Microdata User Guide

Canada Survey of Giving, Volunteering and Participating

2010

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Canada Survey of Giving, Volunteering and Participating (CSGVP) is one component of the Voluntary Sector Initiative (VSI), a collaborative program of the federal government and the voluntary sector. The 2010 CSGVP was conducted by Statistics Canada in the provinces and territories from September 14th to December 10th, 2010. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

Statistics Canada

Client Services
Special Surveys Division
Room 2300
Main Building
150 Tunney's Pasture Driveway
Ottawa, Ontario K1A 0T6
Telephone: 613-951-3321 or call toll-free 1-800-461-9050
Fax: 613-951-4527
E-mail: **ssd@statcan.gc.ca**

2.0 Background

In the course of their busy lives and many commitments, millions of Canadians make a conscious effort to contribute to others and their communities through charitable giving, volunteering their time to charitable and non-profit organizations and by helping individual Canadians directly on their own. In 1997, the National Survey of Giving, Volunteering and Participating (NSGVP) provided the first comprehensive look at the contributions that Canadians made to one another through their gifts of time and money. The NSGVP was developed through a unique partnership of federal government departments and non-profit and voluntary organizations that included the Canadian Centre for Philanthropy (now operating under the name of Imagine Canada), Canadian Heritage, Health Canada, Human Resources and Skills Development Canada, Statistics Canada and Volunteer Canada. Using a similar framework, this survey was conducted again in 2000 as part of the federal government's Voluntary Sector Initiative (VSI). In 2001, the federal government provided funding to establish a permanent survey program at Statistics Canada on charitable giving, volunteering and participating. The survey itself was renamed the Canada Survey of Giving, Volunteering and Participating (CSGVP) to distinguish it from surveys in other countries.

The establishment of a permanent series of surveys provided an opportunity to review the design of the survey instrument to ensure that it would provide the highest quality information on an ongoing basis. Consultations were held with a variety of stakeholders from the charitable and non-profit sector, government, and the academic community to identify ways to improve the survey. In 2010, survey data were collected in the North (Yukon, Northwest Territories and Nunavut) for the third time, where a representative sample of 1,423 respondents aged 15 and older participated in the survey. In 2010, the sample size in the 10 provinces was reduced from the 2007 level in an effort to save costs; for comparison, there were 14,724 respondents in 2000, 20,510 respondents in 2007 and 16,833 respondents in 2010. Over the survey program's lifespan, the questionnaire has been revised in a number of ways, based on experience gained from the earlier survey cycles. Some questions were changed to improve their clarity for respondents. Other questions were added to collect new information of interest. A number of questions were also dropped from the survey. Because the survey is now being conducted on a permanent basis, it may be possible to cycle sets of questions in and out of the survey.

Over time, the survey platform has also changed. The NSGVP was administered to a sub-sample of respondents to Statistics Canada's Labour Force Survey (LFS). Because of concerns about demands being placed on LFS respondents, the provincial component of the 2004, 2007, and 2010 CSGVP were conducted as Random Digit Dialling (RDD) surveys, in which respondents were recruited specifically to participate in the CSGVP.

The 2010 CSGVP continues the method of measuring giving, volunteering and participating established by the 2004 and 2007 CSGVP. It replaces the way these behaviours were measured in the 1997 and 2000 NSGVP. Because of these changes, it is not appropriate to compare the results from the 2010, 2007 or 2004 CSGVP surveys with the previous NSGVP surveys.

3.0 Objectives

The objectives of the Canada Survey of Giving, Volunteering and Participating (CSGVP) are threefold:

1) to collect national data to fill a void of information about individual contributory behaviours including volunteering, charitable giving and civic participation;
2) to provide reliable and timely data to the System of National Accounts; and
3) to inform both the public and voluntary sectors in policy and program decisions that relate to the charitable and voluntary sector.

4.0 Concepts and definitions

This chapter outlines concepts and definitions of interest to the users.

Donor
A donor is a person who made at least one donation of money to a charitable or other non-profit organization in the 12-month reference period preceding the survey.

Financial donation
A financial donation is money given to a charitable or other non-profit organization during the 12-month reference period preceding the survey. Money given to the same organization, on multiple occasions, in response to the same solicitation method, constitutes only one donation. For example, all money donated to a particular religious institution over the 12 months preceding the survey, through a collection at the place of worship, would be considered to be a single donation.

In-kind donation
This is a non monetary donation made to a charitable or other non-profit organization. Examples include donations of clothing or household items and donations of food.

Industry and Occupation
The 2010 Canada Survey of Giving, Volunteering and Participating (CSGVP) provides industry and occupation information for employed persons only (i.e., regarding the job which the individual occupied the week preceding the interview). For industry, statistics have been provided based on the 2007 North American Industry Classification System (NAICS). For occupation, the 2006 National Occupation Classification – Statistics (NOC-S) has been used.

Informal volunteer
The CSGVP defines an informal volunteer (or a direct helper) as a person who helped someone on their own, that is, not through a group or organization, in the 12-month reference period preceding the survey. This includes help given directly to friends, neighbours and relatives, but excludes help given to anyone living in the household. Since these activities are not provided through the structure of an organization, they are not included under the definition of volunteering.

Labour force status
Labour force status designates the status of the respondent vis-à-vis the labour market. For the 2010 CSGVP, estimates of labour force status refer to the survey population aged 15 to 75, as respondents aged 76 and older were not asked the related series of questions.

The three categories of labour force status are “employed”, “unemployed” and “not in the labour force”. For the purposes of the CSGVP, the three categories of labour force status are defined as follows:

Employed
Employed persons are those who, during the week preceding the interview

a) did any work[1] at all at a job or business; or
b) had a job but were not at work due to factors such as own illness or disability, personal or family responsibilities, vacation, labour dispute or other reasons (excluding persons on layoff or between casual jobs).

Unemployed

[1] Work includes any work for pay or profit, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes unpaid family work, which is defined as unpaid work contributing directly to the operation of a farm, business or professional practice owned and operated by a related member of the same household. Such activities may include keeping books, selling products, waiting on tables, and so on. Tasks such as housework or maintenance of the home are not considered unpaid family work.

Unemployed persons are those who, during the week preceding the interview

a) were on temporary layoff (excluding full-time students); or
b) were without work, and had actively looked for work in the past four weeks, (excluding full-time students and retired persons).

Not in the labour force

Persons not in the labour force are those who had not worked during the week preceding the interview and

a) were permanently unable to work; or
b) were full-time students who had a job but were absent from work as a result of a layoff or because they were between casual jobs; or
c) were full-time students or retired persons who did not have a job and had looked for work; or
d) did not have a job and did not look for work.

Mandatory community service

This is unpaid help provided to a group or organization that was mandated, or required, by a school, an employer, a charitable or non-profit organization, or some other authority. The 2010 CSGVP includes mandatory service under the definition of volunteering.

Organization classification

Respondents were asked to provide information on the organizations for which they volunteered and to which they made donations. Respondents were first asked to provide the name of the organization. A pick-list including the most common organizations reported in the 1997 and 2000 surveys was used. If the organization cited by the respondent was not on this pick-list, the respondent was then asked to provide information about what this organization does. This information was then used to group organizations into broad categories.

To classify these organizations, the *International Classification of Nonprofit Organizations (ICNPO)*[2] was used. Although they are classified according to their primary area of activity, some organizations operate in multiple areas. A major advantage of the ICNPO system is that it is used widely by other countries and thus allows for international comparisons. It has also been devised specifically to reflect the range and nature of activities typically undertaken in the non-profit and voluntary sector. The ICNPO system developed by the Johns Hopkins Comparative Nonprofit Sector Project, and modified for use in Canada, groups organizations into 15 Major Activity Groups, including a catch-all "Not Elsewhere Classified" category. These 15 Major Activity Groups are further grouped into 12 categories.

The 15 categories are as follows:

1) *Arts and culture:* includes organizations and activities in general and specialized fields of arts and culture, including media and communications; visual arts, architecture, ceramic art; performing arts; historical, literacy and humanistic societies; museums; and zoos and aquariums.

2) *Sports and recreation:* includes organizations and activities in general and specialized fields of sports and recreation. Two sub-groups of organizations are included in this group: (1) amateur sports (including fitness and wellness centres); and (2) recreation and social clubs (including service clubs).

3) *Education and research:* includes organizations and activities administering, providing, promoting, conducting, supporting and servicing education and research. Three sub-groups are contained in this group: (1) primary and secondary education organizations; (2) organizations involved in other education (i.e., adult/continuing education and vocational/technical schools); and (3) organizations involved in research (i.e., medical research, science and technology, and social sciences). Note that organizations devoted primarily to education and research in the area of

[2] The classification is based on L.M. Salamon and H.K. Anheier, 1997. *Defining the Nonprofit Sector: A Cross-national Analysis.* Manchester University Press.

specific medical conditions (e.g., Heart and Stroke Foundation of Canada, Canadian Cancer Society) are included under category 5, *Health*.

4) *Universities and colleges:* includes organizations and activities related to higher learning. This includes universities, business management schools, law schools and medical schools.

5) *Health:* includes organizations that engage primarily out-patient health-related activities and health support services. Two sub-groups are included in this category: (1) mental health treatment and crisis intervention; and (2) other health services (including public health and wellness education, out-patient health treatment, rehabilitative medical services, and emergency medical services). Also included in this category are organizations devoted primarily to education, research or support services in the area of specific medical conditions (e.g., Heart and Stroke Foundation, Canadian Cancer Society) as well as organizations providing support to the terminally ill (e.g., hospices and other types of palliative care).

6) *Hospitals:* includes organizations that engage primarily in in-patient health care. Two sub-groups are included in this category: (1) hospitals and rehabilitation; and (2) nursing homes.

