



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Regular Article

# Measuring consumption over the phone: Evidence from a survey experiment in urban Ethiopia

Gashaw T. Abate<sup>a</sup>, Alan de Brauw<sup>a,\*</sup>, Kalle Hirvonen<sup>a,b</sup>, Abdulazize Wolle<sup>c</sup>

<sup>a</sup> International Food Policy Research Institute (IFPRI), Washington D.C., United States

<sup>b</sup> The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki, Finland

<sup>c</sup> Economics Department, State University of New York at Albany, Albany, NY, United States

## ARTICLE INFO

## JEL classification:

O10  
I32  
C81

## Keywords:

Survey experiment  
Phone survey  
Survey fatigue  
Food consumption  
Household surveys

## ABSTRACT

The paucity of reliable, timely household consumption data in many low- and middle-income countries have made it difficult to assess how global poverty has evolved during the COVID-19 pandemic. Standard poverty measurement requires collecting household consumption data, which is rarely collected by phone. To test the feasibility of collecting consumption data over the phone, we conducted a survey experiment in urban Ethiopia, randomly assigning households to either phone or in-person interviews. In the phone survey, average per capita consumption is 23 percent lower and the estimated poverty headcount is twice as high than in the in-person survey. We observe evidence of survey fatigue occurring early in phone interviews but not in in-person interviews; the bias is correlated with household characteristics. While the phone survey mode provides comparable estimates when measuring diet-based food security, it is not amenable to measuring consumption using the 'best practice' approach originally devised for in-person surveys.

## 1. Introduction

When it became clear the spread of COVID-19 would become a pandemic in March 2020, many surveys that had been taking place in-person could no longer be fielded due to the concern they would contribute to virus spread. Yet in-person surveys are a key component to many research efforts and monitoring outcomes such as those measuring progress towards the Sustainable Development Goals (SDGs). Without in-person surveys such as the Demographic and Health Surveys (DHS), Household Consumption Expenditure Surveys (HCES), Living Standards Measurement Surveys (LSMS), and other similar surveys conducted by national statistical offices, it is impossible to know what kind of progress is being made towards meeting the SDGs or reducing poverty in general.

The main pivot by many researchers during the early part of the pandemic was to begin conducting phone surveys.<sup>1</sup> There was a veritable explosion of efforts to collect some type of data to monitor situations over the phone, including major coordinated efforts by Innovations for Poverty Action (RECOVER) and the World Bank (Gourlay et al., 2021).

These efforts have played an important role in helping us to understand some of the socioeconomic consequences of the pandemic. In terms of living standards, these surveys generally asked about job loss and loss of income, and they tend to show substantial negative effects (Egger et al., 2021; Josephson et al., 2021; Miguel and Mobarak, 2021). Yet these findings are all based on crude measures, e.g., asking whether household income was lower, the same, or higher than it had been at the same time of the year 12 months ago.

Although these surveys provided valuable information about how living standards were qualitatively changing during the early part of the pandemic, there remain obvious ways that phone surveys cannot replace in-person surveys. Some variables require physical measurement; for example, it is impossible to study how stunting prevalence is evolving among children under 5 years of age without in-person data collection.

Similarly, collecting data on household consumption expenditures to estimate poverty incidence requires complex measurement.<sup>2</sup> The standard household consumption expenditure and poverty measurement involves administering detailed food and non-food consumption

\* Corresponding author.

E-mail address: [a.debrauw@cgiar.org](mailto:a.debrauw@cgiar.org) (A. de Brauw).

<sup>1</sup> We mainly cover the relevant literature in low- and middle-income countries. Over the past 40 years, phone surveys have become the most frequently used data collection method in high income countries. For a review of the key methodological issues in this context, see Chapter 10 in Tourangeau et al. (2000).

<sup>2</sup> Based on the most recent data for each country reported in the World Bank's PovcalNet database, more than 90 percent of the poverty statistics in low and lower-middle income countries originate from household consumption surveys.

modules covering more than 100 items typically consumed in the country (Deaton and Grosh, 2000; Deaton and Zaidi, 2002).<sup>3</sup> Consequently, most phone surveys have not attempted to collect such data, in trying to minimize the time spent on the phone.

As researchers have shied away from collecting complex data over the phone, we lack data on specific trends through the pandemic. In reviewing impacts on incomes, Miguel and Mobarak (2021) do not even attempt to speak directly to trends in poverty incidence. Despite the fact that modelers have predicted large increases in poverty incidence and rising food insecurity due to policies associated with the pandemic (e.g., Laborde, et al., 2021; Lakner et al., 2021; Sánchez-Páramo et al., 2021; Sumner et al., 2020), the lack of data collected in-person means it is difficult to tell whether their predictions have come true.

The surveys that have tried to collect consumption data over the phone during the pandemic suggest the increases in poverty incidence are not as severe as either the crude income measures or models would suggest. Egger et al. (2021) report on phone surveys in Kenya and Sierra Leone that collected data on food consumption in both countries and non-food consumption in Kenya, and find that the value of food consumption increased in both countries, offset by a decline in non-food consumption in Kenya.<sup>4</sup> Janssens et al. (2020) study a sample of households in Kenya collecting financial diaries, and find that households sold assets to maintain food consumption levels. Hirvonen et al. (2021) also find no material change in the value of overall food consumption in a representative sample of Addis Ababa between an in-person survey conducted in 2019 and a phone survey conducted at the same time of year in 2020, though the composition of food consumption changed.

These surveys suggest it might be plausible to conduct phone surveys to measure consumption as it had been before and therefore poverty incidence, particularly if survey efforts first attempt to develop some rapport with households before the long consumption survey, as is true in all the surveys described above. But it is important to quantify differences between phone and in-person measures of consumption before making such conclusions. Therefore, here we test whether consumption data collected over the phone has a comparable distribution to data collected in-person, using a sample that has been asked about food consumption several times in the past. We randomly select half of the sample to be enumerated about consumption in-person, with the other half enumerated over the phone. We do not include other modules in the survey, so we cannot test other differences between phone and in-person surveys. However, note that we can generate other indicators that are often enumerated in phone surveys, such as the household diet diversity score (HDDS) and a food consumption score (FCS) providing alternative measures of the household's food security.

We can then compute poverty incidence using both the consumption measures generated by our phone sample, versus the in-person sample. Note that it is best to at least initially be agnostic about which sample provides closer to a "true" approximation of the distribution of consumption, and therefore poverty incidence. Indeed, an important challenge in survey experiments such as ours is that we do not observe the "true value" against to which to benchmark our estimates (De Weerd et al., 2020). However, when we test for survey fatigue by randomly changing the order in which the food groups appear in the food consumption module, we observe evidence of survey fatigue occurring very early on in the phone interviews but not in the in-person interviews. It seems then that the in-person survey mode does perform better, resulting in less measurement error than the phone survey mode. Our assessment of data quality based on Benford's law also suggest that the

consumption data from the in-person survey are of higher quality than the data from the phone survey. In heterogeneity analysis, we find that bias is attenuated among more educated household heads, and is positively related to household size.<sup>5</sup> This finding implies that the measurement error in phone survey mode is not classical and, as a result, cannot be easily corrected with standard methods used in the literature (Bound et al., 2001).

This paper contributes to the understanding of how variation in survey designs can shape data quality and ensuing analyses (De Weerd et al., 2020; McKenzie and Rosenzweig, 2012; Zezza et al., 2017). Much of the previous work has focused on improving consumption measures used to measure poverty incidence (Abate et al., 2020; Ameye et al., 2021; Backiny-Yetna et al., 2017; Beaman and Dillon, 2012; Beegle et al., 2012; Caeyers et al., 2012; De Weerd et al., 2016; Friedman et al., 2017; Gibson et al., 2015; Gibson and Kim, 2007; Jolliffe, 2001; Kilic and Sohnesen, 2019; Troubat and Grünberger, 2017). We add to this literature by systematically comparing consumption and poverty estimates generated from a phone survey to those from an in-person survey. Finally, many researchers have hypothesized that the phone survey mode is likely to be considerably more vulnerable to response fatigue than the in-person mode, leading to the widespread recommendation to keep phone-based interviews short, and to avoid complex questions (Dabalen et al., 2016; Gourlay et al., 2021). Our results on consumption measurement provide empirical support to this hypothesis. However, in our case, both survey modes result in similar estimates when measuring diet-based food security suggesting that the phone survey mode is appropriate for measuring simpler and cognitively less demanding indicators, as long as the interview time is kept relatively short (Abay et al., 2021a).

## 2. The survey experiment, data and methods

### 2.1. The survey experiment

We designed a survey experiment to understand the implications of using a phone survey mode for household consumption measurement by systematically contrasting responses from computer assisted personal interviews (CAPI, or in-person) and computer assisted telephone interviews (CATI, or phone). The survey instrument in both survey modes were identical and had four sections. The interview began with a brief section containing only three questions needed to construct household size and its dependency ratio. In the first main section, respondents were asked to report on the household's food consumption for each item from a list of 118 food items, grouped into eight food groups. We first went through the list of 118 items asking whether the household consumed the item in the past seven days or not. The survey instrument was programmed to carry forward all items that were consumed in the past seven days to the next sub-section that asked about the consumption frequency ('on how many days was the item consumed') and quantity ('amount consumed') within the 7-day period. The second main section of the questionnaire included a short module asking household's food consumption outside of home within the same 7-day recall period. The final main section of the survey included a non-food consumption module, which asked respondents to recall household expenditures during the last month (e.g., toiletries or electricity expenditures) and during the last 12 months (e.g., school fees or health expenditures). The questionnaire administered for the two groups differed, then, only by the interview mode. For all other aspects, the questionnaire designs for the two groups were identical (Table 1). The full questionnaire is

<sup>3</sup> Although these guidelines were developed more than 20 years ago, they remain relevant and are still widely used to monitor global poverty (see Mancini and Vecchi, 2022).

<sup>4</sup> However, they do find a concurrent rise in some measures of food insecurity.

