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

# Labour Force Survey initiatives under Statistics Canada's Disaggregated Data Action Plan

by Martin Pantel and Yelly Camara

Statistics Canada   Statistique Canada

Canada

# Labour Force Survey initiatives under Statistics Canada's Disaggregated Data Action Plan

Martin Pantel and Yelly Camara[1]

## Abstract

In accordance with Statistics Canada's long-term Disaggregated Data Action Plan (DDAP), several initiatives have been implemented into the Labour Force Survey (LFS). One of the more direct initiatives was a targeted increase in the size of the monthly LFS sample. Furthermore, a regular Supplement program was introduced, where an additional series of questions are asked to a subset of LFS respondents and analyzed in a monthly or quarterly production cycle. Finally, the production of modelled estimates based on Small Area Estimation (SAE) methodologies resumed for the LFS and will include a wider scope with more analytical value than what had existed in the past. This paper will give an overview of these three initiatives.

Key Words: Labour Force Survey, Small Area Estimation, Disaggregation

## 1. Background

The Labour Force Survey (LFS) is a monthly survey measuring the current states of the Canadian labour market. It is a key source of data on employment, unemployment rate, and other labour market indicators. Its sample, composed of approximately 80,000 dwellings, is divided into six rotations groups, one of which is replaced every month with new dwellings called birth units. The sample allocation for this survey is based on quality targets using the coefficient of variation ($CV = Standard\ Error/Estimate$).

This paper gives an overview of three initiatives implemented into LFS in accordance with Statistics Canada's long-term Disaggregated Data Action Plan (DDAP), an agency-wide initiative whose goals include providing users with detailed statistical data that highlights the experiences of specific population groups, such as women, Indigenous peoples, racialized populations and people living with disabilities. Section 2 introduces the Supplement program, a new analytical product which can be adapted to address emerging labour market trends. Section 3 describes how an additional sample was integrated into an existing complex panel design and provides some early analyses of the impact on data quality. The fourth section of this paper presents the small area estimation modelling program that is associated with the LFS. Finally, the last section outlines plan for future work in the Supplement and small area estimation modelling programs.

## 2. LFS Supplement program

Monthly supplementary data started to be collected regularly in April 2020. The regular and dynamic Supplement program that is currently used began collecting monthly data in January 2022, and quarterly data in July 2022. To control overall burden, the questions are asked to a subset of LFS respondents. Indeed, as shown in table 2-1., data for the monthly supplement is collected from the respondents in their second to fifth month in the sample i.e., 4/6 of the sample, and for quarterly supplements, from respondents in their last month in the sample. Eligibility also depends on the labour force status, for example questions on working from home would only be asked to employed respondents. The collected data is processed using appropriate imputation and reweighting strategies.

---

[1]Martin Pantel, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, KIA 0T6 (martin.pantel@statcan.gc.ca); Yelly Camara, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (yelly.camara@statcan.gc.ca).

**Table 2-1: Supplement program topics**

| | 1 | 2 - 5 | 6 |
|---|---|---|---|
| | | **Month in sample** | |
| | **1** | **2 - 5** | **6** |
| **Jan** | x | Working from home (WFH), Intention to leave job | Canadian Income Survey |
| **Feb** | x | WFH, Willingness to move | |
| | | … | |
| **Jun** | x | WFH, Benefits | |
| **Jul** | x | WFH, Training, Employee retention | Dependent contractors, Temporary workers |
| **Aug** | x | WFH, Work ethic | |
| **Sep** | x | WFH, Childcare | |
| **Oct** | x | WFH, Cost of living | (New Quarterly) |
| | | … | |

In the first months of the pandemic, one objective of the LFS was to better measure the impacts of COVID-19 on the labour market with topics such as work absences related to COVID-19, working from home, reception of benefits (e.g., the Canada Emergency Response Benefit, or CERB), and unequal impacts across subpopulations (e.g., racialized groups). Now, this program also provides a trial area for content that could eventually be transferred to the main LFS. For example, the identification of members of racialized groups was integrated into the main questionnaire in January 2022 along with appropriate processing methodologies (imputation and weight calibration). Figure 2-1 shows the impact in Newfoundland and Labrador (N.L.) of moving the question to the main LFS as of January 2022. Until that time (before the current Supplement program was implemented), the identification of racialized groups was done for 5 of the rotations. As expected, the increase from 5 to all 6 rotations improved the quality of the estimates in N.L. The coefficient of variation (CV) of the employment count went from 22% in December 2021 to 8% in January 2022. As shown in Table 2-2, the smaller variance means a narrower confidence interval (CI), making it easier to detect month-to-month changes and differences between groups.

