

Analytical Guide – Portrait of Canadian Society: Experiences during the Pandemic

1.0 Description

The survey series Portrait of Canadian Society (PCS) is a new Statistics Canada initiative. It is a probabilistic web panel that involves asking the same group of participants to complete four brief online surveys over a one-year period. For now, this is an experimental project which is part of a larger effort to modernize our data collection methods and activities. The goal is to collect important data on Canadian society more efficiently, more rapidly and at a lower cost compared to traditional survey methods. We will be able to test this collection method and refine it over time.

The experimental nature of this project and its high degree of non-response have an impact on which estimates should be produced using the web panel. Survey weights were adjusted to minimise potential bias that could arise from panel non-response; non-response adjustments and calibration using available auxiliary information were applied and are reflected in the survey weights provided with the data file. Despite these adjustments, the high degree of non-response to the panel increases the risk of remaining bias, which may impact estimates produced using the panel data. More information about the weighting methods used to adjust for non-response can be found in Section 5. Data quality guidelines and considerations are outlined in Section 6.

Each survey in the series is administered to a sub-sample of General Social Survey - Social Identity (GSS-SI) respondents who agreed to participate in additional surveys when completing the GSS-SI.

From July 19 to August 1, 2021, Statistics Canada conducted the Portrait of Canadian Society: Experiences During the Pandemic (PCS-EP). This survey was the second wave of the PCS.

The purpose of this survey is to help us better understand various aspects of Canadian's life during the pandemic, including access to health care services, perceptions of safety and community. The PCS is designed to produce data at a national level.

This manual has been produced to facilitate the manipulation of the microdata file of the PCS-EP survey results.

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

Statistics Canada

Client Services
Centre for Social Data Integration and Development
Telephone: 613-951-3321 or call toll-free 1-800-461-9050
Fax: 613-951-4527
E-mail: csdid-info-cidds@canada.ca

2.0 Survey methodology

2.1 Target and survey population

The PCS-EP is a sample survey with a cross-sectional design. Each survey in the series is administered to a sub-sample of General Social Survey - Social Identity (GSS-SI) respondents who agreed to participate in additional surveys when completing the GSS-SI.

The target population for the Portrait of Canadian Society (PCS) is the same as that of the GSS-SI. The target population includes all persons 15 years of age and older in Canada, excluding:

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

The frame used for GSS-SI, as well as the sampling strategy, are described in section 5 of the 2020 GSS-SI User Guide.

2.2 Sample Design and Size

To recruit the sample for Portrait of Canadian Society (PCS), recruitment questions were added at the end of General Social Survey – Social Identity (GSS-SI). Approximately 22% of GSS-SI respondents agreed to be approached for future surveys. They formed the sample for PCS.

The table below provides the number of respondents at each stage of the PCS-EP design.

Stages of the Sample	n
Dwellings selected for GSS-SI.	86,804
Individuals who responded to GSS-SI	34,044
Individuals who agreed to be approached for further surveys	7,502
Raw sample for surveys of the PCS	7,502
Panelists who participated in PCS-EP	3,330

The table below provides the number of respondents for PCS-EP by region, age group, and sex.

Area	Domain	n
Geography	Canada	3,330
	Atlantic provinces	502
	Quebec	635
	Ontario	1,125
	Prairies	645
	British-Columbia	423
Age Group	All	3,330
	15-24	131
	25-34	459
	35-44	699
	45-54	671
	55-64	621
	65-74	558
75+	191	
Sex	All	3,330

	Male	1,720
	Female	1,610

3.0 Data collection

PCS: Recruitment

The recruitment for PCS was done by adding two recruitment questions at the end of the GSS-SI questionnaire. GSS-SI was administered from August 17, 2020 to February 8, 2021. The first question asked if respondents would like to participate in a series of short, 15-20 minute surveys about important social topics. The respondents who answered “yes” to this question were asked to provide their email address and cellular phone number. This sub-sample of GSS-SI formed the sample for PCS.

PCS-EP – Experiences During the Pandemic

All respondents from GSS-SI who answered “yes” to the recruitment questions were sent an email invitation with a link to the PCS-EP and a Secure Access Code (SAC) to complete the survey online. Collection for the survey began July 19th, 2021.

