**Proceedings of Statistics Canada Symposium 2022:**
**Data Disaggregation: building a more representative data portrait of society**

# Survey Series on People and their Communities

by Agnes Waye and Cilanne Boulet

Release date: March 25, 2024

Statistics Canada    Statistique Canada

Canada

# Survey Series on People and their Communities

Agnes Waye, Cilanne Boulet[1]

## Abstract

In 2021, Statistics Canada initiated the Disaggregated Data Action Plan, a multi-year initiative to support more representative data collection methods, enhance statistics on diverse populations to allow for intersectional analyses, and support government and societal efforts to address known inequalities and bring considerations of fairness and inclusion into decision making. As part of this initiative, we are building the Survey Series on People and their Communities, a new probabilistic panel specifically designed to collect data that can be disaggregated according to racialized group. This new tool will allow us to address data gaps and emerging questions related to diversity. This paper will give an overview of the design of the Survey Series on People and their Communities.

Key Words:  Disaggregated data; Panel; Diversity; Intersectional analyses.

## 1. Introduction

### 1.1 Background

In recent years, there has been an increasing demand for disaggregated data to support evidence-based policy making. In 2021, Statistics Canada launched the Disaggregated Data Action Plan (DDAP). The DDAP has the goal of supporting data collection to reflect the diversity of the Canadian population and to provide key insights on underrepresented groups. It also aims to provide detailed statistical information about specific population groups, such as women or racialized groups. In this context, having information at this fine level of detail is important for intersectional analyses.

### 1.2 Survey Series on People and their Communities

The Survey Series on People and their Communities (SSPC) is one of the initiatives undertaken as part of the DDAP. The SSPC is a survey series of four social surveys covering a variety of topics such as community engagement, trust in people and confidence in Canadian institutions. The SSPC has several key objectives. First, it aims to measure the experience of racialized groups and immigrants in Canada. Its secondary objective is to produce disaggregated data for targeted groups by geographic region. It also aims to make comparisons between racialized groups as well as with the non-racialized group. The SSPC aims to fill data gaps and needs related to diverse groups to support various Canadian governmental programs and services.

## 2. Survey Design

### 2.1 Survey Overview

The design of the SSPC is a quarterly probabilistic panel. The term "panel" can have different meanings depending on the context. In the framework of the SSPC, a panel refers to recruiting a group of individuals from an initial

[1]Agnes Waye, Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6 (agnes.waye@statcan.gc.ca); Cilanne Boulet, Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6  (cilanne.boulet@statcan.gc.ca)

probability sample who provide their contact information to participate in the subsequent surveys of the series. Not every person selected in the sample stays in the panel; only units who agree and give contact information remain in the panel. These SSPC sample units will be contacted up to three additional times over the course of a year.

The main benefit of creating a panel is to have a readily available group of people who can be contacted quickly when emerging data needs arise. Panels are extremely helpful for reaching populations quickly, which Statistics Canada had experienced first-hand when it carried out its first experimental web panel at the beginning of the COVID-19 pandemic. More details on this first experience with web panels can be found in Baribeau (2020).

The SSPC is Statistics Canada's first time using a probabilistic panel that aims to reach a specific segment of the Canadian population. In order to reflect the experiences of racialized groups and immigrants in Canada, the SSPC sample is stratified by racialized group, immigrant status, and region, with sample sizes determined to ensure that disaggregation objectives can be met for groups that are large enough. Particular attention was paid to avoid excessive oversampling in small strata to ensure that other DDAP initiatives could also be implemented without imposing undue response burden. The sample included a control group made up of people who do not belong to racialized groups. The reason for having this control group is to make comparisons between those in racialized groups and those in non-racialized groups. The control group was also stratified by immigration status with an oversampling of immigrants. This was needed to ensure that comparisons could be made to the experiences of immigrants who make up a substantial portion of the targeted racialized groups.

