2-1-2021

Issues of Demographic Data Collection during Covid-19 and Its Aftermath

Eduardo L.G. Rios-Neto

Follow this and additional works at: https://knowledgecommons.popcouncil.org/series_pdr_essays-covid

How does access to this work benefit you? Let us know!

Issues of Demographic Data Collection during Covid-19 and Its Aftermath

EDUARDO L.G. RIOS-NETO

BEFORE THE COVID-19 PANDEMIC, a revolution in demographic data collection in developing countries` national statistical offices (NSOs) was dawning. This revolution includes topics such as data science, big data, administrative data, satellite images, grid statistics, the use of biomarkers in surveys, and computer-assisted interviews, among others. The Covid-19 outbreak accelerated this revolution. Most advances in data collection achieved during the pandemic are likely to prevail in its aftermath. Based on my own experience, I start this brief statement with a review of Covid-19 in the Brazilian household survey (Pnad Contínua) and a new pulse survey (Pnad Covid19), both conducted by IBGE (the Brazilian National Statistical Office). Then, I analyze the implications of postponing the Brazilian 2020 Demographic Census to 2021. Finally, I discuss the consequences of advances in data collection for future data analysis, including for assessing the impact of Covid-19.

The Brazilian Household Surveys (Pnad Contínua and Pnad Covid19)

Pnad Contínua is a rotating panel household survey. Each new household included in the sample is interviewed five times in months t and t+3, before leaving the panel. Results are published by quarters and monthly using moving quarters. The survey is the official source of the unemployment rate, annual per capita household income, and inputs for calculating the monthly inflation rate.

On the upsurge of the pandemic, IBGE adopted social distancing starting March 17, 2020. Pnad Contínua was at risk of discontinuation. IBGE mitigated this risk by moving from a computer-assisted person interview (CAPI) to a computer-assisted telephone interview (CATI) using a mobile computing device (MCD) or personal digital assistant (PDA). The traditional CATI surveys

Eduardo L.G. Rios-Neto, Instituto Brasileiro de Geografia e Estatistica, The National Statistical Office.

are not compatible with social distancing because they occur in cramped call centers (LAB-CATI). The natural solution was to place the interviewers working from home (WH) using an MCD or portable smartphone (WH-CATI). I consider this reliance on interviewers' WH-CATI as an innovation in Brazil and other Latin American countries such as Argentina, Chile, and Mexico.

A significant challenge for implementing a CATI survey is how to obtain telephone numbers to reach the sampled households. Less than half of the interviewers collected phone numbers in the first round of surveys. All households in the new strata did not have telephone numbers attached. An alternative to obtaining these numbers was to obtain them directly from telephone companies, but the Brazilian Supreme Court ruled against this option. The solution was to obtain the numbers by matching the survey sample with administrative data using the addresses as a block variable. However, not all telephone numbers worked. Some local IBGE offices tried telegrams or relied on motorcycle couriers to obtain the telephone numbers in person. An essential comparative finding contrasting with classical telephone surveys is that the refusal rate in the WH-CATI survey, conducted by our trained interviewer, was much lower, less than 5 percent, than the average refusal rate of around 40 percent. The sample coverage rate comprised more than 60 percent of the households sampled before social distancing, usually leading to the publication of statistics at national and state levels, within a reasonable confidence interval.

Due to the pandemic, IBGE designed another household survey (Pnad Covid19). The new survey provided information on flu symptoms, hygienic habits, reports of personal Covid-19 tests, seeking health care, type of facility visited, labor market experiences (people in and out of the workforce, employed workforce away from work due to social distancing, employed population working from home, number of people working in informal employment, etc.), and income (earnings, retirement income, unemployment insurance, conditional cash transfer receipts, emergency Covid-19 allowance). The option to implement this new survey was also to generate alternative labor statistics in the case of an eventual discontinuation of the traditional Pnad Contínua.

The Pnad Covid19 is also a WH-CATI. It is a fixed panel comprised of the households sampled in the Pnad Contínua during the first quarter of 2019. The sample corresponds to 92 percent of the 211,000 households, comprised of those with a telephone number in the questionnaire. Not all telephone numbers were valid, so matching with administrative data was also necessary. We divided the sample equally into four weeks, providing independent results. Due to its fixed panel nature, we interview the same households each month. IBGE publishes the results both weekly and monthly. The coverage rate of telephone interviews averaged around 70 percent of the sample. The sampling strategy enables us to correct for possible selectivity caused by tele-

phone interviews by comparing them with personal interviews conducted in the first quarter of 2019. The first month covered by Pnad Covid19 was May 2020. It is scheduled to last until November 2020 due to the panel's attrition. IBGE may conduct the panel survey interviews again in the future. In this case, we could perform an analysis before (2019 sample anchor and baseline), during (first wave of Covid-19 from May to November 2020), and after (date to be defined after waves of Covid-19). This temporal analysis could inform an impact evaluation of the pandemic.

In a nutshell, the Pnad Covid19 provided useful data on self-reported flu symptoms that could be a proxy for Covid-19, emphasizing the loss of smell and taste as an essential trait. It showed that most of those with symptoms seeking medical treatment relied on the Brazilian universal and free- of-charge Public Health System (SUS). The number of persons in the formal labor force declined during the Covid-19 pandemic while a significant proportion of those working moved to do so from home. Finally, the Brazilian Government's emergency Covid-19 allowance raised a substantial proportion of low-income households out of extreme poverty during the pandemic's first six months.

