United Nations ECE/CES/2023/28

Economic and Social Council

Distr.: General
7 June 2023

English only

Economic Commission for Europe

Conference of European Statisticians
Seventy-first plenary session
Geneva, 22–23 June 2023
Item 9 of the provisional agenda
Timeliness, frequency and granularity of official statistics

The challenges in remaining a trusted and relevant data provider of timely and granular statistics

Prepared by Canada

Summary

This document provides a Canadian context of recent methodological and program advancements along with a forward-looking perspective of maintaining relevance in a post-pandemic world. Section II outlines the unique pandemic conditions that drove Canada's National Statistical Agency, Statistics Canada, to adapt and respond appropriately with timely data at a more granular level. Section III considers the unique challenges in continuing to remain relevant as a provider of official statistics through the transition from pre- to post-pandemic-data needs, in a context of continued expectations for data driven decision-making. Included is the argument to produce official statistics using a variety of innovative methods and apply a framework of monitoring and inspection that considers the multidimensionality of quality and the need for statistical rigor.

This document is presented to the Conference of European Statisticians session on "Timeliness, frequency, and granularity of official statistics" for discussion.

I. Introduction

1. Prior to the pandemic, the growing concerns of decreased survey data accuracy amidst increased nonresponse, combined with operational constraints due to prolonged survey collection periods drove Statistics Canada to explore new methods for producing more timely and granular data. The methods explored such as non-probability sampling, nowcasting, flash estimation and probability web panels offered the means for agile data responsiveness but often with reduced data accuracy. Research conducted by Statistics Canada focused on the ability to measure and communicate data quality of these methods but instances of their implementation into production were limited. However, in the context of the pandemic, there was an unprecedented need to generate near-real-time data to inform policy and program decisions, which favoured employing new techniques.

2. This document provides the Canadian context for the methods used to provide frequent, timely and granular data before, during and since the pandemic. An overview of the methods employed along with a review of the limitations in both the economic and social context will be provided. Finally, this paper will focus on the challenges expected in the upcoming years with a growing need for data despite a transition back to a more stable social and economic environment. These challenges include the increasingly unfavourable circumstances for traditional surveying that threaten the sustainability of survey-first approaches, an increased demand for more detailed data and the need to validate and communicate the quality of data under a variety of conditions.

II. Advancements in producing more timely and granular data

A. Timely data

3. The COVID-19 pandemic provided an opportunity for the immediate implementation of methods that Statistics Canada had been researching to meet needs for timelier data. Within weeks of the Canadian lockdown, Statistics Canada used a pilot probability web-panel to collect and release data on the social and financial impacts of the lockdown and pandemic on Canadians[1]. While these data relied on surveying approaches, due to the web-only design the response rates were significantly lower than most Statistics Canada surveys. At the same time, the panel was designed in such a way to facilitate more effective nonresponse adjustments and provided the nationally representative data needed for government and researchers to understand the impacts on Canadians within the first weeks of the unexpected lockdown. Additional data was collected and disseminated using crowdsourcing that relied on less rigorous non-probability sampling. This approach inherently brought an increased risk of bias, but it provided a large set of participants in desired subgroups such as healthcare workers, members of racialized groups and First Nation, Métis and Inuit peoples[2] as there was a very large general take-up with approximately 200,000 participants within weeks of it being launched. Additionally flash estimates of weekly family income[3] were produced to allow policy makers the ability to observe and respond to the impacts of pandemic relief benefits provided for middle-to-low-income families. And the monthly Labour Force Survey infrastructure was leveraged to inform Canadians on the pandemic effects through added questions that addressed new issues such as work from home and provided more disaggregation for racialized groups.

