

Recommendations for the publication of official statistics from household surveys in the context of the coronavirus disease (COVID-19) pandemic

April, 2020

Summary

In the context of the global COVID-19 pandemic, ECLAC is preparing a series of short publications with relevant policy recommendations. A number of recommendations are made in this note on gathering labour market information through household surveys, paying particular attention to the sample designs required for this purpose.

A. Introduction

Since the health emergency arising from the global COVID-19 pandemic, which has led most of the countries of the region to declare states of emergency, national statistical offices (NSOs) have had to interrupt abruptly the collection of primary information as part of many of their statistical operations, including household surveys. This note describes possible scenarios for the collection and analysis of household survey data and provides a set of recommendations regarding the capture, publication and dissemination of official statistics from those surveys, in the context of the international situation caused by the COVID-19 pandemic.

In those countries that have declared states of emergency, in order to stop the spread of the disease and protect those people most vulnerable to it, governments have taken difficult decisions that have affected society and the normal functioning of government institutions, educational entities, industry, transportation and commercial establishments, among others. In some cases, the general public has begun to collaborate with voluntary quarantines; but in light of the rapid spread of the disease, governments have had to impose curfews, movement restrictions and compulsory quarantines, among other health measures, and remote learning and working methodologies have been implemented or developed.

In view of this situation, the Statistics Division of ECLAC sent out a short questionnaire to the NSOs and central banks of Latin America and the Caribbean, to learn about the effects that the health emergency was having on the functioning of statistical institutions. Responses from 20 countries indicated that, as at 20 March, they had all declared some form of health emergency as a result of COVID-19. The statistical operations most affected by the situation were, surveys, followed by administrative records and censuses. With regard to household surveys, most respondents said that information collection would be postponed to later in the year in most cases. The main information collection difficulties are movement restrictions and the closure of establishments.

Summary

- A. Introduction
- B. Recommendations for March 2020
- C. Recommendations for the following months
- D. Conclusions

A series of videoconferences have been held since 24 March to assess the impacts of the COVID-19 pandemic on statistical operations.¹ These videoconferences have been convened by ECLAC, in conjunction with the International Labour Organization (ILO) and the National Institute of Statistics (INE) of Chile, in its capacity as coordinator of the virtual talks of the Knowledge Transmission Network (KTN) of the Statistical Conference of the Americas, chaired by the National Administrative Department of Statistics (DANE) of Colombia. Three videoconferences were held between 24 and 26 March: two with the countries of Latin America, on household surveys and prices, respectively, and one with the countries of the Caribbean. Another videoconference, focusing on employment surveys, was held on 1 April.

Movement restrictions have led to the suspension of face-to-face data collection for household surveys until further notice. The NSOs are carrying out surveys over the telephone or the Internet. This change in the collection methodology is necessary if employment and poverty figures are to continue to be produced. These indicators are particularly important in the context of the pandemic, given the profound impact movement restrictions and quarantines have on people's employment.

B. Recommendations for March 2020

To analyse the effect of movement restrictions on household surveys that are conducted on an ongoing basis, the month of March 2020 must be separated from the subsequent months. Data were able to be collected normally in most countries until around the second week of March, when the normal survey activities were interrupted. However, for the subsequent months it will not be possible to collect data in the usual manner, until this unprecedented situation comes to an end.

Depending on the partial coverage of the March surveys, NSOs face three possible scenarios:

- Poor coverage levels, in view of which it would be recommended not to use the information collected for the production of any official statistics.
- Average coverage levels, in view of which it would be recommended to modify the survey's expansion factors to limit the bias levels and to publish national-level official statistics only.
- Acceptable coverage levels, in view of which it would be recommended to continue with the normal publication process.

Taking into account the cross-sectional or rotating sample design of household surveys in the region, and the fact that these designs have been adopted in order to ensure that they are representative at the subnational level (urban and rural, by geographical area, major cities, among other factors), the sample size of a survey in a specific month under normal conditions is larger than the sample size needed to be nationally representative. NSOs must therefore calculate the minimum sample sizes.

For example, the sample for a survey designed to be representative of both urban and rural areas for March comprises 8,000 households across the country. The survey design allows the sample size needed to produce nationally representative results only to be calculated, which in this case could be 1,800 households.

Under the first scenario (poor coverage levels), data collection coverage levels for the first two weeks of March were so low —fewer than 1,800 households— that the available information cannot be used to produce any statistics, not even at the national level, and much less the regional, level. In this context, it is recommended that no information should be published or disseminated and that the situation should be explained to the public. Obviously, the historical series created from these surveys will have a point of discontinuity corresponding to the third month of the year.

In the second scenario (average coverage levels), the information collected —more than 1,800 households in the national sample— would allow some information to be published, subject to certain restrictions and the adjustments described below.

¹ See [online] <https://rtc-cea.cepal.org/en/conectados-rtc/planes-de-contigencia-adoptados-por-las-oficinas-nacionales-de-estadistica-ante-la>.