There were two collection strategies for PCS-EP. The sample was divided into two groups and each group was part of one of the two collection strategies. All potential respondents were sent the invitation email on July 19th. For the first group, the reminder emails were sent on July 20th, July 22nd and July 24th. For the second group, the reminder emails were sent on July 22nd, July 26th, and July 29th. The application remained open until August 1, 2021.

Record Linkage:

To enhance the data from PCS-EP and reduce the response burden, information provided by respondents was combined with information from the General Social Survey - Social Identity. The GSS-SI is the source of socio-demographic variables available on the PCS-EP.

3.1 Disclosure control

Statistics Canada is prohibited by law from releasing any data which would divulge information obtained under the Statistics Act that relates to any identifiable person, business or organization without the prior knowledge or the consent in writing of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data is suppressed to prevent direct or residual disclosure of identifiable data.

4.0 Data quality

Survey errors come from a variety of different sources. They can be classified into two main categories: non-sampling errors and sampling errors.

4.1 Non-sampling errors

Non-sampling errors can be defined as errors arising during the course of virtually all survey activities, apart from sampling. They are present in both sample surveys and censuses (unlike sampling error, which is only present in sample surveys). Non-sampling errors arise primarily from the following sources: non-response, coverage, measurement and processing.

4.1.1 Non-response

Non-response is both a source of non-sampling error and sampling error. Non-response result from a failure to collect complete information from all units in the selected sample. Non-response is a source of non-sampling error in the sense that non-respondents often have different characteristics from respondents, which can result in biased survey estimates if non-response bias is not fully eliminated through weighting adjustments. The lower the response rate, the higher the risk of bias. Non-response is also a source of sampling error; this is discussed further in Section 6.2.

The PCS-EP survey design is carried out in multiple stages, each of which results in some non-response. The table below summarizes the response rate at each of these stages and the resulting cumulative response rate for PCS-EP.

Survey stage	Number of respondents	Response rate
GSS-SI	34,044	40.3%
Opt-in to additional surveys among GSS-SI respondents	7,502	22.0%
Response to PCS-EP among panel participants	3,330	44.4%
Cumulative response rate		3.8%

4.1.2 Coverage errors

Coverage errors consist of omissions, erroneous inclusions, duplications and misclassifications of units in the survey frame. Since they affect every estimate produced by the survey, they are one of the most important types of error. Coverage errors may cause a bias in the estimates and the effect can vary for different sub-groups of the population. This is a very difficult error to measure or quantify accurately.

The PCS-EP data is collected from people aged 15 years and over living in private dwellings within the 10 provinces. Excluded from the survey's coverage are: residents of Yukon, the Northwest Territories, and Nunavut; full-time residents of institutions, and residents of reserves. These groups together represent an exclusion of less than 2% of the Canadian population aged 15 and over.

Since PCS-EP uses the GSS-SI sample and was collected from July 19 to August 1, 2021, there is an undercoverage of residents of the 10 provinces that turned 15 since August 17, 2020, the beginning of GSS-SI collection. There is also undercoverage of those without internet access, since PCS-EP was collected entirely online. This undercoverage is greater amongst those age 65 years and older.

4.1.3 Measurement errors

Measurement errors (or sometimes referred to as response errors) occur when the response provided differs from the real value; such errors may be attributable to the respondent, the questionnaire, the collection method or the respondent's record-keeping system. Such errors may be random or they may result in a systematic bias if they are not random.

4.1.4 Processing errors

Processing errors are the errors associated with activities conducted once survey responses have been received. They include all data handling activities after collection and prior to estimation. Like all other errors, they can be random in nature, and inflate the variance of the survey's estimates, or systematic, and introduce bias. It is difficult to obtain direct measures of processing errors and their impact on data quality especially since they are mixed in with other types of errors (nonresponse, measurement and coverage).

4.2 Sampling errors

Sampling error is defined as the error that results from estimating a population characteristic by measuring a portion of the population rather than the entire population. For probability sample surveys, methods exist to calculate sampling error. These methods derive directly from the sample design and method of estimation used by the survey.

