## Surveys Under Lockdown; a pandemic lesson

The Covid-19 crisis is affecting major data collection activities of national statistical systems around the world. This paper provides an overview of maximum impact of the pandemic crisis on different data collection operations and possible responses by the national statistical offices. It also proposes a mixed-mode panel survey design to enable continuing data collection under lockdown and after pandemic situation with minimum impact on quality of survey results.

## I. A blessing in disguise

An invisible enemy has forced us to "change" in many ways, some of which we have always wished for but never dared to (or needed to). Covid-19 is not only claiming lives, but also changing them by exposing the world's resilience and adaptability. Since the early days of 2020, our lives and work have been transformed in different ways over the course of few short weeks. Sports fans are turning to video games and concerts, museums and zoos are increasingly live streamed. The most important meetings of the world are being convened via multiple IT applications with zero cost and carbon emissions; trainings are being conducted using e-learning and remote teaching; schools have gone online; quarantined lovers are distance-dating; robots are delivering Big Macs and drive-thru virus testing has become commonplace. The most liberal citizens are demanding martial law enforcement by armies who have put their guns down to build hospitals, and the average person reads about hygiene, listens to scientific news, knows about exponential function and bell-shape distribution more than ever.

How is this lockdown situation going to change the production of official statistics? Every one of us must have experienced changes in at least one of the above manners. But how about one of the main activities in our line of work: data collection? For years, the statistical community has discussed new modalities for data collection to replace or complement direct, face-to-face measurements and to improve the efficiency and timeliness of statistical processes and products. Nevertheless, the practice of innovative data collection methods for major, nationwide operations are sparse, with face-to-face interviews remaining the primary modality in most statistical systems for censuses and large household surveys. We now have an unprecedented opportunity to think differently and make decisions that can transform our statistical operations for the future.
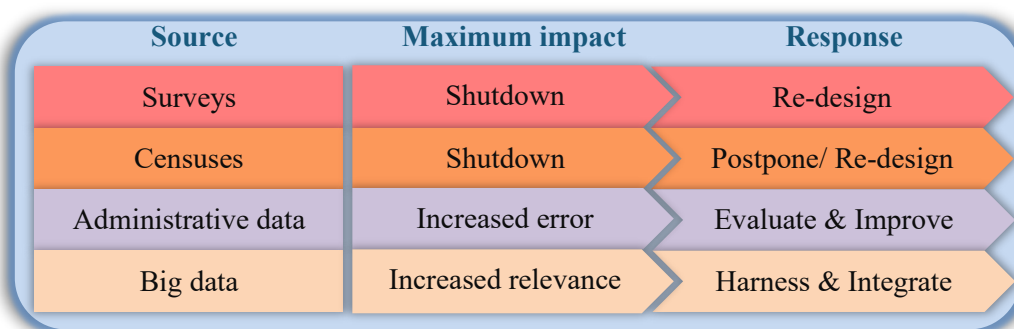
This paper does not aim to provide a comprehensive account of all survey design aspects that can be applied under normal circumstances. For instance, under normal conditions, I would not recommend changing data collection mode in the middle of any survey operation or adjusting estimates for non-sampling errors after survey/census completed. But crisis situation requires approaches that are flexible, but at the same time have minimum impact on quality. This paper aims to propose ideas for introducing rapid change in survey design due to crisis and taking necessary actions to mitigate possible quality impacts. The goal is to adopt an approach which may be inferior to the original survey but fills data gap with reasonable accuracy.

## II.    Official statistics under lockdown

The lockdown situation impacts data collection procedures differently and each requires a different response. Diagram 1 divides major data collection operations into four categories by source of data and describes the maximum impact that a lockdown situation may have and the potential response by statistical agencies:

Diagram 1- Pandemic maximum impact on statistical data collection

| Source | Maximum impact | Response |
|---|---|---|
| Surveys | Shutdown | Re-design |
| Censuses | Shutdown | Postpone/ Re-design |
| Administrative data | Increased error | Evaluate & Improve |
| Big data | Increased relevance | Harness & Integrate |

- Surveys: When sample surveys are conducted through face-to-face interviews, the maximum impact of lockdown could potentially be a complete shutdown of field operations. For some ad-hoc surveys, statistical agencies may still afford to postpone surveys to a post-pandemic time, without any significant disruption to statistical time series. Nonetheless, for regular (annual and more frequent) surveys that measure short-term socio-economic variables (such as labour force, income and expenditure, and price surveys) that are considered vital for decision making, the postponement of survey operations may not be an option. In addition to their importance to decision making in ordinary situations, three incentives make running of such surveys under lockdown even more essential:

  o *Agility in policy response:* rapid policy response without timely statistics during a pandemic situation is akin to solving a Rubik's cube blindfolded.

  o *Evidence for post-pandemic policies:* government and businesses need to analyze lockdown impact on population and economy in order to better design long-term resilient policies.

  o *Continuity of time series:* missing data for the pandemic period (or any important shock) will distort time series analysis in the future.

