Analytical Guide - Canadian Perspectives Survey Series 6: Substance Use and Stigma during the Pandemic

1.0   Description

The Canadian Perspectives Survey Series (CPSS) is a set of short, online surveys beginning in March 2020 that will be used to collect information on the knowledge and behaviours of residents of the 10 Canadian provinces. All surveys in the series will be asked of Statistics Canada's probability panel. The probability panel for the CPSS is a new pilot project initiated in 2019. An important goal of the CPSS is to directly collect data from Canadians in a timely manner in order to inform policy makers and be responsive to emerging data needs. The CPSS is designed to produce data at a national level (excluding the territories).

The survey program is sponsored by Statistics Canada. Each survey in the CPSS is cross sectional. Participating in the probability panel and the subsequent surveys of the CPSS is voluntary.

The sixth survey of the CPSS is CPSS6 – Substance Use and Stigma during the Pandemic. It was administered from January 25, 2021 until January 31, 2021.

Any questions about the survey, the survey series, the data 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: statcan.csdidclientservice-ciddsservicealaclientele.statcan@canada.ca

2.0   Survey methodology

Target and survey population

The target population for the Canadian Perspectives Survey Series (CPSS) is residents of the 10 Canadian provinces 15 years of age or older.

The frame for surveys of the CPSS is Statistics Canada's pilot probability panel. The probability panel was created by randomly selecting a subset of the Labour Force Survey (LFS) respondents. Therefore the survey population is that of the LFS, with the exception that full-time members of the Canadian Armed Forces are included. Excluded from the survey's coverage are: persons living on reserves and other Aboriginal settlements in the provinces; the institutionalized population, and households in extremely remote areas with very low population density. These groups together represent an exclusion of less than 2% of the Canadian population aged 15 and over.

The LFS sample is drawn from an area frame and is based on a stratified, multi-stage design that uses probability sampling. The LFS uses a rotating panel sample design. In the provinces, selected dwellings remain in the LFS sample for six consecutive months. Each month about one-sixth of the LFS sampled dwellings are in their first month of the survey, one-sixth are in their second month of the survey, and so on. These six independent samples are called rotation groups.

For the probability panel used for the CPSS, four rotation groups from the LFS were used from the provinces: the rotation groups answering the LFS for the last time in April, May, June and July of 2019.

From these households, one person aged 15+ was selected at random to participate in the CPSS - Sign-Up. These individuals were invited to Sign-Up for the CPSS. Those agreeing to join the CPSS were asked to provide an email address. Participants from the Sign-Up that provided valid email addresses formed the probability panel. The participation rate for the panel was approximately 23%. The survey population for all surveys of the CPSS is the probability panel participants. Participants of the panel are 15 years or older as of July 31, 2019.

Sample Design and Size

The sample design for surveys of the CPSS is based on the sample design of the CPSS – Sign-Up, the method used to create the pilot probability panel. The raw sample for the CPSS – Sign-Up had 31,896 randomly selected people aged 15+ from responding LFS households completing their last interview of the LFS in April to July of 2019. Of these people, 31,626 were in-scope at the time of collection for the CPSS - Sign-Up in January to March 2020. Of people agreeing to participate in the CPSS, that is, those joining the panel, 7,242 had a valid email address. All panel participants are invited to complete the surveys of the CPSS.

Stages of the Sample	n
Raw sample for the CPSS – Sign-Up	31,896
In-scope Units from the CPSS – Sign-Up	31,628
Panelists for the CPSS (with valid email addresses)	7,242
Raw sample for surveys of the CPSS	7,242

3.0   Data collection

CPSS – Sign-Up

The CPSS- Sign-Up survey used to create Statistics Canada's probability panel was conducted from January 15th, 2020 until March 15th, 2020. Initial contact was made through a mailed letter to the selected sample. The letter explained the purpose of the CPSS and invited respondents to go online, using their Secure Access Code to complete the Sign-Up form. Respondents opting out of joining the panel were asked their main reason for not participating. Those joining the panel were asked to verify basic demographic information and to provide a valid email address. Nonresponse follow-up for the CPSS-Sign-Up had a mixed mode approach. Additional mailed reminders were sent to encourage sampled people to respond. As well, email reminders (where an email address was available) and Computer Assisted Telephone Interview (CATI) nonresponse follow-up was conducted.