<sup>5</sup> This finding is in line with growing literature documenting non-classical measurement error in household surveys conducted in low- and middle-income countries (e.g., Abay, et al., 2019; Abay et al., 2021b; Carletto et al., 2013; Desiere and Jolliffe, 2018; Gibson et al., 2015; Gibson and Kim, 2010; Gourlay et al., 2019).

**Table 1**  
Comparison of in-person and phone data collection.

	In-person	Phone
<b>Method of data capture</b>	Computer-assisted personal interviewing (CAPI)	Computer-assisted telephone interviewing (CATI)
<b>Recall period in the food consumption modules</b>	7 days	7 days
<b>Recall period in the non-food consumption module (*)</b>	1 month or 12 months	1 month or 12 months
<b>Designated respondent</b>	Household member who decides on food purchase and/or preparation	Household member who decides on food purchase and/or preparation
<b>Consumption measurement</b>	118 food items (frequency and quantity consumed)	118 food items (frequency and quantity consumed)

Note: (\*) 1 month for non-food expenditures such as toiletries and utilities and 12 months for expenditures such as school fees and health expenses.

included in the Online Appendix.

## 2.2. Household sample

The household sample for this survey experiment originates from a randomized control trial (RCT) conducted to assess the impact of video-based behavioral change communication on fruit and vegetable consumption in Addis Ababa, Ethiopia (Abate et al., 2021). The baseline and endline surveys for the RCT took place in September 2019 and February 2020, respectively.<sup>6</sup> The sample of 930 households randomly selected from six sub-cities, 20 *woredas* (districts), and 40 *ketenas* (neighborhoods; or clusters of households) within Addis Ababa.<sup>7</sup> Comparison of household characteristics against those reported in other surveys from Addis Ababa suggest that the sample is representative of the households residing in the city (Hirvonen et al., 2020).

The endline survey was administered just before the COVID-19 pandemic was declared in 2020, a setup that was highly optimal for launching COVID-19 phone surveys. Phone numbers were collected from 887 households of the 895 households (99%) that took part in the February 2020 survey. To monitor the food security situation in Addis Ababa during the pandemic, we selected a random subsample of 600 households for monthly phone surveys (Hirvonen et al., 2021). In total, four phone survey rounds were carried out between June and August 2020. In the August 2020 phone survey round, we administered the same food consumption module described above for all households selected for the phone surveys (Hirvonen et al., 2021). Table A1 in the Appendix summarizes the various surveys with the sample of households used in this study.

The survey experiment contrasting consumption data collected via in-person and phone modes was administered over a 10-day period in September 2021 (i.e., one year after the last COVID-19 phone survey).<sup>8</sup> The sampling frame for this study was based on 895 households that were interviewed during the in-person survey conducted in February 2020, the endline survey of the video RCT. Out of the 895 households, 448 were randomly selected for an in-person interview and 447 for a

<sup>6</sup> The endline survey also included a survey experiment to quantify the degree of telescoping bias in recalled food consumption by experimentally varying the recall method, see Abate et al. (2020) for more details.

<sup>7</sup> Melesse et al. (2019) provide a detailed description of the sampling strategy.

<sup>8</sup> The exact dates were 31 August to 9 September 2021.

phone interview.<sup>9</sup> A total of 797 households were interviewed; 421 in the in-person group and 376 in the phone group.<sup>10</sup> Administering the consumption modules over the phone took 41 min on average (median) and while the average (median) interview duration was 43 min for an in-person visit. The quality of the connection was generally good for the phone interviews, and based on enumerators' assessment, rarely affected the interview quality.<sup>11</sup>

The survey team tasked with the in-person surveys followed recommended COVID-19 preventive measures when visiting the households. First, both the enumerators and respondents were provided with facemasks that they were required to wear during the interview. Second, the enumerators were required to thoroughly wash their hands with soap for 20 s or use disinfectant (containing more than 70% alcohol) before entering and when leaving the respondent's premises. Third, the survey coordinator conducted daily check-ups with enumerators regarding any COVID-19 related symptoms. Finally, the interview was conducted outdoors with at least 2-m distance between the enumerator and the respondent.

Ethical approval for the survey experiment was obtained from the institutional review boards (IRB) of the International Food Policy Research Institute (IFPRI) and the College of Medicine and Health Sciences at Hawassa University in Ethiopia. Informed oral consent was obtained from all participants at the start of the interview. Enumerators provided respondents a brief overview of the study objectives and informed them that their participation in the study was entirely voluntary.

## 2.3. Data

Food consumed at home was reported in terms of quantities consumed, which we converted into local currency units (Ethiopian birr) using retail price data collected by the Central Statistical Agency (CSA) of Ethiopia. We used the retail price data for Addis Ababa from February 2020 (the latest month available to us) and then used a food-specific consumer price index for Addis Ababa to express our food consumption data in September 2021 prices. Food consumption outside the home as well as non-food expenditure were collected in birr terms, thus requiring no price adjustments.

Each household's total consumption was calculated by first converting all consumption expenditure data to weekly terms and then adding up the three consumption components: food consumption at home; food consumption expenditures outside the home; and non-food expenditures. The official poverty data in Ethiopia come from the Household Consumption Expenditure Survey (HCES) collected every five years. The HCES survey is conducted throughout the Ethiopian calendar year to address consumption seasonality and covers nearly 400 food items and more than 850 non-food items. The latest HCES was administered in 2015/16, after which food prices and prices of non-food items have both been rising annually at a double-digit rate. Considering the high inflation rate and the considerable methodological differences between our survey and the HCES, we do not attempt to update the HCES poverty line for September 2021. Instead, we calibrate our poverty

<sup>9</sup> To ensure balance between the two groups, we block-randomized using the following variables: sex, age and education of the household head, household size, and an asset index. The data for these variables were collected in the previous in-person visits.

<sup>10</sup> Out of the 70 households in the phone survey group that were not interviewed, 16 did not answer the call, 37 had their phone switched off or not working, 10 had wrong numbers, and 5 had no phone numbers. Only 2 households refused to take part in the phone survey.

<sup>11</sup> At the end of each phone interview, we asked enumerators to rate the quality of the connection during the call. 74 percent of the phone interviews were rated as 'very good' ('we heard each other very well'), 19 percent as 'good', 5 percent as 'OK/average' and only 2 percent (5 interviews) as 'bad' or 'very bad'.

line for the in-person sample to match the 16.8 percent poverty headcount based on the national poverty line and reported for Addis Ababa using the 2015/16 HCES (FDRE, 2018).

We also use our food consumption data to study how the phone survey mode affects household dietary diversity, an indicator of household food security (Hoddinott and Yohannes, 2002). First, we computed the HDDS of Swindale and Bilinsky (2006) by grouping the 118 food items in our consumption module into 12 food groups: cereals; roots and tubers; vegetables; fruits; meat, poultry and offal; eggs; fish and seafood; pulses, legumes and nuts; milk and milk products; oil and fats; sugar and honey; and miscellaneous foods. The HDDS is a sum of all food groups from which the household consumed food items during the 7-day recall period, with a minimum of one and maximum of 12. Second, we constructed the food consumption score (FCS) developed by the WFP (2008). The FCS combines dietary diversity and consumption frequency by grouping the consumed food items into nine groups and allocating more weight to protein rich foods.<sup>12</sup> The weighted FCS index ranges between zero and 112, with higher scores indicating a better food security situation.

After dropping two households with implausible consumption values, the final sample of 795 households is formed, out of which 421 are from the in-person group and 374 are from the phone group. Table 2 shows that the in-person and phone groups are similar in terms of basic household characteristics. Moreover, the households in the two subsamples are balanced in terms of the number of times they had been interviewed since September 2019. We also see no meaningful differences in the household per capita food consumption collected in September 2019, whether we examine means (Table 2) or full distributions (Figure A1 in the Appendix).

**Table 2**  
Household characteristics, by survey mode.

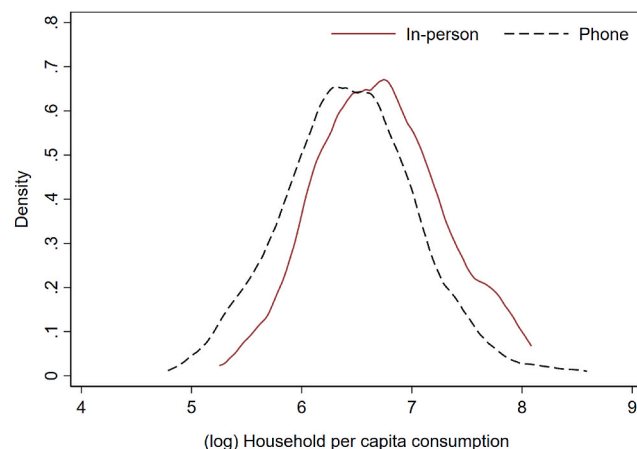
Variable	In person	Phone	Difference	t-test
	Mean/ [SE]	Mean/ [SE]		p-value
Female respondent	0.922 [0.017]	0.917 [0.018]	0.005	0.843
Household size	4.800 [0.110]	4.832 [0.092]	-0.032	0.792
Male-headed household (*)	0.568 [0.029]	0.572 [0.036]	-0.004	0.898
Head's education in years (*)	6.675 [0.297]	6.543 [0.310]	0.132	0.655
Household asset index (*)	-0.035 [0.124]	-0.009 [0.161]	-0.026	0.828
Number of times the household has been interviewed since September 2019	5.684 [0.086]	5.805 [0.082]	-0.121	0.315
(log) Household per capita food consumption in September 2019 (*)	5.570 [0.037]	5.534 [0.042]	0.036	0.416
Number of households:	421	374		
Clusters:	40			

Note: Unit of observation is household. Standard errors (SE) are clustered at enumeration area level. Difference in means between the groups tested with a t-test (null-hypothesis: difference in means = 0).

Note: N = 795 households.

(\*) Based on data collected in previous survey rounds.