7) *Social Services:* includes organizations and institutions providing human and social services to a community or target population. Three sub-groups are contained in this category: (1) social services (including organizations providing services for children, youth, families, the handicapped and the elderly, and self-help and other personal social services); (2) emergency and relief; and (3) income support and maintenance.

8) *Environment:* includes organizations promoting and providing services in environmental conservation, pollution control and prevention, environmental education and health, and animal protection. Two sub-groups are included in this category: (1) environment; and (2) animal protection.

9) *Development and housing:* includes organizations promoting programs and providing services to help improve communities and promote the economic and social well-being of society. Three sub-groups are included in this category: (1) economic, social and community development (including community and neighbourhood organizations); (2) housing; and (3) employment and training.

10) *Law, Advocacy and Politics:* includes organizations and groups that work to protect and promote civil and other rights, advocate the social and political interests of general or special constituencies, offer legal services or that promote public safety. Three sub-groups are contained in this category: (1) civic and advocacy organizations; (2) law and legal services; and (3) political organizations.

11) *Grant-making, fundraising and voluntarism promotion:* includes philanthropic organizations and organizations promoting charity and charitable activities including grant-making foundations, voluntarism promotion and support, and fund-raising organizations.

12) *International:* includes organizations promoting cultural understanding between peoples of various countries and historical backgrounds as well as those providing relief during emergencies and promoting development and welfare abroad.

13) *Religion:* includes organizations promoting religious beliefs and administering religious services and rituals (e.g., churches, mosques, synagogues, temples, shrines, seminaries, monasteries and similar religious institutions), in addition to related organizations and auxiliaries of such organizations.

14) *Business and professional associations, unions*: includes organizations promoting, regulating and safeguarding business, professional and labour interests.

15) *Groups not elsewhere classified.*

The correspondence between the 12 category classification and the 15 category classification is as follows:

12 Category ICNPO	15 Category ICNPO
1) Culture and recreation	1) Arts and culture 2) Sports and recreation
2) Education and research	3) Education and research 4) Universities and colleges
3) Health	5) Health 6) Hospitals
4) Social services	7) Social services
5) Environment	8) Environment
6) Development and housing	9) Development and housing
7) Law, advocacy and politics	10) Law, advocacy and politics
8) Philanthropic intermediaries and voluntarism	11) Grant-making, fundraising and voluntarism promotion
9) International	12) International
10) Religion	13) Religion
11) Business and professional associations, unions	14) Business and professional associations, unions
12) Groups not elsewhere classified	15) Groups not elsewhere classified

Participant

The CSGVP defines a participant as a person who was a member of at least one group, organization or association in the 12-month reference period preceding the survey. This includes professional organizations or unions; service clubs or fraternal organizations; political groups; cultural, educational, or hobby related organizations; sports or recreation organizations; religious organizations; seniors' or youth groups; support or self-help programs; environmental groups; and community or school related associations.

Reference period

For most questions in the CSGVP questionnaire, the reference period was the 12 months preceding the interview. For the provincial component, interviews were conducted from September 14th to December 10th, 2010. The territorial or northern component interviews took place during same time period as the provincial component.

Volunteer

This is a person who volunteered, that is, who performed a service without pay, on behalf of a charitable or other non-profit organization, at least once in the 12 month reference period preceding the survey. This includes any unpaid help provided to schools, religious organizations, sports or community associations.

5.0 Survey methodology for the provincial component

In the 10 provinces, the 2010 Canada Survey of Giving, Volunteering and Participating (CSGVP) was administered between September 14th and December 10th, 2010 as a Random Digit Dialling (RDD) survey, a technique whereby telephone numbers are generated randomly by computer. Interviews were conducted by telephone.

5.1 Population coverage

The target population consisted of the population 15 years of age or older residing in Canada's 10 provinces, with the exception of the institutionalized population.

The surveyed population excluded persons living in households without a land phone line, i.e., those living in households with no phone or with only cell phones were excluded. It is estimated that in 2010 approximately 16%[3] of households in the 10 provinces had no land line telephone. It is important to realize that although these persons were excluded from the population surveyed, the estimates were weighted to account for them. The underlying assumption is that the people in these households have the same characteristics and behaviours as those surveyed.

5.2 Sample design

The sample design is a stratified simple random sample of telephone numbers. The sample design is sometimes referred to as a Random Digit Dialing survey.

5.2.1 Stratification

The sample for the provincial component of CSGVP is based on a stratified design employing probability sampling. The stratification was done at the province / census metropolitan area (CMA) level. Twenty-seven strata were formed. Each province was divided into a number of CMA strata (ranging from zero in Prince Edward Island to four in Ontario) and one additional residual "non-CMA" stratum comprising the remainder of the province.

5.2.2 Sample allocation

To cut costs, the sample size was decreased in 2010 to approximately two thirds of the 2007 sample size. Even so, the allocation of this smaller sample was driven by the same general criteria used in the 2007 sample design, namely to be able to produce:

1) cross-sectional estimates for volunteers provincially and for the three largest CMAs;
2) cross-sectional estimates for non-volunteers provincially and for the three largest CMAs; and
3) national cross-sectional estimates for immigrants

It was determined that approximately 16,800 responses would be required to meet these objectives. A Kish allocation (alpha = 0.2)[4] was used to distribute the total expected responses among the three large CMAs, the remainder of the province where these three CMAs occurred, and all other provinces. The sample was then allocated proportionally within the province to the remaining strata. A response rate of 55% was assumed, thus a sample size of approximately 30,500 would be required to obtain the 16,800 responses. With an RDD design it is necessary to take into account that not all telephone numbers

[3] Residential Telephone Services Survey, Statistics Canada, 2010.

[4] In 2007, a power allocation (power = 0.2) was used.

will be valid residential numbers. An RDD sample will include a significant number of business and non-working numbers. In addition, during the data collection process there will inevitably be some numbers which will not be able to be resolved as being a business or a residential number. The sample size was increased to take all these occurrences into account based on the experience of Statistics Canada's General Social Survey (GSS). The resulting sample size was 83,778 telephone numbers.

5.3 Sample selection

The sample for the provincial component of the CSGVP was generated using a refinement of RDD sampling called the Elimination of Non-Working Banks (ENWB). Within each stratum, a list of working banks (area code + next five digits) was compiled from telephone company administrative files. A working bank, for the purposes of social surveys, is defined as a bank which contains at least one working residential telephone number. Thus, all banks with only unassigned, cell phone, non-working, or business telephone numbers are excluded from the survey frame.

A systematic sample of banks (with replacement) was selected within each stratum. For each selected bank, a two-digit number (00 to 99) was generated at random. This random number was added to the bank to form a complete telephone number. This method allowed listed and unlisted residential numbers, as well as business and non-working (i.e., not currently or never in-service) numbers, to have a chance of being in the sample. An automated pre-dialling screening activity, aimed at removing not-in-service and known business numbers, was performed prior to sending the sample to the computer-assisted telephone interviewing (CATI) unit. The final sample sent to the CATI unit consisted of 57,957 telephone numbers.

Each telephone number in the CATI sample was dialled to determine whether or not it reached a household. If the telephone number was found to reach a household, the person answering the telephone was asked to provide information on the individual household members. One person in the household aged 15 or older was selected at random to complete the survey. Proxy interviews were not accepted.

The selected respondent was asked a series of 15 questions which determined their volunteer status. If the respondent was found to be a volunteer, they continued through the rest of the questionnaire. On the other hand, non-volunteers were sub-sampled at a rate of 68% and only the sub-sample continued through the remaining relevant sections of the questionnaire (this rate varied by province in 2010 to better reflect the varying volunteer rates seen across the provinces)[5]. At the time the sample file was created, a flag was included which was randomly set so that it had a 68% chance of being set to one and a 32% chance of being set to zero. If a respondent was a non-volunteer and the randomly set flag on the sample file had been set to one, then they continued; if the flag had been set to zero, the interview ended after the series of 15 questions.

[5] In 2007 the rate was 50% and was constant across all provinces.

5.4 Sample size by province

The following table shows the number of telephone numbers generated for the provincial component of the 2010 CSGVP, as well as the number of respondents before and after sub-sampling of non-volunteers.

Province	Number of telephone numbers generated	Number of responses before sub-sampling non-volunteers	Number of responses after sub-sampling non-volunteers
Newfoundland and Labrador	6,953	1,250	1,067
Prince Edward Island	5,399	1,083	1,083
Nova Scotia	5,430	1,086	1,086
New Brunswick	6,753	920	815
Quebec	14,210	3,480	2,209
Ontario	18,658	3,227	2,480
Manitoba	4,947	1,137	1,137
Saskatchewan	4,295	966	966
Alberta	5,826	1,318	1,190
British Columbia	11,307	2,366	2,026
Total Provinces	**83,778**	**16,833**	**14,059**

6.0 Survey methodology for the territorial (northern) component

In the three territories, the Canada Survey of Giving, Volunteering and Participating (CSGVP) was administered between September 14th and December 10th, 2010 to a sub-sample of dwellings taken from three months of the Labour Force Survey (LFS) sample combined. The sample design of the CSGVP in the territories is therefore closely tied to that of the LFS. The CSGVP was not collected as a true live LFS supplement since the collection was not done at the same time as the LFS collection. As a result, the CSGVP had to repeat the collection of the roster information as well as any LFS variables of interest.

6.1 Population coverage

The target population consisted of the population 15 years of age and older residing in Canada's three territories with the following exceptions:

- institutionalized population
- residents of Indian Reserves (with one exception, residents of the Hay River Reserve in the Northwest Territories are included in the target population)
- full-time members of the Canadian Armed Forces

In the Yukon and Northwest Territories, only the population in selected communities is surveyed by the LFS. For operational and cost reasons, very small communities are excluded. It is estimated that the communities covered represent over 90% of the population aged 15 and over in the Yukon and Northwest Territories. In Nunavut, the communities eligible for sampling cover less than 70% of the population aged 15 and over. The estimates are also calibrated to these coverage totals.

