Attrition in mobile phone panel surveys

Instead of using long questionnaires administered in person, researchers are increasingly turning to phone surveys, which require shorter instruments but can be administered over multiple, shorter interviews. A limitation of high-frequency phone surveys is study attrition, where individuals enrolled in a baseline survey may not be reachable or willing to complete follow-up interviews.

This brief shares some evidence on phone survey attrition calculated from existing data collected in the early 2010s in Tanzania and Senegal. In these cases, the researchers distributed devices to respondents, ensuring the best-known conditions for minimizing attrition. In addition to presenting attrition rates calculated over multiple survey waves, the brief explores whether there is *differential* attrition by respondent type, examining changes to the sample composition. Differential attrition can lead to bias in the parameters that researchers are trying to estimate.

The results show that attrition was low after the initial drop-off from in-person baseline to first phone follow-up, but there were small but statistically significant differences between the characteristics of attriters and non-attriters in multiple follow-ups. These data provide a useful benchmark and some cautions for planning future studies.

Motivation

Attrition has always been a concern in longitudinal research. A *high* attrition rate reduces the available sample size for analysis. A *differential* attrition rate, where differences are related to variables that the researcher wants to study, produces bias as the sample becomes less representative of the original target population. There is renewed interest in measuring and reducing attrition for phone surveys as the COVID-19 crisis has extended into its second year.

A second important concern, which is specific to phone surveys and especially problematic studying poverty in low- and medium-income countries (LMICs) is coverage bias. Phone surveys exclude those who lack devices, connectivity, calling plans, and reliable electricity for charging. As with differential attrition, this may systematically exclude marginalized populations of interest and therefore introduce troubling bias. Mobile phone penetration is not widely reported, but evidence on device ownership suggests that cell phone owners are not representative of the population, even though ownership rates have been rapidly increasing.[1]

Existing Evidence

Much of the existing evidence on response rates for mobile phone panel surveys in LMICs comes from the World Bank's tests of high-frequency phone surveys in the early 2010s and work done during the 2013-2016 Ebola crisis, where face-to-face surveying was infeasible.[2] The World Bank used sampling frames built in partnership with national statistical offices and NGOs. They started with face-to-face data collection for the baseline and provided phones and phone credit incentives to all respondents. This was followed by several rounds of computer-assisted telephone interviewing (CATI), where interviewers tried to reach the same respondents on the phone, offering phone credit as an incentive.

[1] International Telecommunications Union, 2020; Lau et. al., 2019

[2] Dabalen et. al., 2016

IPA's evidence briefs are part of a series reviewing existing evidence on implementing surveys using computer-assisted telephone interviewing (CATI) and other remote survey modes. This document was prepared by Janina Roemer and Michael Rosenbaum with helpful input from Steven Glazerman. It is based on data analysis performed by Janina Roemer.

These World Bank high frequency surveys resulted in low attrition rates, nearly 20 percent of the baseline face-to-face survey after two years in some sites. Other attempts in Ghana, Peru, as well as Liberia and Sierra Leone during the Ebola crisis, yielded mixed results with attrition as low as 3 percent per round reported in some cases and high rates of attrition in others.[3] Some researchers have found that protocol modifications may substantively change attrition rates in phone surveys.[4]

However, there is a dearth of reported evidence on how sample composition changes due to nonresponse, and panel attrition especially. The World Bank's "Listening to Tajikistan" study investigated how a limited set of demographic variables predicted sample composition 24 following monthly phone survey rounds compared to the first phone survey round. They report that wealthier households, smaller households, and households headed by women were sometimes more likely to respond, but there are substantive differences by round and by region.[5]