**Figure 2-1: Quality of estimate of employed (CV), for racialized groups, in Newfoundland and Labrador**

**Table 2-2: Quality of estimate of employed (CI), for racialized groups, in Newfoundland and Labrador**

|  | Sample size | Estimate | 95% LB | 95% UB |
|---|---|---|---|---|
| December 2021 | 55 | 8,600 | 4,600 | 12,000 |
| January 2022 | 92 | 8,200 | 6,800 | 9,500 |

# 3. Sample increase

## 3.1 Objectives

The sample increase aimed for further demographic and geographic disaggregation. Indeed, the goal was to improve the quality of estimates for racialized groups as well as city-level estimates. However, the small population size of most of the racialized groups, their regional concentration and the absence of a person-level register made it impossible to do it efficiently within the current LFS framework. In fact, obtaining a CV of 25% for a three-month moving average unemployment rate of 5% for the Filipino, Arab, Latin American, and Southeast Asian groups would have required doubling the Ontario sample size.

Thus, the focus was on oversampling the largest Census Subdivisions (CSD), which are approximately equivalent to municipalities. Not only would this allow geographic disaggregation, but it would also indirectly oversample members of racialized groups because many of them are overrepresented in the largest CSDs.

## 3.2 Implementation

The LFS is carefully designed to rotate through its frame for 10 years, while also allowing other surveys to select non-overlapping samples. To fit within this design, the following two-target approach was used in calculating the additional sample size:
- Target CV of 16.5%, for a three-month moving average unemployment rate at a level of 5%, for the 20 largest CSDs.
- Target CV of 25%, for a three-month moving average unemployment rate at a level of 5%, for CSDs 21 to 23 and the largest two in each province (if not already part of the top 23).

Using data from August 2020 to January 2021 to carry out the simulations, it was determined that 13,300 dwellings needed to be added to the sample. They were all allocated to the provinces and gradually integrated to our existing sample, one rotation group at a time, from November 2021 to April 2022 by increasing the number of birth units (units in their first month of collection).

Because the sample design was not modified, the original LFS quality targets are automatically met.

## 3.3 Results

### 3.3.1 Impact on the quality of employment count

The sample increase improved the quality of estimates for the targeted CSDs. Figures 3-1 and 3-2 show the quality of estimates for two of the targeted CSDs as the additional sample was phased in over the November 2021 to April 2022 period. The employment count CV for Burnaby went from 14% to 8%, a decrease of 6 percentage points. In Toronto, where there was already a large sample (and a small CV) before the DDAP initiative, the CV of employment count decreased by one percentage point to 3%.
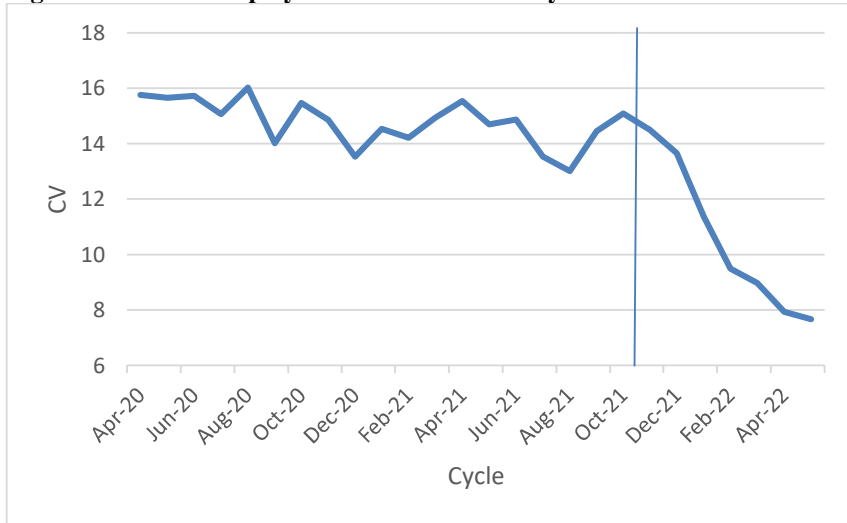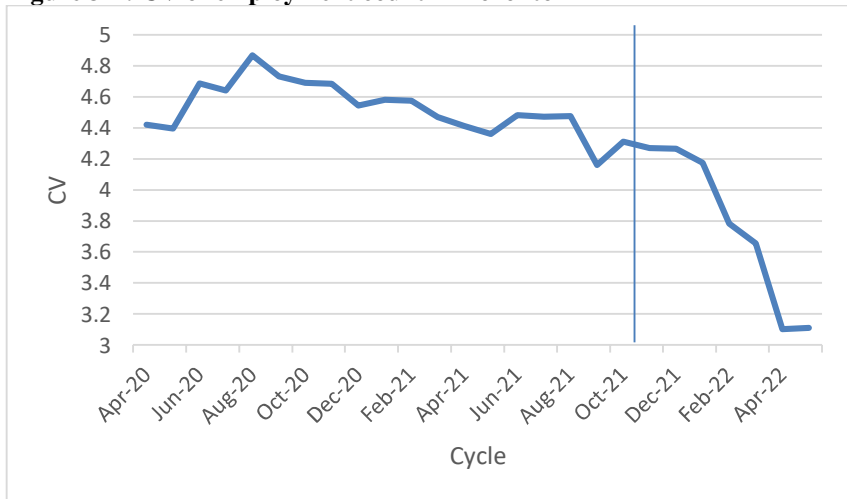
**Figure 3-1: CV of employment count in Burnaby**



**Figure 3-2: CV of employment count in Toronto**



Because many of the CSDs targeted had a high proportion of members of racialized groups, for example 51% in Toronto and 63% in Burnaby compared to 22% for Canada overall (estimates from 2016 Census), the sample increase contributed to the improvement of quality of the estimates for these groups.

