unpaid
help, cultural activities and organized sport questions.   For a description of the file layouts, contents and
correct interpretation of data on the microdata tape, users should refer to Appendices D, E, F, G, H and I.

The weighting factor on the Main File (FWGHT) was placed on each record to indicate the number of persons
that the record represents. This weighting factor refers to the number of times a particular record should
contribute to a population estimate. For example, the estimate of the number of Canadians 15 years and older
who feel trapped in a daily routine (i.e. E2G = 1) is 7,329,963.   The value of FWGHT is summed over all
records with this characteristic. The weighting process is described in Section 7.1.

Similarly, the Time Use Summary File, has a weighting factor (TIMEWGT) which was placed on each record
to indicate the number of persons that the record represents. The Time Use Summary File weighting process
is the same as the one for the Main File and is described in Section 7.1.

Records on the Time Use Episode File have the same weight as the Time Use Summary File. This file is
structured differently from the Main and the Time Use Summary Files and users should refer to Appendix Q
for the correct methods of using this file.

7.1   Weighting

A self-weighting sample design is one for which the weights of each unit in the sample are the same. The GSS
sample for Cycle 7 was selected using the Elimination of Non-Working Banks (ENWB) sampling technique,
which has such a design, with each household within a stratum having an equal probability of selection.

This probability is equal to:
                    Number of telephone numbers
                    sampled within the stratum

Total number of possible
telephone numbers within the stratum

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

Where possible, each survey month was weighted independently. This was done in an attempt to ensure that
each survey month contributes equally to estimates. If monthly sample sizes were not large enough, two or
more survey monthswere combined in certain steps of the weighting.

1)   Basic Weight Calculation

Each household (responding and non-responding) in the RDD 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:

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

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:

$$\text{Factor 1} = \frac{\text{Total of the household weights of all households in each stratum-month group}}{\text{Total of the household weights of responding households in each stratum-month group}}$$

Non-responding households were then dropped.

3)   Multiple Telephone Adjustment

Weights for households 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.

$$\text{Factor 2} = \frac{1}{\text{Number of non-business telephone numbers}}$$

4)   Person Weight Calculation

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

5)   Regional Office (RO) - Stratum - Month Adjustment

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

Projected Census population count
for the RO-stratum-month

Sum of the person weights for
 the RO-stratum-month

6)   Province - Age - Sex Adjustment

♀The next weighting step was to ratio adjust the weights of to agree with Census projected province-age
group-sex distributions. Census projected population counts were obtained for males and females within the
following seven age groups:

         15-19,   20-24,   25-34,   35-44,
         45-54,   55-64,   65-69   70 +

For each of the resulting classifications the person weights for records within the classification were adjusted
by multiplying by the following ratio:

Projected Census population count
for the province-age group-sex

Sum of the person weights of records
   for the province-age group-sex

where,

                                                    Dec
Projected population count   =   Projected Census population count for
province-age group-sex
                                                    Jan

                                                                        12

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

Also the sample size of some cells did not meet the minimum size requirement. These cells were collapsed
with an adjacent age group cell to meet the requirement.

7)   Province - Day of the week (Designated Day) Adjustment

Time use information was collected from respondents for a selected day of the week so that each day would
have an equal number of respondents. An adjustment was made to the person weights on records within each
province and the selected day of the week, ensuring that population estimates would be consistent with Census

projected population counts. The projected counts for each province should have had an equal number of
respondents. The adjustment was done by multiplying the person weight for each record within the province
- day of the week combination by the following ratio:

Projected Census population count
for the province-day

Sum of the person weights for the province-day

where,

$$\text{Projected population count}^{\text{Dec}}_{\text{Jan}} = \text{Projected Census population count for province}$$

12 * 7

8)   Raking Ratio Adjustments

The weights of each respondent were adjusted several times using a raking ratio procedure. This procedure
ensured that estimates produced for RO-Stratum-Month, Province-Age Group-Sex and Province-Day of the
week totals would agree with the Census projections. This adjustment was made by repeating steps 5), 6) and
7) of the weighting procedures.

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.

Contact was made or attempted with 12,765 households during the survey.   Of these, 1,577 (12.4%) were
non-responding households. The non-responding households included 927 household refusals, 459 households
that could not be reached during the entire survey period ("ring-no-answer" households) and 191 cases where
a response could not be obtained due to language difficulties or other problems. An interview was attempted
with a person randomly selected from the eligible household members of the 11,188 responding households.
From these households, 9,815 usable responses were obtained.   The difference consists of 509 person-level
refusals and 864 cases where the interview could not be completed for some other reason. A response rate of
76.9% 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).