Given the critical role of regular surveys that are normally conducted through face-to-face data collection, they need to be immediately re-designed during lockdown and continued with the application of new or combined modes of data collection. Broadly speaking, two modes of data collection are commonly applied for sample surveys: standard interviews (mobile, telephone, and computer assisted interviewing (face-to-face, telephone or mobile)), and self-administered questionnaires (web-based or postal). Conducting regular surveys also provides an opportunity to collect new data on the needs and behavior of people and businesses during lockdown. This will address the data demand to inform rapid policy response, as well as future analysis of the situation. Section III aims to illustrate re-design of a hypothetical monthly survey under a lockdown situation.

- Censuses: With the exception of few rare cases, censuses are the costliest statistical operation that any national statistical office (NSO) operates. The cost of fighting Covid-19 together with economic recession will most likely divert resources away from the 2020 round of population and housing censuses around the world. When conducted through interviews (of any form), census preparations take much longer than any sample survey. Therefore, it is not practical, neither

efficient to redesign censuses during a pandemic period. Unless mode of data collection is not face-to-face, this leaves NSOs with no choice but to postpone a 2020 round of population and housing census or any other census scheduled for this year.

- <u>Administrative data</u>: The impact of a pandemic on administrative data is completely different from censuses and surveys. The production of administrative data and register statistics may continue (with limited functionalities) during lockdown. However, there are many ways in which quality of such data/statistics can be undermined. Under-reporting and misclassification in registering events and activities are more likely to occur when both government services and mobility of people are restricted. In particular, when registrations are not online, health systems are overburdened, administrative staff are mobilized for emergency responses, and there is lack of incentive and/or skill for accurate recording of events (such as cause of death in the case of vital statistics), the

chance of error is likely to increase. Data custodians and NSOs (as coordinators of statistical systems) need to tighten quality checks, regularly assess, and transparently communicate quality of admin data and registers with users.

- <u>Big data</u>: Finally, there is one area that Covid-19 can provide a valuable opportunity for data collection with minimal to no risk. As the majority of our lives move to a virtual world, big data provides the most comprehensive coverage of human activities in entire history. From education, to work, shopping, health and sentiments; the internet has never collected this volume of data with such a large coverage of the global population. Mobile and location data that are being used nowadays to keep individuals away from high-risk areas, can be utilized in the future to reach the most vulnerable groups. This is an unprecedented time for statistical systems to harness big data for the production of "real time" official statistics.