### 3.3.2 Impact on number of immigrant respondents

With over 15,000 additional responding persons on a monthly basis, the sample increase also had an impact on the proportion of immigrant respondents in the collected sample (unweighted). Prior to the sample increase, around 18% of the collected sample consisted of people born outside of Canada. However, the same group accounts for approximately 37% of the additional 15,000 respondents. The weighting methodology takes into account the additional respondents, so the weighted estimates themselves are not affected by the sample increase. However, since there are more immigrants contributing to the estimates, there is an increase in the statistical quality of the estimates related to the immigrant population and of those related to racialized groups.

The CV of employment count for immigrants improved from 1.5% in October 2021, to 1.2% in May 2022 at the national level. In Markham (one of the targeted CSDs), the CV of employment count for immigrants improved from 24% in October 2021, to 11% in May 2022.

# 4. Small area estimation modelling program

The LFS is designed to produce high-quality estimates for key geographical aggregations. SAE uses auxiliary data and statistical models (Fay-Herriot area-level models) to produce reliable estimates for smaller geographical regions that may not have been targeted by the LFS design.
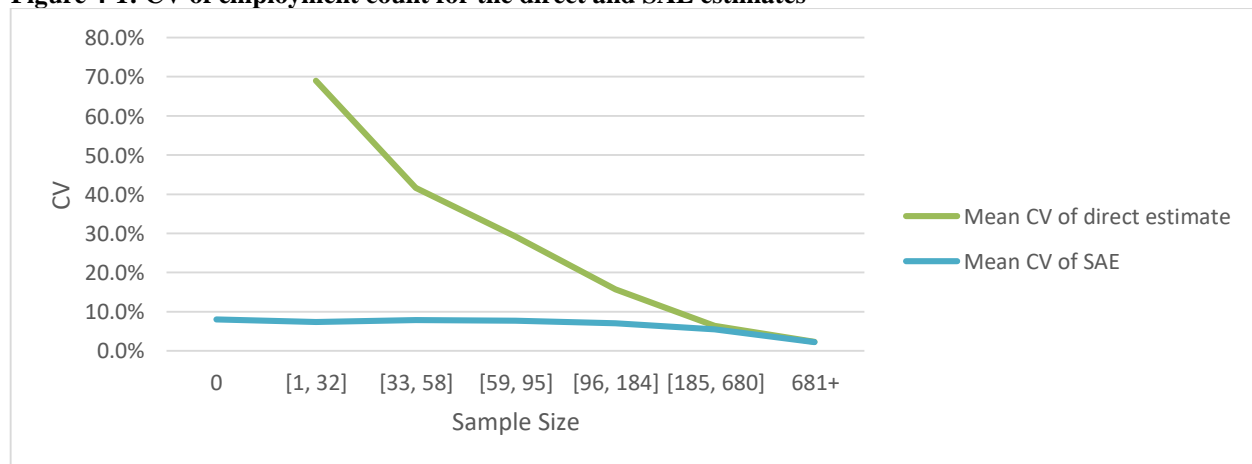
Until February 2020, these models had been used to produce:
- Estimates of unemployment rate and employment count, for each of the 149 Census Metropolitan Areas and Census Agglomerations
- Estimates of hourly wages for Economic Region / Occupation combinations.

The auxiliary variable used in the models was the number of employment insurance (EI) beneficiaries. During the pandemic, new benefits were introduced by the government and some people were uncertain about their own labour force status. This led to some concern whether the EI data represented unemployment in the same way as before the pandemic. The SAE program was suspended until reliable auxiliary data could once again be obtained. The recent availability of the EI data, along with the objectives set out by the DDAP initiative, have provided an opportunity to resume this modeling program.

The value added of SAE can be seen by examining the quality of the employment count estimates, as shown in Figure 4-1 below. We observe similar results for the unemployment rate and hourly wage estimates. When the collected sample size is small (1 to 32 individuals), the CV of the direct estimates $\left(\frac{\sqrt{direct\ variance\ estimate}}{direct\ estimate}\right)$ can be so large, close to 70%, that usage or dissemination of the estimates may not be advisable. With SAE modelling, it is possible to generate estimates of good quality for small collected samples, even in areas with a sample size of zero. As with any application of SAE using Fay-Herriot models, the gains diminish as the collected sample size increases. With a large sample, the SAE CV $\left(\frac{\sqrt{mean\ square\ error\ of\ SAE}}{direct\ estimate}\right)$ approaches the direct CV, and the SAE estimate approaches the direct estimate.

**Figure 4-1: CV of employment count for the direct and SAE estimates**



# 5. Future work

Future work for the supplements includes new topics including financial stress, training, digital platform work, gig (informal) work, support payments and childcare expenses. For the small area estimation modelling program, new geographies, sources of administrative data and labour market indicators will be explored.

# Acknowledgement