<sup>12</sup> The FCS food groups are: main staples (weight: 2); pulses (3); vegetables (1); fruits (1); meat, eggs, fish (4); dairy products (4); sugar (0.5); oil/butter (0.5); and condiments (0).



Note: N=795 households

**Fig. 1.** Distribution of (ln) weekly consumption per capita (in birr), by survey mode.

### 2.4. Estimation methods

We quantify the difference in reported household per capita consumption values across the two groups using ordinary least squares (OLS). In the most basic model, we regress both the per capita consumption value and its logarithm on a binary treatment variable valued one if the household was randomly selected into the phone group, and zero if into the in-person group. In subsequent models, we control for differences in basic household characteristics (household size, and household head's sex and level of education in years) as well as sub-city fixed effects. Finally, when we discuss percentage differences derived from the coefficients in semi-log regressions they are based on the approximate unbiased variance estimator proposed by van Garderen and Shah (2002):  $100 \times (e^{\hat{\beta} - 0.5\hat{V}(\hat{\beta})} - 1)$ , where  $\hat{\beta}$  refers to the estimated coefficient and  $\hat{V}$  to the estimated variance. Finally, the standard errors in all household level regressions are clustered at the enumeration area (*ketena*) level.

## 3. Results

### 3.1. Household total per capita consumption

Fig. 1 contrasts the full distributions of (log) household weekly per capita consumption measured in birr between households that received an in-person visit and households that were interviewed over the phone. The estimated household consumption distribution for the phone group lies to the left of the distribution estimated for the in-person group, indicating that the whole distribution of total consumption values resulting from the phone survey resulted in lower values than that of the in-person survey.

The regression estimates reported in Table 3 quantify the difference in household weekly food consumption when the data were collected over the phone relative to when the in-person survey mode was used. In columns 1 and 2, the dependent variable is the natural logarithm (ln) of household per capita consumption value in birr, whereas non-logged values are used in columns 3 and 4. Unadjusted estimates are reported in odd columns, whereas estimates in even columns are adjusted for differences in basic household characteristics as described above. Because the differences between the unadjusted and adjusted regressions are negligible, we focus our reporting and discussion on the adjusted regression results.

**Table 3**  
Impact of phone survey mode on household weekly per capita consumption.

	(1)	(2)	(3)	(4)
Dependent variable:	(ln) Household per capita consumption (birr)		Household per capita consumption (birr)	
Phone survey mode	-0.271*** (0.059)	-0.262*** (0.054)	-207.69*** (58.16)	-200.61*** (52.65)
Household level controls?	No	Yes	No	Yes
Sub-city fixed effects?	No	Yes	No	Yes
Observations	795	795	795	795
R <sup>2</sup>	0.051	0.288	0.031	0.232
In-person group mean of the dependent variable	n/a	n/a	966.27	966.27

Note: Ordinary least squares regression. Unit of observation is household. Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with \*  $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Relative to the in-person survey, on average the phone survey mode decreases the reported household per capita consumption expenditures by 23 percent (Table 3, column 2).<sup>13</sup> The 95% confidence interval (CI) for this estimate ranges between -14.2 and -31.1. The estimates based on non-logged per capita consumption variable are similar. Considering that the mean per capita consumption in the in-person group is 966 birr, the 201 birr difference reported in Column 4 of Table 3 translates into 21 percent lower average per capita consumption in the phone survey group.

### 3.2. Components of consumption

Food consumed at home represents 50.3 percent of the total consumption among the in-person group and 55.8 percent among the phone survey group.<sup>14</sup> The regression estimates reported in Column 1 of Table 4 indicate that the reported per capita food consumption values are 13 percent lower on average when the phone survey mode is used (95%-CI: -5.5; -20.7). However, we do not find strong evidence to suggest that some food groups were more affected than others. We re-estimated the main regression using the value of food consumption for each of seven categories of food as the dependent variable; in Figure A2 in the Appendix, we observe that all the coefficient estimates are negative and suggest 5 to 25 percent lower consumption, with overlapping confidence intervals.

About 60 percent of the households in our sample report to have consumed food items outside of their home in the past 7 days. This reporting incidence varies by survey mode with households in the phone survey group being 13 percentage points less likely to report to have consumed foods outside their home (Table 4, column 2). A regression based on a non-logged outcome variable shows that the food expenditures outside of the home are 40.2 percent lower in the phone group relative to the in-person group (Table 4, column 3).<sup>15</sup>

All the households in our sample report positive (non-zero) non-food consumption values. Column 4 in Table 4 shows the impact of the phone survey mode when the dependent variable is logged weekly per capita non-food consumption. On average, the phone survey mode lowers the

<sup>13</sup> Recall that we use the equation reported at the end of Section 2 to interpret the coefficients in semi-log regressions. As a result, the numbers reported in the text will differ slightly from the commonly used interpretation of  $100 * \hat{\beta}$ , where  $\hat{\beta}$  is the coefficient estimate reported in the regression tables.

<sup>14</sup> The difference is statistically significant ( $p = 0.003$ ).

<sup>15</sup> Considering that the mean value in the in-person group is 53.91 birr, the difference of 21.66 birr estimated with OLS translates to 40.2 percent (21.66/53.91).

**Table 4**  
Impact of phone survey mode on components of household consumption.

	(1)	(2)	(3)	(4)
Dependent variable:	(ln) Household per capita food consumption at home	Household consumed food outside home (0/1)	Household per capita food consumption outside home	(ln) Household per capita non-food consumption
Phone survey mode	-0.143*** (0.043)	-0.129** (0.056)	-21.66** (8.34)	-0.35*** (0.09)
Household level controls?	Yes	Yes	Yes	Yes
Sub-city fixed effects?	Yes	Yes	Yes	Yes
Observations	795	795	795	795
R <sup>2</sup>	0.221	0.079	0.062	0.226
In-person group mean of the dependent variable	n/a	0.660	53.92	n/a

Note: Ordinary least squares regression. Unit of observation is household. 0/1 = binary variable. Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with \*  $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

reported non-food consumption by 30.1 percent (95%-CI: -15.5; -42.1).

### 3.3. Poverty estimates

Next, we estimate the impact of using phone survey mode on poverty estimates. Since poverty is defined at the individual level, we need to convert our data from household to individual level. To do so, we use a weighted least squares regression method where the weights are frequency weights based on household size. Using our calibrated poverty line, in Table 5 we estimate that poverty rate is 17 percentage points higher when phone survey mode is used compared to when consumption data are collected through in-person visits (95%-CI: 9.99; 24.1). Since the poverty rate in the in-person sample is calibrated at 16.8 percent, using the phone survey mode effectively doubles the poverty rate in this context.

**Table 5**  
Impact of phone survey mode on poverty rate.

	(1)	(2)
Dependent Variable:	Consumption Below Poverty Line (0/1)	
Phone survey mode	0.168*** (0.036)	0.170*** (0.035)
Household level controls?	No	Yes
Sub-city fixed effects?	No	Yes
Observations (weighted)	3828	3828
Households	795	795
R <sup>2</sup>	0.038	0.181
In-person group mean of the dependent variable	0.168	0.168

Note: Weighted least square regression with household size used as a frequency weight. After applying the weight, the unit of observation is individual. Dependent variable obtains value 1 if the household's per capita consumption is below the poverty line, zero otherwise. 0/1 = binary variable. Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with \*  $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**Table 6**  
Impact of phone survey mode on household dietary diversity indicators.

	(1)	(2)	(3)	(4)
Dependent variable:	Household diet diversity score (HDDS)		Food consumption score (FCS)	
Phone survey mode	0.060 (0.132)	0.058 (0.135)	-2.120 (1.629)	-2.055 (1.646)
Household level controls?	No	Yes	No	Yes
Sub-city fixed effects?	No	Yes	No	Yes
Observations	795	795	795	795
R <sup>2</sup>	0.000	0.121	0.003	0.111
In-person group mean of the dependent variable	9.07	9.07	63.97	63.97

Note: Ordinary least squares regression. Unit of observation is household. Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 3.4. Measures of food security

In Table 6, we report the impacts of using the phone survey mode on two widely used diet-based food security measures, HDDS and FCS. Both can be computed from the food consumption survey data. All four reported impact estimates are relatively small in magnitude and not statistically different from zero. The HDDS and FCS do not require respondents to estimate quantities consumed, only whether the food item was consumed in the past 7 days (HDDS) or the consumption frequency in terms of number of days in the past 7 days (FCS). In contrast, collecting data for food consumption measures is cognitively more demanding because it requires respondents to also estimate quantities consumed in the household during the recall period. Our results therefore indicate that the phone survey mode appears to lead to similar estimates when measuring diet-based food security to in-person surveys but leads to much lower estimates of the value of household food or non-food consumption.

## 4. Mechanisms, extensions, and robustness

### 4.1. Survey fatigue

Our survey experiment shows that the phone survey mode leads households to underestimate their food and non-food consumption expenditures. As a result, if we trusted the phone survey mode and tried to use it in the same manner that we had used in-person surveys to measure poverty prior to the pandemic, we would conclude that the poverty headcount is twice as high using the phone survey data than the data collected in-person. Here, we study whether survey fatigue can help explain differences between results of the two survey modes.

The large difference in consumption and poverty incidence estimates between the two survey modes could result from respondent or enumerator fatigue. For example, fatigued respondents pay less attention when responding to cognitively demanding questions (e.g., amount or value of consumption), increasing the risk of measurement error. Survey experts have hypothesized that the risk of respondent fatigue is considerably higher in phone surveys than in in-person surveys (Dabalén et al., 2016; Gourlay et al., 2021). Consequently, it has been widely recommended to keep the phone survey duration short to minimize the risk of survey fatigue (Glazerman et al., 2020; Hoogeveen et al., 2014; Hughes and Velyvis, 2020; Jones and von Engelhardt, 2020; Kopper and Sautmann, 2020). While it is certainly intuitive that the risk of survey fatigue is higher in phone surveys, to the best of our knowledge, no studies have attempted to compare survey fatigue between phone and in-person modes using the same survey form.