6.2 Sample design

The LFS in the north employs a multi-stage design. In the north, communities form the primary sampling units (PSU). Sampling of PSUs is followed by sampling of households.

6.2.1 Sample rotation

The LFS design in the north employs a rotating panel design in which the sample consists of eight panels, or rotation groups. The households in the panel are contacted once every three months and remain in the sample for eight quarters. This results in the household being in the sample for almost two years. The survey is conducted monthly. One third of the quarterly sample is contacted each month, thus, 1/24th of the sample is rotated each month.

6.2.2 Modifications to the Labour Force Survey design in the territories for the Canada Survey of Giving, Volunteering and Participating

The CSGVP sample included all households in the July, August and September 2010 LFS sample excluding households that were in the LFS sample for the first time.

The CSGVP used seven of the eight rotation groups in the July, August and September LFS sample. The birth rotation group was excluded. Roster information was collected for all members of the household and then one household member 15 years of age or older was selected at random to complete the remainder of the CSGVP questionnaire. Proxy responses were not permitted. Unlike the provincial component, in the territorial component there was no sub-sampling of the non-volunteers. All non-volunteers were asked to complete all relevant sections of the questionnaire.

6.3 Sample size

The sample consisted of the non-birth rotation groups of the July, August, September quarterly sample of the LFS. The initial sample size was 2,289. The following table gives the breakdown by territory:

Territory	Initial sample size	Number of respondents
Nunavut	835	543
Northwest Territories	665	457
Yukon	789	423
Total territories	**2,289**	**1,423**

7.0 Data collection

7.1 Questionnaire design

The 2010 Canada Survey of Giving, Volunteering and Participating (CSGVP) measured giving, volunteering and participating in the same way as in 2007. It extended the methods used in 2004 in order to generate a repeated time series, and replaced the way these behaviours were measured in the 1997 and 2000 National Survey of Giving, Volunteering and Participating. Experiences gained from the 2004 CSGVP suggested that a number of small adjustments were required relating to the questionnaire content.

In preparation for the 2004 CSGVP, extensive consultations were held with key federal, provincial and territorial government representatives, as well as representatives from the voluntary sector and academics. These consultations were focused primarily on survey content and were held from January through April 2002. Following the consultations, the steering committee members met to discuss priorities and content issues. This meeting resulted in the development of a draft questionnaire to be used in focus-group testing and one-on-one interviews. Qualitative testing of content was conducted during the summer months across Canada. Changes to the survey subsequent to the qualitative testing resulted in a pilot test in April 2003. This allowed adjustment for any errors in the computer application, and also provided an opportunity to refine the survey procedures.

In 2010, a few questions were added to address specific situations that occurred in 2010. First, two questions were added to determine the portion of reported volunteer hours that was associated with the 2010 Olympics. Another three questions were added to ensure that respondent's included their donations towards natural disaster relief efforts such as those affected in Haiti or Chile. These questions were qualitatively tested in the Ottawa area in the spring of 2010. Apart from these questions, since both the survey design and majority of the questionnaire content for the 2010 CSGVP were virtually the same as the 2004 and 2007 CSGVP, external qualitative testing was not undertaken during this survey cycle.

The types of questions included in the CSGVP are divided into two major categories: those that measure behaviours and indicate what individuals are doing in terms of their giving, volunteering and participating, and those that measure correlates of these behaviours. This latter category includes attitudes and motivations, as well as factors that potentially constrain or facilitate giving and volunteering.

7.2 Supervision and quality control

All Statistics Canada interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures of the surveys to which they are assigned. Senior interviewers are also responsible for periodically monitoring the interviewers.

Interviewers were trained on the survey content and the computer-assisted telephone interviewing (CATI) application. In addition to classroom training, the interviewers completed a series of mock interviews to become familiar with the survey and its concepts and definitions.

7.3 Data collection methodology

7.3.1 Provincial component

For the 10 provinces, all data were collected using computer-assisted telephone interviewing. The CATI system has a number of generic modules which can be quickly adapted to most types of surveys. A front-end module contains a set of standard

response codes for dealing with all possible call outcomes, as well as the associated scripts to be read by the interviewers. The survey introduction used a standard approach which introduces the agency, informs the respondent of the name and purpose of the survey and the names of the survey sponsors, outlines how survey results will be used and provides an estimated interview duration.

The random selection of one person per household was carried out at the time of the interview. The interviewer first obtained the age, sex and relationships of everyone in the household. Once this information was completed, the CATI application randomly selected one individual to be the CSGVP respondent. Respondents were informed that their participation in the survey was voluntary, and that their information would remain strictly confidential.

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

The cases were distributed to five Statistics Canada regional offices. The workload and interviewing staff within each office was managed by a project manager. The automated scheduler used by the CATI system ensured that cases were assigned randomly to interviewers. There were a maximum of 25 call attempts per case identified as a residential phone number; once the maximum was reached, the case was reviewed by a senior interviewer who determined if additional calls would be made.

7.3.2 Territorial component

Collection of the CSGVP in the territories was very similar to the collection in the provinces with the following exceptions:

- All data were collected using a computer-assisted personal interview (CAPI) application which allowed responses to be captured directly by the interviewer at the time of the interview; and
- While most interviews were collected by telephone (74%), for households without landlines, interviews were conducted in person (26%).

7.4 Non-response

Interviewers were instructed to make all reasonable attempts to obtain a completed interview with the randomly selected member of the household. Those who at first refused to participate were re-contacted up to two more times to explain the importance of the survey and to encourage their participation. For cases in which the timing of the interviewer's call was inconvenient, an appointment was arranged to call back at a more convenient time. For cases in which there was no one home, numerous call backs were made.

8.0 Data processing

The main output of the Canada Survey of Giving, Volunteering and Participating (CSGVP) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

8.1 Data capture

Responses to survey questions are captured directly by the interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission.

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

8.2 Editing

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

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

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

8.3 Coding of open-ended questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format, and coded at head office. The computerized questionnaire contained a pick-list of common organizations which was used to assist the interviewer when entering information regarding the type of organization for which the respondent volunteered (VD_Q01) or to which the respondent donated (GS_Q01). If the organization cited by the respondent was not on this pick-list, the respondent was asked to provide some information regarding what the organization does. This information was used to code the type of organization using the International Classification of Nonprofit Organizations (ICNPO), Revision 1 (see Chapter 4.0 for further information on this classification system).

Coding of the industry (2007 North American Industry Classification System) and occupation (2006 National Occupational Classification System – Statistics) relating to the job which the respondent had the week preceding the interview was performed based on responses to questions LF_Q05 to LF_Q08.

For the following six questions on the CSGVP questionnaire, the text in the "Other – specify" write-in category was examined at head office and, where possible, coded into an existing category:

- FV_Q16: other volunteer activities;
- FG_Q15: other methods in which donations were made to a charitable or non-profit organization;
- SK_Q07: acquired skills;
- SD_Q01: religion;
- SD_Q03: country of birth; and
- SD_Q08: ancestral ethnicity.

8.4 Imputation

Imputation is the process that supplies valid values for those variables that have been identified for a change either because of invalid information or because of missing information. The new values are supplied in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits. In other words, the objective is not to reproduce the true microdata values, but rather to establish internally consistent data records that yield good aggregate estimates.

We can distinguish between three types of non-response. Complete non-response is when the respondent does not provide the minimum set of answers. These records are dropped and accounted for in the weighting process (see Chapter 13.0). Item non-response is when the respondent does not provide an answer to one question, but goes on to the next question. These are usually handled using the “not stated” code or are imputed. Finally, partial non-response is when the respondent provides the minimum set of answers but does not finish the interview. These records can be handled like either complete non-response or multiple item non-response.

In the case of the CSGVP, donor imputation was used to fill in missing data for some item and partial non-response. Further information on the imputation process is given in Section 9.2.3.

8.5 Creation of derived variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. Most derived variable names have a ”D” in the fourth character position of the name. Some derived variables may have a ”G” in the fourth character position of the name. In most cases, these are variables which have been grouped for ease of use.

Examples of derived variables include:

- total number of hours volunteered (VD1DHRS);
- total number hours volunteered for the 15 organization types (VD1DTX01 to VD1DTX15 on the master file, VD1GTX01 to VD1GTX15 on the public use microdata file (PUMF));
- total amount of donations (GS1DATOT on the master file, GS1GATOT on the PUMF);
- total amount of donations for the 15 organization types (GS1DAX01 to GS1DAX15 on the master file, GS1GAX01 to GS1GAX15 on the PUMF); and
- total amount of donations by solicitation method (FG1DA03 to FG1DA15 on the master file, FG1GA03 to FG1GA14 on the PUMF).

Derived variables for donations were derived from the Giving (GS) file and placed on the MAIN file (see Chapter 14.0 for further information on the file structure.)

In general, a derived variable was not calculated if any part of the equation was not answered (i.e., don’t know, refused or not stated.) In these cases, the code assigned to the derived variable was usually “not stated”.

8.6 Weighting

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

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example if the number of people who had volunteered in the preceding 12 months is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 13.0.

8.7 Suppression of confidential information

It should be noted that the "Public Use" Microdata Files may differ from the survey "master" files held by Statistics Canada. These differences usually are the result of actions taken to protect the anonymity of individual survey respondents. The most common actions are the suppression of file variables, grouping values into wider categories, and coding specific values into the "not stated" category. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 10.0 of this document.

Some of the general features of the PUMF for the 2010 CSGVP include:

The survey master file includes geographic identifiers that are generally more explicit than the PUMF, e.g. census metropolitan areas and population centres. The PUMF provides data only at the provincial level. The master file also includes some demographic variables which are excluded from the PUMF. These include ancestral ethnicity and immigration status.

