Household surveys and COVID-19: Q&A

To monitor the socio-economic impacts of COVID-19, is it better to add questions to an existing survey, or launch a new survey?

There is no single right answer to this question. The approach which is preferable will depend on national circumstances. Factors which will need to be considered include:

- Is fieldwork for the existing survey able to be carried out at the current time?
- How long is the questionnaire? Will adding the COVID-19 questions make it too long, leading to lower response?
- Will the COVID-19 questions affect people's responses to the existing survey questions?
- Does the design and content of the existing survey allow for rapid processing of the results?
- Is there a sampling frame available for selecting a sample for a new (telephone) survey?

One potential approach may be to run a short COVID-19 impacts survey as a follow-up to a recent (or ongoing) survey, using either the full sample or a sub-sample as appropriate. This approach has a number of advantages. For example, unlike adding questions to an existing survey, it does not disrupt or add length to the core questionnaire of that survey, but it should still be possible to combine the data from the two, allowing analysts to make use of the variables collected on the existing survey alongside the COVID-19 questions. The respondents are also already familiar with the NSO, likely leading to higher response rates than for a standalone survey, and there is also a lot of information about the characteristics of respondents and non-respondents available, which can be used to enhance the survey weighting.

How can you maintain comparability of time series if moving existing surveys to a new design? (e.g. moving from face-to-face to telephone)

The current situation creates a risk of discontinuities in statistics regardless of changes in survey design. Gaps in data collection and considerably lower response rates mean considerable care is needed in interpreting apparent changes in long-running time series, even without a switch in survey mode.

Under normal circumstances, a period of parallel running is generally recommended when making substantial changes to survey design such as a change of mode, in order to assist in quantifying the effects. However, this is unlikely to be possible at this time. Discontinuities can also be dealt with through backcasting and time series models, though these too are likely to prove challenging when the underlying variables of interest may also be changing considerably at the current time. Approaches to imputation, weighting and the potential use of auxiliary variables (e.g. from administrative data) should be reviewed where possible.

Given this, there is a need to focus on transparent communication about quality with users. If it is believed that a break in series has occurred, this should be clearly indicated in reporting, including the use of appropriate flags in databases. Releases should also highlight where the precision of estimates are lower than usual or other dimensions of quality are affected. It may also no longer be appropriate to publish as many detailed breakdowns while data is affected by the impacts of the pandemic.

The ILO have published some guidance for labour market statistics, which covers these issues in more detail. The guidance is also broadly relevant for other data collected using household surveys. https://ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_743156.pdf

Sampling and population coverage:

How can we collect data on vulnerable groups, who are not likely to have phones? What is the best approach if CATI is not feasible for them?

How can you deal with the challenge of low numbers of individuals owning telephones in rural areas?

There are potentially two different ways of dealing with low coverage of individuals on a telephone frame.

One is to use a supplementary frame for rural areas or vulnerable groups so the two frames can be combined and have better coverage. The frame and sample for these groups can potentially be constructed with support from individuals such as village leaders and health workers. Depending on circumstances, one approach may to be use 'action-based sampling', that is selecting a random sample from service users (e.g. users of a health centre). If this or standard list-based sampling is not deemed possible, non-probability sampling methods may need to be considered.

It may be possible for access to a telephone to be arranged for the purpose of the survey. If not, then individuals without telephones will still have to be approached through face-to-face interview. If this is needed, protective equipment for interviewers should be considered).

The second approach is to use post-stratification to adjust the results, so the estimates reflect somewhat the situation of those without a telephone number.

Ultimately it would be important to have a good understanding of the structure and coverage of the telephone frame and decide how to proceed further.

Is it possible to get support from telephone companies in obtaining a sampling frame?

Yes, it is possible though is not yet very common. It requires negotiations and agreement signed between National Statistical Offices for sharing such information.

In many countries, there is typically is a national organization that determines which carrier gets to use which telephone numbers in which part of the county. While there is generally no centralized repository for telephone numbers that are actually connected to subscribers, the use of this information narrows down the list of potential numbers which then need to be used for Random Digit Dialling (RDD) sampling, reducing the proportion of numbers in the sample which are not assigned to subscribers (or belong to businesses).

It can be challenging to get the cooperation of private company carriers to provide full lists of phone numbers for landline or mobile phones in service. However, a few

countries have statistical agencies that do undertake sampling of subscriber telephone numbers, and they are able to use random stratified sampling. If it is not possible to agree access to the full list of active non-business numbers, alternative agreements may be possible (for example details of the ranges of numbers currently in use or even the phone company providing a sample of numbers themselves).