The application included a standard set of response codes to identify all possible outcomes. The application was tested prior to use to ensure that only valid question responses could be entered and that

all question flows would be correctly followed. These measures ensured that the response data were already "clean" at the end of the collection process.

Interviewers followed a standard approach used for many StatCan surveys in order to introduce the agency. Selected persons were told that their participation in the survey was voluntary, and that their information would remain strictly confidential.

CPSS6 – Substance Use and Stigma during the Pandemic

All participants of the pilot panel for the CPSS, minus those who opted out after previous iterations of CPSS, were sent an email invitation with a link to the CPSS6 and a Secure Access Code to complete the survey online. Collection for the survey began on January 25th, 2021. Reminder emails were sent on January 26th, January 28th and January 30th. The application remained open until January 31, 2021.

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 are 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: nonresponse, coverage, measurement and processing.

4.1.1 Nonresponse

Nonresponse errors result from a failure to collect complete information on all units in the selected sample.

Nonresponse produces errors in the survey estimates in two ways. Firstly, non-respondents often have different characteristics from respondents, which can result in biased survey estimates if nonresponse bias is not fully corrected through weighting. Secondly, it reduces the effective size of the sample, since fewer units than expected answered the survey. As a result, the sampling variance increases and the precision of the estimate decreases. The response rate is calculated as follows:

[Responding units / (Selected units – out-of-scope units)] x 100%

The following table summarizes the response rates experienced for the CPSS6 – Substance Use and Stigma during the Pandemic. Response rates are broken down into two stages. Table 4.1.1a shows the take-up rates to the panel in the CPSS- Sign-Up and Table 4.1.1b shows the collection response rates for the survey CPSS6 – Substance Use and Stigma during the Pandemic.

Table 4.1.1a Participation in the Pilot Probability Panel for the CPSS – Sign-Up

	Stages of the Sample for the CPSS – Sign-Up			
	Raw sample for the CPSS – Sign-Up	In-scope Units from the CPSS – Sign-Up	Panelists for the CPSS (with valid email addresses)	Participation Rate for the Panel for CPSS
n	31,896	31,628	7,242	22.9%

Table 4.1.1b Response Rates for the CPSS6 – Substance Use and Stigma during the Pandemic

	Stages of the Sample for the CPSS6 – Substance Use and Stigma during the Pandemic			
	Panelists for the CPSS (with valid email addresses)	Respondents of CPSS6 – Substance Use and Stigma during the Pandemic	Collection Response Rate for CPSS6 – Substance Use and Stigma during the Pandemic	Cumulative Response Rate
n	7,242	3,941	54.4%	12.5%

As shown in Table 4.1.1b, the collection response rate for the CPSS6 – Substance Use and Stigma during the Pandemic is 54.4%. However, when nonparticipation in the panel is factored in, the cumulative response rate to the survey is 12.5%. This cumulative response rate is lower than the typical response rates observed in social surveys conducted at Statistics Canada. This is due to the two stages of nonresponse (or participation) and other factors such as the single mode used for surveys of the CPSS (emailed survey invitations with a link to the survey for online self-completion), respondent fatigue from prior LFS response, the inability of the offline population to participate, etc.

Given the additional nonresponse experienced in the CPSS6 – Substance Use and Stigma during the Pandemic, there is an increased risk of bias due to respondents being different than non-respondents. For this reason, a small bias study was conducted. Please see Section 6.0 for the results of this validation.

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; in the case of a census they may be the main source 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.

For the CPSS, the population covered are those aged 15+ as of July 31, 2019. Since collection of the CPSS6 – Substance Use and Stigma during the Pandemic was conducted from January 25th – 31st, 2021, there is an undercoverage of residents of the 10 provinces that turned 15 since July 31, 2019. There is also undercoverage of those without internet access. This undercoverage is greater amongst those age 65 years and older.