The survey master file includes certain detailed information which is included on the PUMF only in grouped form. These include:

- precise age of respondent;
- number of children aged 0 to 5 in the household (grouped to a yes/no variable indicating presence of children aged 0 to 5 in the household);
- a detailed 43 category North American Industry Classification (NAICS) which has been collapsed to an 18 category grouping;
- country of birth, which has been grouped to "Canada" and "Outside Canada".

As well, for certain variables that are susceptible to identifying individuals, the PUMF has some data suppressed locally, i.e. some of the values in the master file have been coded as "not stated" on the PUMF.

9.0 Data quality

9.1 Response rates

9.1.1 Response to the provincial component

The telephone resolved rate and telephone hit rate, by province, are provided in the following table.

The telephone **resolved rate** is defined as the proportion of telephone numbers confirmed, either in the pre-screening process or in the field, as being either residential or out-of-scope (e.g., business or non-working numbers, numbers for cell phones, non-residences or collective dwellings) as a proportion of the total number of telephone numbers generated.

$$resolved\ rate = \frac{number\ of\ resolved\ telephone\ numbers}{number\ of\ telephone\ numbers\ generated}$$

Telephone resolved rate by province

Province	Telephone numbers generated	Telephone numbers sent to collection	Telephone numbers resolved in the field	Total resolved	Resolved rate (%)	Confirmed residential telephone numbers	Respondents
Newfoundland and Labrador	6,953	4,027	3,319	6,245	89.8	1,812	1,250
Prince Edward Island	5,399	4,052	3,520	4,867	90.1	1,571	1,083
Nova Scotia	5430	3,620	3,183	4,993	92.0	1,691	1,086
New Brunswick	6,753	3,723	3,197	6,227	92.2	1,716	920
Quebec	14,210	11,338	10,077	12,949	91.1	5,495	3,480
Ontario	18,658	13,942	12,406	17,122	91.8	5,819	3,227
Manitoba	4,947	2,952	2,521	4,516	91.3	1,637	1,137
Saskatchewan	4,295	3,127	2,615	3,783	88.1	1,456	966
Alberta	5,826	4,026	3,447	5,247	90.1	1,987	1,318
British Columbia	11,307	7,150	6,256	10,413	92.1	3,561	2,366
Total Provinces	**83,778**	**57,957**	**50,541**	**76,362**	**91.1**	**26,745**	**16,833**

Response rates are given for the provincial component of the Canada Survey of Giving, Volunteering and Participating (CSGVP) in the following table. A respondent is defined as a sampled person who completed the 15 questions in the Formal Volunteering (FV) module of the questionnaire that determine whether or not the person was a volunteer. For Random Digit Dialling (RDD) methodology, a new method to calculate response rate was developed during the 2007 cycle of the survey so that continuing iterations of the survey would be comparable to 2004 survey data. Hence, for comparison purposes, the 2010 response rate was calculated using the old method (same calculation as in 2004) in addition to the new method.

The **old response rate** is defined as the number of sampled persons who completed at least this minimum requirement divided by the number of *confirmed residential telephone numbers* (those that have been resolved).

$$response\ rate - old\ method = \frac{number\ of\ respondents}{number\ of\ confirmed\ residential\ telephone\ numbers}$$

The **old hit rate** is defined as the proportion of resolved telephone numbers that were confirmed to be residential telephone numbers.

$$hit\ rate - old\ method = \frac{number\ of\ confirmed\ residential\ telephone\ numbers}{number\ of\ resolved\ telephone\ numbers}$$

Response rate by province – Old Method

Province	Hit rate (%)	Response rate (%)
Newfound land and Labrador	29.0	69.0
Prince Edward Island	32.3	68.9
Nova Scotia	33.9	64.2
New Brunswick	27.6	53.6
Quebec	42.4	63.3
Ontario	34.0	55.5
Manitoba	36.2	69.5
Saskatchewan	38.5	66.3
Alberta	37.9	66.3
British Columbia	34.2	66.4
Total Provinces	**35.0**	**62.9**

The **new response rate** used since 2007 is defined as the number of sampled persons who completed at least the minimum requirement divided by the number of *estimated residential telephone numbers*. Estimated residential telephone numbers include all telephone numbers resolved as residential as well as a portion of unresolved telephone numbers that are estimated to be households.

$$response\ rate = \frac{number\ of\ respondents}{estimated\ number\ of\ residential\ telephone\ numbers}$$

The **new hit rate** is defined as the estimated number of residential telephone numbers divided by the number of telephone numbers generated.

$$hit\ rate = \frac{estimated\ number\ of\ residential\ telephone\ numbers}{number\ of\ telephone\ numbers\ generated}$$

The **collection hit rate** is defined as the estimated number of residential telephone numbers divided by the number of telephone numbers sent to collection.

$$collection\ hit\ rate = \frac{estimated\ number\ of\ residential\ telephone\ numbers}{number\ of\ telephone\ numbers\ sent\ to\ collection}$$

Response rate and hit rates by province – New Method

Province	Estimated residential telephone numbers	Hit rate (%)	Collection hit rate (%)	Response rate (%)
Newfound land and Labrador	2,136	30.7	53.0	58.5
Prince Edward Island	1,790	33.2	44.2	60.5
Nova Scotia	1,881	34.6	52.0	57.7
New Brunswick	1,953	28.9	52.5	47.1
Quebec	6,116	43.0	53.9	56.9
Ontario	6,452	34.6	46.3	50.0
Manitoba	1,861	37.6	63.0	61.1
Saskatchewan	1,716	40.0	54.9	56.3
Alberta	2,300	39.5	57.1	57.3
British Columbia	4,029	35.6	56.3	58.7
Total Provinces	**30,234**	**36.1**	**52.2**	**55.7**

9.1.2 Response to the territorial component

Response rates are given for the territorial component of the CSGVP in the following table. The same definition of respondent applies in the territories as in the provinces.

$$response\ rate = \frac{number\ of\ respondents}{number\ of\ respondents + number\ of\ non\text{-}respondents}$$

Response rate by territory

Territory	Total sample	Out-of-scope	Respondents	Non-respondents	Response rate (%)
Yukon	789	179	423	187	69.3
Northwest Territories	665	117	457	91	83.4
Nunavut	835	137	543	155	77.8
Total Territories	**2,289**	**433**	**1,423**	**433**	**76.7**

9.2 Survey errors

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

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

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

9.2.1 Data collection

Interviewer training consisted of a self-study of the CSGVP Interviewer's Manual and a review of the summary publication *Caring Canadians, Involved Canadians: Highlights from the 2007 Canada Survey of Giving, Volunteering and Participating*, followed by two days of classroom training. The manuals included a description of the background and objectives of the survey, as well as a glossary of terms and a set of questions and answers. The classroom sessions included a presentation of survey objectives, a review of key concepts and practice time with training cases (mock interviews) using the computer-assisted telephone interviewing (CATI) application. They also provided an opportunity for interviewers to ask questions before the start of collection.

9.2.2 Data processing

Data processing of the CSGVP was done in a number of steps including verification, coding, editing, imputation, estimation, and confidentiality. At each step a picture of the output files was taken and verification was performed by comparing the files at the current and previous step.

9.2.3 Non-response and imputation

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 item or partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred either because the interviewer was unable to contact the respondent, because no member of the household was able to provide the information, or because the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of individuals who responded to the survey to compensate for those who did not respond.

In most cases, item or 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. In item and partial non-response cases, for certain variables donor imputation was performed. Most of these imputations were done in order to provide complete data enabling the calculation of totals (e.g., total number of hours and total amount of donations). Also, the imputation helped to keep records in the sample, even if part of the required information was not filled in by the respondent.

All imputations involved donor records that were selected using a score function. For each item non-response or partial non-response record (also called a recipient record), certain characteristics were compared to those from all potential donor records. When a characteristic was the same for a donor record and the recipient record, a value was added to the score of that donor. The donor record with the highest score was deemed the "closest" donor and was chosen to fill in missing pieces of information of the non-respondent. If there was more than one donor record with the highest score, a random selection occurred. The pool of donor records was made up in such a way that the

imputed value assigned to the recipient, in conjunction with other non-imputed items from the recipient, would still pass the edits.

Imputation was done in six steps. The first three steps related to imputation of variables on the Main file (see Chapter 14.0 for file structure). The first step was to impute both personal and household income. Due to an application error, the personal income variable was not collected for this cycle nor was it imputed. The second step was to impute the hours volunteered by activity for the main organization. The third step was to impute the total hours volunteered for the second and third organizations and the total hours volunteered for all other organizations combined.

The fourth step was to impute variables on the Giving (GS) file related to amount donated. This step also included creating additional GS file records for cases where a value for GS_Q07, *Did you make any other donations in response to this solicitation method*?, was imputed as “yes”.

The fifth step was to impute, on the Main file, missing data in any of the variables indicating whether the respondent made a donation in response to each of the 13 methods of solicitation (FG_Q03 to FG_Q15 from the Financial Giving to Charitable Organizations (FG) section of the questionnaire). At this stage, imputation was performed only for cases which were already known to be givers (i.e., cases which already had a value of "yes" in at least one of FG_Q03 to FG_Q15). This step also included creating additional GS records for cases where one or more of FG_Q03 to FG_Q15 was imputed as "yes".

The sixth step was to impute partially completed records where the donor status could not be determined because of missing values in FG_Q03 to FG_Q15. A total of 88 variables were imputed. This last step again included creating additional GS file records for cases where any of FG_Q03 to FG_Q15 was imputed as “yes”.

The following table shows the number of records imputed for some of the key variables of the survey. The rates for the income variables are high but in 46% of the cases where the household income value was imputed, the respondent had reported an income range. This is comparable to the 2007 survey when the rate was 44%.