## 2.2 Target Population and Frame

The target population of the SSPC is the population of Canadians who are age 15 years or older and living in one of the ten Canadian provinces. In practice, the surveyed population is slightly smaller than the target population. Individuals living in institutions and on reserves are excluded from the sample.

The sampling frame of this survey consists of the respondents to the 2021 Canadian Census of Population long-form. (The short-form Canadian Census of Population was conducted for the entire Canadian population, whereas the long-form Census was conducted on approximately 25% of the Canadian population.) The long-form Census asked detailed questions on many topics, including racialized and cultural group identity, education, marital status, etc. In addition, immigration status is available for long-form respondents as it was integrated via linkage to administrative data as part of Census data processing. As such, the Census long-form provides a natural frame for the SSPC as it allows for targeted oversampling in order to be able to disaggregate the data by racialized groups, geographic region, immigrant status (non-immigrant, recent immigrant, and non recent immigrant).
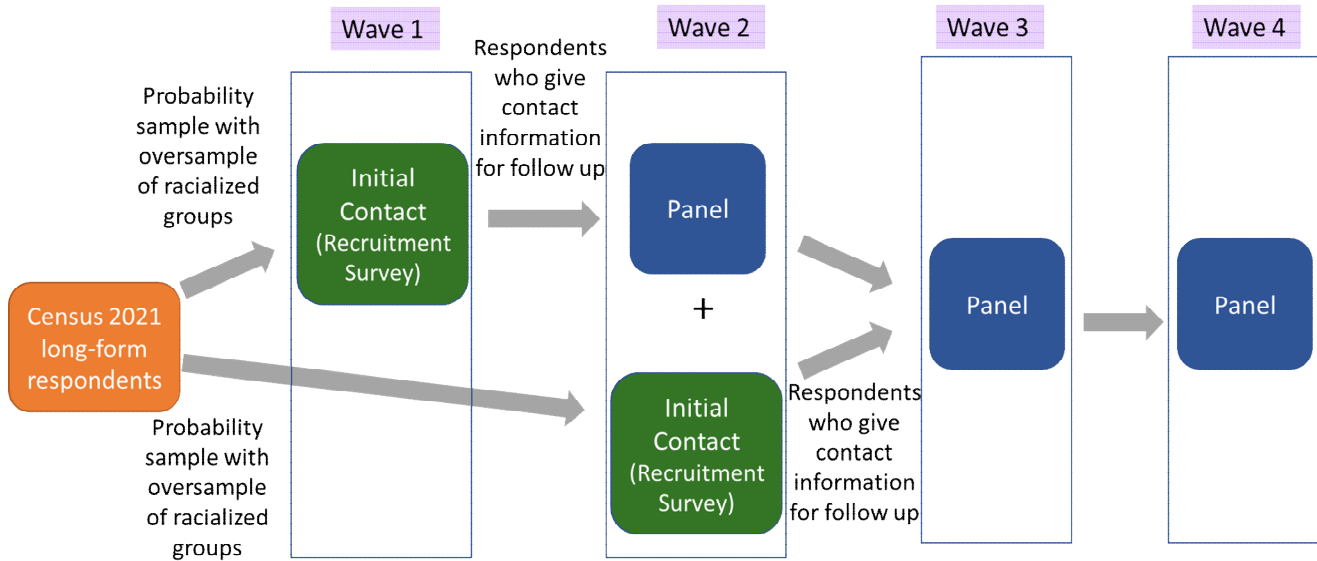
## 2.3 Waves of SSPC

There are four waves in the SSPC, each with a different questionnaire. Initial contact occurs at either Wave 1 or 2 and is used to recruit the panelists to be contacted in subsequent waves. Due to operational constraints, the recruitment survey for SSPC was split over two waves, instead of having only one wave for recruitment. This split of recruitment over two waves will also allow for adjustments to the sample design at Wave 2, if needed.

The sample at Wave 1 is drawn from the frame with an oversample of racialized groups. Initial contact is made with the sample units, they are asked to complete a survey and, at the end of the survey, they are asked for their contact information (email and phone number) to participate in the future surveys of the series. The Wave 1 respondents who give contact information to participate in follow-up questionnaires will be recontacted at Waves 2, 3, and 4.