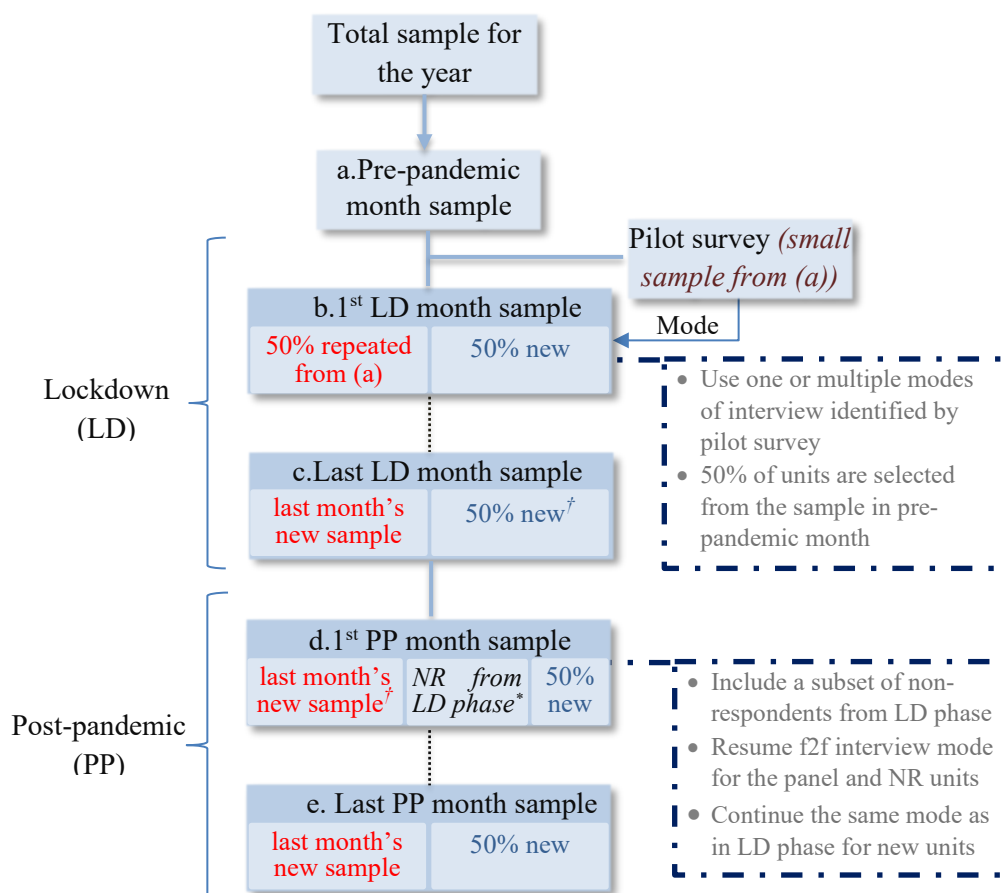## III.    Mixed-mode panel survey under lockdown

This section aims to provide ideas for re-designing a hypothetical monthly household survey under a lockdown situation. The assumption is made here that a household survey is conducted monthly by an NSO through face-to-face interviews and the results are combined every three months to produce quarterly publications. The hypothetical country is under lockdown for an unknown period and face-to-face interviews can therefore not be conducted. The ideal response by NSO is to quickly re-design the survey, aiming to:

a.    collect the most important survey information during lockdown and continue original design after lockdown without interrupting the quarterly publications;

b.    minimize effect of non-sampling errors introduced to the results due to the use of multiple modes of data collection;

c.    collect data on people's needs and behavior (in addition to survey data) necessary for rapid response to the pandemic situation.

Diagram 2 illustrates this new design in two phases: during lockdown and after lockdown. In order to satisfy the need for assessing and reporting on non-sampling errors, as well as tracing the temporal changes (at least during lockdown), the new design is a panel survey in which every month's sample contains a subset of sampling units carried over from the previous month.

Diagram 2- Mixed-mode panel survey under lockdown; schematic design

Total sample for the year

a. Pre-pandemic month sample

Pilot survey *(small sample from (a))*

Mode

**b. 1st LD month sample**

| 50% repeated from (a) | 50% new |

Lockdown (LD)

- Use one or multiple modes of interview identified by pilot survey
- 50% of units are selected from the sample in pre-pandemic month

**c. Last LD month sample**

| last month's new sample | 50% new† |

**d. 1st PP month sample**

| last month's new sample† | NR from LD phase* | 50% new |

Post-pandemic (PP)

- Include a subset of non-respondents from LD phase
- Resume f2f interview mode for the panel and NR units
- Continue the same mode as in LD phase for new units

**e. Last PP month sample**

| last month's new sample | 50% new |

†these two subsamples contain the same units and interviewed using two different modes to contribute to measuring mode effects. It is important that 50% repeated sample in box d is interviewed face-to-face to allow observing mode effect.

*the purpose of this subsample is to estimate non-response effect by comparing results of their interview with the results of face-to-face interview of repeated subsample in the same month.

## Pilot survey

Assuming that the survey was running in a face-to-face interview mode prior to pandemic crisis, the re-design begins with a pilot survey consisting of a small sample from sampling units who already responded to a survey questionnaire in the previous month (pre-pandemic). This allows us to assess different modes of interviewing by comparing answers to pilot survey and results of the face-to-face interview in previous month with the same units. The pilot should test response rate, accuracy of information, respondent burden, different incentive patterns, as well as type of questionnaire (single questionnaire, modular questionnaire- with and without optional modules, skipping questions, etc.).