Number and percentage of records imputed for selected variables

Variable	Records imputed	Total records	% imputed
Household income	6,077	15,482	39.3
Hours for organization 1	297	15,482	1.9
Hours for organization 2	170	15,482	1.1
Hours for organization 3	137	15,482	0.9
Donations to organizations 1 to 10	11,001	59,032	18.6
Donations to organizations 11+	3,601	59,032	6.1

The following table shows the resulting impact on the actual estimates.

Percentage of estimate originating from imputed values

Variable	Imputed estimate (millions)	Total estimate (millions)	% imputed
Household income	922,872	2,339,139	39.5
Hours for organization 1	61.5	1,623.9	3.8
Hours for organization 2	12.2	326.4	3.7
Hours for organization 3	4.1	117.7	3.5
Amount of total donations	5,056.1	11,863.9	42.6
Number of donors	1.1	23.4	4.5

The CSGVP imputation process worked well and helped to fill incomplete responses with the experience of other respondents with similar or identical characteristics. This adds to the number of units used in any analysis performed by researchers.

Note that the public use microdata file does not contain any of the imputation flags. The impact of this is an additional layer of confidentiality.

9.2.4 Measurement of sampling error

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

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

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

For example, suppose that, based upon the 2004 CSGVP results, one estimates that 54.8% of Canadians aged 15 to 24 had done some volunteering in the preceding year, and this estimate is found to have a standard error of 0.012. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.012}{0.548}\right) X\ 100\% = 2.2\%$$

There is more information on the calculation of coefficients of variation in Chapters 11.0 and 12.0.

10.0 Guidelines for tabulation, analysis and release

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

10.1 Rounding guidelines

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

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

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

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

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

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

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

10.2 Sample weighting guidelines for tabulation

The sample design used for the Canada Survey of Giving, Volunteering and Participating (CSGVP) was not self-weighting. Users must apply the proper survey weight when producing simple estimates, including the production of ordinary statistical tables.

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

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

10.3 Definitions of types of estimates: categorical and quantitative

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

10.3.1 Categorical estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of Canadians who volunteered and the number of Canadians who made financial donations are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of categorical questions:

Q: In the past 12 months, did you do any of the following activities without pay on behalf of a group or an organization? This includes any unpaid help you provided to schools, religious organizations, sports or community associations. Did you do any … teaching, educating or mentoring?
R: Yes / No

Q: In the past 12 months, did you make a charitable donation … by responding to a request through the mail?
R: Yes / No

10.3.2 Quantitative estimates

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

An example of a quantitative estimate is the average number of hours contributed by volunteers. The numerator is an estimate of the total number of hours volunteered and its denominator is the number of persons who volunteered.

Examples of quantitative questions:

Q: In the past 12 months, how many hours did you spend on unpaid activities for this organization?
R: |_|_|_|_| hours

Q: What was the amount of the donation to this organization?
R: |_|_|_|_|_| dollars

10.3.3 Tabulation of categorical estimates

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

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

10.3.4 Tabulation of quantitative estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of hours volunteered by persons aged 65 and over, multiply the value reported in VD1DHRS (hours volunteered) by the final weight for the record, then sum this value over all records with DH1GAGE = 6 (age group 65 and over).

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of hours volunteered by those aged 65 and over,

a) estimate the total number of hours volunteered ($\hat{X}$) as described above,
b) estimate the number of people in this category ($\hat{Y}$) by summing the final weights of all records with DH1GAGE = 6, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

10.4 Guidelines for statistical analysis

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

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures may differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

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

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

1) select all respondents from the file who reported RESPSEX = male;

2) calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported RESPSEX = male;

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

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

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

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

10.5 Coefficient of variation release guidelines

Before releasing and/or publishing any estimates from the CSGVP, users should first determine the quality level of the estimate. The quality levels are *acceptable, marginal* and *unacceptable.* Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 9.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 9.0 to be more fully aware of the quality characteristics of these data.

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

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

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

Quality Level Guidelines

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

10.6 Release cut-offs for the Canada Survey of Giving, Volunteering and Participating

The following table provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 12.0.

For example, the table shows that the quality of a weighted estimate of 10,000 people possessing a given characteristic in Newfoundland and Labrador is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 12.1 and Example 4 in Section 12.1.1 explain the correct procedure to be used for ratios.

Province / Territories	Acceptable CV 0.0% to 16.5%	Marginal CV 16.6% to 33.3%	Unacceptable CV > 33.3%
Newfoundland and Labrador	20,000 & over	5,000 to < 20,000	under 5,000
Prince Edward Island	5,500 & over	1,500 to < 5,500	under 1,500
Nova Scotia	39,000 & over	10,000 to < 39,000	under 10,000
New Brunswick	40,000 & over	10,000 to < 40,000	under 10,000
Quebec	170,000 & over	41,000 to < 170,000	under 41,000
Ontario	240,000 & over	60,000 to < 240,000	under 60,000
Manitoba	48,000 & over	12,000 to < 48,000	under 12,000
Saskatchewan	45,000 & over	12,000 to < 45,000	under 12,000
Alberta	130,000 & over	33,000 to < 130,000	under 33,000
British Columbia	100,000 & over	26,000 to < 100,000	under 26,000
Provinces	**170,000 & over**	**42,000 to < 170,000**	**under 42,000**
Territories	3,500 & over	1,000 to < 3,500	under 1,000
Canada	**170,000 & over**	**42,000 to < 170,000**	**under 42,000**

11.0 Variance estimation for master and share files

In order to determine the quality of the estimate and to calculate the coefficient of variation (CV), the standard deviation must be calculated. Confidence intervals also require the standard deviation of the estimate. The Canada Survey of Giving, Volunteering and Participating (CSGVP) uses a stratified simple random sample design that is subjected to non-response adjustments, imputation and calibration, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, an approximate method was needed, the bootstrap method. With the use of the bootstrap weights and the BOOTVAR program, CV's and other variance estimates can be derived with accuracy.

Bootstrap method for variance estimation

Independently, in each stratum, a simple random sample of $(n-1)$ of the n units in the sample is selected with replacement. Note that since the selection is with replacement, a unit may be chosen more than once. This step is repeated R times to form R bootstrap samples. An average initial bootstrap weight based on the R samples is calculated for each sample unit in the stratum. The entire process (selecting simple random samples, recalculating weights for each stratum) is repeated B times, where B is large, yielding B different initial bootstrap weights.

These weights are then adjusted according to the same weighting process as the regular weights: non-response adjustment, calibration and so on. The end result is B final mean bootstrap weights for each unit in the sample. The variation among the B possible estimates based on the B bootstrap weights are related to the variance of the estimator based on the regular weights and can be used to estimate it.

For the CSGVP, a set of 250 mean bootstrap weights was prepared. Using these weights to estimate the variance is recommended over the approximate variance estimation methods discussed in the next chapter.

Statistical packages for variance estimation

Bootvar

Users should note that bootstrap weights are provided and should be used for variance estimation. BOOTVAR is a macro program that can be used to do the variance calculation using the bootstrap weights. The Bootvar program is available in SAS or SPSS format. It is made up of macros that compute variances for totals, ratios, differences between ratios and for linear and logistic regression.

Bootvar may be downloaded from Statistics Canada's Research Data Centre (RDC) website. Users must accept the Bootvar Click-Wrap Licence before they can read the files. There is a document on the site explaining how to adapt the system to meet users' needs.

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-eng.htm
SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-eng.htm

For the CSGVP, the Bootvar parameters B = 250 and R = 20 must be used since there are 250 bootstrap weights based on a mean of 20 observations.

Other packages

Other than Bootvar, there are different commercial software packages that can carry out some design-based analysis for variance estimation; Stata 9 or 10, SUDAAN and WesVar.

These methods can be adapted for the CSGVP from a paper by Owen Phillips *"Using bootstrap weights with Wes Var and SUDAAN"* (Catalogue no. 12-002-X20040027032) in *The Research Data Centres Information and Technical Bulletin, Chronological index*, Fall 2004, vol.1 no. 2 Statistics Canada, Catalogue no. 12-002-XIE.

12.0 Variance estimation for public use microdata files

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

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

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province, which were used to produce the Approximate Sampling Variability Tables for the 2010 Canada Survey of Giving, Volunteering and Participating (CSGVP).

Province/Territories	Design effect	Sample size	Population
Newfoundland and Labrador	1.39	1,067	433,258
Prince Edward Island	1.37	1,083	118,639
Nova Scotia	1.51	1,086	792,337
New Brunswick	1.47	815	636,868
Quebec	1.55	2,209	6,591,090
Ontario	1.52	2,480	10,941,744
Manitoba	1.57	1,137	989,999
Saskatchewan	1.50	966	837,047
Alberta	1.48	1,190	2,989, 697
British Columbia	1.51	2,026	3,875,082
Provinces	**2.33**	**14,059**	**28,205,761**
Territories	1.96	1,423	79,676
Canada	**2.55**	**15,482**	**28,285,437**

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

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

12.1 How to use the coefficient of variation tables for categorical estimates

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

Rule 1: Estimates of numbers of persons possessing a characteristic (aggregates)

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

Rule 2: Estimates of proportions or percentages of persons possessing a characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of volunteers who provided health care or support including companionship is more reliable than the estimated number of volunteers who provided health care or support including companionship. (Note that in the tables the coefficients of variation decline in value reading from left to right).

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

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

Rule 3: Estimates of differences between aggregates or percentages

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

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

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

Rule 4: Estimates of ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of persons with a university degree and the numerator is the number of volunteers with a university degree.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of volunteers with a university degree as compared to the number of volunteers without a university degree, the standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is

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

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

Rule 5: Estimates of differences of ratios

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

12.1.1 Examples of using the coefficient of variation tables for categorical estimates

The following examples based on the 2004 Canada Survey of Giving, Volunteering and Participating (Master File) are included to assist users in applying the foregoing rules. Please note that the data for these examples are different than the results obtained from the current survey and are only to be used as a guide.