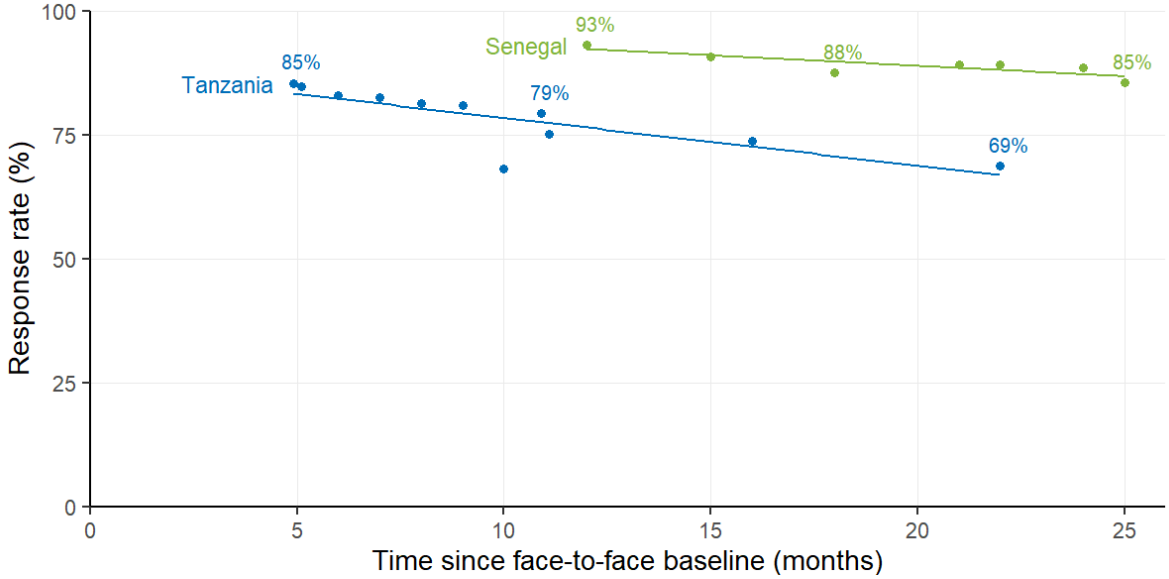
Response Rates and Attrition Rates

To complement the information already published by the World Bank about phone survey attrition, we analyzed data from surveys in two of the six countries in the World Bank's "Listening to Africa" series that were publicly available in December 2020.[6]

As shown in Figure 1, response rates were high. 93 percent responded to the first round in Senegal and 85 percent in Tanzania. In the last CATI rounds, almost two years from baseline, the response rates were 85 percent in Senegal and 69 percent in Tanzania. In terms of sample attrition from the first phone round to the last phone round, this corresponds to 8 percentage points and 17 percentage points respectively. This is equivalent to approximately 1.3 and 1.7 percentage points per round on average, or 0.6 and 1.0 percentage points per month.

[3] Ballivian et. al., 2015; Etang & Himelein, 2019; Heath et. al., 2019; Maffioli, 2019
[4] Özler & Cuevas, 2019
[5] Broccolini, Seitz & Azevedo, 2017
[6] Listening to Africa

Figure 1: Response rates in both panels



Data: Survey - Listening to Senegal, Year 2014-2017, National Agency for Statistics and Demography (ANSD) of the Republic of Senegal, www.ansd.sn; Twaweza, Sauti za Wananchi survey (Rounds 1-24, October 2012-September 2014), www.twaweza.org. Note: This reports the response rates for the individual phone survey rounds as a percentage of respondents who had initially agreed to participate in the phone survey. The face-to-face survey was conducted in month 0. The outlier in month 10 in the Tanzania data corresponds to a survey about political corruption. Further results and code to reproduce this analysis can be found here.

Composition Changes

Attrition, in and of itself, is only a problem for the accuracy of estimates if the attrition does not occur at random. If the sample composition changes, then attrition may result in biased estimates of a population parameter. This is especially a concern over long periods of time between recruitment and surveying and many rounds, as respondents have multiple opportunities to stop participating in the survey.

The two "Listening to Africa" surveys with open data provide an opportunity to test if attrition introduced biases in these two panel surveys. To do so we use the face-to-face baseline surveys which allows us to analyze demographic characteristics of respondents and nonrespondents of the CATI survey rounds alike. For each CATI round we conduct two types of tests: (1) We test for differences between respondents and nonrespondents, analyzing several shared demographic characteristics in both surveys,[7] namely age and gender identity of the respondents, whether they live in an urban area and whether they have completed secondary education, and whether the respondent works in agriculture.[8] (2) We test whether the analysis sample after attrition differs in terms of any of these same variables, from the original sample interviewed in person at baseline. The first test asks, "Are attriters different from non-attriters?" The second test asks, "Are these differences large enough to

[7] Similar to a covariate balance test, we estimated a linear model using OLS for each round with the discrete characteristic response status as the outcome and a set of demographic variables as explanatory variables. We then tested that the coefficients of all demographics were equivalent to 0 to test if these samples were similar on these observable characteristics. Code to reproduce these data and the full results are available on Github here.

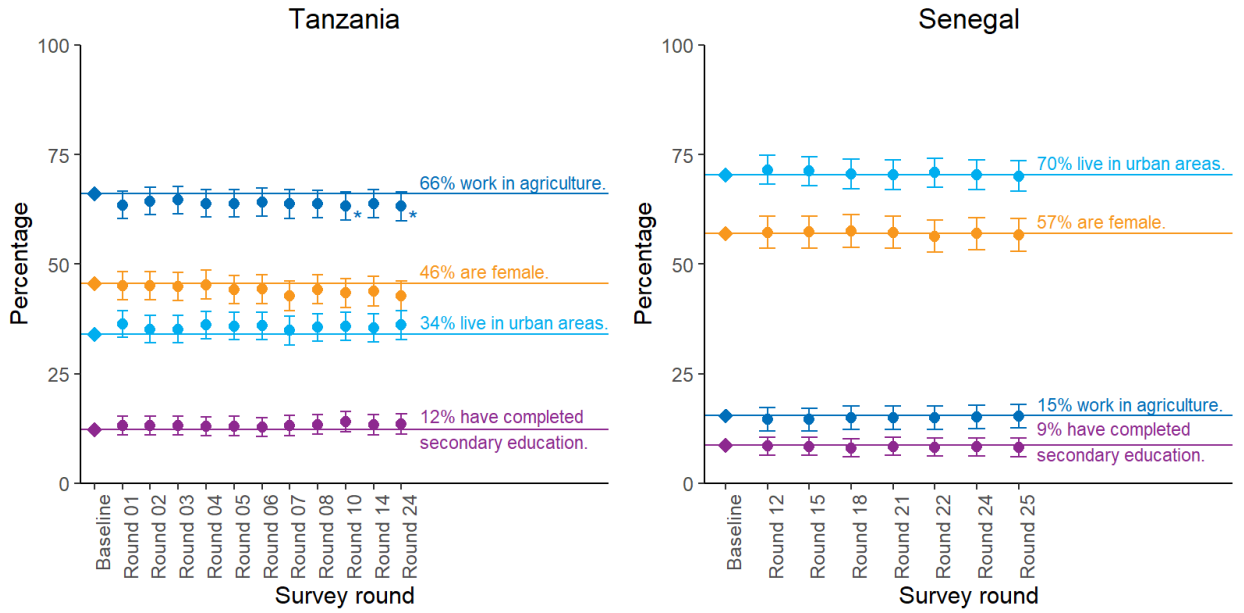
[8] We include an indicator for whether the respondent works in agriculture based on high rates of agricultural labor and feedback from IPA survey teams that different contact protocols are more effective during planting and harvesting seasons.