After the pilot, we should be able to identify one or a combination of modes to replace face-to-face

interviews during lockdown. In practice, we may end up providing more than one option to the respondent. For instance, they may have to choose between a self-administered web-based questionnaire and a telephone or mobile interview. The options depend heavily on

- o *type of survey questions:* not every mode is suitable for all types of questions. For instance, telephone may be applicable to price survey, but not suitable for expenditure surveys. One may consider mobile assisted web-based questionnaire or other internet-based survey for expenditure surveys.

- o *frame information:* mode of survey depends on the information available in sampling frame. For instance, postal questionnaire requires address, whereas telephone survey needs telephone number.

- o _public access to technology:_ this is especially important when NSOs wish to consider internet or mobile surveys. When internet penetration rate is very low, it will significantly bias the results. More importantly, many disadvantaged population groups are likely to be missed from the sampling frame when such modes are applied.

The survey will run monthly for the rest of the year in two phases: lockdown phase and post-pandemic phase. Data collection mode (whatever is chosen) will remain same for entire period of each phase but changes from lockdown to post-pandemic phase. Sample size will remain approximately same for each month and consists of two sub-samples, one from previous month (repeated sample) and one newly selected from sampling frame (new sample). We assume that overlap is 50%. In other words, half of the sample in each month comes from previous month (new sample in the previous month). The survey design (e g. stratification, clustering, etc.) will remain the same as the original design. In cases where representativeness of epicenter areas is important, this may need a tweak in the design by applying purposive (non-probability) sampling methods that give a definite chance of selection to epicenter areas.

## Alternative sources of data

The lockdown situation provides an opportunity to re-think the standard survey approach and modernize statistical data collection operations to incorporate alternative data sources; administrative records or big data sources (remote sensing, mobile phone records, credit card and scanner data, etc.). In addition to loosing sample due to inability to conduct face-to-face interviews, responses from other survey modes such as online and phone may be reduced under lockdown conditions as well as due to sensitive nature of questions. The New Zealand's 2018 Census of Population and Housing provides a good example of combining survey (in this case Census) data with

administrative data[1]. Therefore, it is strongly recommended to investigate possibility of using alternative sources of data and combine them with data collection modes other than face-to-face to be used under lockdown phase.

## Overlaps

Three types of overlaps of samples are proposed in this design:

- month to month: Overlapping samples serve two purposes:

  1) enable estimating temporal change. It is essential specifically during crisis. If you are redesigning a survey with rotation or longitudinal design, this overlap is not necessary and original panels suffice the temporal change estimation.

  2) enable measuring mode effect. This can be achieved by comparing two sub-samples of each month's sample at post-pandemic phase (the 50% sample repeated from last month interviewed face-to-face and 50% new sample interviewed using the same mode(s) used at the lockdown phase), plus comparing one sub-sample from the last lockdown month and the first post-pandemic month (shown by † on diagram 2). This will increase number of observations for mode effect. Diagram 3 illustrates a case in which we assume lockdown situation ends at the month of August and survey has another four months to continue (given the overlap pattern, five more sub-samples left to be surveyed during September to December). We aim to survey each sub-sample using both face-to-face and the lockdown mode in two subsequent months. This gives us eight possible observations for mode effects: four observations on mode effects on the same sub-sample and four observations on mode effect at each month.

---

[1]     https://www.stats.govt.nz/privacy-impact-assessments/creating-the-2018-census-dataset-by-combining-administrative-data-and-census-forms-data-our-privacy-considerations.

- <u>non-response</u>: since we are not using face-to-face interview in lockdown phase, it is likely to encounter a significant non-response rate. This overlap is introduced to help measuring non-

response effect. A sample of non-response samples from lockdown phase is added to sample in the first month of post-pandemic (shown by * in diagram 2).