It is known that non-respondents are more likely to be males and more likely to be younger.   In the responding

sample, 3.7% were males between the ages of 15 and 19, while in the overall population, approximately 4.4%
were 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.

7.3 Types of Estimates

The following sections deal with producing estimates from either the Main File or the Time Use Summary
File. For simplicity, only the Main File is referenced, although the techniques can also be applied to the Time
Use Summary File.

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 people (5,522,390) who describe their
state of health as excellent (Question K21) is an example of this kind of estimate. These estimates are readily
obtained by summing the final weights (FWGHT) of the records possessing the characteristic in question.

7.3.2 Quantitative Estimates

Some variables on the General Social Survey microdata file are quantitative in nature (e.g. age). From these
variables, it is possible to obtain such estimates as the average number of weeks worked in the last 12 months
(H13) for males 15 years or older living in Ontario, having worked between 1 and 52 weeks during the last
12 months.   These estimates are of the following ratio form:

$$\text{Estimate (average)} = \frac{X}{Y}$$

The numerator (X) is a quantitative estimate of the total of the variable of interest (say, H13) for a given
sub-population (say, males in Ontario i.e. DVSEX=1 and PROV=5). X would be calculated by multiplying
the final weight (FWGHT) by the variable of interest (H13) and summing this product over all records for
males.
♀The denominator (Y) is the qualitative estimate of the number of participants (males in Ontario with H13)
within that sub-population. Y would be calculated by summing the final weight (FWGHT) over all male
respondents in Ontario with 1   H13   52.

The two estimates X and Y are derived independently and then divided to provide the

quantitative estimate.
For the example given X (the weighted sum of weeks) equals XXXX and Y (the number of males in the
subpopulation) equals YYYY. The average number of weeks is then calculated to be:

$$\frac{XXXX}{YYYY} = \frac{128,918,398.16}{3,861,075.33} = 33.4$$

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 complex
surveys presents problems to analysts because the survey design and the selection probabilities affect the
estimation and variance calculation procedures that should be used.

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 an analysis of all respondents who consider themselves as "workaholics", the following steps are required:

- Select all respondents from the file who considered themselves as a workaholic (E2B = 1);
- Calculate the Average Weight for these records;
- For each of these respondents calculate a "working" weight equal to FWGHT / Average Weight;
- Perform the analysis for these respondents 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

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 1992.

$$\sum_{i=1}^{9,815} FWGHT = 21,294,313$$

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

the population.

When estimates of the number of persons are desired, while using the Main File, FWGHT is
to be used.

Examples & Interpretation:

(i) In 1992, nearly 48% of female (DVSEX = 2) Canadians 15 years of age and older (5.2 million)
     stated they felt more rushed (A5 = 1) than compared to five years ago.

(ii) 51% of Canadians 25 to 44 years of age (DVAGEGR GE 04 and DVAGEGR LE 07) tend to cut back
     on their sleep, when they need more time for other activities (E2C = 1).

(iii)   78% of males (DVSEX = 1) aged 15 to 24 (01   DVAGEGR   03) stated that during the past 12
         months they regularly participated in sports (G1 = 1) while only 57% of females (DVSEX = 2) in
         the same age category took part regularly.

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 files.

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

8.1 Minimum Sample Size For Estimates

Users should determine the number of records on the particular microdata file which contribute to the
calculation of a given estimate. This number should be 15 or more. When the number of contributors to the
weighted estimate is less than this, 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 households.
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 which are not related to sampling may occur at almost every phase of a
survey operation. Interviewers
may misunderstand instructions, respondents may make errors in answering questions,
the answers may be
incorrectly entered on the questionnaire and errors may be introduced in the
processing
and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little
effect on estimates derived
from the survey. However, errors occurring systematically will contribute to biases
in the survey estimates.
Considerable time and effort 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 results. The
extent of non-response varies from partial non-response (failure to answer just one
or some questions) to total
non-response. Total non-response occurred because the interviewer was either
unable to contact the
respondent, no member of the household was able to provide the information, or the
respondent refused to
participate in the survey. Total non-response was handled by adjusting the weight
of households 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, 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 virtually with certainty that the differences would be less than three standard
errors.

Because of the large variety of estimates that can be produced from a survey, the standard deviation 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 on the
following page.

♀ Type of Estimate	Coefficient of Variation	Policy Statement
1. Unqualified unrestricted	0.0 to 16.5%	Estimates can be considered for general release. No special notation is required.
2. Qualified release but	16.6 to 33.3%	Estimates can be considered for general unrestricted 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 such statistical tables, such estimates should be excluded.