The most commonly used measure to quantify sampling error is sampling variance. Sampling variance measures the extent to which the estimate of a characteristic from different possible samples of the same size and the same design differ from one another. For sample designs that use probability sampling, the magnitude of an estimate's sampling variance can be estimated.

Factors affecting the magnitude of the sampling variance include:

1. The variability of the characteristic of interest in the population: the more variable the characteristic in the population, the larger the sampling variance.
2. The size of the population: in general, the size of the population only has an impact on the sampling variance for small to moderate sized populations.
3. The response rate: the sampling variance increases as the sample size decreases. Since non-respondents effectively decrease the size of the sample, non-response increases the sampling variance.
4. The sample design and method of estimation: some sample designs are more efficient than others in the sense that, for the same sample size and method of estimation, one design can lead to smaller sampling variance than another.

The standard error of an estimator is the square root of its sampling variance. This measure provides an indication of sampling error using the same scale as the estimate whereas the variance is based on squared differences.

The coefficient of variation (CV) of an estimate is a relative measure of the sampling error. It is defined as the estimate of the standard error divided by the estimate itself. It is very useful for measuring and comparing the sampling error of quantitative variables with large positive values. However, it is not recommended for estimates such as proportions, estimates of change or differences, and variables that can have negative values.

It is considered a best practice at Statistics Canada to report the sampling error of an estimate through its 95% confidence interval. The 95% confidence interval of an estimate means that if the survey were repeated over and over again, the confidence interval would cover the true population value 95% of the time (or 19 times out of 20).

5.0 Weighting

The principle behind estimation in a probability sample is that each unit selected in the sample represents, besides itself, other units that were not selected in the sample. For example, if a simple random sample of size 100 is selected from a population of size 5,000, then each unit in the sample

represents 50 units in the population. The number of units represented by a unit in the sample is called the survey weight of the sampled unit.

The weighting phase is a step that calculates, for each person, an associated sampling weight. This weight appears on the microdata file, and must be used to derive estimates representative of the target population from the survey. For example, if the number of individuals who smoke daily is to be estimated, it is done by selecting the records referring to those individuals in the sample having that characteristic and summing the weights entered on those records. The weighting phase is a step which calculates, for each record, what this number is. This section provides the details of the method used to calculate sampling weights for the PCS-EP.

The weighting of the sample for the PCS-EP has multiple stages to reflect the stages of sampling, participation and response to get the final set of respondents. The following sections cover the weighting steps to create the survey weights for PCS-EP.

5.1 Design weights

The initial panel weights are the final calibrated GSS-SI weights. These are the GSS-SI design weights adjusted for out-of-scope units and GSS-SI nonresponse, and then calibrated to population control totals. More information on these weights is available in section 8.1 of the GSS-SI user guide.

5.2 Nonresponse/Nonparticipation Adjustment

During collection of the PCS-EP, responses are obtained only from a proportion of sampled units. Individuals who responded to GSS-SI may decide not to opt-in to additional surveys and therefore not participate in the panel. Additionally, some individuals who opted into the panel, do not respond during PCS-EP collection. Weights of the nonresponding and nonparticipating units were redistributed to participating units. Units that did not participate in the panel had their weights redistributed to the participating units with similar characteristics within response homogeneity groups (RHGs).

The variables available for building the RHGs were available for both responding and non-responding units. These included personal characteristics (such as age, gender, education, population group, sexual orientation, employment information, voting behaviour, and personal income), household characteristics (such as home ownership and household income), and variables related to GSS-SI collection (such as the month of GSS response and whether response was online or interviewer-assisted). An adjustment factor was calculated within each response group as follows:

$$\frac{\text{Sum of weights of respondents and nonrespondents}}{\text{Sum of weights of respondents}}$$

The weights of the respondents were multiplied by this factor to produce the non-response adjusted weights. The nonparticipating units were dropped from the weighting process at this point.