Evidence from in-person surveys suggests that survey fatigue can

lead to under reporting and overall deterioration of data quality in some settings (Ambler et al., 2021; Baird et al., 2008; Schündeln, 2018), but not always (Laajaj and Macours, 2021).<sup>16</sup> In a recent phone survey conducted in rural Ethiopia, Abay et al. (2021a) estimate that delaying the timing of a dietary diversity module by 15 min increased the likelihood that the respondents reported not to have consumed from certain food groups, resulting in an 8 percent decline in the mothers' dietary diversity score.<sup>17</sup>

To explore the role of survey fatigue, we cross-randomized the order in which the food groups appeared in the first main section of the survey, the "food consumed at home" module.<sup>18</sup> Specifically, we implemented two versions of this food consumption module, ordering the food groups differently (see Appendix Table A2). For example, in version 1, mango appeared as the 5th item while in version 2, it appeared as the 73rd item. Similarly, in version 1, rice was the 52nd item on the list while in version 2, it was the 11th item on the list. Exploiting this variation, we use the food item level data to construct a variable that takes on the value of 1 when each food appears later in the questionnaire relative to the other version, and 0 otherwise.<sup>19</sup> Using the example above, this variable would be 1 when mangoes appear as the 73rd item, and when rice appears as the 52nd item. Using our food item level data, we then regressed the weekly household per capita consumption of the food item on this binary variable capturing the item's relative position in the questionnaire, and the indicator variable for the phone survey mode. To assess whether the impact of delaying when the item is asked in the module differs between phone and in-person survey modes, we interact the two variables and include the interaction term in the regression. In these regressions we control for food item fixed effects, meaning that our estimates are identified from variation in the survey mode or relative position in the questionnaire for the same food items. As additional controls, we include household size, an indicator variable for male-headed households, the head's years of education, and sub-city fixed effects.

Table 7 provides the results. In column 1, we estimate the model without the interaction term. Moving the item later in the questionnaire results in a report that is, on average, 5.8 percent lower for the item than if it takes on its earlier position.<sup>20</sup> Using the phone survey mode, the average report suggest the value of consumption is 15.5 percent lower than found with the in-person survey mode. In column 2, we estimate the model with the interaction term. The basic variable now captures the effect of placing the item later in the questionnaire in the in-person survey; this coefficient is close to zero and not statistically significant. The CI is relatively tight around zero (95-% CI: -0.0167; 0.0016) indicating that survey fatigue does not play a role in the in-person survey

<sup>16</sup> Evidence from survey experiments conducted in high-income countries have documented respondent fatigue in phone survey mode (e.g., Eckman et al., 2014), Roberts, et al. (2010).

<sup>17</sup> Garlick et al. (2020) randomly assigned small firms to weekly phone and in-person surveys finding that phone survey respondents systematically under-reported labor supply, stock, and inventory relative to in-person respondents. However, the authors did not explicitly test whether these differences could be driven by survey fatigue.

<sup>18</sup> Laajaj and Macours (2021), Ambler, et al. (2021), and Abay et al. (2021a) also randomize the order in which questions are asked in their surveys to study survey fatigue.

<sup>19</sup> As can be seen from Appendix Table A2, we administered two different versions of the food consumption module by simply changing the ordering of the food groups. As a result, we do not have sufficient variation in our data to test this with a 'distance variable' that captures the number of items between the version 1 and version 2.

<sup>20</sup> The calculations in this paragraph are as follows: 5.8 percent lower is calculated as  $-0.230/3.97$  and 15.5 percent lower is calculated as  $-0.615/3.97$ , using the estimates reported in Table 7, column 1, and 11.9 percent lower is calculated as  $[-0.014+(-0.458)]/3.97$ , using the estimates reported in Table 7, column 2.

**Table 7**

Impact of item's relative position in the questionnaire and phone survey mode on reported per capita food consumption value measured in birr.

	(1)	(2)
Dependent Variable:	(ln) Household per capita consumption of the food item	
Item appeared later in the questionnaire	-0.230** (0.101)	-0.014 (0.159)
Phone survey mode	-0.615*** (0.203)	-0.368 (0.239)
Item appeared later in the questionnaire * Phone survey mode		-0.458** (0.222)
Household level controls?	Yes	Yes
Sub-city fixed effects?	Yes	Yes
Food item fixed effects?	Yes	Yes
Observations	93,810	93,810
In-person group mean of the dependent variable	3.97	3.97

Note: Ordinary least squares regression. Unit of observation is food item consumed (or not) in each household. Number of food items is 118 and number of households is 795 resulting in 93,810 observations. Dependent variable is household per capita consumption of the food item measured in birr. Standard errors are clustered at the food item level and reported in parentheses. Statistical significance denoted with \*  $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

mode, at least in this relatively early part of the questionnaire. In contrast, the coefficient on the interacted variable is negative, relatively large in magnitude, and statistically different from zero; it suggests that delaying an item in the phone survey mode leads to a report that is 11.9 percent lower on average than an item occurring later in the in-person survey. This finding is strongly suggestive that the in-person mode leads to less survey fatigue than the phone survey mode.

In Appendix Table A3 we replicate this analysis, only considering the responses to the Yes/No questions regarding whether the household consumed the item or not during the 7-day period. Interestingly, all coefficients in the interacted model appear insignificant implying that only consumption quantity reports are affected, but not responses on whether the household consumed the item or not. This finding is in line with our earlier result according to which diet-based food security measures do not seem to be affected by variation in survey mode.

#### 4.2. Data quality

We next use Benford's law as a benchmark for assessing data quality. According to Benford (1938), the distribution of first-digits in many numerical data sets approximately follow the probability (P):

$$P(d) = \log_{10}(d+1) - \log_{10}(d)$$

where  $d \in \{1, \dots, 9\}$  refers to the first-digit of the observation.

It is unlikely that survey data perfectly conform to the Benford's law distribution (Kaiser, 2019), but previous work (Abate et al., 2020; Garlick et al., 2020; Schündeln, 2018) has used the distance between the observed distribution and the predicted distribution under Benford's law as a measure of data quality. Here, we calculate this distance separately for the data collected by phone and for the data collected by in-person visits. Following Schündeln (2018), we compute normalized Euclidean distances between the observed first-digit distribution and the one predicted by Benford's law.<sup>21</sup>

We use the digits of the quantities consumed as reported by the households in the food consumption module. The specific question asks for the quantity consumed and the unit (e.g., kg, litre, cup, or a locally

<sup>21</sup> The Euclidean distance is calculated as the square root of the sum of squared differences between the observed percentage and the percentage predicted by the Benford's law. We further normalize the calculated distances by taking a Z-score: subtracting the mean distance and dividing this by the standard deviation calculated using the pooled data.

**Table 8**

Regression results from interaction models.

	(1)	(2)	(3)	(4)
Dependent variable:	(ln) Household food consumption per capita		(ln) Household non-food consumption per capita	
Phone survey mode	-0.223*** (0.060)	0.073 (0.117)	-0.427*** (0.143)	-0.224 (0.183)
Phone survey mode * Head's education in years	0.015** (0.007)		0.011 (0.014)	
Phone survey mode * Household size		-0.041* (0.022)		-0.027 (0.027)
Household level controls?	Yes	Yes	Yes	Yes
Sub-city fixed effects?	Yes	Yes	Yes	Yes
Observations	795	795	795	795
R <sup>2</sup>	0.595	0.595	0.227	0.227

Note: Ordinary least squares regression. Unit of observation is household. Household level controls include household size (number of members), indicator variable for male-headed households, and head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with \*  $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

used unit such as *tassa*). Of note is that Benford's law is scale-invariant; the law holds irrespective of the unit in which the consumed quantities were reported.

Figure A3 in the Appendix reports the observed first-digit distributions in our data and compares them to the distribution predicted by Benford's law.<sup>22</sup> The null hypothesis that the observed distributions follow Benford's law is rejected for both groups ( $p < 0.001$ ). However, relative to the in-person group, the phone group is much more likely to report the smallest possible value (i.e., value 1) as the first digit, possibly indicating limited cognitive engagement with the question.

Next, we calculate the Euclidean distances separately for each of the 33 consumption units reported by the households and for both survey mode groups. We then test whether the consumption unit specific average Euclidean distances for the two groups are statistically different by regressing the mean distance on our binary treatment variable. Table A4 in the Appendix shows that the coefficient on the treatment variable is positive and statistically different from zero, indicating that the data collected via the phone survey deviate more from the Benford's law than data collected via the in-person survey. This finding suggests that the consumption data from the in-person survey are of higher quality than data from the phone survey.

#### 4.3. Heterogeneity

The results show that using the phone survey mode leads to substantial underestimation of household consumption expenditures. It is tempting to think that it could be possible to devise relatively simple adjustment factors to correct for this attenuation bias. Unfortunately, evidence from previous survey experiments suggests that because the measurement error is usually not independent of household characteristics (i.e., non-classical), such adjustment factors do not exist (De Weerd et al., 2020). To explore the possibility that the phone survey mode varies by household type, we interacted the phone survey indicator variable with the household head's level of education and household size. Table 8 provides the results when household per capita food consumption (Columns 1–2) and non-food consumption (Columns 3–4) is used as the dependent variable. For household food consumption, we observe that the bias decreases with household head's education and increases with household size.<sup>23</sup> The former result suggests that

<sup>22</sup> We calculated these distributions using a user-written Stata routine devised by Jann (2007).

<sup>23</sup> Table 2 reports that the difference in household size between the two household groups is not statistically different from zero.



respondents from more educated households better overcome survey fatigue in phone surveys. In contrast, the cognitive burden increases with household size as the number of consumption events is higher within the recall period (Fiedler and Mwangi, 2016; Gibson and Kim, 2007). Larger households are bound to have more consumption events than smaller households, making them more vulnerable to survey fatigue. For non-food consumption, the coefficients are of the same sign and magnitude but not statistically different from zero, possibly because of the larger variation in the data relative to the food consumption data. Overall, these heterogeneous impacts imply that adjustment factors to account for the bias caused by the phone survey mode cannot be easily developed.