If cooperation with telephone companies is not possible at all, another approach could be to explore the feasibility of working with line ministries who may hold contact details for large proportions of households or individuals. Additionally, in some countries, marketing companies may have developed, through agreements with telephone companies and approaches including automated dialling, lists of numbers which are in service, though there will typically be a cost associated with accessing such lists.

For those without access to any lists, RDD is considered the best sampling method to use, based on the virtual frame of all possible phone numbers. However, the main challenge with this is dealing with the large proportion of numbers which are not currently assigned, which can be time consuming unless automated dialling is used.

More information on this topic is available in an interview with Professor Jim Lepkowski on telephone frames available on the UNSD COVID-19 response hub: https://covid-19-response.unstatshub.org/statistical-programmes/using-telephone-interview-for-household-surveys/

Survey design and operations:

How do you let people know about interviews in advance among populations where most people cannot read?

SMS messages sent in advance can be a good way of improving response rates. However, they will naturally be less effective for populations with low rates of literacy.

Besides SMS messages, there are still multiple ways to boost response rates. Telephone calls to schedule interviews at convenient times can obviously be used, but ideally these should be supplemented with other approaches including information campaigns, involving, for example, messages on the radio, community leaders or health workers.

In situations where respondents might need help in distinguishing phone calls of national statistical offices from scam calls, guidance can be provided to respondents on ways for verification. See one example of such guidance from Statistics Estonia at https://www.stat.ee/artikkel-24-04-2020-oluline-teada-kuidas-veenduda-telefonikusitluse-turvalisuses

How do you consider the cost implications of telephone interviewing?

Telephone surveys are generally lower cost than face-to-face surveys of a comparable size, due to removing the need for travel time and associated costs. However, there will obviously be costs associated with making the necessary telephone calls. Investment on constructing the telephone frame will also add to the overall cost. For many other elements, such as data processing, the costs should be comparable.

How long should a telephone interview take? What's people's experience with levels of cooperation?

There is clear evidence that shorter questionnaires lead to higher response rates and a lower proportion of respondents cutting the interview short. However, the empirical evidence on the optimum interview length is limited. It is likely that this will vary according to a number of factors, including how engaging the topic is for respondents. Generally, 30 minutes is seen as an absolute upper limit. It should be remembered that a number of minutes of this total will need to be dedicated to collecting core socio-demographic information.

To avoid respondents cutting short the interview, it is generally also recommended that any questions which are considered to be particularly difficult or sensitive are left towards the end of the interview, where the rapport with the interviewer should be stronger and the impact on item non-response if the interview is cut short should be less.

What is the estimated number of questions for an interview taking 20-30 minutes?

This will depend on the type of questions asked. It should be possible to conduct a battery of 10 straightforward questions with the same response options for each within 2-3 minutes. Conversely, a single very complex question may require up to two minutes to ask and secure an answer.

As the data quality for such complex questions is also likely to be lower via telephone, where the use of showcards for multiple response options is not possible, it is generally recommended to keep questions as short and simple as possible.

Collection of household-level information and use of proxies:

It can be challenging to use telephone surveys for household-level information. What's the best way of dealing with these challenges?

How do proxy responses affect CATI results?

It is necessary to distinguish between telephone surveys (individual respondent selected from a telephone frame) and using telephone interviews to carry out household surveys.

In telephone surveys, you would typically like to get information about individual respondents as well as additional information at the household-level. Depending on whom you approach, the respondent may or may not be a good informant of household information. Common approaches such as adding probing questions can help improve accuracy of the collected information. For example, if you are interested in establishing the household structure, you may ask probe questions such as "Did you forget to include anyone who was with you on [date]?"

Proxy responses about other members of the household for the main survey topic are generally not recommended.

For household surveys that use telephone as the collection mode, the best informant for the particular topics covered by the survey could potentially be asked to complete the survey. This may not be the person who initially answered the phone, and can be established in an initial call to schedule the main interview.

What experiences do people have of using CATI for surveys with modules for different household members (e.g. Module 1 for household head and Module 2 for caregivers of children, to be followed up with a different member of the same household)

If using telephones to carry out interviews of households selected based on area sampling – you can interview different respondents within the households. But if it is a telephone frame and you are interviewing individuals – using a proxy is not desirable unless absolutely necessary.

When interviewing multiple individuals within a household, the risk of non-cooperation or unavailability of an individual is naturally higher, meaning that this approach is likely to lead to an increased number of partially completed interviews.