4. On the economic side, Statistics Canada released their first ever advance indicator estimates[4] for March 2020. Flash estimates were also produced for retail sales, wholesale

[1] The Daily — Canadian Perspectives Survey Series 1: COVID-19 and working from home, 2020 (statcan.gc.ca)
[2] The Daily — Impacts of COVID-19 on Canadians: First results from crowdsourcing (statcan.gc.ca) & Surveys and statistical programs - Impacts of COVID-19 on Health Care Workers: Infection Prevention and Control (ICHCWIPC) (statcan.gc.ca)
[3] Experimental estimates of family weekly income, January 2020 to March 2021 (statcan.gc.ca)
[4] The Daily — Gross domestic product by industry: Nowcast, March 2020 (statcan.gc.ca)

trade and manufacturing beginning with the April 2020 reference month. These advance indicators were calculated based on the incomplete set of survey and administrative data which previous Statistics Canada studies had shown performed better than nowcasting methods relying on aggregate modelling approach. These methods reduced the lag for an initial estimate from 50 days to 20 days following the reference month. In both the social and in the economic contexts, the methods used were not expected to produce estimates to the same degree of accuracy as typical Statistics Canada releases. For collected data, the economic flash estimates were sometimes based on response rates 30 percentage points lower than final response rates. However, the unique situation of the COVID-19 pandemic and the immediate demand for trusted information on the impact of the pandemic provided an unprecedented opportunity for near-real-time data signals. From the multi-dimensional perspective of quality[5] this trade-off of lower accuracy for increased timeliness was accepted by data users and Statistics Canada and enabled governments to react promptly through subsidies to support Canadians.

B. Granular data

5. While there were many adverse effects of the pandemic on Canadians and the Canadian economy these effects were far from uniform. The pandemic put a spotlight on the growing socioeconomic inequities. In response, Statistics Canada put into action the Disaggregated Data Action Plan to provide data that reflected the diversity of the Canadian population and economy for policy makers. Disaggregated data showed disparities for many specific population groups such as amongst racialized groups and in low-income neighbourhoods. But even amongst equity seeking groups, the impacts were uneven. For example, it took extremely granular data to further disaggregate racialized groups to observe the increased disadvantage in terms of COVID rates and COVID-related mortality[6] and increased economic hardship[7] that Black Canadians faced. Insights such as these were possible through data integration methods such as probabilistic data linkages of administrative data sources and other respondent collected data combined with modelling and imputation to provide data at the lowest levels.

6. Economically, different sectors of the markets were much harder hit than others[8] and ongoing detailed data was provided to aid governments in effectively target and support the economy and Canadians through a variety of business[9] and support programs[10].

III. The issues and challenges in remaining relevant post-pandemic

7. Some economic and social pressures may have lessened in recent months, but there remain many challenges requiring quality data to make timely data-driven decisions that reflect the nuances of varied disaggregated experiences. Data users have become accustomed to frequent, detailed, and timely social and economic insights and we need to leverage this opportunity moving forward. To remain relevant, there are several issues and challenges that need to be considered and addressed. These include (but are not limited to) communicating

[5] The six dimensions of quality are relevance, accuracy, timeliness, accessibility, interpretability and coherence (Statistics Canada's Quality Guidelines: Guidelines for ensuring data quality (statcan.gc.ca)).

[6] COVID-19 mortality among racialized populations in Canada and its association with income (statcan.gc.ca)

[7] The Daily — Study: A labour market snapshot of Black Canadians during the pandemic (statcan.gc.ca)

[8] Impacts of the COVID-19 pandemic on productivity growth in Canada (statcan.gc.ca)

[9] Impact of COVID-19 on businesses majority-owned by various sub-population groups and visible minorities, third quarter of 2021 (statcan.gc.ca)

[10] An analysis of Canadian business support programs in response to the global COVID-19 pandemic (statcan.gc.ca)

quality and fitness-for-use, the statistical and operational challenges of providing a plethora of data and the need for a sustainable statistical framework.

A. Communicating quality

8. An attribute of NSOs that distinguishes them from some other data providers is the focus on quality in the production of official statistics. Statistics Canada assesses quality based on six dimensions: relevance, accuracy, timeliness, accessibility, interpretability and coherence. Trade-offs have always existed between these dimensions, but the methods used to produce flash estimates and probability web-panels in the early months of the pandemic prioritized timeliness, which traditionally may have been viewed as secondary to data accuracy. From the multidimensional quality perspective, data, such as the flash economic estimates, were timelier, which increased their relevance, but the methods relied on lower response rates or partial data inputs, which affected their accuracy. The coherence of the data was also affected as revisions were released based on later and more complete data. This in turn may have also impacted their interpretability.

