Innovative Collection Methods During The COVID-19 Period:

Sharing experience with three surveys: *the survey on the impact of covid-19 on companies in the industrial sector*, *the survey on the impact of covid-19 on households and informal enterprises* and the survey of pos-covid-19 investment intentions of formal sector enterprises.

by **Ibrahima Tall,** Economist Statistician,
Head of Business Statistical Bureau at National agency of Statistic and demography of Senegal

---

Over the past two decades, the world has experienced health disruptions that have shown the need to reform country's national statistical systems. For West African countries, SARS-2002, H1N1-2009 and Ebola-2014 have finished to show the limits of data collection and compilation habits. Even beyond the statistical system, it is the entire economy that is brought to its knees due to the external dependence of the countries of this region. Advances in known computer science over the past two decades have helped to make the statistical system increasingly resilient. This therefore offers opportunities for innovation, particularly in this period of the covid-19 pandemic, allowing statistical production to meet expectations in terms of monitoring the economy by producing information in real time. The national agency of statistics and demography of Senegal has undertaken innovative data collection strategies through online and phone interviews. The processing of the data collected is done in real time, thus contributing to the output of results over a record period of less than one month. However, a carefully designed collection and monitoring system was first put in place to ensure proper collection. In this paper, we will first present the collection methods used during previous health crisis periods. Next, the data production procedure used during the three covid-19 surveys will be described: *(1)-the survey of the impact of covid-19 on companies in the industrial sector*, (2)-*the survey on the impact of covid-19 on households and informal enterprises* and (3)-the survey of pos-covid-19 investment intentions of formal sector enterprises. Finally, it will be exposed the constraints and difficulties encountered during these three survey operations as well as the prospects for innovation.

Introduction

Data collection is a costly and time-consuming process. In fact, the design of the questionnaire, although being the main operation, follows an empirical and methodological review which constitutes the starting point of the study. This is followed by the training of interviewers during period which the questionnaire and manuals are presented.

After the collection step, the data are compiled and transformed into a format that can be used by statistical software (SPSS, STATA). Data processing sessions are then carried out to ensure data consistency before moving on to output`s tables and graphs for the analysis.

This process is usually sequential (step by step), and involves moving to a face-to-face interview.

However, this usual procedure is no longer appropriate in the event of a health emergency, in particular an epidemic or even a pandemic. It is important to use innovative approaches adapted to the current situation.

Thus, in what follows, we will first present the approaches used during periods of past health crises, then we will show the techniques used in the present context of covid19. Finally, the main constraints and difficulties encountered during collection operations carried out since the start of the pandemic will be discussed.

I.   Approaches used during past health crises

The periods of crisis discussed in this section are those corresponding to SARS 2002, H1N1-2009 and Ebola-2014.

The health crisis of 2002 had instilled psychosis among the African population, although Africa, with the exception of South Africa, was not affected. Indeed, the country had put in place an alert system as well as a set of measures aimed at preventing the occurrence of cases.

The preventive measures thus adopted contributed to modifying the approach used for collecting and producing data until 2004. The modifications mainly concerned the profile of the interviewers. In addition to a required qualification and experience in data collection, interviewers are now subject to an age limit of 50 years.

In addition to this age criterion, some measures were observed, since the collection was done face to face on a paper questionnaire.

This is how interviewers are required not to shake hands, to wash their hands regularly and to keep a distance from interviewees.

Apart from these instructions, the collection procedure has not changed. Once completed, the paper questionnaires are codified and then entered on computers via an input mask developed with Cspro.

Unlike SARS-2002, influenza A has affected Senegal with more than 300 confirmed cases and 0 deaths. The disease has spread across the country. The first measure taken in data collection operations is momentary shutdown. However, it was quickly

necessary to find a collection procedure adapted to the situation because of the publication urgency thus as to the monthly indices (consumer price index, construction cost index) as well as to information relating to unemployment, poverty and economic indicators on businesses.

Two approaches were used depending on whether the survey was demographic or economic.

For demographic surveys, the interviewers were installed in a large room equipped with telephone sets. This room is called "call-center". The operation is therefore carried out by telephone call and the households surveyed are drawn from the base of the general population and demographic census (RGPH) of 2002.

The telephone survey device remained even after influenza A and is used for surveys with small sample sizes (less than 1000 households) with a light questionnaire (less than 5 pages).

For economic surveys, the field of units is reduced to all companies with a valid contact: a phone number or an e-mail. Interviewers call units and send them letters of introduction from the agency to inform them about the investigation. The interviewers submit the questionnaires to the enterprises selected in the CUCI database, which is a database containing information on enterprises that have submitted a financial statement at least once. Interviewers provide reminders when enterprises are late in responding to the questionnaire by calling or sending them an email. Investigators also answer questions from enterprises about understanding the questioner. The survey finished until the response rate is satisfactory (over 80%).

