WIDER Working Paper 2022/93

Measuring consumption over the phone

Evidence from a survey experiment in urban Ethiopia

Gashaw T. Abate,[1] Alan de Brauw,[2] Kalle Hirvonen,[3] and Abdulazize Wolle[4]

August 2022

Abstract: The paucity of reliable, timely household consumption data in many low- and middle-income countries has 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 done 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 was 23 per cent lower than in the in-person survey, and the estimated poverty headcount was twice as high. There is 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] International Food Policy Research Institute (IFPRI), Washington, DC, USA; [2] IFPRI; [3] IFPRI and UNU-WIDER, Helsinki, Finland, corresponding author: kalle@wider.unu.edu; [4] Economics Department, State University of New York at Albany, Albany, NY, USA.

1   Introduction

When it became clear the spread of COVID-19 would become a pandemic in March 2020, many surveys that had previously taken place in person could no longer be fielded due to the concern that they would contribute to the spread of the virus. Yet in-person surveys are a key component of many research efforts and the monitoring of outcomes such as those measuring progress towards the Sustainable Development Goals (SDGs). Without in-person surveys such as 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.[1] There was a veritable explosion of efforts to collect socioeconomic data over the phone, including major co-ordinated efforts by Innovations for Poverty Action (Research for Effective COVID-19 Response/RECOVR) and the World Bank (Gourlay et al. 2021). These efforts 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 based on crude measures, e.g., asking whether household income was lower, the same, or higher than it had been 12 months previously.

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 five years of age without in-person data collection.

Similarly, collecting data on household consumption expenditures to estimate poverty incidence requires complex measurement.[2] The standard method of household consumption expenditure and poverty measurement involves administering detailed food and non-food consumption modules covering more than 100 items typically consumed in a country (Deaton and Grosh 2000; Deaton and Zaidi 2002).[3] Most phone surveys did not attempt 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

---

[1] We cover mainly 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).

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

[3] 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 2021).

modellers have predicted large increases in poverty incidence and 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 that it is difficult to tell whether their predictions have come true.

The surveys that tried to collect consumption data over the phone during the pandemic suggest that 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.[4] 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 from 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 that it might be plausible to conduct phone surveys to measure consumption, and therefore poverty incidence, in the same way as it had been previously, particularly if the survey involves an attempt to develop some rapport with households before the long consumption survey, as was the case in all the surveys described above. But it is important to quantify differences between phone and in-person measures of consumption before reaching such conclusions. Therefore, here we test whether consumption data collected over the phone have a comparable distribution to data collected in person, using a sample that had been asked about food consumption several times in the past. We randomly select half of the sample to be surveyed about consumption in person, with the other half surveyed 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 a household's food security.

We can then compute poverty incidence using 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 Weerdt 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 suggests 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 that it is positively related to household size.[5] This finding implies that the measurement

---

[4] However, they do find a concurrent rise in some measures of food insecurity.

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

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 Weerdt 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 Weerdt 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 with 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 for 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. 2021b).

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 was 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 had consumed the item in the past seven days or not. The survey instrument was programmed to carry forward all items that had been consumed in the past seven days to the next subsection, which asked about consumption frequency ('on how many days was the item consumed?') and quantity ('amount consumed') within the seven-day period. The second main section of the questionnaire included a short module asking about the household's food consumption outside of the home within the same seven-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).

3

Table 1: Comparison of in-person and phone data collection

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

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

Source: authors' construction based on own data.

The full questionnaire is included in Appendix B.

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 behavioural change communication on fruit and vegetable consumption in Addis Ababa, Ethiopia (Abate et al. 2021a). The baseline and endline surveys for the RCT took place in September 2019 and February 2020, respectively.[6] The sample of 930 households was randomly selected from six sub-cities, 20 *woredas* (districts), and 40 *ketenas* (neighbourhoods or clusters of households) within Addis Ababa.[7] Comparison of household characteristics against those reported in other surveys from Addis Ababa suggests 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 set-up that was highly optimal for launching COVID-19 phone surveys. Phone numbers were collected from 887 of the 895 households (99 per cent) 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 Appendix A 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 ten-day period in September 2021 (i.e., one year after the last COVID-19 phone survey).[8] The sampling frame for this study was based on 895 households that had been 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

---

[6] 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.

[7] Melesse et al. (2019) provide a detailed description of the sampling strategy.

[8] The exact dates were 31 August to 9 September 2021.

and 447 for a phone interview.[9] A total of 797 households were interviewed; 421 in the in-person group and 376 in the phone group.[10] Administering the consumption modules over the phone took 41 minutes on average (median), while the average (median) interview duration was 43 minutes for an in-person visit. The quality of the connection was generally good for the phone interviews and enumerators' assessment was that it rarely affected the interview quality.[11]

The survey team tasked with the in-person surveys followed recommended COVID-19 preventative 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 seconds or use disinfectant (containing more than 70 per cent alcohol) before entering and when leaving the respondent's premises. Third, the survey co-ordinator conducted daily check-ups with enumerators regarding any COVID-19-related symptoms. Finally, the interview was conducted outdoors with at least two metres' 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 with 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 and 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 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, since which food prices and prices of non-food items have both risen annually at a double-digit rate. Considering the high inflation rate and the considerable methodological differences between our survey and

---

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

[10] 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, ten had wrong numbers, and five had no phone numbers. Only two households refused to take part in the phone survey.