9. Early in the pandemic early economic signals with reduced accuracy were seen as an acceptable trade-off for the increased timeliness. However, as the economy progressively returns to stability, this trade-off should be reassessed using "fit-for-use" principles[11]. While flash estimate performance has been monitored since their introduction, Statistics Canada recently began a more thorough review of the flash estimates for Monthly Industry Surveys. The purpose is to assess the timeliness, accuracy, quality indicators for these estimates.

10. Using fit-for-use principles, a subsequent step will be to understand the intended use of the flash estimates by stakeholders and whether early indicators such as flash estimates can sufficiently meet their data quality needs. For example, in some cases, the monthly direction of the trend estimate could be sufficient for key stakeholders. Whereas in other cases, flash estimates that differ even slightly from the preliminary estimates (that rely on complete data) may not be useful. With a more stable economy and three years' experience of releasing flash estimates, the timing is right for taking stock of their use, their historical accuracy, potential advancements in the accuracy of the methods, and the provision of new and better-quality indicators to aid users to assess their own fit-for-use purposes.

B. Statistical and operational challenges to providing frequent, timely and granular data of high quality

11. There are many significant statistical and operational challenges to producing frequent, timely and granular data.

12. Traditional surveying approaches used to produce granular or disaggregated data involve more surveys, larger samples (to facilitate greater disaggregation) and longer questionnaires (to collect more detailed information about respondents). These approaches create operational challenges related to workforce capacity and costs, and they are counter-productive to delivering timely results. They also pose respondent relation challenges as increased surveying leads to additional response burden. The weight of this burden is especially heavy for minority populations that are frequently oversampled. Surveying has experienced notable downward trends in response rates that predate the pandemic. This decrease is likely irreversible and especially pronounced in social surveys and has only accelerated since the onset of the pandemic. The traditional approach to counteract nonresponse is to increase sample sizes. This does not address nonresponse bias and, as the surveying context moves further from the context of negligible nonresponse and response missing at random, other approaches should be considered to supplement or replace surveys.

13. Administrative data offer opportunities to address issues such as response burden while promoting a best practice of using the "ask once" principle[12]. Administrative data may

[11] 1.3 Data quality (statcan.gc.ca)

[12] A concept set by governments that aims to ensure that citizens, institutions, and companies need only provide information once to government agencies.

also provide larger "samples" for more disaggregation. However, these samples may be incomplete, non-random, and will often have an unknown participation mechanism. Furthermore, administrative data may not be timely if collected from a variety of jurisdictions or if requiring significant processing and clean-up procedures. In addition, access to administrative data sources may be limited and rely on collaborations with stakeholders and partners in both the government and private sector as well as the buy-in of Canadians.

C. Creating a sustainable statistical framework to meet data needs

14. To create a sustainable statistical framework that meets data needs in the coming years there are many significant challenges; three are presented below. The first is the need for increased access to, and use of, admin data. The second is the advancement of methods to integrate and adjust the administrative or other data to produce timely, granular, and frequent estimates. The third is to create an advanced framework that regularly validates and assesses data quality of these new sources and techniques.

1. Increased access and use of administrative data

15. In Canada, the need to build a national statistical ecosystem that provides high-quality data and information to meet the needs of Canadians through increased data stewardship and greater data flow and sharing between jurisdictions[13] has been identified. This will require stronger cooperation between the federal, provincial, and territorial governments for data sharing and to use standard concepts. If implemented effectively, combining data sources from various jurisdictions will greatly reduce data gaps and serve Canadians.

2. Advancement of methods

16. Additional data sources will also require an increased use of data integration methods to combine these data sources, such as was performed in the COVID mortality among racialized populations in Canada and its association with income example[6]. Furthermore, there is a need to continue research and advancement into modelling to address data gaps and data incoherences. Statistics Canada has successfully employed examples of these techniques to produce more disaggregated data addressing issues such as intersectionality or the social connection to business data[14] but these methods will need to be expanded across programs to address data needs. Statistics Canada is currently conducting Blank Page Reviews for certain programs to assess how data needs can be addressed through a rethink of available data sources and integration to meet these needs. More research into model-based estimation as an alternative to produce timely data also needs to be pursued. And along with the advancement of methods a strong stakeholder engagement to gain their acceptance and support is required.