Example 1: Estimates of numbers of persons possessing a characteristic a (aggregates)

Suppose that a user estimates that 5,615,215 men were volunteers during the reference period. How does the user determine the coefficient of variation of this estimate?

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

2) The estimated aggregate (5,615,215) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 6,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 1.8%.

4) So the approximate coefficient of variation of the estimate is 1.8%. The finding that there were 5,615,215 (to be rounded according to the rounding guidelines in Section 10.1) male volunteers during the reference period is publishable with no qualifications.

Canada Survey of Giving, Volunteering and Participating, 2004 - Master File

Approximate Sampling Variability Tables for Canada including the Territories

NUMERATOR OF PERCENTAGE	ESTIMATED PERCENTAGE													
('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	161.9	161.1	160.3	157.9	153.6	149.3	144.9	140.3	135.5	130.6	125.5	114.5	88.7	51.2
2	114.5	113.9	113.4	111.6	108.6	105.6	102.4	99.2	95.8	92.3	88.7	81.0	62.7	36.2
3	93.5	93.0	92.6	91.1	88.7	86.2	83.6	81.0	78.2	75.4	72.4	66.1	51.2	29.6
4	80.9	80.6	80.2	78.9	76.8	74.7	72.4	70.1	67.8	65.3	62.7	57.3	44.4	25.6
5	72.4	72.1	71.7	70.6	68.7	66.8	64.8	62.7	60.6	58.4	56.1	51.2	39.7	22.9
6	66.1	65.8	65.5	64.4	62.7	61.0	59.1	57.3	55.3	53.3	51.2	46.8	36.2	20.9
7	61.2	60.9	60.6	59.7	58.1	56.4	54.8	53.0	51.2	49.4	47.4	43.3	33.5	19.4
8	57.2	57.0	56.7	55.8	54.3	52.8	51.2	49.6	47.9	46.2	44.4	40.5	31.4	18.1
9	54.0	53.7	53.4	52.6	51.2	49.8	48.3	46.8	45.2	43.5	41.8	38.2	29.6	17.1
10	51.2	51.0	50.7	49.9	48.6	47.2	45.8	44.4	42.9	41.3	39.7	36.2	28.1	16.2
11	48.8	48.6	48.3	47.6	46.3	45.0	43.7	42.3	40.9	39.4	37.8	34.5	26.7	15.4
12	46.7	46.5	46.3	45.6	44.4	43.1	41.8	40.5	39.1	37.7	36.2	33.1	25.6	14.8
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
300	**********	**********	9.3	9.1	8.9	8.6	8.4	8.1	7.8	7.5	7.2	6.6	5.1	3.0
350	**********	**********	8.6	8.4	8.2	8.0	7.7	7.5	7.2	7.0	6.7	6.1	4.7	2.7
400	**********	**********	8.0	7.9	7.7	7.5	7.2	7.0	6.8	6.5	6.3	5.7	4.4	2.6
450	**********	**********	7.6	7.4	7.2	7.0	6.8	6.6	6.4	6.2	5.9	5.4	4.2	2.4
500	**********	**********	7.2	7.1	6.9	6.7	6.5	6.3	6.1	5.8	5.6	5.1	4.0	2.3
750	**********	**********	**********	5.8	5.6	5.5	5.3	5.1	4.9	4.8	4.6	4.2	3.2	1.9
1000	**********	**********	**********	5.0	4.9	4.7	4.6	4.4	4.3	4.1	4.0	3.6	2.8	1.6
1500	**********	**********	**********	**********	4.0	3.9	3.7	3.6	3.5	3.4	3.2	3.0	2.3	1.3
2000	**********	**********	**********	**********	3.4	3.3	3.2	3.1	3.0	2.9	2.8	2.6	2.0	1.1
3000	**********	**********	**********	**********	**********	2.7	2.6	2.6	2.5	2.4	2.3	2.1	1.6	0.9
4000	**********	**********	**********	**********	**********	**********	2.3	2.2	2.1	2.1	2.0	1.8	1.4	0.8
5000	**********	**********	**********	**********	**********	**********	2.0	2.0	1.9	1.8	1.8	1.6	1.3	0.7
6000	**********	**********	**********	**********	**********	**********	**********	1.8	1.7	1.7	1.6	1.5	1.1	0.7
7000	**********	**********	**********	**********	**********	**********	**********	**********	1.6	1.6	1.5	1.4	1.1	0.6
8000	**********	**********	**********	**********	**********	**********	**********	**********	**********	1.5	1.4	1.3	1.0	0.6
9000	**********	**********	**********	**********	**********	**********	**********	**********	**********	1.4	1.3	1.2	0.9	0.5
10000	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	1.3	1.1	0.9	0.5
12500	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	1.0	0.8	0.5
15000	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	0.7	0.4
20000	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	**********	0.4

NOTE: FOR CORRECT USAGE OF THESE TABLES, PLEASE REFER TO MICRODATA DOCUMENTATION.

Example 2: Estimates of proportions or percentages of persons possessing a characteristic

Suppose that the user estimates that 1,605,006 / 5,615,215 = 28.6% of men who volunteered did some teaching, educating or mentoring. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., men who were volunteers), it is necessary to use both the percentage (28.6%) and the numerator portion of the percentage (1,605,006) in determining the coefficient of variation.

3) The numerator, 1,605,006 does not appear in the left-hand column (the “Numerator of Percentage” column) so it is necessary to use the figure closest to it, namely

1,500,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 30.0%.

4) The figure at the intersection of the row and column used, namely 3.5% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 3.5%. The finding that 28.6% of men who volunteered did some teaching, educating or mentoring can be published with no qualifications.

Example 3: Estimates of differences between aggregates or percentages

Suppose that a user estimates that 1,979,228 / 6,193,361 = 32.0% of women who volunteered did some teaching, educating or mentoring, while 1,605,006 / 5,615,215 = 28.6% of men who volunteered did some teaching, educating or mentoring. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for women as 3.0%, and the CV of the estimate for men as 3.5%.

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

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

where $\hat{X}_1$ is estimate 1 (women), $\hat{X}_2$ is estimate 2 (men), and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d} = 0.320 - 0.286 = 0.034$ is

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.320)(0.030)]^2 + [(0.286)(0.035)]^2} \\ &= \sqrt{(0.0000921) + (0.0001002)} \\ &= 0.014\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.014 / 0.034 = 0.412$

4) So the approximate coefficient of variation of the difference between the estimates is 41.2%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter F (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimate.

Example 4: Estimates of ratios

Suppose that the user estimates that 1,979,228 women who volunteered did some teaching, educating or mentoring on behalf of an organization, while 1,605,006 men who volunteered did some teaching, educating or mentoring. The user is interested in comparing the estimate of women versus that of men in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of female volunteers who did some teaching, educating or mentoring on behalf of an organization. The denominator of the estimate ($\hat{X}_2$) is the number of male volunteers who did some teaching, educating or mentoring on behalf of an organization.

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

3) The numerator of this ratio estimate is 1,979,228. The figure closest to it is 2,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 3.4%.

4) The denominator of this ratio estimate is 1,605,006. The figure closest to it is 1,500,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 4.0%

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

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

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

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0.034)^2 + (0.040)^2} \\ &= \sqrt{0.001156 + 0.0016} \\ &= 0.052 \end{aligned}$$

6) The obtained ratio of female versus male volunteers who did some teaching, educating or mentoring on behalf of an organization is 1,979,228 / 1,605,006 which is 1.23 (to be rounded according to the rounding guidelines in Section 10.1). The coefficient of variation of this estimate is 5.2%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of differences of ratios

Suppose that the user estimates that the ratio of female volunteers to male volunteers is 1.039 for ages 15 to 24 while it is 1.169 for ages 55 and over. The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First calculate the approximate coefficient of variation for the 15 to 24 age group ratio ($\hat{R}_1$) and the 55 and over age group ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the 15 to 24 age group ratio is 7.07% and 5.66% for ages 55 and over.

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

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

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 1.039 – 1.169 = -0.13 is

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(1.039)(0.0707)\right]^2 + \left[(1.169)(0.0566)\right]^2} \\ &= \sqrt{(0.005396) + (0.004378)} \\ &= 0.099\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.099 / (-0.13) = -0.762.

4) So the approximate coefficient of variation of the difference between the estimates is 76.2%. The estimate of the difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter F (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

12.2 How to use the coefficient of variation tables to obtain confidence limits

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

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

of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval ($CI_{\hat{x}}$):

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

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

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

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

12.2.1 Example of using the coefficient of variation tables to obtain confidence limits

A 95% confidence interval for the estimated proportion of male volunteers who did some teaching, educating or mentoring (from Example 2, Section 11.1.1) would be calculated as follows:

$\hat{X}$ = 28.6% (or expressed as a proportion 0.286)

t = 2

$\alpha_{\hat{x}}$ = 3.5% (0.035 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.286 – (2) (0.286) (0.035), 0.286 + (2) (0.286) (0.035)}

$CI_{\hat{x}}$ = {0.286 – 0.020, 0.286 + 0.020}

$CI_{\hat{x}}$ = {0.266, 0.306}

With 95% confidence it can be said that between 26.6% and 30.6% of male volunteers did some teaching, educating or mentoring.

