Northwestern
GLOBAL POVERTY
RESEARCH LAB

ipa
INNOVATIONS FOR
POVERTY ACTION

What does the evidence say about mode effects on data quality?

Comparing Phone Surveying to Face-to-Face Interviewing

We reviewed evidence on whether asking the same questions via different survey modes—over the phone versus face to face—produces different answers in low- and middle-income countries (LMICs). While there is limited evidence from LMIC settings on these differences, known as mode effects, those studies are summarized here, which turned up examples of meaningful mode effects. In particular, there is evidence suggesting that respondents may be more likely to give socially desirable responses over the phone than in person.

Motivation

A key concern across all modes of data collection is that survey data accurately reflects the world. This is especially relevant for remote data collection modes, as the COVID-19 pandemic has forced many researchers to pivot from established modes of data collection, such as face-to-face surveys, to remote surveys where there are open questions about the accuracy of data collected. For surveys to accurately measure the intended information, researchers need to understand if and how mode effects impact survey responses.

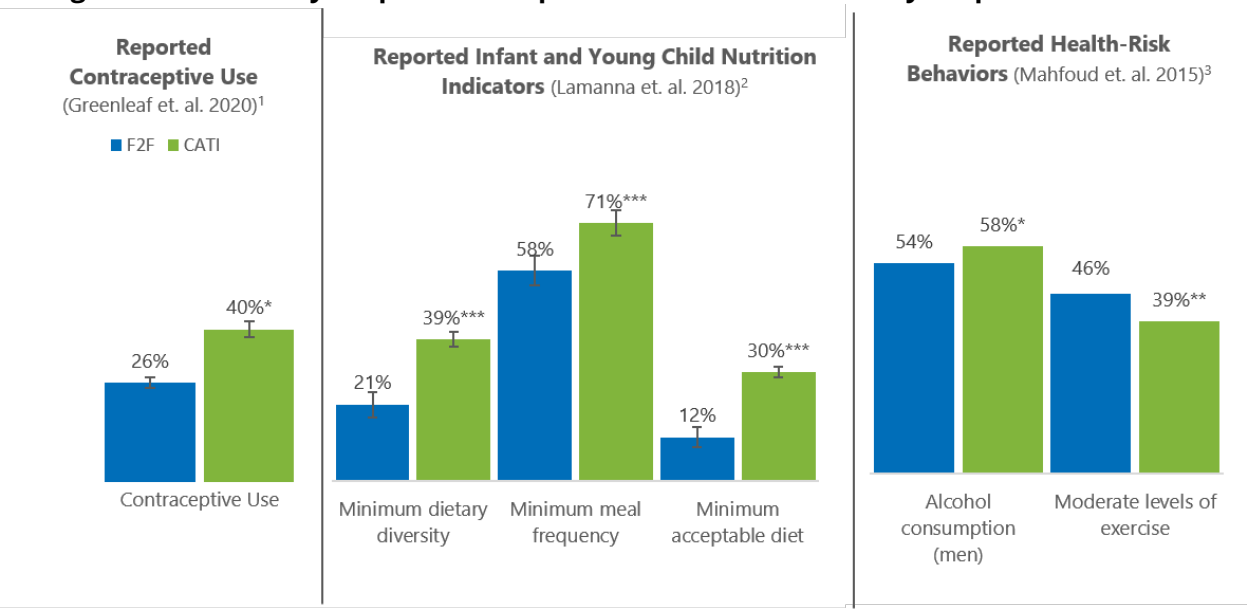
Existing Evidence

Evidence on validity of remote data collection in LMICs is limited but does exist for a variety of modes, sampling techniques and populations. Gibson et al. (2017) aimed to review and synthesize studies from LMICs to identify mode effects but were unwilling to draw general conclusions due to the limited number of studies. Instead, one can point to a few individual studies which did find meaningful differences in responses to the phone survey mode (known as Computer-Administered Telephone Interviews, or CATI) compared to responses from face-to-face interviewing, abbreviated here as F2F.[1] These examples suggest that survey mode may affect respondents' tendency to give exaggerated, socially desirable answers, but it was not always the same mode that produced this bias. At the same time, it should be noted that there were survey questions in these studies for which mode effects were not found, and one study where no evidence of mode effects was found.

[1] There is some evidence comparing other modes (IVR and SMS) but we have focused on CATI in this brief.

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. These briefs are made possible with the generous support from and collaboration with Northwestern University's Global Poverty Research Lab (GPRL).

Figure 1: CATI Survey Responses Compared to Face-to-Face Survey Responses



*[1] n=2,379; [2] n=578 ; [3] n=630; Mahfoud et. al. used McNemar's test for dependent samples to compare response differences. Therefore, we do not report confidence intervals in this figure to ensure comparability across sites. *p<.05; **p<.01; ***p<.001****

Figure 1 shows six survey responses from three studies with the F2F average compared to the CATI average for each. Differences in the heights of adjacent bars suggest mode effects, with all six being statistically significant. The first study, conducted in Burkina Faso, used a national sample of women of reproductive age surveyed F2F about contraceptive use (Greenleaf et al., 2020). For the CATI survey, random digit dialing was used to generate another representative sample of women. The CATI sample reported higher contraceptive use, 40% versus 26% in F2F, a statistically significant difference of 14 percentage points. The other two studies used test-retest designs, where respondents were randomly assigned to be interviewed with one mode first, and re-interviewed in the other mode shortly thereafter. In one of those studies, female caregivers in Kenya reported 12 to 18 percentage point higher scores for infant nutrition indicators in the CATI survey than F2F (Lamanna et al., 2018). In the last study presented in Figure 1, conducted in Lebanon, respondents reported higher rates of exercising regularly when speaking to an in-person interviewer than over the phone, 46% versus 39%, while men reported lower rates of alcohol consumption, 54% versus 58% (Mahfoud et al., 2015). These differences were statistically significant.