make a difference in the sample composition?" We find that in these samples the answers to these questions are a qualified "yes" and a "no," respectively.

We find significant differences between the two groups—namely respondents and nonrespondents—for three out of seven CATI rounds of the survey conducted in Senegal and all of the 11 rounds in Tanzania. However, these differences do not share a meaningful pattern between rounds and surveys for most variables. Working in agriculture is the only characteristic where all coefficients are consistently negative over both surveys and all rounds, with 11 out of 18 rounds being significant at $p < 0.05$. Some variables show substantial differences in many rounds. For example, in the final round in Tanzania we find that 52 percent of attriters but only 43 percent of non-attriters were female.

Although these differences in sample composition between respondents and nonrespondents are statistically significant, meaning they are unlikely to arise from pure chance or sampling variation, they are not large enough to drive significant changes between the composition of each round and the baseline. We compare the baseline sample to each phone survey round on all of the covariates tested for attrition individually, finding that no characteristics are statistically significantly different from the baseline at $p < 0.05$ in any round of the two surveys. Figure 2 shows the results of these tests with the baseline average of each characteristic marked with a horizontal line.

Figure 2: Composition of the phone samples compared to the face-to-face baseline sample



*Note: This reports the results of four linear models estimated by OLS, each comparing the average value of one characteristic between the respondents of the different survey rounds. Horizontal lines indicate the face-to-face baseline values. Rounds correspond approximately to months after baseline data collection, however, in the survey conducted in Tanzania, some months have multiple rounds. Therefore, we use the round numbering from the data descriptions in this case. Only rounds with response rate data available are shown. 95 percent confidence intervals of robust standard errors are reported for each survey round. * p < 0.10 ** p < 0.05 *** p < 0.01.*

Implications

When planning new longitudinal surveys, researchers must anticipate the likely response rates (percentage of the sample frame who complete a given survey) as well as sample attrition (percentage of those enrolled in the study who are lost to follow-up) from different study designs,

and gauge the impact of such attrition on the quality of inferences. Associated study design decisions include the mode (e.g. in-person, phone, or a combination), frequency, and duration of surveys. Our ability to predict response rates and attrition rates has implications for not only quality of inference but also sample size and cost. As this brief shows, studies like the World Bank "Listening to" series provide useful data points for making such predictions.

The data presented here should be interpreted as a near-best case scenario for phone surveys in LMICs. The baseline was conducted in person. The data collectors were known to respondents. The data collectors distributed mobile phones at baseline to all respondents and provided airtime credits as completion incentives. As such, these response rates of 85 and 93 percent to the first round of phone surveys should be considered an upper bound. In random digit dial (RDD) surveys in nine LMICs conducted in 2020, response rates to the first round of interviews ranged from 6 to 59 percent.[9] Similarly, average attrition rates reported here, 1.3 and 1.7 percentage points per round, or 0.6 and 1.0 percentage points per month, in Senegal and Tanzania respectively, might also be considered an upper bound, because of the high frequency of the surveys and the reasons listed above, including device distribution and the data collector being known to respondents.

Keeping in mind that these were favorable conditions for phone surveys, the results on differential attrition can still be instructive for researchers planning future efforts. With distribution of devices, the follow-up samples closely resemble the initial sample, suggesting that this strategy may be successful at keeping both coverage bias and differential attrition bias within tolerable limits, but two caveats apply: (1) it is not known which contextual factors besides the distribution of phones may need to be present to reproduce these results in other settings and (2) while the sample changes were small and not statistically significant, the evidence suggests that there is a true change in the type of respondent who drops out of the sample over time. In surveys with less frequent follow-ups or with more sensitive questionnaire content, the nonrespondent group could be larger, resulting in sample changes that could be more salient. Given all of these qualifications, we note that researchers should be encouraged to analyze survey paradata such as response status or at least release public versions for re-analysis of such data so that the field can benefit from a richer understanding of how to collect high quality data with the lowest cost and respondent burden.

[9] (Dillon, Glazerman, and Rosenbaum, 2021)