5.3 Calibration

Control totals were computed using demography projection data. During calibration, an adjustment factor is calculated and applied to the survey weights. This adjustment is made such that the weighted sums match the control totals. Three sets of population control totals were used for PCS-EP:

- 1) Geographic region, age group, and sex. The geography and age groupings selected for calibration took into account the sometimes small number of respondents in different categories. The five geographic regions used for calibration were the Atlantic Provinces, Quebec, Ontario, the Prairie Provinces, and British Columbia. The age groups used were 15-34 year olds, 35-64 year olds, and those aged 65 years or more.
- 2) Sub-regional geographies. Respondent weights were also calibrated so that the sum within each province, as well as within the CMAs of Montreal, Toronto, and Vancouver, match population control in those sub-regional geographies.
- 3) Age group at a national level. Respondent weights were calibrated to population totals (nationally) within more granular age groupings. These groupings were defined as 15-24 year olds, 25-34 year olds, etc. up to respondents aged 75 years or more.

5.4 Bootstrap weights

Bootstrap weights were generated for the PCS-EP survey respondents. Each bootstrap replicate was generated based on the initial PCS-EP design weights, and then adjusted for non-response and calibrated as described above.

6.0 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.

6.1 Rounding guidelines

Users are urged to adhere to the following rounding guidelines when producing estimates and statistical tables computed from these microdata files:

- a) Estimates in the main body of a statistical table are to be rounded 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.
- 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 using normal rounding. Averages, rates, percentages, proportions and ratios are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves using normal rounding. Sums and differences are to be

derived from their corresponding unrounded components and then are to be rounded themselves using normal rounding.

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

6.2 Sample weighting guidelines for tabulation

The PCS-EP uses a complex sample design and estimation method, and the survey weights are therefore not equal for all the sampled units. When producing estimates and statistical tables, users **must** apply the proper survey weights. 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.

6.3 Release guidelines for quality

Before releasing and/or publishing any estimates, analysts should consider the quality level of the estimate. Given the experimental nature of the PCS-EP and its high degree of non-response, all estimates produced using the web panel should be accompanied by a quality warning to use the estimates with caution.

While data quality is affected by both sampling and non-sampling errors, this section covers quality in terms of sampling error. It is considered a best practice at Statistics Canada to report the sampling error of an estimate through its 95% confidence interval (CI). The confidence interval should be released with the estimate, in the same table as the estimate. In addition to the confidence intervals, PCS-EP estimates are categorized into one of two release categories:

Category E

The estimate and confidence interval should be flagged with the letter E (or some similar identifier) and accompanied by a quality warning to use the estimate with caution. Data users should use the 95% confidence interval to assess whether the quality of the estimate is sufficient.

Category F

The estimate and confidence interval are not recommended for release. They are deemed of such poor quality, that they are not fit for any use; they contain a very high level of instability, making them unreliable and potentially misleading. If analysts insist on releasing estimates of poor quality, even after being advised of their accuracy, the estimates should be accompanied by a disclaimer. Analysts should acknowledge the warnings given and undertake not to disseminate, present or report the estimates, directly or indirectly, without this disclaimer. The estimates should be flagged with the letter F (or some similar identifier) and the following warning should accompany the estimates and confidence intervals: "Please be warned that these estimates and confidence intervals [flagged with the letter F] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and may be invalid."

The rules for assigning an estimate to a release category depends on the type of estimate.

Release Rules for Estimated Proportions and Estimated Counts

Estimated proportions and estimated counts are computed from binary variables. Estimated counts are estimates of the total number of persons/households with a characteristic of interest; in other words, they are the weighted sum of a binary variable (e.g., estimated number of immigrants). Estimated proportions are estimates of the proportion of persons/households with a characteristic of interest (e.g., estimated proportion of immigrants in the general population). Estimated counts and proportions can also be computed from categorical variables: that is, estimates of the number or proportion of persons/household who belong to a category.

The release rules for estimated proportions and estimated counts are based on sample size. Table 1 provides the release rules for the PCS-EP, for all estimated proportions and counts except estimates for visible minorities.