3. Validating and assessing quality

17. To effectively communicate quality and allow fit-for-use assessments in future years there will need to be an ecosystem to assess and validate the quality of new methods and outputs along with easily digestible quality indicators. Each source (survey, administrative or model-based) has inherent limitations related to the certain dimensions of quality such as timeliness and accuracy. Furthermore, new sources and an advancement in methods could create a fragmented data system that impacts the coherence and interpretability of the data. Therefore, to remain relevant and provide high-quality statistical information that matters to Canadians, Statistics Canada considering a framework of monitoring and inspection to support data quality assessments with these innovations.

18. Each data source and each statistical method comes with assumptions. For surveys, an assumption is that nonresponse is missing at random. For administrative data an

[13] Recommendations from the Canadian Statistics Advisory Council Canadian Statistics Advisory Council 2022 Annual Report - Trust, Governance and Data Flows in the National Statistical System (statcan.gc.ca)

[14] Identifying Indigenous Business Owners and Indigenous-Owned Businesses (statcan.gc.ca)

assumption is that the data does not have under or over coverage. For model-based estimates, an assumption is that the model is up-to-date or has no model-drift. To assess and improve quality, in each scenario, these assumptions must be validated.

19. For surveys this may involve techniques such as small-scale non-respondent surveys. Examples currently in practice at Statistics Canada are the Dwelling Classification Survey[15] and the Census Undercoverage Survey[16] which select sample to validate and adjust estimates to dwelling and population estimates. A pilot nonresponse survey will be conducted to better understand nonrespondents and learn how to prevent or at least correct for nonresponse bias. Data linkages to adjust estimates or assess quality are other options. An example is the Census Overcoverage Study[17] which uses deterministic and probabilistic record linkages to identify census overcoverage counts, used for determining the Census Net Undercoverage rate used in the Demographic Estimates Program.

20. For administrative data, a framework that supports monitoring and inspection could include surveys to (i) complement missing data, (ii) increase data outputs or (iii) strengthen model training and validation. An example currently employed by Statistics Canada are the pseudo-census approaches employed in various economic surveys such as the Annual Survey of Service Industries: Advertising and Related Services[18]. All units in the survey population are selected but only a subset receive the questionnaire, survey responses of respondents are used to impute responses for units not selected for the survey. Another example is the Survey of Employment, Payroll and Hours[19] that uses administrative data for two key variables and a relatively small sample to compute model parameters.

IV. Conclusions

21. NSOs have never been more relevant than in the recent years of the pandemic. In Canada, significant gains were made to produce more timely data through methods such as flash estimates and a probabilistic web panel or crowdsourcing. Data integration combined with methods such as imputation provided the granular data required to identify the unique experiences and varied hardships faced by Canadians and industries such as increased COVID mortality and labour market impacts across specific population groups. To maintain this relevance, the momentum must continue to meet the ever-growing needs for timely, granular, and frequent data. Accomplishing this will require increased access to administrative data sources and significant work across governments and stakeholders at all levels to build collaborations and increase data flow. Statistical innovation is required where traditional approaches may rely on outdated assumptions to stay relevant in this digital age where data information is readily available. Finally, in parallel a modified/updated/evolved system to monitor and inspect new data sources and new statistical methods, along with the traditional methods and sources, is a necessary complement to validate and correct data quality, within the fit-for-purpose model.

[15] Surveys and statistical programs - Census of Population - Dwelling Classification Survey (DCS) (statcan.gc.ca)
[16] Surveys and statistical programs - Census Undercoverage Survey (CUS) (statcan.gc.ca)
[17] Coverage Technical Report, Census of Population, 2016 (statcan.gc.ca)
[18] Surveys and statistical programs - Annual Survey of Service Industries: Advertising and Related Services (ADVT) (statcan.gc.ca)
[19] Surveys and statistical programs - Survey of Employment, Payrolls and Hours (SEPH) (statcan.gc.ca)