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

the HCES, we do not attempt to update the HCES poverty line for September 2021. Instead, we calibrate our poverty line for the in-person sample to match the 16.8 per cent 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 using the phone survey mode affects the results obtained on 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 the sum of all food groups from which the household consumed food items during the seven-day recall period, with a minimum of 1 and maximum of 12. Second, we constructed the food consumption score (FCS) developed by the World Food Programme (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.[12] The weighted FCS index ranges between 0 and 112, with higher scores indicating a better food security situation.

Table 2: Household characteristics, by survey mode

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

Source: authors' construction based on own data.

After dropping two households with implausible consumption values, the final sample of 795 households was formed, out of which 421 were from the in-person group and 374 from the phone

---

[12] 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).

group. Table 2 shows that the in-person and phone groups were similar in terms of basic household characteristics. Moreover, the households in the two subsamples were 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 Appendix A).

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 1 if the household was randomly selected into the phone group and 0 if into the in-person group. In subsequent models, we control for differences in basic household characteristics (household size, and household head's gender 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): $100 \times \left(e^{\hat{\beta}-0.5\hat{V}(\hat{\beta})} - 1\right)$, 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

Figure 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.

Figure 1: Distribution of (ln) weekly consumption per capita (in birr), by survey mode
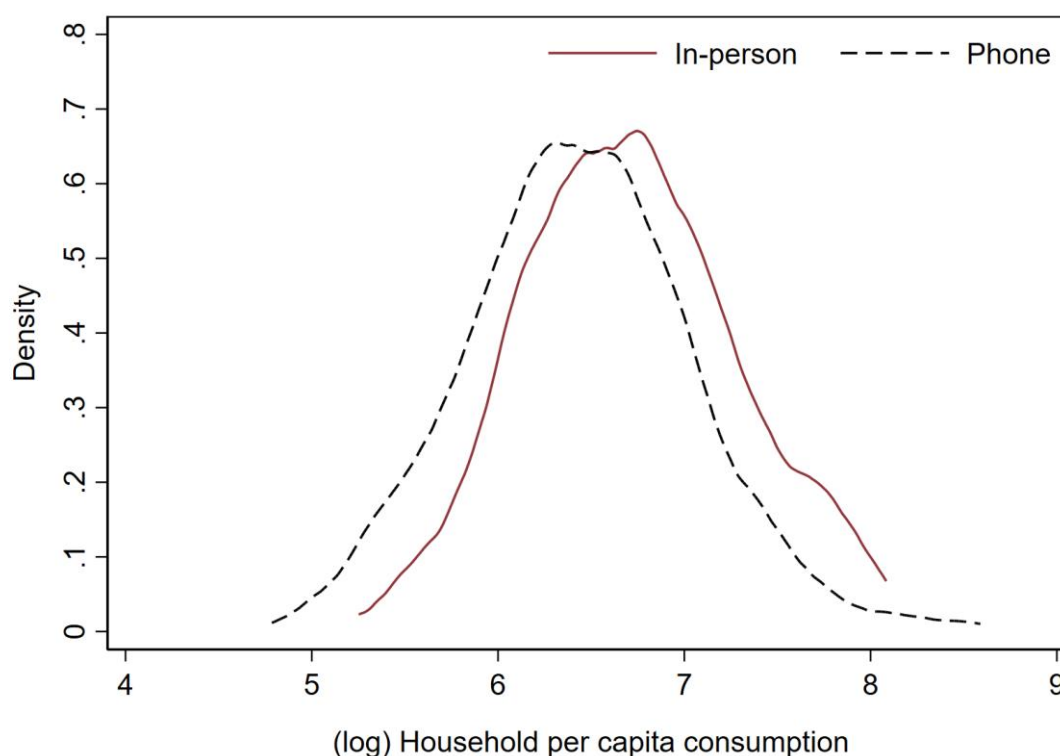
Note: N = 795 households.

Source: authors' illustration based on own data.

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***	−0.262***	−207.69***	−200.61***
	(0.059)	(0.054)	(58.16)	(52.65)
Household level controls	No	Yes	No	Yes
Sub-city fixed effects	No	Yes	No	Yes
Observations	795	795	795	795
R^2	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: OLS 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' illustration based on own data.

Relative to the in-person survey, on average using the phone survey mode results in a decrease in reported household per capita consumption expenditures by 23 per cent (Table 3, column 2).[13] The 95 per cent 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 per cent lower average per capita consumption in the phone survey group.