#### 4.4. Enumerator effects

The survey team of 21 enumerators were all trained together and supervised by the same survey coordinator. To simplify survey logistics, the enumerators were tasked with conducting either phone interviews or in-person interviews. This collinearity between enumerator assignment and survey mode raises a concern that the estimated survey mode effects could be completely driven by enumerator effects.<sup>24</sup> To address this concern, we conduct three robustness checks. First, we show that our main findings are robust to controlling for enumerator characteristics: age, level of education, and past survey experience (see Column 2 in Table A5 in the Appendix). Second, to explore whether one poorly performing enumerator in the phone survey group could explain our results, we assess the sensitivity of our result to omitting one enumerator at a time from the sample. Results are remarkably robust to running the main regression across these 21 sub-samples (see Figure A4 in the Appendix). Third, we show that our results are robust to the controlling for enumerator random effects (Table A5, column 3 in the Appendix) as well as Mundlak (1978) correlated random effects (Table A5, column 4 in the Appendix).<sup>25</sup> Though we cannot use enumerator fixed effects, the combination of this evidence suggests that we can conclude enumerator effects could not have had much influence on the difference between in-person and phone survey results.

#### 4.5. Cost considerations

Compared to in-person surveys, phone surveys are typically considerably less costly to administer (Gourlay et al., 2021). In this case, the cost per interview was approximately 3 times lower for phone surveys than in-person surveys. The cost differences are mainly due to survey logistical costs (which are marginal for the phone survey but represent about a third of the total cost of the in-person survey) and survey personnel costs due to differences in the number of interviews per day. While there was not much difference in the time phone and in-person surveys took, phone enumerators were able to conduct about three times as many interviews in a day than in-person enumerators because

<sup>24</sup> Previous work in this area has found that the enumerator effects play a negligible role in shaping survey responses, unless the questions are sensitive in nature (Di Maio and Fiala, 2020).

<sup>25</sup> The random effects estimator controls for enumerator heterogeneity by decomposing the unobserved heterogeneity to variance occurring between enumerators and within enumerators (i.e., across different interviews conducted by the same enumerator). The key assumption of the random effect estimator is that the correlation between the treatment status and the random effects is zero, or in the correlated random effects model, that it takes on a specific parameter. We acknowledge that, in our application, this assumption may not hold. However, simulation studies suggest that the 'heterogeneity bias' stemming from the violation of this assumption is relatively small (see Bell and Jones, 2015). Considering this point and the fact that the estimated coefficient based on the random effects estimator is very close to the coefficient reported in column 2 of Table 3, we believe that unobserved enumerator effects are not driving our results.

the survey mode allows them to make the next call as soon as they were ready, while the in-person survey requires enumerators to travel to the next household. However, there are a few ways that the in-person costs were minimized in this urban context. For instance, travel costs were relatively low, as enumerators could travel to the neighborhoods on their own, so vehicle rental was limited to supervisory vehicles. Had households been more spread out (e.g., in a rural survey), the cost difference would have been much larger.

The cost difference suggests that with the same resources, using phone surveys would allow for a sample size roughly three times larger than in-person surveys, in the same type of urban setting. Increasing the sample size that much implies a sizable gain in statistical power and thus improvement in the precision of consumption and poverty estimates.<sup>26</sup> However, as we have shown above, the phone survey mode comes with a systematic downward bias. Consequently, survey experts interested in measuring household consumption using the standard method face a trade-off between precision and accuracy when deciding between in-person and phone survey mode. In our view, the bias introduced by the phone survey mode in this context is too large to be ignored over potential gains in precision. If poverty incidence is to be measured with phone surveys, different methods of doing so consistent with current methods of poverty estimation are necessary.

## 5. Conclusions

Pre-pandemic, development economists and practitioners were using phone surveys in only a few contexts. In research, they were used when projects required high-frequency data or in contexts that were difficult to reach (Dabalén et al., 2016; Dillon, 2012; Hoogveen et al., 2014). Meanwhile, WFP (2017) was building up knowledge about how to use phone surveys to monitor food insecurity. As the pandemic began, phone surveys suddenly became the only option for many types of data collection, and research on living standards and food insecurity shifted rapidly to phone surveys, to understand the socioeconomic implications of the pandemic.

The subsequent COVID-19 phone surveys have provided important information about the socioeconomic consequences of the pandemic in many low- and middle-income countries with limited infrastructure to provide real-time economic or employment data to inform policy decisions. However, the economic information collected at the household level has been largely restricted to subjective indicators measuring income or employment losses, offering limited information about the severity or depth of the crisis (De Weerd, 2008; Hirvonen et al., 2021).<sup>27</sup> Indeed, there have been only few attempts to measure household consumption to inform how the progress toward meeting the first Sustainable Development Goal of 'No Poverty' has been affected by the pandemic. Finally, there remains considerable uncertainty on the implications on the use of the phone survey mode on data quality, particularly in low- and middle-income country contexts where the pre-pandemic roll out of phone survey technology and testing had been relatively slow (Gourlay et al., 2021).

Our research begins to address some of these important methodological knowledge gaps. To measure the extent of bias on household consumption measures in phone surveys, we conducted a survey

<sup>26</sup> There is another channel through which phone surveys can be more efficient than in-person surveys. In-person surveys typically require cluster sampling to simplify logistics and reduce potentially sizable transportation costs (particularly in rural areas). As the same logistical concerns are absent in phone surveys, they permit applying a simple random sampling through random direct dial techniques that is more efficient than cluster sampling.

<sup>27</sup> At the same time, with imperfect and non-random mobile phone access in rural areas, the data may not be representative as the poor and people in more remote areas may have less access to phones or be outside of coverage areas when phone surveys are fielded (Ambel et al., 2021; Brubaker et al., 2021).

experiment in Addis Ababa, Ethiopia, randomly assigning a balanced and representative sample either to a phone or an in-person interview mode. We find the phone survey mode leads to a statistically significant and large underestimation of household consumption. Relative to the in-person survey mode, the phone survey mode decreases the reported household per capita consumption expenditures by 23 percent, on average. Consequently, the estimated poverty rate is twice as high when the phone survey mode is used.

We therefore should reinterpret results in Hirvonen et al. (2021), which used the same household sample to show that the total value of food consumption expenditures had not changed much between August–September 2019 and August 2020. The former survey was collected in-person, and the latter by phone; if we use the results here to re-interpret that paper, it seems that if anything the average value of food consumption rose by August 2020; moreover, that paper shows that the value of relatively nutritious foods might have declined; that concern is far lower given those results likely underestimate all categories of food consumption.

The mechanism appears to be linked to survey fatigue that results in phone survey respondents to greatly under-estimate consumption quantities, but not whether they consumed the item during the recall period. Our heterogeneity analysis suggests the bias increases when more people eat within the household, possibly because of the increased cognitive burden in remembering larger number of consumption events. In contrast, the bias is attenuated by education, suggesting that more educated individuals can overcome issues of attention.

Our study has some important limitations. First, our sample is not nationally representative and importantly does not cover rural households that are typically poorer and consume fewer food and non-food consumption items. Consumption surveys in rural areas could take less time to complete than in urban areas, making the phone survey mode more feasible.<sup>28</sup> Another external validity concern relates to the fact that the household sample used in this study had responded to two or three food consumption surveys prior to this survey experiment (see Table A1 in the Appendix). Consequently, the household in our sample may have become more attuned to recalling consumption events than a new, randomly selected sample of households. Finally, while we hypothesize that the documented survey fatigue is driven by respondents, the design of our experiment does not allow us to distinguish whether the fatigue is driven by fatigue among respondents or fatigue among enumerators.

These limitations aside, our findings suggest that while phone surveys can provide large cost savings, they cannot replace in-person surveys for *standard* household consumption and poverty measurement, as outlined in Deaton and Grosh (2000). However, the phone survey mode does appear to be useful for monitoring diet-based food security indicators that do not require information about the quantities consumed, as used by the WFP (2017) in their Vulnerability Analysis and Mapping surveys.

Given the prevalence of cell phone ownership, figuring out how to use phone survey data to best contribute to accurate consumption and poverty measurement in low- and middle-income countries forms an important future research agenda. One option is to substantially shorten the consumption modules to accommodate the greater risk of survey fatigue in phone surveys. However, the available evidence from low- and middle-income country contexts suggest that shorter modules systematically underestimate consumption levels and thus overestimate poverty headcounts (Beegle et al., 2012; Jolliffe, 2001; Pradhan, 2009). Therefore, when adjusting the consumption module length, survey

<sup>28</sup> However, a limited and unequal access to phones can be a major obstacle to administering representative phone surveys in rural areas. For example, in Ethiopia, only 40 percent of rural households have access to a phone, and those that have, tend to be more educated and wealthier (Wieser et al., 2020). Furthermore, rural households tend to be larger than urban households, potentially exacerbating bias related to household size.

practitioners need to balance between accuracy and survey fatigue. Finding a balance in which accuracy is maximized and risk of survey fatigue minimized in phone surveys constitutes an important task for future survey methodology research.<sup>29</sup>

Another option is to rely on cross-survey imputation methods. In recent years, these methods have become popular among poverty economists to estimate poverty in contexts and periods lacking consumption survey data (e.g., Dang, et al., 2021; Doudich et al., 2016; Stifel and Christiaensen, 2007). These types of imputation methods typically begin by using a household consumption survey and by regressing household consumption expenditures on a set of household characteristics, such as household demographics, employment status, and asset and education levels. Then another survey that collected data on the same characteristics is used, as the estimated model parameters can be applied to these household characteristics to predict household consumption expenditures and poverty rates. Phone surveys could be used to (relatively inexpensively) collect data on these household characteristics, link these data to a previous household consumption expenditure survey, and estimate poverty using cross survey imputation methods. However, the validity of this approach rests on some important assumptions. First, the relationship between household consumption expenditures and its predictors should remain stable over time (Christiaensen et al., 2012). Considering relative price changes occurring as a consequence of the COVID pandemic and the conflict between Russia and Ukraine, it remains an open question about where and when this assumption would hold. Second, linking parameters estimated from in-person consumption survey to household characteristics obtained from a phone survey assumes that survey mode effects do not matter (Kilic and Sohnesen, 2019). Considering the evidence presented here and other emerging work testing survey mode effects (e.g., Garlick, et al., 2020), this assumption is clearly a strong assumption requiring further validation. Third, one must always be cognizant that phone ownership is correlated with income, and lower income people with phones may be less likely to keep them turned on (and therefore answer calls), to preserve their batteries.