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

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 qualitative 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 Variance Tables"). These tables were produced using the
coefficient of variation formula based on a simple random sample. Since estimates for Cycle 7 of the General
Social Survey are based on a complex sample design, a factor called the Design Effect has been introduced
into the variance formula.

The Design Effect for an estimate is the actual variance for the estimate (taking into account the design that
was used) divided by the variance that would result if the estimate had been derived from a simple random
sample. 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.

                     Design Effects

          Geographic Area   Design Effect

          Canada   1.53

          Newfoundland   1.16

          Prince Edward Island   1.14

          Nova Scotia   1.17

          New Brunswick   1.12

          Quebec   1.21

          Ontario   1.29

          Manitoba   1.16

          Saskatchewan   1.23

          Alberta   1.18

          British Columbia   1.21

          Atlantic Region   1.23

          Prairie Region   1.27

Approximate variance tables are provided for each province, the Atlantic Region, the Prairie Region and
Canada. 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
coefficient of variation in the "Not for Release" range (see the policy regarding the release of the survey
estimates on preceding pages).

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 (e.g., 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 files correspond to those produced
by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such
estimates. It may be misleading 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 where, due to technical or other limitations, a different rounding technique is used
      resulting in estimates different from Statistics Canada estimates, 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.   FILE STRUCTURE

In view of the nature of the time use data and the difference in the sample size,

the microdata file consists of
the three subfiles described below.

The Main File is composed of 341 variables covering general background, cultural participation, unpaid help
measurement and organized sport variables.   There are 9,815 records.

The Time Use Summary File consists of one record per respondent and summarizes the total time spent on
each of 167 activities, the 10 major categories, the 24 subcategories, total time spent at each location and total
time spent with various persons.   In addition, it contains a subset of characteristics found on the main file.
This is the most widely used file for time use analysis.   There are 8,996 records.

The Time Use Episode File consists of all episodes reported by respondents.   Each respondent generated a
variable number of records depending on the number of episodes reported.   For each episode, there is
information on the activity, start and end time, duration, location and an indication of who the respondent was
with for that episode.   There are 190,327 records.

There is some duplication across the three files, however, this is done to facilitate the use of the files.   The
variables SEQNUM can be used for linking the files.

Special Notes

1.   The variables on the Main File are generally in the following order:
      general identification information and weight for each record (variables 1 to 4);
      as they appear on the GSS 7-2 questionnaire (variables 5 to 327); for most of these fields, a derived
      variable was created to assist the user with the data analysis;
      derived variables with information obtained on the GSS 7-1 questionnaire (variables 328 to 341).

      Due to the large number of variables on the Main File, an index is provided in Appendix C.

2.   Variable Acronyms - Numerous variable names directly link the data to the questionnaire.   For
      example, the acronym DVD3, refers the user to question D3 of the questionnaire, the source of the
      data provided by this particular variable.

3.   Not Stated Categories - Generally a code 9 for a one digit field, a code 99 for a 2 digit field, etc.
      indicate that the respondent did not answer a question and therefore the answer is not stated.   As the
      following example indicates, two types of "Not Stated" categories may appear.

            PLACE   Where were you?/Were you still....

            01   Respondent's home
            02   Respondent's work place
            03   Someone else's home
            04   Other place
            05   Car (Driver)
            06   Car (Passenger)
            07   Walking
            08   Bus and subway

```
09          Bicycle
10          Other form of transit
88          Not stated (activity code is 001 or 002)
98          Respondent is in transit, form of transit is not stated
99          Not stated
```

Code 9, 99, etc. is the "true" not stated category for all variables on the file.

In certain questions, however, a second `Not Stated' category appears. Although the respondent may not have marked a response, the information was actually partially available. Because of the branching pattern of a particular response, related information which followed, allows imputations of the original question. Other responses within the question were truly not stated. These cases are thus identified separately.

4. The sample and population counts and the mean values for each variable in the data dictionaries are calculated from all respondents not only the ones specified in the coverage component of the description of the variable.

♀10. 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 $750.00. Tabulations can be obtained at a cost that will reflect the resources required to produce the tabulation.

Sample Selection Procedures, Weighting and Estimation
David Paton
Development and Analysis Section
Informatics and Methodology Field
(613) 951-1467

Subject Matter, Data Collection and Data Processing
Ghislaine Villeneuve
General Social Survey
Housing, Family and Social Statistics Division
(613) 951-4995