Table 1: General rules for proportions and counts, except visible minority estimates

Sample Size (n)	Release Category	Action
$n \geq 200$	E	Release with quality warning; users should use CI as quality indicator
$n < 200$	F	Suppress the estimate and its CI for quality reasons

For estimated proportions, n is defined as the unweighted count of the number of respondents in the denominator (not the numerator) of the proportion. For estimated counts, n is defined as the unweighted count of the number of respondents with nonzero values that contribute to the estimate.

Special rules for estimates by visible minority

Table 2 provides special release rules that are to be used whenever estimates are produced for a visible minority group (i.e., using VISMIN or VISMINFL). Special rules are required because of the GSS-SI sample design that included an oversample of certain visible minority groups.

Table 2: Special rules for proportions and counts for visible minority estimates

Sample Size (n)	Release Category	Action
$n \geq 350$	E	Release with quality warning; users should use CI as quality indicator
$n < 350$	F	Suppress the estimate and its CI for quality reasons

Given the number of respondents to the PCS-EP, these rules imply that individual visible minority groups cannot be used as domains for analysis based on the PCS-EP but that analysis by VISMINFL is permissible. On the other hand, given that the experiences of different visible minority groups can be very different from each other, it may not be suitable to produce an estimate for all visible minority groups together (VISMINFL = 1). It is therefore recommended that, even though these estimates should not be disseminated, estimates by the more disaggregated VISMIN categories be compared between them before deciding to group all visible minority groups together.

Release Rules for Means and Totals of Quantitative Variables

The release rules for the estimated means and totals of quantitative variables or amounts are based on the sample size and on the CV of the estimate. Table 3 provides the release rules for the PCS-EP, except visible minority estimates.

Table 3: General rules for means and totals

Sample Size (n)	Release Category	Action
$n \geq 200$ and $CV \leq 50\%$	E	Release with quality warning; users should use CI as quality indicator
$n < 200$ or $CV > 50\%$	F	Suppress the estimate and its CI for quality reasons

For estimated means, n is defined as the unweighted count of the number of respondents that contribute to the estimate including values of zero. For estimated totals, n is defined as the unweighted count of the number respondents with nonzero values that contribute to the estimate.

Special rules for estimates by visible minority

Table 4 provides special release rules that are to be used whenever estimates are produced for a visible minority group (i.e., using VISMIN or VISMINFL). Special rules are required because of the GSS-SI sample design that included an oversample of certain visible minority groups.

Table 4: Special rules for means and totals for visible minority estimates

Sample Size (n)	Release Category	Action
$n \geq 350$ and $CV \leq 50\%$	E	Release with quality warning; users should use CI as quality indicator
$n < 350$ or $CV > 50\%$	F	Suppress the estimate and its CI for quality reasons

Given the number of respondents to the PCS-EP, these rules imply that individual visible minority groups cannot be used as domains for analysis based on the PCS-EP but that analysis by VISMINFL is permissible. On the other hand, given that the experiences of different visible minority groups can be very different from each other, it may not be suitable to produce an estimate for all visible minority groups together (VISMINFL = 1). It is therefore recommended that, even though these estimates should not be disseminated, estimates by the more disaggregated VISMIN categories be compared between them before deciding to group all visible minority groups together.

Release Rules for Differences

In order to assign a release category for an estimated difference between two estimates, the analyst must first determine the release category of each of the two estimates using the rules described above. Next, the release category of the estimated difference or the estimate of change is assigned the lower release category of the two estimates; this can be specified as follows:

- If one or both estimates are category F estimates, then assign the estimated difference to category F and suppress it.
- Otherwise, assign the estimated difference to category E and release with a quality warning.

Additional Rules Regarding Confidence intervals

The above release rules should suppress most estimates and confidence intervals of poor quality. There are also two additional conditions that indicate that a confidence interval is of poor quality. An estimate and its confidence interval should be assigned to release category F if either of the following two conditions are true:

- The lower bound of the 95% confidence interval is equal to the upper bound of the interval; in other words, the confidence interval is of length zero. (Exceptions are if the estimate corresponds to a calibration control total.)
- The lower bound or upper bound of the 95% confidence interval is not a plausible value for the estimate. For example, the lower bound for an estimated proportion is negative.