Diagram 3- subsample rotation for mode effect during post-pandemic phase
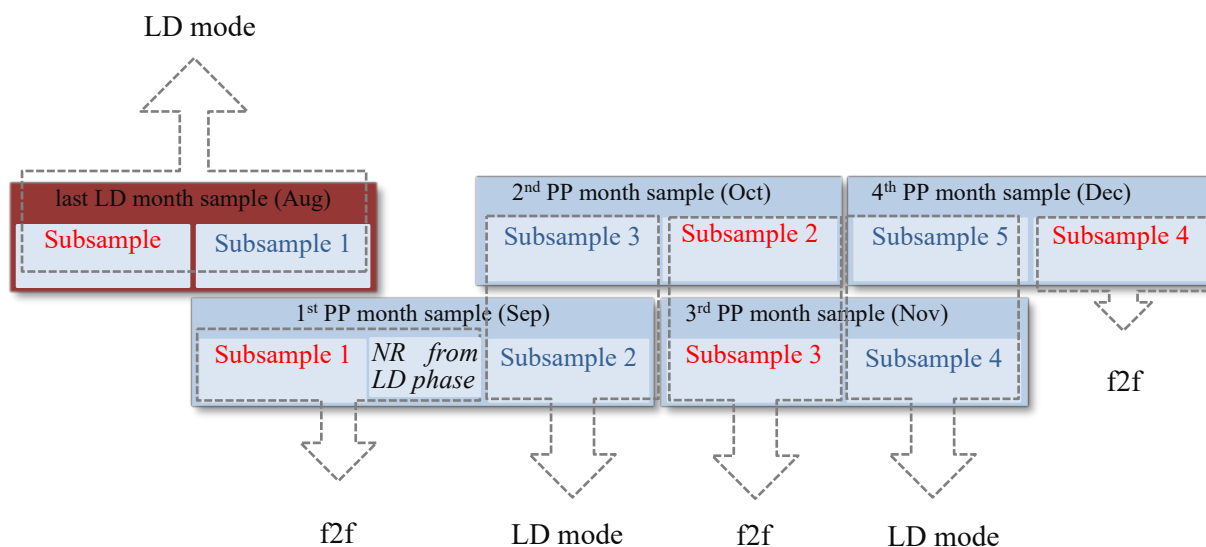


**Important note on adjustments:**

1.      Non-response adjustment is normally done on survey weights when non-response effect is at random (same effect on all subpopulations). This overlap will help us to assess whether the effect is completely random and we can safely adjust sampling weights for non-response rate. When effect is not random (it is bias for certain population groups), it should be reported to users.

2.      In measuring mode effect, we need to remember during lockdown situation survey results drastically change and therefore not every change is due to mode effect. Under ideal conditions, to estimate mode effect with confidence, we need parallel run of a treatment sample with 50% of the control sample for over a year. Clearly, under lockdown situation such experiment is not possible. The proposed design attempts to maximize observations for measuring the mode effect. If the post-pandemic duration is not long enough to provide us with sufficient observations, it is strictly recommended not to use the mode effect measures for adjusting survey estimates and only report mode effect assessment results together with non-response effects to the users.

## Lockdown phase

- Same mode of data collection (except face-to-face) should be applied for the entirety of this period. NSOs may decide to use mixed modes in this phase (including use of alternative sources) as appropriate.

- Questionnaire may be redesigned for this period to suit the mode of data collection, increase timeliness, reduce cost and accommodate for specific questions that are necessary for pandemic rapid response. The latter may be done in consultation with respective departments of the government such as the ministry of health, education, labour and social welfare.

- The repeated sample (50% sample overlap) mainly serves the purpose of estimating temporal change and understanding how the socio-economic situation evolves during a pandemic.

## Post-pandemic phase

- For the first month following a pandemic, when NSO is able to resume face-to-face interviews, the sample consists of three parts: a 50% overlap from the previous month, a subset of non-respondents during the lockdown phase, and new sampling units from the frame.

- The repeated sample in this month serves two purposes at the same time; to estimate temporal changes, and measure mode effect (a non-sampling error caused by changing mode of data collection).

- The non-response sub-sample will help us to measure non-response effect (second source of non-sampling error).

- Sample in the second post-pandemic month will only consist of repeated sample and new sample.

- At this phase, the repeated sample (and non-response sample in the first month) will be interviewed face-to-face and the new sample will be surveyed with the same mode as lockdown phase. This will help assessing mode and non-response effects (diagram 3).

---

**Important note:**

1. This note aims to stimulate ideas for adapting a feasible survey design during pandemic or any other crisis situation. Proposed ideas are not tested and should be taken on board with caution and only after successful piloting.

2. I have not addressed several issues that I believe standard sampling and survey methods can address: weighting, estimating non-sampling errors (non-response and mode effects), accounting for non-sampling errors in estimation process, questionnaire design.

3. Some components of the method are flexible, and survey designers may consider different alternatives when more appropriate: rate of sampling overlap, panel rotation, and modes of interview.

---

For more information regarding ESCAP's work in statistics development please visit: http://www.unescap.org/our-work/statistics

Previous issues of Stats Brief: http://www.unescap.org/resource-series/stats-brief

Contact us: stat.unescap@un.org

DECADE OF ACTION