Finally, it would be useful to experiment with split questionnaire designs in a phone survey setup. In this method, respondents are randomly assigned fractions of the full questionnaire and the missing data are then imputed using multiple imputation techniques (Raghu-nathan and Grizzle, 1995). Recent applications of a split questionnaire design with in-person surveys suggest that the approach can produce reliable consumption and poverty estimates with considerably shorter interview durations (Pape, 2021; Pape and Mistiaen, 2015). It remains an open question about whether split designs could be used to generate low bias estimates of poverty incidence with phone surveys.

#### Credit author statement

Gashaw Abate: Conceptualization, Formal Analysis, Methodology, Investigation, Writing and Editing; Alan de Brauw: Conceptualization, Methodology, Writing and Editing, Funding Acquisition; Kalle Hirvonen: Conceptualization, Formal Analysis, Methodology, Writing and Editing; Abdulazize Wolle: Investigation, Formal Analysis, Data Curation.

#### Data availability

Part of the data are publicly available; the remainder are in process of being made available. Code will be made available once review process is complete.

<sup>29</sup> It is important to note, however, that such major adjustments to survey design compromise the comparability to earlier consumption and poverty statistics that were based on different methodologies.

## Acknowledgements

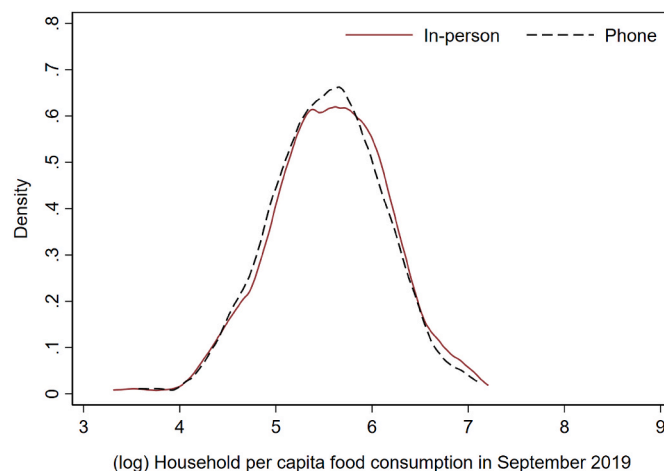
We thank Kibrom Abay, Joachim De Weerd, two anonymous reviewers, and conference participants in the Methods and Measurement Conference organized by the IPA-Northwestern Research Methods Initiative and seminar participants at the Helsinki GSE for useful comments. We are grateful to the households that participated in this study. We acknowledge NEED NUTRITIONAL and their survey staff (Abraha Weldegerima, Alemayehu Deme, Nadiya Kemal, Fikirte Sinkineh, Meskerem Abera, Rediet Dadi, Fitsum Aregawi, Asnakech Yosef, Ekiram

Shimelis, Biruktawit Abebe, Yared Tilahun, Tesfaye Eana, Yosef Regasa, Ashenafi Hailemariam, Selamawit Genene, Shambel Asefa, Teshale Hirpesa, Getachew Buko, Habtamu Ayele, Ayinalem Reta, Mohamed Meka, Kibrom Tadesse, and Huluhager Endashaw) for collecting the data. Any remaining errors are the sole responsibility of the authors. This work was undertaken as part of, and funded by, the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH). The opinions expressed here belong to the authors, and do not necessarily reflect those of A4NH or CGIAR. This study is registered in the AEA RCT Registry with a unique identifying number AEARCTR-0008097.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jdeveco.2022.103026>.

## Appendix



**Fig. A1.** Distribution of (ln) weekly food consumption per capita (in birr) in September-2019, by survey mode in August-2021. Note: N = 795 households.

**Table A1**

Surveys administered to the household sample used in this study

Survey	Date	N	Relevant questionnaire features
Baseline survey (in-person)	September 2019	900	Food consumption module + video screening
Endline survey (in-person)	February 2020	900	Food consumption module + bounded recall experiment
Phone surveys	May, June, and July 2020	600	Food security modules
Phone survey	August 2020	600	Food consumption module
Phone & in-person survey	August 2021	800	Food and non-food consumption modules

**Table A2**

Order of the food groups in the two versions of the 'food consumed at home' module

Food group	Order in version 1	Order in version 2
Fruits	1	6
Vegetables	2	7
Cereals	3	1
Pulses	4	2
Meat and fish	5	3
Eggs and dairy	6	4
Oils and butter	7	5
Spices and beverages	8	8

Note: Both phone and in-person surveys included two types of food consumption modules with different order in which the food groups appeared in the questionnaire. This table shows the order of food groups in both questionnaire types.

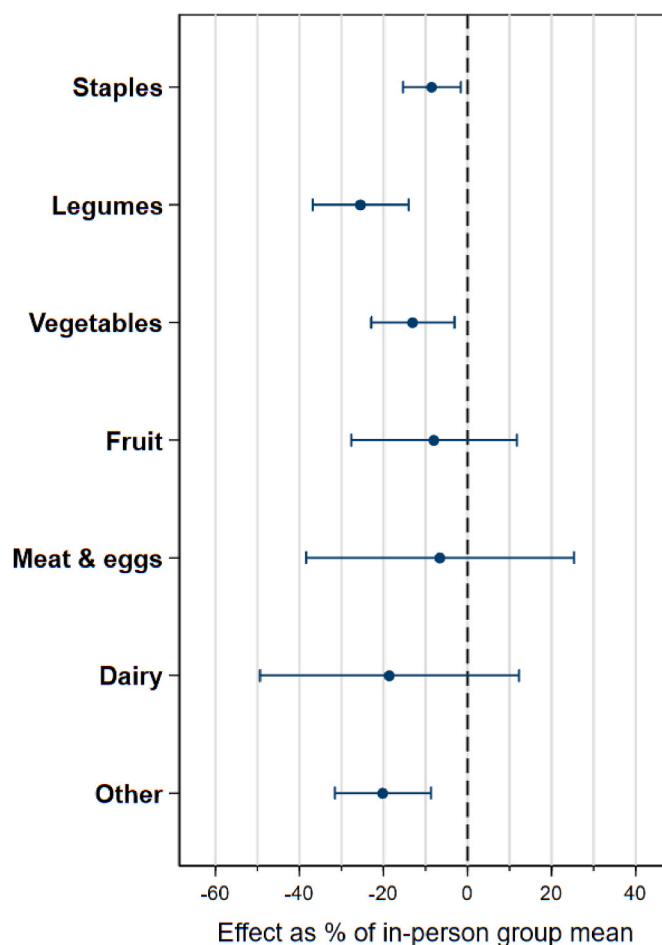


Fig. A2. Impact of phone survey mode on household consumption of different food groups.

Note: Based on ordinary least squares regression. Unit of observation is household; N = 795. All regressions included household level controls (household size, indicator variable for male-headed households, and head’s education in years) and sub-city fixed effects. Dots quantify the difference in household per capita consumption-expenditure (in birr) when the phone survey method is used relative to when the in-person method is used. The difference is measured as a percent of the mean household per capita consumption-expenditure value reported in the in-person group. Capped bars are 95-% confidence intervals, calculated from standard errors clustered at the enumeration area level.

Table A3  
Replicating Table 7, but using binary consumption variable as the dependent variable

	(1)	(2)
Dependent Variable:	Household consumed the food item (0/1)	
Item appeared later in the questionnaire	-0.007*** (0.003)	-0.004 (0.004)
Phone survey mode	-0.008* (0.005)	-0.004 (0.005)
Item appeared later in the questionnaire * Phone survey mode		-0.008 (0.005)
Household level controls?	Yes	Yes
Sub-city fixed effects?	Yes	Yes
Food item fixed effects?	Yes	Yes
Observations	93,810	93,810
In-person group mean of the dependent variable	0.211	0.211

Note: Ordinary least squares regression. Unit of observation is food item consumed (or not) in each household. Number of food items is 118 and number of households is 795 resulting in 93,810 observations. Dependent variable obtains a value 1 if the household reported to have consumed the item in the past week, zero otherwise. 0/1 = binary variable. Standard errors are clustered at the food item level and reported in parentheses. Statistical significance denoted with \* p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01.