4.1.3 Measurement errors

Measurement errors (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. It is very costly to accurately measure the level of response error and very few surveys conduct a post-survey evaluation.

4.1.4 Processing errors

Processing errors are the errors associated with activities conducted once survey responses have been received. It includes 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 errors are defined as the errors that result 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 for a given sample size 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, nonresponse 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 is easier to interpret since it 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) is a relative measure of the sampling error. It is defined as the estimate of the standard error divided by the estimate itself, usually expressed as a percentage (10% instead of 0.1). 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, then 95% of the time (or 19 times out of 20), the confidence interval would cover the true population value.

5.0 Weighting

The principle behind estimation in a probability sample such as those of the CPSS, 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. In the terminology used here, it can be said that each person has a weight of 50.

The weighting phase is a step that calculates, for each person, his or her 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 CPSS6 – Substance Use and Stigma during the Pandemic.

The weighting of the sample for the CPSS6 – Substance Use and Stigma during the Pandemic 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 first create the panel weights, then the weighting steps to create the survey weights for CPSS6 – Substance Use and Stigma during the Pandemic.

5.1 Creating the Panel Weights

Four consecutive rotate-out samples of households from the LFS were the starting point to form the panel sample of the CPSS. Since households selected from the LFS samples are the starting point, the household weights from the LFS are the first step to calculating the panel weights.

5.1.1   Household weights

Calculation of the Household Design Weights – HHLD_W0, HHLD_W1

The initial panel weights are the LFS subweights (SUBWT). These are the LFS design weights adjusted for nonresponse but not yet calibrated to population control totals. These weights form the household design weight for the panel survey (HHLD_W0).

Since only four rotate-outs were used, instead of the six used in a complete LFS sample, these weights were adjusted by a factor of 6/4 to be representative. The weights after this adjustment were called HHLD_W1.

Calibration of the Household Weights – HHLD_W2

Calibration is a step to ensure that the sum of weights within a certain domain match projected demographic totals. The SUBWT from the LFS are not calibrated, thus HHLD_W1 are also not calibrated. The next step is to make sure the household weights add up to the control totals by household size. Calibration was performed on HHLD_W1 to match control totals by province and household size using the size groupings of 1, 2, or 3+.

5.1.2   Person Panel weights

Calculate Person Design Weights – PERS_W0

One person aged 15 or older per household was selected for the CPSS – Sign-Up, the survey used to create the probability panel. The design person weight is obtained by multiplying HHLD_W2 by the number of eligible people in the dwelling (i.e. number of people aged 15 years and over).

Removal of Out of Scope Units – PERS_W1

Some units were identified as being out-of-scope during the CPSS – Sign-Up. These units were given a weight of PERS_W1 = 0. For all other units, PERS_W1 = PERS_W0. Persons with a weight of 0 are subsequently removed from future weight adjustments.

Nonresponse/Nonparticipation Adjustment – PERS_W2

During collection of the CPSS – Sign-Up, a certain proportion of sampled units inevitably resulted in nonresponse or nonparticipation in the panel. Weights of the nonresponding/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).

Many variables from the LFS were available to build the RHG (such as employment status, education level, household composition) as well as information from the LFS collection process itself. The model was specified by province, as the variables chosen in the model could differ from one province to the other.

The following variables were kept in the final logistic regression model: education_lvl (education level variable with 10 categories), nameissueflag (a flag created to identify respondents not providing a valid name), elg_hhldsize (number of eligible people for selection in the household), age_grp (age group of the selected person), sex, kidsinhhld (an indicator to flag whether or not children are present in the household), marstat (marital status with 6 categories), cntrybth (an indicator if the respondent was born in Canada or not), lfsstat (labour force status of respondent with 3 categories), nocs1 (the

first digit of National Occupational Classification code of the respondent if employed, with 10 categories), and dwelrent (an indicator of whether the respondent dwelling is owned or rented). RHGs were formed within provinces. 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 PERS_W2 weights, adjusted for panel nonparticipation. The nonparticipating units were dropped from the panel.