12.3 How to use the coefficient of variation tables to do a t-test

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

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

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

12.3.1 Example of using the coefficient of variation tables to do a t-test

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of female volunteers who did some teaching, educating or mentoring and the proportion male volunteers who did some teaching, educating or mentoring. From Example 3, Section 11.1.1, the standard error of the difference between these two estimates was found to be 0.015. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.320 - 0.286}{0.014} = \frac{0.034}{0.014} = 2.43$$

Since t = 2.43 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

12.4 Coefficients of variation for quantitative estimates

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

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the number of hours volunteered for arts and culture organizations would be greater than the coefficient of variation of the corresponding proportion of volunteers who volunteered for arts and culture organizations. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

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

12.5 Coefficient of variation tables

Refer to CSGVP2010_CVTabsE.pdf for the coefficient of variation tables.

13.0 Weighting

A statistical weight was placed on each record of the data file. This weight indicates the number of persons in the population represented by the sampled unit.

Since the 2010 Canada Survey of Giving, Volunteering and Participating (CSGVP) was conducted as a Random Digit Dialling (RDD) survey in the 10 provinces while it used a sub-sample of the Labour Force Survey (LFS) sample in the three territories, two different sets of weighting procedures were used.

13.1 Weighting for the provincial component

The weighting for the provincial component consisted of several steps:

- calculation of the basic telephone weight;
- an adjustment for unresolved telephone numbers;
- dropping out-of-scope records;
- an adjustment for the number of telephone lines in the household;
- adjustments for non-response (household level and person level);
- an adjustment for selecting only one person from the household;
- an adjustment for sub-sampling non-volunteers;
- an adjustment for outliers;
- an adjustment of provincial charitable donation quartiles (replacing the previous cycle's personal income adjustment);and
- an adjustment to make the population estimates consistent with known province-age-sex totals from the Census projected population counts for persons 15 years of age and over.

The details of these steps follow.

1. Calculation of the basic telephone weight

The initial weight is the inverse of the probability of selection of the telephone number, calculated as follows within each stratum:

$$w_1 = \left(\frac{total\ number\ of\ \ possible\ \ telephone\ numbers\ \ \ from\ \ working\ \ banks}{number\ of\ \ sampled\ telephone\ numbers} \right)$$

There were 83,778 phone numbers selected in the sample.

2. Adjustment for unresolved telephone numbers

Before data collection, the 83,778 phone numbers underwent a screening process, leaving 57,992 telephone numbers. Of these 57,992 phone numbers, 35 were LFS respondents and were dropped before entering the field (in order to prevent respondent burden) leaving 57,957 phone numbers for data collection. However, these 35 cases were not dropped in the weighting process until after the first household non-response adjustment. This is due to the fact that these are in-scope units and should not be treated in the same way that out-of-scope or unresolved units are. However, since the 35 phone numbers were not sent to the field, they should not enter into the response rate calculation. Thus, the numbers in this section will not match exactly with the numbers reported in Chapter 9.0.

Each of the remaining 57,957 records sent to collection either had an initial status equal to residential or the initial status was unknown.

At the end of the data collection period, call history information obtained during collection was used to determine the final status of each record. Each unit was identified as out-of-scope, in-scope or unresolved. The weights of the resolved and out-of-scope records were adjusted to

account for the unresolved records and the unresolved records were dropped. The adjustment was performed at the stratum level separately for those with initial status of residential and those with initial status unknown (see Section 5.2 for description of strata).

A total of 7,416 unresolved records were dropped, leaving 50,576 records.

The weights were adjusted as follows within each stratum and initial status:

$$w_2 = w_1 * \left(\frac{\sum w_1 \ \ for\ resolved\ telephone\ numbers + \sum w_1 \ \ \ for\ unresolved\ telephone\ numbers}{\sum w_1 \ \ for\ resolved\ telephone\ numbers} \right)$$

3. Dropping out-of-scope telephone numbers

Phone numbers that were resolved after collection to be non-working or otherwise out-of-scope (businesses, cell phones, non-residences, collective dwellings, etc.) were dropped. A total of 26,745 records remained at this point.

4. Adjustment for missing number of telephone lines (first household level non-response adjustment)

In order to convert the telephone level weight calculated in Step 2 into a household level weight, it was necessary to divide the telephone weight by the number of telephone lines associated with the household. There are cases where the number of lines cannot be derived because of either item non-response or total household non-response. In the case of item non-response, the number of lines was imputed to one. The remaining cases where the number of telephone lines could not be derived were dropped (including the 35 cases that were on the LFS sample) and the weights of the retained units were inflated to compensate for the dropped records.

As a result of a non-response study, it was discovered that those cases who eventually responded, but had at least one refusal or in-progress language barrier code in the history of calls, had much lower volunteer rates than other cases. Adjustment groups were formed by splitting each stratum into groups based on the presence of a refusal and/or language barrier.

The weights were adjusted as follows within each stratum and refusal / language barrier group:

$$w_3 = w_2 * \left(\frac{\sum w_2 \ \ for \ \ households \ \ with\ number\ of \ \ lines + \sum w_2 \ \ for \ households \ \ missing \ \ number\ of \ \ lines}{\sum w_2 \ \ for\ households \ \ with \ \ number\ of \ \ lines} \right)$$

A total of 20,921 records remained.

5. Adjustment for number of telephone lines in the household

Weights for households with more than one telephone line (with different telephone numbers) were adjusted downwards to account for the fact that such households have a higher probability of being selected. The telephone weight was divided by the number of lines in the household. The maximum adjustment was capped at four to prevent outliers. At this stage the telephone weight becomes the household weight.

The weights were adjusted as follows:

$$w_4 = \left(\frac{w_3}{number\ of \ \ in\text{-}scope \ \ telephone \ \ lines \ \ in \ \ the \ \ household} \right)$$

6. Adjustment for household non-response (second household level non-response adjustment)

This step accounts for the remaining non-responding households, i.e., those for whom the number of telephone lines in the household could be derived. The weights were inflated, within

stratum, to compensate for non-responding households. Non-responding households were dropped at this step, leaving 19,862 records.

The weights were adjusted as follows within each stratum:

$$w_5 = w_4 * \left(\frac{\sum w_4 \ for\ household\ respondents + \sum w_4 \ for\ household\ non\text{-}respondents}{\sum w_4 \ for\ household\ respondents} \right)$$

7. Adjustment for sampling only one person in the household (aged 15 or over)

The household weight calculated in Step 6 was multiplied by the number of members in the household aged 15 or over. This adjustment was capped at five to prevent outliers. After this step, the weight changes from representing households to representing persons.

The weights were adjusted as follows:

$$w_6 = w_5 * (number\ of\ household\ members\ aged\ 15 or over)$$

8. Adjustment for person level non-response

The weights were then inflated to compensate for non-responding persons. This adjustment was done within non-response groups built using a logistic regression model. Subsequently, non-responding persons were dropped, leaving 16,833 records.

The weights were adjusted as follows within each stratum, age group and sex:

$$w_7 = w_6 * \left(\frac{\sum w_6 \ for\ person\ respondents + \sum w_6 \ for\ person\ non\text{-}respondents}{\sum w_6 \ for\ person\ respondents} \right)$$

9. Adjustment for sub-sampling non-volunteers

The weighted sub-sampling rate for non-volunteers was calculated within each stratum, as follows, using the weighted counts from the previous step:

$$Weighted\ sub-sampling\ rate = \left(\frac{\sum w_7 \ for\ selected\ non-volunteers}{\sum w_7 \ for\ selected\ non\text{-}volunteers + \sum w_7 \ for\ non\text{-}selected\ non-volunteers} \right)$$

The inverse of this rate was multiplied by the weights for the selected non-volunteers and the non-selected non-volunteers were dropped.

For non-volunteers, the weights were adjusted as follows within each stratum:

$$w_8 = \left(\frac{w_7}{weighted\ sub\text{-}sampling\ rate} \right)$$

For volunteers, $w_8 = w_7$

The final number of records was 14,059.

10. Calibration to known population totals

An adjustment was made to the weights in order to make population estimates consistent with external population counts for persons 15 years and older. The following external control totals were used:

- Population totals for each province/census metropolitan area (CMA) stratum, and

- Population totals by province, sex and the following age groups: 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69 and 70 and over.

This calibration step was performed merely as a temporary adjustment before identifying outliers. Once outliers were identified, this calibration step was ignored.

11. Identification and treatment of outliers

The treatment of outliers is a process which diminishes the impact of outlying weighted values. Outliers were identified for four variables: the total hours volunteered (VD1DHRS), the total amount of donations (GS1DATOT), total household income (IN1_03) and total personal income (IN1_04). Once the outliers were identified, their impact on the total estimates was diminished by reducing the weight (w_8) from Step 10 or by reducing the value of outliers, using a winsorization technique. The weight (or value) of the outlier was reduced such that the adjusted weighted value of the outlier was more in line with the non-outlier distribution.

The resulting weight from this step was w_9.

12. Adjustment of charitable donation quartiles

After producing the Master file person weights and comparing the weighted estimates to other available data sources, it was decided that an adjustment should be applied as a proxy for the personal income adjustment that was done for 2007. The variable for personal income was not collected for 2010, so an adjustment was done based on matching the weighted population of 2010 to 2007 total charitable donations quartiles. First, the 2007 file was used to define boundaries in the charitable donations such that one quarter of the 2007 weighted population fit into each category. Next those boundaries were adjusted by the consumer price index between 2007 and 2010. Finally the weights of the records in the corresponding categories in 2010 were adjusted so that again each category contained one quarter of the weighted population.

The weight resulting from this step was w_{10}.

13. Calibration to known population totals

The calibration at this step was performed in the same manner as in Step 11, the only difference being the weights input into the calibration process. The input to this calibration was the set of weights, w_{10}, output from Step 12, after adjusting for outliers. After the calibration was complete, the outlier detection was performed again to ensure there were no outliers remaining.

The weight, w_{11}, produced at this step, is the final weight, WTPM, on the Master Microdata File and WTPP on the Public Use Microdata File.

13.2 Weighting for the territorial component

The following steps describe how the weights for the territorial component were calculated.