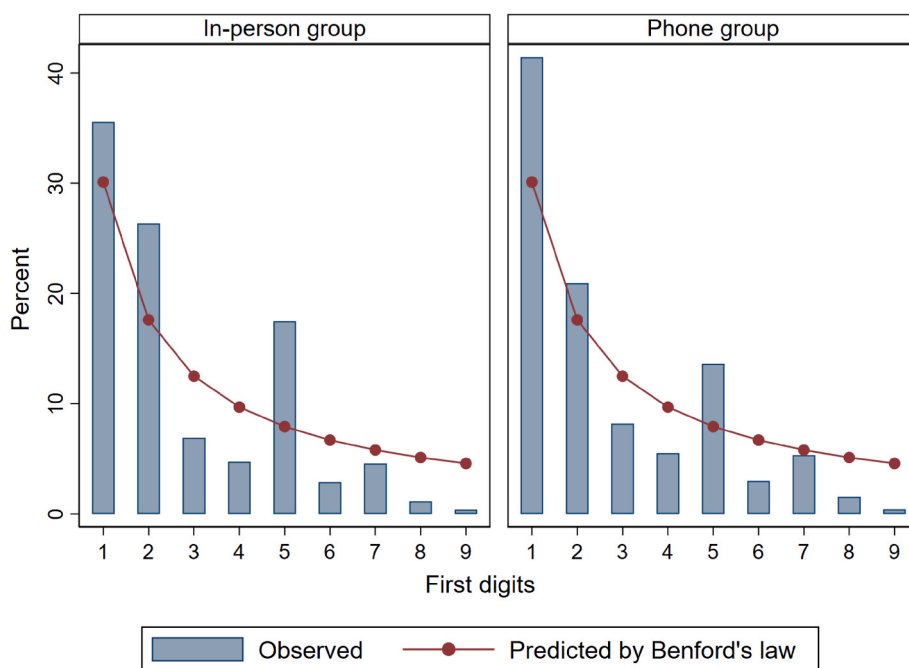


Fig. A3. Predicted and observed first-digit distributions, by survey mode.  
 Note: N = 10,526 for 'In-person group' and N = 9042 for the 'Phone group'.

Table A4

Testing differences in Euclidean distance to the distribution predicted by Benford's law

	(1)	(2)
Phone survey mode	0.328** (0.156)	0.328** (0.156)
Consumption unit fixed effects?	No	Yes (N = 33)
Observations:	66	66

Note: Dependent variable is Euclidean distance to the distribution predicted by Benford's law. Unit of observation is unit in which the quantity consumed was reported in (one for each group). Coefficients measure Z-scores. Standard errors clustered at food item level and they are reported in parentheses. Statistical significance denoted with \* p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01.

Table A5

Robustness to controlling for enumerator characteristics

	(1)	(2)	(3)	(4)
Dependent Variable:	(ln) Household per capita consumption			
Phone survey mode	-0.262*** (0.054)	-0.269*** (0.054)	-0.263*** (0.059)	-0.263*** (0.062)
Household level controls?	Yes	Yes	Yes	Yes
Sub-city fixed effects?	Yes	Yes	Yes	Yes
Enumerator characteristics?	No	Yes	No	No
Enumerator random effects?	No	No	Yes	Yes
Enumerator means of household level controls?	No	No	No	Yes
Observations	795	795	795	795
R <sup>2</sup>	0.288	0.290	n/a	n/a
R <sup>2</sup> within	n/a	n/a	0.224	0.224
R <sup>2</sup> between	n/a	n/a	0.600	0.652
R <sup>2</sup> overall	n/a	n/a	0.286	0.290

Note: Ordinary least squares regression. Unit of observation is household. Dependent variable is (ln) household total per capita consumption (in birr). Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Enumerator characteristics include enumerator's age, level of education, and survey experience (number of surveys involved in since September 2019). Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with \* p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01.

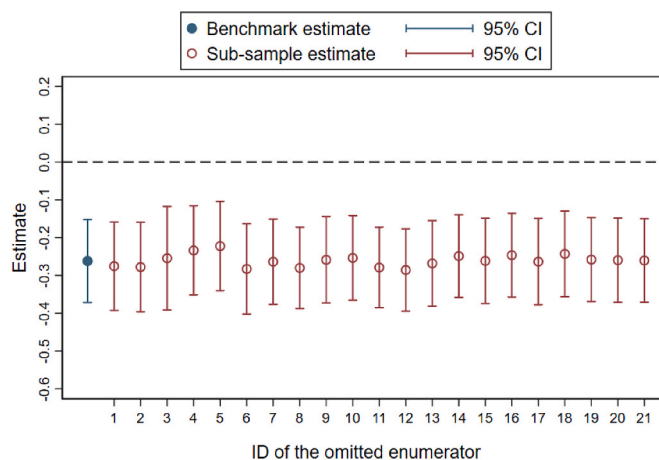


Fig. A4. Robustness of leaving one enumerator out of the dataset at a time.

Note: The blue solid dot represents the benchmark OLS estimate for the full sample reported in column 2 of Table 3. The maroon hollow dots are equivalent OLS estimates for 21 different sub-samples when one enumerator is dropped from the dataset. The capped vertical lines represent the corresponding 95% confidence intervals.

## References

- Abate, G.T., de Brauw, A., Gibson, J., Hirvonen, K., Wolle, A., 2022. Telescoping causes overstatement in recalled food consumption: Evidence from a survey experiment in Ethiopia. *World Bank Econ. Rev.* 36 (4), 909–933.
- Abate, G.T., Baye, K., de Brauw, A., Hirvonen, K., Wolle, A., 2021. Video-based behavioral change communication to change consumption patterns. In: IFPRI Discussion Paper 02052. International Food Policy Research Institute, Washington D. C. (IFPRI).
- Abay, K.A., Abate, G.T., Barrett, C.B., Bernard, T., 2019. Correlated non-classical measurement errors, 'Second best' policy inference, and the inverse size-productivity relationship in agriculture. *J. Dev. Econ.* 139, 171–184.
- Abay, K.A., Berhane, G., Hoddinott, J.F., Tafere, K., 2021a. Assessing Response Fatigue in Phone Surveys: Experimental Evidence on Dietary Diversity in Ethiopia. The World Bank, Washington, DC. World Bank Policy Research Working Paper no. 9636.
- Abay, K.A., Bevis, L.E., Barrett, C.B., 2021b. Measurement error mechanisms matter: agricultural intensification with farmer misperceptions and misreporting. *Am. J. Agric. Econ.* 103, 498–522.
- Ambel, A., McGee, K., Tsegay, A., 2021. Reducing bias in phone survey samples: effectiveness of reweighting techniques using face-to-face surveys as frames in four african countries. In: Policy Research Working Paper 9676. The World Bank, Washington D.C.
- Ambler, K., Herskowitz, S., Maredia, M.K., 2021. Are we done yet? Response fatigue and rural livelihoods. *J. Dev. Econ.* 153, 102736.
- Ameje, H., De Weerd, J., Gibson, J., 2021. Measuring macro-and micronutrient intake in multi-purpose surveys: evidence from a survey experiment in Tanzania. *Food Pol.* 102, 102042.
- Backiny-Yetna, P., Steele, D., Djima, I.Y., 2017. The impact of household food consumption data collection methods on poverty and inequality measures in Niger. *Food Pol.* 72, 7–19.
- Baird, S., Hamory, J., Miguel, E., 2008. Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1). Center for International and Development Economics Research, Berkeley, CA, UC Berkeley.
- Beaman, L., Dillon, A., 2012. Do household definitions matter in survey design? Results from a randomized survey experiment in Mali. *J. Dev. Econ.* 98, 124–135.
- Beegle, K., De Weerd, J., Friedman, J., Gibson, J., 2012. Methods of household consumption measurement through surveys: experimental results from Tanzania. *J. Dev. Econ.* 98, 3–18.
- Bell, A., Jones, K., 2015. Explaining fixed effects: random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* 3, 133–153.
- Benford, F., 1938. The law of anomalous numbers. *Proc. Am. Phil. Soc.* 551–572.
- Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: *Handbook of Econometrics*. Elsevier, pp. 3705–3843.
- Brubaker, J., Kilic, T., Wollburg, P., 2021. Representativeness of individual-level data in COVID-19 phone surveys. In: Policy Research Working Paper 9660. The World Bank, Washington D.C.
- Caeyers, B., Chalmers, N., De Weerd, J., 2012. Improving consumption measurement and other survey data through CAPI: evidence from a randomized experiment. *J. Dev. Econ.* 98, 19–33.
- Carletto, C., Savastano, S., Zezza, A., 2013. Fact or artifact: the impact of measurement errors on the farm size-productivity relationship. *J. Dev. Econ.* 103, 254–261.
- Christiaensen, L., Lanjouw, P., Luoto, J., Stifel, D., 2012. Small area estimation-based prediction methods to track poverty: validation and applications. *J. Econ. Inequal.* 10, 267–297.
- Dabalén, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y., von Engelhardt, J., 2016. *Mobile Phone Panel Surveys in Developing Countries: a Practical Guide for Microdata Collection*. The World Bank, Washington D.C.
- Dang, H.-A.H., Kilic, T., Carletto, C., Abanokova, K., 2021. Poverty imputation in contexts without consumption data. In: Policy Research Working Paper 9838. The World Bank, Washington D.C.
- De Weerd, J., 2008. Field notes on administering shock modules. *J. Int. Dev.* 20, 398–402.
- De Weerd, J., Beegle, K., Friedman, J., Gibson, J., 2016. The challenge of measuring hunger through survey. *Econ. Dev. Cult. Change* 64, 727–758.
- De Weerd, J., Gibson, J., Beegle, K., 2020. What can we learn from experimenting with survey methods? *Annual Review of Resource Economics* 12, 431–447.
- Deaton, A., Grosh, M., 2000. Consumption. In: Grosh, M., Glewwe, P. (Eds.), *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of Living Standards Measurement Study*. World Bank, Washington D.C., pp. 91–133.
- Deaton, A., Zaidi, S., 2002. Guidelines for Constructing Consumption Aggregates for Welfare Analysis. World Bank Publications.
- Desiere, S., Jolliffe, D., 2018. Land productivity and plot size: is measurement error driving the inverse relationship? *J. Dev. Econ.* 130, 84–98.
- Di Maio, M., Fiala, N., 2020. Be wary of those who ask: a randomized experiment on the size and determinants of the enumerator effect. *World Bank Econ. Rev.* 34, 654–669.
- Dillon, B., 2012. Using mobile phones to collect panel data in developing countries. *J. Int. Dev.* 24, 518–527.
- Doudich, M., Ezzrari, A., Van der Weide, R., Verme, P., 2016. Estimating quarterly poverty rates using labor force surveys: a primer. *World Bank Econ. Rev.* 30, 475–500.
- Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., Presser, S., 2014. Assessing the mechanisms of misreporting to filter questions in surveys. *Publ. Opin. Q.* 78, 721–733.
- Egger, D., Miguel, E., Warren, S.S., Shenoy, A., Collins, E., Karlan, D., Parkerson, D., Mobarak, A.M., Fink, G., Udry, C., 2021. Falling living standards during the COVID-19 crisis: quantitative evidence from nine developing countries. *Sci. Adv.* 7, eabe0997.
- FDRE, 2018. Poverty and Economic Growth in Ethiopia 1995/96-2015/16." Addis Ababa, Planning and Development Commission of the Federal Democratic Republic of Ethiopia (FDRE).
- Fiedler, J.L., Mwangi, D.M., 2016. Improving household consumption and expenditure surveys' food consumption metrics: developing a strategic approach to the unfinished agenda. In: IFPRI Discussion Paper 1570. International Food Policy Research Institute, Washington D.C. (IFPRI).
- Friedman, J., Beegle, K., De Weerd, J., Gibson, J., 2017. Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania. *Food Pol.* 72, 94–111.
- Garlick, R., Orkin, K., Quinn, S., 2020. Call me maybe: experimental evidence on frequency and medium effects in microenterprise surveys. *World Bank Econ. Rev.* 34, 418–443.
- Gibson, J., Kim, B., 2007. Measurement error in recall surveys and the relationship between household size and food demand. *Am. J. Agric. Econ.* 89, 473–489.
- Gibson, John, Kim, Bonggeun, 2010. Non-classical measurement error in long-term retrospective recall surveys. *Oxf. Bull. Econ. Stat.* 72 (5), 687–695. <https://doi.org/10.1111/j.1468-0084.2010.00599.x>.
- Gibson, J., Beegle, K., De Weerd, J., Friedman, J., 2015. What does variation in survey design reveal about the nature of measurement errors in household consumption? *Oxf. Bull. Econ. Stat.* 77, 466–474.
- Glazerman, S., Rosenbaum, M., Sandino, R., Shaughnessy, L., 2020. *Remote Surveying in a Pandemic: Handbook*. Newark, DE: Innovation for Poverty Action (IPA).