5.2   Creating the CPSS6 weights

Surveys of the CPSS start with the sample created from the panel participants. The panel is comprised of 7,242 individuals, each with the nonresponse adjusted weight of PERS_W2.

Calculation of the Design Weights – WT_DSGN

The design weight is the person weight adjusted for nonresponse calculated for the panel participants (PERS_W2). No out-of-scope units were identified during the survey collection of CPSS6 – Substance Use and Stigma during the Pandemic. Since all units were in-scope, WT_DSGN =PERS_W2 and no units were dropped.

Nonresponse Adjustment – WT_NRA

Given that the sample for CPSS was formed by people having agreed to participate in a web panel, the response rates to the survey were relatively high. Additionally, the panel was designed to produce estimates at a national level, so sample sizes by province were not overly large. As a result, nonresponse was fairly uniform in many provinces. The RHGs were formed by some combination of age group, sex, education level, rental status, LFS status, whether or not children are present in the household, marital status and the first digit of the National Occupational Classification (NOC) code for respondents who are employed. 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 WT_NRA weights, adjusted for survey response. The nonresponding units were dropped from the survey.

Calibration of Person-Level Weights – WT_FINL

Control totals were computed using LFS 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. Most social surveys calibrate the person level

weights to control totals by sex, age group and province. For CPSS6, calibration by province was not possible, since there were very few respondents in some categories in the Atlantic and Prairie Provinces. In addition, there were very small counts for male respondents aged 15 to 24 in the Atlantic Provinces. For this reason, the control totals used for CPSS6 – Substance use and stigma during the pandemic were by age group and sex by geographic region, where the youngest age group for males in the Atlantic region was collapsed with the second youngest age group. The next section will include recommendations for analysis by geographic region and age group.

5.3   Bootstrap Weights

Bootstrap weights were created for the panel and the CPSS6 – Substance Use and Stigma during the Pandemic survey respondents. The LFS bootstrap weights were the initial weights and all weight adjustments applied to the survey weights were also applied to the bootstrap weights.

6.0   Quality of the CPSS and Survey Verifications

The probability panel created for the CPSS is a pilot project started in 2019 by Statistics Canada. While the panel offers the ability to collect data quickly, by leveraging a set of respondents that have previously agreed to participate in multiple short online surveys, and for whom an email address is available to expedite survey collection, some aspects of the CPSS design put the resulting data at a greater risk of bias. The participation rate for the panel is lower than typically experienced in social surveys conducted by Statistics Canada which increases the potential nonresponse bias. Furthermore, since the surveys of the CPSS are all self-complete online surveys, people without internet access do not have the means to participate in the CPSS and therefore are not covered.

When the unweighted panel was compared to the original sample targeted to join the panel, in particular there was an underrepresentation of those aged 15-24, those aged 65 and older, and those with less than a high school degree. These differences were expected due to the nature of the panel and the experience of international examples of probability panels. Using LFS responding households as the frame for the panel was by design in order to leverage the available LFS information to correct for the underrepresentation and overrepresentation experienced in the panel. The nonresponse adjustments performed in the weighting adjustments of the panel and the survey respondents utilised the available information to ensure the weights of nonresponding/nonparticipating units went to similar responding units. Furthermore, calibration to age and sex totals helped to adjust for the underrepresentation by age group.

Table 6.1 shows the slippage rates by certain domains post-calibration of CPSS6 – Substance Use and Stigma during the Pandemic. The slippage rate is calculated by comparing the sum of weights in the domain to that of the control total based off of demographic projections. A positive slippage rate means the sample has an over-count for that domain. A negative slippage rate means the survey has an under-count for that domain. Based on the results shown in Tables 6.1 and 6.2, it is recommended to only use the data at the geographical levels and age groups where there is no slippage. That is nationally, by geographic region (Maritime Provinces, Quebec, Ontario, Prairie Provinces, and British Columbia), and by the four oldest age groups.