At Wave 2, a new non-overlapping sample is selected from the frame. This sample is asked to complete a survey consisting of Wave 1 and Wave 2 questions and, at the end of the survey, they are also asked for their contact information (email and phone number) to participate in the future surveys of the series. The Wave 2 respondents who give contact information to participate in follow-up questionnaires, will be recontacted at Waves 3 and 4. At Wave 2, in addition to this new sample, we will also contact units from Wave 1 who had given us phone or email and agreed to participate in follow up questionnaires; this group is only asked the Wave 2 questions.

For Waves 3 and 4, we will contact everyone who gave contact information for follow-up from both waves. Figure 2.3-1 below summarises the various waves of the SSPC.

**Figure 2.3-1**
**Waves of the Survey Series on People and their Communities**



## 2.4 Mode of Collection

Collection from Wave 1 takes place from October 2022 to January 2023. The mode of collection for Waves 1 and 2 is primarily a self-administered web questionnaire, along with some CATI (computer assisted telephone interviewing) non-response follow-up. Collection at Waves 3 and 4, from panel participants having provided personal contact information at the recruitment survey, is expected to be a lot quicker. It will also include the use of both web and CATI collection.

There is some support for collection in non-official languages, such as Spanish and Arabic, which is important for reaching respondents who may not be able to communicate in English or French.

## 2.5 Survey Content

For Wave 1 of the SSPC, we asked various socio-demographic questions (such as education level and marital status) as well as subject matter specific questions. Some examples of subject matter questions include participation in groups (such as clubs and associations), sense of belonging to the local community, and confidence in institutions.

It is important to note that there is some overlap between questions that were asked on the long-form Census and the SSPC. In the SSPC, we asked questions for variables where the answer could have changed since the Census. For example, education or marital status could have changed in the one and a half years between the Census and the SSPC collection periods. The question about racialized group identity was asked on the SSPC even though it was already asked on the long-form Census because the concept of racialized group identity can be fluid and can change over time or in different situations.

In addition to the time elapsed between the long-form Census and the SSPC collection periods, another reason for asking some questions again on SSPC is to get more accurate answers. For the long-form census, one person could answer questions on behalf of all household members, so it is possible that the sampled individual for the SSPC may give a response that is different than what was answered on the long-form Census.

Table 2.5-1 below shows the variables coming from the long-form Census and the SSPC Wave 1. For the future waves, the survey content has not been finalized yet, but topics such as discrimination, political engagement and participation in sports are being considered.

**Table 2.5-1**
**Comparison of Variables from the Census and SSPC**

| Variable | Census | SSPC Wave 1 |
|---|---|---|
| Age, Sex, Gender | x | x |
| Marital Status | x | x |
| Income | x | |
| Language | x | |
| Racialized groups | x | x |
| Immigration | x | |
| Education | x | x |
| Labour | x | x |
| Sexual Orientation | | x |
| Disability | x | x |

## 2.6 Sample Design at Wave 1

At Wave 1, 40,000 units were sampled for the SSPC. There were various constraints to consider at sampling because of other Statistics Canada surveys that were also interested in surveying the same target population (racialized groups). Therefore, it was important to keep this in mind when determining the SSPC sample size.

Since the objective of the SSPC is to have data for very specific population groups, we needed to create a high number of strata. Some of these strata sizes are quite small. The limited overall sample size, together with small strata sizes, meant that we had very few units to sample in some of these strata. As a result, it may not be possible to produce reliable estimates for all strata. For example, in certain regions, such as the Atlantic, there are very few members of some specific racialized groups. Many estimates at the regional level can be produced for the recruitment waves, but for the panel waves, where the number of participants will be lower, only national level estimates are targeted by the sample design. The constraints above made sampling quite challenging for the SSPC since we had to adopt a sample design that guaranteed a certain number of sampling units for every domain of interest where disaggregated data must be produced.