Covid-19 and the Demographic Censuses

Covid-19 hit the planned 2020 round of demographic censuses hard. In those countries where the pandemic census operations were already underway (e.g., Mexico and the United States), it was difficult to stop. But, like many other countries that had not initiated the census before the onset of Covid-19, Brazil faced the postponement of census operations to 2021. In the Brazilian case, coverage difficulties, length of the questionnaire, and budgetary ceiling were already weakening national census operations. However, Covid-19 raised census operation difficulties to another level.

In the 2010 Brazilian Demographic Census, the interview mode was CAPI, with the use of MCDs for the first time. IBGE was planning to maintain the same strategy in 2020. With the outbreak of the pandemic, four factors reinforced IBGE's caution regarding the CAPI mode: uncertainty regarding the duration of the first wave of contagion, the possible upsurge of second or further waves, the lack of vaccines for census collectors, and the likely absence of herd immunity. A risk-averse strategy is to adopt a mixed and concurrent data-collection strategy, CAPI, and CATI. CATI requires telephone numbers, which can be obtained using the match of household lists with administrative data. Alternatively, telephone numbers can be collected at the beginning of the census operation, during the field verification of household listings in the census tracts, with minimum personal interaction protocols. A third data-collection alternative is CAWI (Computer Assisted Web Interviewing).

This self-enumeration strategy enumerates hard-to-reach households such as gated residences and wealthy families. CAWI is also the last resource for unsuccessful CAPI and CATI interview attempts.

As mentioned earlier, IBGE had to deal with complex issues associated with census coverage, even before Covid-19. The uncertain environment due to Covid-19 aggravated these challenges. A risk-averse strategy is the application of technical tools to monitor the census operation. The estimation of population and dwelling density at the grid- or census-tract level is a control strategy to the three phases of a demographic census. It helps to evaluate the quality of the households` listings during the pre-enumeration phase. It also contributes to the daily measurement of coverage in the census tracts during the enumeration phase. Finally, it is an alternative technique to evaluate census coverage and enable imputation during the post-enumeration phase. Another technical tool is the real-time application of demographic indicators such as sex ratio, age reporting, etc., displayed in dashboards so that the operation can correct problems during the census enumeration phase.

It is tempting to include Covid-19 questions in the census, but we have to consider that the census operation is planned well in advance to the enumeration stage. The Brazilian census consists of two questionnaires, a basic (short) questionnaire applied to all households and a long one devoted to a household sample. If a demographic census provides precise population enumeration, broken down by socioeconomic attributes, it will fulfill its primary role. Regarding Covid-19, it is crucial to have denominators (population at risk) to analyze several indicators, especially the attack rate.

Other Data Collection Innovations and Methodologies Contributing to the Analysis of Covid-19

The academic community will consider the Covid-19 pandemic as a marker for several natural experiments with potential to analyze causality. At the micro-level, a longitudinal dataset would be crucial for these studies. Unfortunately, longitudinal surveys are not typical in developing countries. The sampling strategy in the Pnad Covid19 consisted of selecting sampled households previously chosen in the first quarter of 2019 in the Pnad Contínua. It is possible to conduct the "before" (first quarter of 2019) and "during" (from May to November 2020) analysis, using the Pnad Covid19 data. If IBGE recovers the Pnad Covid19 panel in the future, then the "after" component will enable even more robust causal analyses.

The use of biomarkers in surveys is becoming more diffused, although still not prevalent in developing countries. Testing for Covid-19 has become essential for tracking herd immunity and other aspects. This testing information linked to a longitudinal dataset is crucial for causal analysis. The Brazilian

Health Ministry considers the possibility of testing the whole Pnad Covid19 sample, which would enrich this type of research.

IBGE conducted two national health household surveys (Pesquisa Nacional de Saude—PNS) with independent samples in 2013 and 2019. These surveys are comprehensive on lifestyle habits such as smoking, alcohol consumption, physical exercise practices, etc. They also include the prevalence of chronic diseases among household members. Even before the pandemic, IBGE planned to match the individuals in these households with the vital registration of deaths. The datasets linkage will entail a mortality analysis. With the onset of Covid-19, it will be possible to perform a before-and-after study, including the mediating role of lifestyles and comorbidity factors.

If the demographic census measured household deaths due to Covid-19, there would be relevant applications. This census question complements vital statistics in terms of measuring socioeconomic differentials. Another critical factor not available in vital registration is the role of household composition, especially the presence of the elderly and the measurement of household-level transmission, which might translate into more than one death. The role of the high bedroom density of individuals is a micro-level factor in the diffusion of contagion. Deaths in the census also entail analyzing ecological (macro-level) differentials such as a cluster of segregated residential areas comprised of dwellings with low sanitation coverage, residence in slums, the proximity of houses, etc.

In a nutshell, the pandemic forced national statistical offices to innovate to keep producing statistics minimizing time-series break. Some of these innovations proved so useful that they should stay in post-pandemic times without social distancing. Also, statistical offices should strive to produce data that could support the new research agenda derived from Covid-19.

Acknowledgments