Another study, not shown in the figure (Gallup, 2012) found very little difference between responses by mode in Honduras, and also similar Cronbach's alpha reliability statistics for construct validity using CATI versus F2F responses.

Mechanisms

For the three out of four studies that found evidence of mode effects, several mechanisms could explain the findings. These include non-response bias, measurement error, and social desirability bias. We can probably rule out measurement error because the studies used the same questionnaires across modes and conducted the CATI and F2F measurements within a maximum of four months of time[2]. Data collection mode can impact survey responses through a variety of sources of error, such as non-response bias and measurement error, but social desirability bias is of particular interest given the results of these studies.

While research suggests that respondents are more likely to misreport sensitive information in a phone or in-person interview in order to present themselves in a more positive light (referred to as social desirability bias), less is known about how these different modes may affect the degree and directionality of misreporting (Tourangeau and Yan, 2007). The study in Lebanon (Mahfoud et al., 2015) included questions about health and health-risk behaviors, which may be sensitive to this kind of response bias. Both men and women reported more exercise, generally recognized as a socially desirable behavior (Tourangeau and Yan, 2007), in the F2F interviews compared to the phone, while questions about cigarette smoking and health conditions such as diabetes and heart disease did not show differences. Along the same lines, men reported lower alcohol consumption in F2F interviews than they did over the phone, which is more evidence that phone interviews may reduce socially desirable response bias.

The relevant norms informing socially desirable response bias may vary across populations and cultures. For example, Lebanon is ranked third in the world for highest cigarette consumption per capita[3], so the Lebanese may not regard smoking as undesirable like many other cultures. This pattern implies respondents were likely more truthful about their behaviors in the phone interview.

In Kenya (Lamanna et al., 2018), the evidence went in the opposite direction. Female caregivers were interviewed about the quality of their own diets and those of their infants and, while diet scores for women did not differ across modes, infant diet scores were as much as 18 percentage points higher in the CATI interviews. These results suggest mode effects, such as social desirability. Adult diets may be subject to disapproval but are hardly as sensitive to judgment as infant diets and feeding practices, which could portray a caregiver in a bad light. As such, asking about infant diets may be subject to social desirability bias, resulting in overreporting of the quality of diets. The higher infant diet scores in CATI interviews may reflect a population-specific bias effect as overreporting would more likely be observed in the F2F interview.

[2] Time between surveys ranged from 9 days (Lamanna et al., 2018), a minimum of 10 weeks (Gallup, 2012), approximately 2 months (Mahfoud et al., 2015) to 4 months (Greenleaf et al., 2020)
[3] The Tobacco Atlas, 2018

Between the two survey modes, F2F interviews likely exert more pressure to provide socially acceptable answers than CATI interviews, given the interaction also includes social cues perceived through eye contact, facial expressions and body language. In a CATI interview, social cues are reduced to only those recognized in phone conversations and may be easier to ignore. This was observed in the Lebanon study but the opposite was seen in the Burkina Faso and Kenya studies. Unlike the Lebanon study that surveyed men and women, the other two studies focused on women of reproductive age, which may explain the unexpected bias in CATI responses.

Women in African countries are less likely to own a phone than men, and some have suggested that concerns with security, harassment, agent and operator trust and confidence safety limit how women use phones.[4] General inexperience or the noted concerns with mobile phones may have resulted in respondents feeling suspicious and uncomfortable towards the phone-based survey and thus, more susceptible to social desirability bias. The same pattern of social desirability bias may have played a role in Burkina Faso (Greenleaf et al., 2020), where women reported more contraceptive use, likely a sensitive question, over the phone. Even after adjusting for potential frame and composition bias, the difference remains significant, indicating another form of error such as social desirability.

Unlike the other studies, the Ballivian et al. (2015) study did not appear to provide evidence of similar bias. This could be in part to the survey questions, which inquired about perceptual information on socioeconomic status and factual information on household infrastructure and access to the internet, which may have been less sensitive and susceptible to social-desirability bias. It may also be due to the nationally representative survey sample, or to the study design, which included monetary and psychological incentives.[5] We cannot conclusively tell which of these differences explain the presence of mode effects, but these possibilities suggest that researchers designing questionnaires should consider that mode effects could be present and go in either direction, depending on the nature of the questions and possibly the characteristics of respondents and interviewers. Further testing in the future would be valuable for adding to this literature on mode effects.

[4] GSMA, Connected Women, 2015

[5] Psychological incentive participation in the survey was presented as "an interesting and engaging activity which would give panelists the opportunity to have their voice be heard by national and global leaders, and to represent their country nationals."