- Gourlay, S., Kilic, T., Lobell, D.B., 2019. A new spin on an old debate: errors in farmer-reported production and their implications for inverse scale-Productivity relationship in Uganda. *J. Dev. Econ.* 141, 102376.
- Gourlay, S., Kilic, T., Martuscelli, A., Wollburg, P., Zezza, A., 2021. High-frequency phone surveys on COVID-19: good practices, open questions. *Food Pol.* 105, 102153.
- Hirvonen, K., Abate, G.T., de Brauw, A., 2020. Food and nutrition security in Addis Ababa, Ethiopia during COVID-19 pandemic: May 2020 report. In: IFPRI-ESSP Working Paper 143. Washington D.C., Ethiopia Strategy Support Program. ESSP) of the International Food Policy Research Institute (IFPRI).
- Hirvonen, K., de Brauw, A., Abate, G.T., 2021. Food consumption and food security during the COVID-19 pandemic in Addis Ababa. *Am. J. Agric. Econ.* 103, 772–789.
- Hoddinott, J., Yohannes, Y., 2002. Dietary diversity as a food security indicator. In: IFPRI-FCND Discussion Paper 136. International Food Policy Research Institute, Washington, DC (IFPRI).
- Hoogeveen, J., Croke, K., Dabalén, A., Demombynes, G., Giugale, M., 2014. Collecting high frequency panel data in Africa using mobile phone interviews. *Canadian Journal of Development Studies/Revue canadienne d'études du développement* 35, 186–207.
- Hughes, S., Velyvis, K., 2020. Tips to quickly switch from face-to-face to home-based telephone interviewing. *Mathematica*. (Accessed 11 December 2021). Accessed.
- Jann, B., 2007. DIGDIS: Stata module to analyze the distribution of digits. In: *Statistical Software Components S456853*. Boston College Department of Economics revised 18 Jan 2021.
- Janssens, W., Pradhan, M., de Groot, R., Sidze, E., Donfouet, H.P.P., Abajobir, A., 2020. The short-term economic effects of COVID-19 on low-income households in rural Kenya: an analysis using weekly financial household data. *World Dev.* 138, 105280.
- Jolliffe, D., 2001. Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. *J. Econ. Soc. Meas.* 27, 1–23.
- Jones, L., von Engelhardt, J., 2020. Insights into the utility and management of mobile phone panel surveys: evidence from surveys of household resilience in Myanmar. Available at: <https://ssrn.com/abstract=3732529>. (Accessed 14 December 2021). Accessed.
- Josephson, A., Kilic, T., Michler, J.D., 2021. Socioeconomic impacts of COVID-19 in low-income countries. *Nat. Human Behav.* 5, 557–565.
- Kaiser, M., 2019. Benford's law as an indicator of survey reliability—can we trust our data? *J. Econ. Surv.* 33, 1602–1618.
- Kilic, T., Sohnesen, T.P., 2019. Same question but different answer: experimental evidence on questionnaire design's impact on poverty measured by proxies. *Rev. Income Wealth* 65, 144–165.
- Kopper, S., Sautmann, A., 2020. Best practices for conducting phone surveys. In: Abdul Latif Jameel Poverty Action Lab. J-PAL). (Accessed 14 December 2021). Accessed: Laajaj, R., Macours, K., 2021. Measuring skills in developing countries. *J. Hum. Resour.* 56, 1254–1295.
- Laborde, D., Martin, W., Vos, R., 2021. Impacts of COVID-19 on global poverty, food security, and diets: insights from global model scenario analysis. *Agric. Econ.*
- Lakner, C., Yonzan, N., Mahler, D.G., Aguilar, R.A.C., Wu, H., 2021. Updated Estimates of the Impact of COVID-19 on Global Poverty: Looking Back at 2020 and the Outlook for 2021. The World Bank. (Accessed 11 December 2021). Accessed.
- Mancini, G., Vecchi, G., 2022. On the Construction of a Consumption Aggregate for Inequality and Poverty Analysis. International Bank for Reconstruction and Development and The World Bank, Washington D.C.
- McKenzie, D., Rosenzweig, M., 2012. Preface for symposium on measurement and survey design. *J. Dev. Econ.* 1, 1–2.
- Melesse, M.B., van den Berg, M., de Brauw, A., Abate, G.T., 2019. Understanding urban consumers' food choice behavior in Ethiopia: promoting demand for healthy foods. In: IFPRI-ESSP Working Paper 131. International Food Policy Research Institute, Washington D.C. (IFPRI).
- Miguel, E., Mobarak, A.M., 2022. The Economics of the COVID-19 Pandemic in Poor Countries. *Ann. Rev. Econ.* 14, 253–285.
- Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica: J. Econom. Soc.* 69–85.
- Pape, U., 2021. Measuring poverty rapidly using within-survey imputations. In: The World Bank Policy Research Working Paper 9530. The World Bank, Washington D.C.
- Pape, U., Mistiaen, J., 2015. Measuring household consumption and poverty in 60 minutes: the Mogadishu high frequency survey. In: The World Bank Policy Research Working Paper 8430. The World Bank, Washington D.C.
- Pradhan, M., 2009. Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia. *Fisc. Stud.* 30, 391–417.
- Raghunathan, T.E., Grizzle, J.E., 1995. A split questionnaire survey design. *J. Am. Stat. Assoc.* 90, 54–63.
- Roberts, C., Eva, G., Allum, N., Lynn, P., 2010. Data quality in telephone surveys and the effect of questionnaire length: a cross-national experiment. In: ISER Working Paper Series No. 2010-36. Institute for Social and Economic Research, University of Essex, UK.
- Sánchez-Páramo, C., Hill, R., Mahler, D.G., Narayan, A., Yonzan, N., 2021. COVID-19 Leaves a Legacy of Rising Poverty and Widening Inequality. The World Bank. (Accessed 11 December 2021). Accessed.
- Schündeln, M., 2018. Multiple visits and data quality in household surveys. *Oxf. Bull. Econ. Stat.* 80, 380–405.
- Stifel, D., Christiaensen, L., 2007. Tracking poverty over time in the absence of comparable consumption data. *World Bank Econ. Rev.* 21, 317–341.
- Sumner, A., Hoy, C., Ortiz-Juarez, E., 2020. Estimates of the impact of COVID-19 on global poverty. In: WIDER Working Paper 2020/43. Helsinki, UNU-WIDER.
- Swindale, A., Bilinsky, P., 2006. Household dietary diversity score (HDDS) for measurement of household food access: indicator guide. In: Food and Nutrition Technical Assistance Project, Academy for Educational Development, vol. 360. FANTA FHI, Washington, DC. Washington D.C.
- Tourangeau, R., Rips, L.J., Rasinski, K., 2000. *The Psychology of Survey Response*. Cambridge University Press, New York.
- Troubat, N., Grünberger, K., 2017. Impact of survey design in the estimation of habitual food consumption: a study based on urban households of Mongolia. *Food Pol.* 72, 132–145.
- van Garderen, J.K., Shah, C., 2002. Exact interpretation of dummy variables in semilogarithmic equations. *Econom. J.* 5, 149–159.
- WFP, 2017. Conducting mobile surveys responsibly - a field book for WFP staff. In: World Food Programme (WFP). Vulnerability Analysis & Mapping (VAM), Rome.
- Wieser, C., Ambel, A.A., Bundervoet, T., Tsegay, A.H., 2020. Monitoring COVID-19 Impacts on Households in Ethiopia: Results from a High-Frequency Phone Survey of Households. The World Bank. Report #1.
- World Food Programme (WFP), 2008. Food Consumption Analysis: Calculation and Use of the Food Consumption Score in Food Security Analysis. Technical Guidance Sheet. WFP Vulnerability Analysis & Mapping, Rome.
- Zeza, A., Carletto, C., Fiedler, J.L., Gennari, P., Jolliffe, D., 2017. Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES). Introduction to the special issue. *Food Pol.* 72, 1–6.