The Ebola epidemic in 2014 severely affected West African countries, particularly Guinea, where the only case recorded in Senegal came from. From then on, collection operations were suspended for a few weeks. When the surveys start, for this time, it was conducted face to face, although the collection method has evolved since influenza A, moving from a paper questionnaire to a CAPI (Computer Assisted Personal Interviewer).

II. Collection methods during covid19

The outbreak of covid-19 in Senegal has placed the country in deep psychosis due to the loss of life seen in other countries. Thus, in March, the state of emergency was declared, including a curfew, restriction of travel between regions, the closure of schools, the limitation of the number of passengers in the vehicles of public transport etc.

The direct consequence of these measures on statistical production is the stopping of collection operations. However, the price survey continued by telephone. After three months, the government decides to lift the state of emergency and allow travel between regions. Several surveys were then initiated in relation to covid-19: the impact of covid-19 on households and informal production units, the survey on the impact of covid-19 on companies in the industrial sector and the survey of companies' post-covid-19 investment intentions. All of these surveys were organized using the same approach.

To do this, an online form has been developed using the World Bank's Survey Solution platform. For each survey, the units are taken from the base of the general census of enterprises (RGE) of 2016, which covered all economic units with furnished premises. The RGE database has a telephone number of the owner or manager of the enterprises. This contact is used to make enterprises aware of the survey's objectives and also to update the email addresses contained in the database.

In drawing the sample, a non-response rate of 50% was considered, which increases the value of the theoretical size by half. This was justified by a possible unavailability of enterprises to respond to the form due to the reduction of staff (homeworking, worker overall, etc.) as well as the reduction in working hours. The drawing plan is standard for surveys carried out in the division. This plan concerns a proportional selection by following the structure by size and by sector of activity of the sampling frame. In each stratum, a simple random selection is carried out using a random function which assigns a value from the normal distribution to each unit in the base. These random values are then sorted in descending order within each stratum. The first units of each stratum are selected.

Once the awareness and email update step are completed, the sample of units is distributed among the interviewers. They are responsible for sending the form links to the enterprises assigned to them, providing them the specific access parameters. When a unit is fired and does not have a valid contact in the database, the name of that enterprise is kept in the sample along with other information's (address, manager or owner's name, legal regime etc.). Interviewers are called upon to search for an enterprise's contact on the internet by typing the enterprise's name. When the interviewers could not find a contact of the enterprises, this enterprise is replaced.

In addition, the link may not be well received. interviewers should ask enterprises to check for spam or junk mail, and if necessary, the link is resent. In some cases, interviewers offer a telephone interview when the respondent is available for that.

When the enterprise does not respond 24 hours after sending the link, the interviewers calls the manager or owner by phone for a reminder and resend the link. After four days of reminder, when the form has still not been started, the investigator moves to the workplace of the unit, taking certain health precautions (wearing a mask, hydro-alcoholic gel). He also has a paper version of the questionnaire to allow the respondent who does not want to use the web application to complete the paper questionnaire and the interviewer can enter this paper questionnaire on the web application.

When the unit cannot be found via the address available in the database, it is then removed from the sample and replaced by another unit from the same stratum following the order of selection.

This collection procedure proved to be satisfactory in so far as the three surveys yielded a response rate of at least 70%, thus making it possible to move on to the processing and reporting. However, this collection procedure has many shortcomings and could hinder the veracity of the figures at national level.

III. Difficulties encountered during collection

The health crisis posed a constraint of non-mobility in the collection procedure. The applicable type of data collection is CATI (Computer Assisted Telephone Interview) or CAWI (Computer Assisted Web Interview). This assumes that all units in the population of interest have a contact which can be either a phone number or an email. This is not always the case. In fact, in the survey on the impact of covid-19 on informal production units, several units do not have any contacts (neither a telephone number, nor an email address) and were not found on the Internet.

In addition, the possibility of replacing a unit that does not respond to the form is sometimes abused by interviewers who, instead of carrying out recalls, prefer to replace the unit. This could lead to exhausting the list of replacement units in a stratum.

Conclusion

Collecting data during an epidemic is a delicate operation. It requires ingenuity and proven expertise in order to reach a representative sample of the target population. In the surveys carried out during the period of covid-19, the effective size was overestimated by taking a non-response rate of 50%.

Also, the health crisis led to the closure or relocation of several units which, when drawn into the sample, were not found. A replacement procedure was thus adopted to ensure the representativeness of the sample. However, difficulty arises for informal units due to the unavailability of valid contact.