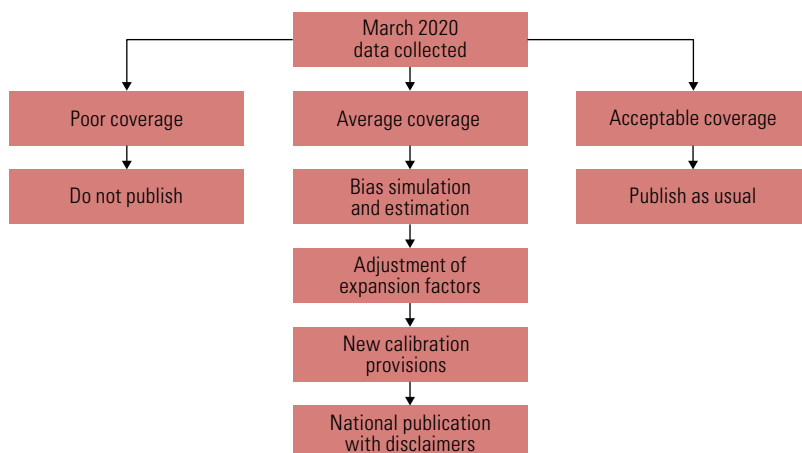
First of all, imputation methods should not be used to complete the missing information (which may be as high as 50% of the original sample). Even if sufficient information is available for imputations, they would assume that the phenomenon being studied behaved in the same way throughout the month of March, and would not reflect its true behaviour in the context of the COVID-19 pandemic. Meanwhile, generating valid predictive models to complete all the information as initially planned would be a major undertaking in terms of the measurement of both the imputation error and the sampling error associated with the selection of the sampling units. Furthermore, using the imputation model and adjusting predictive models to provide official statistics for a country would be a departure from the expanded sample paradigm and the principle of representativeness which is fundamental for NSOs.

In the second scenario, it is recommended to use the data collected as the main input to obtain national estimates, following the same logic used by NSOs when publishing official statistics. To that end, the expansion factors may be substantially modified to correct the bias generated by the lack of responses as a result of the states of emergency in the countries.

The simulation to estimate bias should try to mimic the final structure of the primary sampling units (PSUs) used in March 2020 (in terms of both geography and socioeconomic status, in cases where the PSU framework is stratified) and seek to replicate the sample used in previous months (for example, in March 2019). This will produce a trimmed database that can be used to obtain an estimate for March 2019 as a counterfactual to March 2020. The usual estimation procedures must be carried out using this trimmed database to obtain counterfactual estimates for March 2019. The difference between the published estimates and the counterfactual estimates will give an idea of the magnitude of bias in each of the constructs of interest in the household surveys. In the light of these biases, the expansion factors will have to be adjusted accordingly for March 2020 to correct this difference. This procedure could be iterative and could require new provisions in the calibration of the expansion factors, by adjusting the constraints to national totals.

It is recommended that only national level information should be published and disseminated. It should not be disaggregated by the sampling design categories. This is because the phenomenon associated with non-response cannot be classified as random; therefore, the existing bias will grow as the estimates are disaggregated. It is recommended that NSOs communicate in a timely manner that they will not publish disaggregated statistics. They should also explain that the national figures do not correspond to a regular survey and may have lost comparability with previous records in the statistical series. Lastly, it is recommended to specify the exact period in March during which data were collected, which in most cases will be the first two weeks of the month, as well as national coverage rates, by subnational domains and by sex. Diagram 1 clarifies the suggested recommendations for the three scenarios described in this section.

Diagram 1
Recommendations for three information gathering scenarios for March 2020



Source: Prepared by the authors.

C. Recommendations for the following months

As long as movement restrictions remain in force, it will be almost impossible to carry out face-to-face data collection for household surveys. Modifying how these surveys are carried out, through telephone or online interviews, poses several challenges.

One of them is the inability to access the households in the sample. Household surveys in most countries are based on repeated visits to rotating samples. For each period (month or quarter), a number of households that have been visited on previous occasions are replaced by new ones.

Given that the telephone numbers of the households in the survey sample are collected on the first visit, these data would not be available for the households to be included in the April sample. In other words, there would be no way to contact the probabilistically selected households. Decisions must therefore be made about the sample of households that will be used to carry out the surveys while continuing to limit field operations.

A second challenge is the length and content of the questionnaire. If the interview is conducted over the telephone, considerably less time will be available to complete it than in a face-to-face visit. The length of the questionnaire must therefore be reduced so that it focuses on the subset of most relevant questions in this context. New questions must also be included to determine whether people who are not working are still employed by their employer and to classify their employment status properly.²

Such modifications create a scenario in which national authorities must accept the loss of comparability in the series of employment and income indicators. This loss of comparability can be partially remedied later, by carrying out surveys to collect the information retrospectively. However, this does not mean that the collection of information on employment and other social statistics should be suspended. Even with the aforementioned constraints, these figures are crucial to inform public policy and decision-making in a highly uncertain context.

When gathering information in these unprecedented times, top priority should be given to publishing official statistics based on the probabilistic selection of a sample of households or persons. Given the impossibility of contacting the new households planned for the April sample, it is recommended that NSOs adopt a temporary change in the design of the household surveys and choose to follow up on a sample panel in the following months. Panel rotation and the application of the usual sample designs should be frozen until normalcy is restored and face-to-face surveys can be carried out.