Table 6.1 Slippage rates by geographic region

Area	Domain	n	Slippage Rate
Geography	Canada*	3,941	0%
	Newfoundland and Labrador	111	-9.6%

	Prince Edward Island	86	18.3%
	Nova Scotia	224	2.2%
	New Brunswick	169	0.1%
	Quebec	653	0%
	Ontario	1,170	0%
	Manitoba	315	-3.0%
	Saskatchewan	270	-3.0%
	Alberta	424	1.7%
	British Columbia	519	0%

Based on the 10 provinces; territories are excluded

Table 6.2 Slippage rates by age group

Area	Domain	n	Slippage Rate
Age Group	15-24	193	3.1%
	25-34	445	-2.7%
	35-44	629	0%
	45-54	654	0%
	55-64	917	0%
	65+	1,103	0%

After the collection of CPSS6 – Substance Use and Stigma during the Pandemic, a small study was conducted to assess the potential bias due to the lower response rates and the undercoverage of the population not online. The LFS data was used to produce weighted estimates for the in-scope sample targeted to join the probability panel (using the weights and sample from PERS_W1). The same data was used to produce weighted estimates based on the set of respondents from the CPSS6 survey and the weights WT_FINL. The two set of estimates were compared and are shown in Table 6.3. The significant differences are highlighted.

Table 6.3 Changes in estimates due to nonparticipation in the CPSS and the COVID-19 survey

Subject	Recoded variables from 2019 LFS	Estimate for in-scope population (n=31,628)	Estimate for W6 of CPSS (n=3,941)	% Point Difference
Education	Less than High School	15.5%	13.4%	2.1%
	High School no higher certification	25.9%	25.5%	0.4%
	Post-secondary certification*	58.6%	61.1%	-2.4%
Labour Force Status	Employed	61.2%	61.7%	-0.6%
	Unemployed	3.4%	3.3%	0.1%
	Not in Labour Force	35.3%	34.9%	0.4%
Country of Birth	Canada*	71.7%	75.4%	-3.7%
Marital Status	Married/Common-law	60.4%	60.6%	-0.2%
	Divorced, separated, widowed	12.8%	12.1%	0.6%
	Single, never married	26.9%	27.2%	-0.4%
Kids	Presence of children*	31.7%	34.5%	-2.8%
Household Size	Single person	14.4%	14.4%	0.1%
	Two person HH	34.8%	36.6%	-1.8%
	Three or more people	18.4%	19.4%	-1.1%
Eligible people for panel	One eligible person aged 15+	15.9%	16.1%	-0.2%
	Two eligible people*	49.3%	52.1%	-2.8%
	Three or more eligible people*	34.8%	31.8%	3.1%
Dwelling	Apartment	12.1%	12.3%	-0.2%
	Rented	24.8%	24.1%	0.7%

Occupational Code	Management occupations (NOC0)	6.0%	5.8%	0.2%
	Business Finance and Administration (NOC1)	10.7%	11.3%	-0.6%
	Natural and Applied Sciences and related occupations (NOC2)*	5.2%	7.0%	-1.8%
	Health Occupations (NOC3)	4.7%	4.3%	0.4%
	Occupations in education, law and social, community and government services (NOC4)	7.6%	8.3%	-0.7%
	Occupations in art, culture, recreation and sports (NOC5)	2.5%	3.2%	-0.7%
	Sales and service occupations (NOC6)	16.6%	16.1%	0.5%
	Trades, transport and equipment operators and related occupations (NOC7)	9.6%	9.0%	0.6%
	Natural resources, agriculture and related production occupations (NOC8)*	1.6%	0.9%	0.7%
	Occupations in manufacturing and utilities (NOC9)	2.9%	2.6%	0.4%

Estimates that are significantly different at α= 5%.

While many estimates do not show significant change, the significant differences show that some bias remains in the CPSS6 – Substance use and stigma during the pandemic. There is an underrepresentation of those where there were three or more eligible participants for the panel and of people working in NOCS8. And there is an overrepresentation of those with a post-secondary certification, of people born in Canada, of people working in NOC2, of households where there were two eligible participants for the panel, and of households with children. These small differences should be kept in mind when using the CPSS6 – Substance use and stigma during the pandemic survey data. Investigation about differences in estimates is ongoing, and as evidence of differences are identified, strategies are being tested to improve the methodology from one wave of the survey to the next.