A portion of the SSPC sample was selected using another method, which will be described in section 3.2.

## 2.7 Expected Response Rate

We were able to calculate an expected response rate for SSPC based on our experience with other similar surveys at Statistics Canada. When we initially contact the sample units, we expect 50% of these units to respond to the recruitment survey. Of those who respond, we only expect 50% to give contact information for subsequent waves. For the panel waves, we expect a 40% response rate. Therefore, the cumulative response rate is expected to be around 50%*50%*40% = 10%. This cumulative response rate is expected for those who we recruit at Wave 1 as well as those who we recruit at Wave 2. This expected response rate is quite low and can be cause for concern with regards to nonresponse bias. However, we have a very wide range of variables available from the long-form Census which can help us significantly when it comes to our nonresponse adjustment model at the weighting step. Despite this low response rate, we still want to use panels because of its extremely important advantage of providing timely information, as mentioned earlier. For the Wave 1 respondents of the SSPC, we see that the percentage of people who have given their contact information so far is a lot higher than 50%, meaning that the estimated cumulative response rate above is likely conservative.

## 3. Challenges and Lessons Learned

## 3.1 Contact Information

A major challenge for Wave 1 of the SSPC was adding in recent contact information for sample units. The contact information is available from the Census, but since collection for the SSPC took place more than one year after the census reference date, we tried updating the contact information by integrating new information from different sources.

For addresses, we prioritized various sources according to the quality of the data, the date when the information was last updated, and we checked for coherence between the different sources. We prioritized addresses that matched between multiple sources. Recent immigrants may also tend to move more frequently, which may make it more difficult to have reliable contact information for them. For future waves, we may explore the possibility of consulting address sources from special programs that provide support for new immigrants.

We cleaned and standardized addresses, phone numbers, names, and email addresses, which took a considerable amount of effort since we combined many different sources of data. We had to take special considerations into account when working with names that were not of English or French origin. We also used these sources to identify sample units who had died or emigrated since the census date.

## 3.2 Proxy Ethnicity Indicator

A portion of the SSPC sample units was based on the Proxy Ethnicity Indicator (PEI). The PEI is an experimental research and development initiative under the Longitudinal Social Data Development Program at Statistics Canada. The PEI is an algorithm used to assign, either directly, indirectly or via modelling racialized group identity at the individual level for the entire Canadian population. This indicator is based on administrative data and historical census data.

Recall that the main SSPC sample is selected from the long-form Census, which covers 25% of Canadians. One advantage of using the PEI is it can be defined for the entire Canadian population. Also, the PEI can be produced on an annual basis, whereas the Census is conducted every 5 years. Therefore, in terms of coverage and timeliness, there are obvious advantages to using the PEI to create a frame that includes information about likely racialized group membership of individuals.

The SSPC integrated a sample based on the PEI as an experiment for Wave 1. At Wave 1, 12.5% of the sample consisted of units that were selected from the short-form Census and were assigned a racialized group identity based on the PEI. Since the racialized group identity question will be asked on SSPC, this will help validate the PEI and evaluate the possibility of using PEI for targeted sampling on future surveys. If we see that the PEI is effective at assigning racialized group identities, in the future, we will be able to sample from the entire Canadian population instead of only 25% of Canadians (which is currently the case for the SSPC frame at Wave 1). Covering the entire Canadian population will help to minimize the problems mentioned earlier at sampling, especially for cases where the strata sizes are small.

## 3.3 Conclusion

Panel surveys can be very powerful tools for collecting data in a timely manner and the SSPC is adapting this approach to gather information about the experiences of Canadians belonging to racialized groups. Implementing such surveys can come with special challenges and limitations. There are many possible improvements and changes we could make to future waves of SSPC and Wave 1 will serve as an exploratory phase from which lessons will be drawn for future iterations.

## References

Baribeau, B. (2020), "Breaking with Tradition: Piloting a Probability Web Panel", Internal Document, Statistics Canada.