This panel should be composed of those who were surveyed in the months immediately preceding March 2020; for example, January and February 2020, and the last months of 2019. Obviously, March 2020 would be ruled out as it does not correspond to a usual or complete survey.

If the NSO is not able to carry out mass telephone surveys covering the entire sample for the chosen month (for example, January 2020), a national subsample could be designed, using two-stage sampling. Any modification to this panel sample follow-up is considered valid if the telephone sample corresponds to a probability sample, for which the selection probability is known in advance.

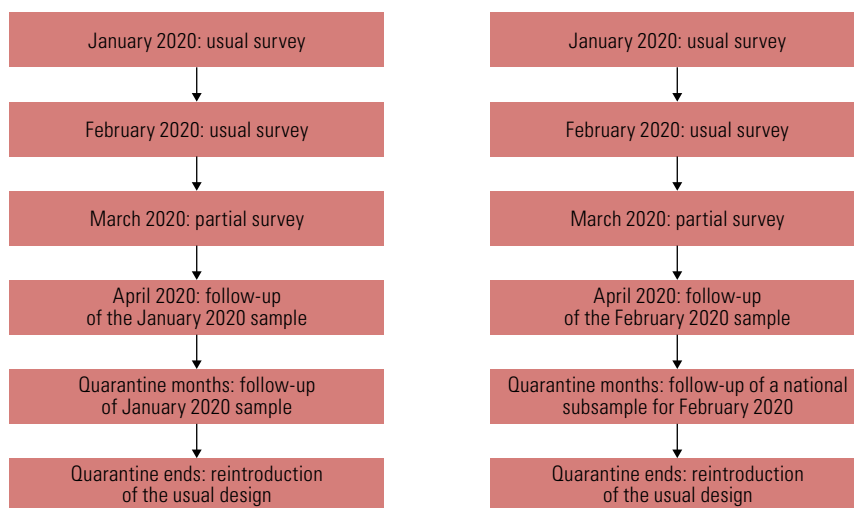
Following up on a panel is a simple and effective way to have a probabilistic sample to calculate the necessary statistics. This strategy is free from selection biases, like any representative sample, but will allow the effect of non-sampling bias resulting from the change in collection modality to be calculated. As a result, an unbiased estimate of gross flows can be made (monthly monitoring of the employment situation of household members), and the effect of the pandemic on the statistics derived from the survey can be assessed.

Although a telephone survey is exposed to selection bias, since not all the households contacted in January have a telephone or will answer the call from the NSO, this can be corrected using statistical models, calibration techniques and adjusted expansion factors.

² See the general description of the impact of the COVID-19 pandemic on the collection of labour market statistics in International Labour Organization (ILO), "COVID-19 impact on the collection of labour market statistics" [online] <https://ilostat.ilo.org/topics/covid-19/covid-19-impact-on-labour-market-statistics/>.

Diagram 2 presents two examples of how a country can follow up on a panel from previous surveys. In the scenario on the left, a country is assumed to have the capacity as of April 2020 to conduct a telephone (or online) survey of the same households that were selected in January 2020. The second scenario illustrates a case in which the NSO follows up a subsample of households (or people) based on the February 2020 probabilistic sample. In both cases, the panel is subject to follow up only for the months that the movement restrictions are in place.

Diagram 2
Scenarios for follow up of panel samples based on previous surveys



Source: Prepared by the authors.

At the end of this confinement period, the usual sampling design can be reintroduced. This reintroduction can be gradual if the country wants to analyse in depth the biases generated by the change in collection modality. For example, a percentage of the information may be collected in person and another percentage collected over the telephone (or online). This information will allow the levels of comparability of the time series to be defined and will determine how these data can be combined with the medium-term series.

D. Conclusions

This note proposes a strategy to address the impact of movement restrictions on the execution of household surveys that are being conducted on an ongoing basis. A distinction is made between the survey of March 2020, for which part of the information was collected in the usual manner, and the surveys of the subsequent months, which require the adoption of a different collection modality.

For the month of March, it is proposed to use the collected part of the sample to publish national-level statistics only, without disaggregations. For the month of April and for as long as this unprecedented situation lasts, it is recommended that NSOs change the collection modality, stop using the usual sampling designs, follow up a sample panel while quarantines (or states of emergency) are in force and reintroduce the original design in the months following the return to normalcy.

This strategy allows unbiased statistics on the labour market to be generated, based on the probabilistic selection of a sample of households or individuals. These statistics provide crucial information for public policy and decision-making on the labour market.





This document is part of a series of reports prepared by the Economic Commission for Latin America and the Caribbean (ECLAC) on the evolution and effects of the COVID-19 pandemic in Latin America and the Caribbean. It was prepared by the Statistics Division, directed by Rolando Ocampo, under the general coordination of Alicia Bárcena, Executive Secretary of ECLAC.

Copyright © United Nations, 2020



Economic Commission for Latin America and the Caribbean (ECLAC)
Comisión Económica para América Latina y el Caribe (CEPAL)
www.eclac.org