1. Calculation of initial weights

Because the sample for the territorial component of the CSGVP was selected from the Labour Force Survey (LFS) sample, the initial weight, w_1, was calculated based on the subweight from the Labour Force Survey tabs file. The subweight is the initial design weight adjusted for out-of-scope units and non-responding units on the LFS. Subweights are unique for each month, province, and stratum combination so a subweight could be attributed to each unit in the CSGVP sample.

2. Dropping out-of-scope units

Some units that were on the CSGVP sample file did not match to the LFS tabs file, so the outcome (respondents, out-of-scope, non-respondent, unresolved) of these non-matching units was determined based on outcome codes. Units resolved as out-of-scope for CSGVP were dropped since these would have been already adjusted for in the LFS subweight (which is used as CSGVP's initial weight).

3. Weight share adjustment to subweight totals

An adjustment was made to the weights in order to make the subweight totals on the CSGVP sample file match to subweight totals on the LFS tabs file. This was done because there were missing units on the CSGVP sample file that did not match to the LFS tabs file and vice versa and the total number of households should remain consistent with LFS totals. The adjustment is done within month, stratum and type as follows:

$$w_2 = subweight * \frac{\sum_{tabs} w_1}{\sum_{sample} w_1}$$ within month, stratum, and type

4. Adjustment of subweights for not sampling birth rotation

In the LFS sample, there are eight rotations. In the CSGVP, the sample was drawn from seven of the eight rotations (it excludes the birth rotation). Thus, the subweights are adjusted to account for the fact that we did not sample from the birth rotation by multiplying each subweight by 8/7:

$$w_3 = w_2 * \frac{8}{7}$$

5. Adjustment of subweights for sampling from three months

Since each month on the tabs file adds up to approximately the total number of households in the population, an adjustment needs to be made because the CSGVP samples from three months. The adjustment is made by dividing each weight by 3 as follows:

$$w_4 = w_3 * \frac{1}{3}$$

6. Adjustment of weights for non-response

The CSGVP sample can be considered as being comprised of four groups:

1) respondents;
2) units determined to be out-of-scope;
3) non-respondents, resolved to be in-scope; and
4) non-respondents whose in-scope/out-of-scope status is unresolved.

Each of the sample units in the territorial component were assigned a status defined by these four groups based on the outcome code of the collection application. Since the final weights of the 2,289 respondents should reflect the entire in-scope population, the weights of the in-scope respondents should be inflated to account for the non-respondents. The weights should also be adjusted to account for the fact that the fourth group contains both in-scope and out-of-scope units. Assuming that the proportion of units that are out-of-scope among the unresolved units is the same as the proportion of out-of-scope among the resolved units, the weights of the respondents can be adjusted for non-response using the following formula:

$$w_{5gi} = w_{4gi} * \frac{(\sum_{Resp} w_{4gi} + \sum_{NR,res} w_{4gi})}{\sum_{Resp} w_{4gi}} * \frac{(\sum_{Resp} w_{4gi} + \sum_{NR,res} w_{4gi} + \sum_{OOS} w_{4gi} + \sum_{NR,unres} w_{4gi})}{(\sum_{Resp} w_{4gi} + \sum_{NR,res} w_{4gi} + \sum_{OOS} w_{4gi})}$$

where

g represents the level at which the adjustment is performed

w_{4gi} equals the initial weight of unit i in adjustment group g

$\sum_{Resp} w_{4gi}$ equals the sum of the initial weights of all respondents in adjustment group g

$\sum_{NR,res} w_{4gi}$ equals the sum of the initial weights of all resolved non-respondents in adjustment group g

$\sum_{NR,unres} w_{4gi}$ equals the sum of the initial weights of all unresolved non-respondents in adjustment group g

$\sum_{OOS} w_{4gi}$ equals the sum of the initial weights of all out-of-scope units in adjustment group g

This adjustment was performed within each stratum, provided that the total number of respondents plus non-respondents was greater than 30 and the adjustment factor was approximately less than two. If these conditions did not hold, strata were combined for adjustment purposes. There were four strata were the sample size warranted a collapsing of strata.

7. Adjustment for sampling one person per household

The weight calculated in Step 6 was multiplied by the number of people in the household 15 years of age or older. In order to avoid problems with outliers and to be consistent with the weighting procedure for the provincial component, a maximum of five household members aged 15 and over, was placed on this adjustment. There were 19 cases where the number of persons in the household aged 15 or older was greater than five.

$$w_6 = w_5 * (number\ of\ household\ members\ aged\ 15 or older)$$

8. Calibration to known population totals

The calibration step ensures that the sum of the weights of the respondents is equal to known population counts. The calibration was performed using age group/sex control totals by territory, with the three age groups being ages 15 to 24, 25 to 54, and 55 and over. In addition, in Nunavut the calibration also included a control total for the Inuit population aged 15 and over. The control totals used were for the October 2010 reference month.

This calibration step was performed merely as a temporary adjustment before identifying outliers. Once outliers were identified, this calibration step was ignored.

9. Identification and treatment of outliers

The treatment of outliers is a process which diminishes the impact of outlying weighted values. Outliers were identified for four variables: the total hours volunteered (VD1DHRS), the total amount of donations (GS1DATOT), total household income (IN1_03) and total personal income (IN1_04). Once the outliers were identified their impact on the total estimates was diminished by reducing the weight (w_6), from Step 7, or changing their value using a winsorization technique.

The weight of the outlier was reduced such that the adjusted weighted value of the outlier was more in line with the non-outlier distribution.

The resulting weight from this step was w_7.

10. Calibration to known population totals

The calibration at this step was performed in the same manner as in Step 8, the only difference being the weights input into the calibration process. The input to this calibration was the set of weights (w_7), output from Step 9, after adjusting for outliers. After the calibration was complete, the outlier detection was performed again to make sure that there were no outliers remaining.

The weight w_8 produced at this step is the final weight WTPM on the Master microdata file and WTPP on the Public Use Microdata File.

14.0 Structure of the files

There are two data files for the 2010 Canada Survey of Giving, Volunteering and Participating: the main answer file (MAIN.TXT), and the giver file (GS.TXT). To link between the MAIN and GS Master files use the variable MASTERID and to link between the two Public Use Microdata Files use the variable PUMFID.

MAIN.TXT

This is the main answer file and contains one record per respondent. All questions except for those on the GS file are located here. In addition, summary derived variables have been created from the GS file and placed on the MAIN file.

GS.TXT

This is the "giving" or charitable donation answer file. It contains one or more records for each person who made a financial donation: one record for each of up to 10 charitable organizations to which the respondent donated, over the 12 month reference period, in response to a particular solicitation method. For each of the 13 methods of solicitation itemized in the questionnaire, a donor may therefore have up to 10 records, each containing information regarding the type of organization, as well as the total amount of all donations made to that organization in response to that method of solicitation. In cases where the respondent donated to more than 10 organizations in response to a given method of solicitation, the total amount of all donations made to the remaining organizations is present on the 10th record as derived variable GS1D08.

15.0 Variable naming conventions

The 2010 Canada Survey of Giving, Volunteering and Participating (CSGVP) has adopted a standard eight character variable naming convention for variables on the microdata files.

Variable name component structure

- The **first two** characters are a combination of letters that identify the section of the questionnaire in which the variable was collected or from which the data used to derive the variable came.

Positions 1 and 2	Questionnaire section name
FV	Formal Volunteering
HV	History of Volunteering
VS	Volunteer Specifics
VD	Volunteer Details
MV	Main Volunteer Activities
RV	Reasons for Volunteering
GV	Volunteering in General
SK	Skills Gained from Volunteering
NV	Reasons for Not Volunteering (more)
IV	Informal Volunteer Activity
FG	Financial Giving to Charitable Organizations

Positions 1 and 2	Questionnaire section name
GS	Giving Specifics
DG	Decisions on Giving
RG	Reasons for Giving
NG	Reasons for Not Giving (more)
OG	Other Giving
EA	Youth Experiences and Attitudes
HG	Health in General
ED	Education
LF	Labour Force Status
SD	Socio-demographics
IN	Income

- The **third** character of the variable name is an identifier of the "wave" or iteration of a longitudinal survey. This is always equal to "1" on the 2010 CSGVP.

- The **fourth** character of the variable name refers to the variable type.

Position 4	Variable type	Description
–	Collected variable	A variable that appeared directly on the questionnaire
C	Coded variable	A variable coded from one or more collected variables (e.g., National Occupational Classification – Statistics)
D	Derived variable	A variable calculated from one or more collected or coded variables, usually calculated during head office processing (e.g., total hours volunteered)
F	Flag variable	A variable calculated from one or more collected variables (like a derived variable), but usually calculated by the computer application for later use during the interview (e.g., volunteer flag).
G	Grouped variable	Collected, coded, suppressed or derived variables collapsed into groups (e.g., age groups)
I	Imputation flag	A flag indicating whether a particular variable has been imputed (not present on the Public Use Master File).

- The **fifth, sixth, seventh and eighth** characters identify the variable or the question number from the questionnaire. In general, the last four positions follow the naming on the questionnaire. Numbers are used where possible (e.g., Q01 becomes 01). "Mark-all that apply" type questions use letters for each possible answer category (e.g., Q01 (Mark all that apply) becomes 01A, 01B, 01C, etc.).

Examples of variable names

MV1_02A: Number of hours spent canvassing for the main volunteer organization	
MV	Main Volunteer Activities section of the questionnaire
1	2010 CSGVP
_	Collected variable
02	Question number from questionnaire
A	First category in a "Mark all that apply" type question

FV1FVOL: Volunteer flag	
FV	Formal Volunteering section of the questionnaire
1	2010 CSGVP
F	Flag
VOL	Variable name

Note: A few important variables do not follow the naming convention (e.g., MASTERID, PUMFID, PROVCODE, WTPM and WTPP).