3.2   Components of consumption

Food consumed at home represents 50.3 per cent of the total consumption among the in-person group and 55.8 per cent among the phone survey group.[14] The regression estimates reported in column 1 of Table 4 indicate that the reported per capita food consumption values are 13 per cent lower on average when the phone survey mode is used (95 per cent 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 Appendix A, we observe that all the coefficient estimates are negative and suggest 5 to 25 per cent lower consumption, with overlapping confidence intervals.

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***	−0.129**	−21.66**	−0.35***
	(0.043)	(0.056)	(8.34)	(0.09)
Household level controls	Yes	Yes	Yes	Yes
Sub-city fixed effects	Yes	Yes	Yes	Yes
Observations	795	795	795	795
R^2	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: OLS 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' construction based on own data.

About 60 per cent of the households in our sample reported having consumed food items outside of their home in the past seven days. This reporting incidence varies by survey mode, with households in the phone survey group being 13 percentage points less likely to report having consumed foods outside their home (Table 4, column 2). A regression based on a non-logged

---

[13] 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.

[14] The difference is statistically significant (p = 0.003).

outcome variable shows that food expenditures outside of the home are 40.2 per cent lower in the phone group relative to the in-person group (Table 4, column 3).[15]

All the households in our sample reported 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 sees reported non-food consumption that is lower by 30.1 per cent (95 per cent CI: −15.5; −42.1).

3.3 Poverty estimates

Next, we estimate the impact of using the 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 square regression method where the weights are frequency weights based on household size. Using our calibrated poverty line, in Table 5 we estimate that the poverty rate is 17 percentage points higher when the phone survey mode is used compared with when consumption data are collected through in-person visits (95 per cent CI: 9.99; 24.1). Since the poverty rate in the in-person sample is calibrated at 16.8 per cent, 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***	0.170***
	(0.036)	(0.035)
Household level controls	No	Yes
Sub-city fixed effects	No	Yes
Observations (weighted)	3,828	3,828
Households	795	795
R^2	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 of 1 if the household's per capita consumption is below the poverty line, 0 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' construction based on own data.

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

---

[15] 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 per cent (21.66/53.91).

consumed, only whether the food item was consumed in the past seven days (HDDS) or the consumption frequency in terms of number of days in the past seven 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 to in-person surveys when measuring diet-based food security but leads to much lower estimates of the value of household food or non-food consumption.

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	0.058	−2.120	−2.055
	(0.132)	(0.135)	(1.629)	(1.646)
Household level controls	No	Yes	No	Yes
Sub-city fixed effects	No	Yes	No	Yes
Observations	795	795	795	795
R^2	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: OLS 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' construction based on own data.

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 as we had used in-person surveys to measure poverty prior to the pandemic, we would conclude that the poverty headcount was twice as high using the phone survey data as using 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 between the two survey modes in consumption and poverty incidence estimates 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 (Dabalen et al. 2016; Gourlay et al. 2021). Consequently, it has been widely recommended to keep 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).[16] In a recent phone survey conducted in rural Ethiopia, Abay et al. (2021b) estimate that delaying the timing of a dietary diversity module by 15 minutes increased the likelihood that the respondents will report not having consumed items from certain food groups, resulting in an 8 per cent decline in the mothers' dietary diversity score.[17]

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.[18] Specifically, we implemented two versions of this food consumption module, ordering the food groups differently (see Appendix A Table A2). For example, in version 1, mango appeared as the fifth 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 the value of 1 when each food appears later in the questionnaire relative to the other version, and 0 otherwise.[19] 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 the point in the module at which the item is asked about differs between the 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 to later in the questionnaire results in a report that is, on average, 5.8 per cent lower for the item than if it takes on its earlier position.[20] The average report in the phone survey mode suggests that the value of consumption is 15.5 per cent lower than that 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 per cent CI: −0.0167; 0.0016), indicating that survey fatigue does not play a role in the in-person survey mode, at least in this relatively early part of the questionnaire. In contrast,

---

[16] Evidence from survey experiments conducted in high-income countries has documented respondent fatigue in phone survey mode (e.g., Eckman et al. 2014), Roberts et al. (2010).

[17] Garlick et al. (2020) randomly assigned small firms to weekly phone and in-person surveys, finding that phone survey respondents systematically under-reported labour 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.

[18] Abay et al. (2021b), Ambler et al. (2021), and Laajaj and Macours (2021) also randomize the order in which questions are asked in their surveys to study survey fatigue.

[19] As can be seen from Appendix A 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.

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

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 per cent 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.

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**	−0.014
	(0.101)	(0.159)
Phone survey mode	−0.615***	−0.368
	(0.203)	(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: OLS 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' construction based on own data.

In Appendix A Table A3 we replicate this analysis, considering only the responses to the yes/no questions regarding whether the household consumed the item or not during the seven-day period. Interestingly, all coefficients in the interacted model appear insignificant, implying that only consumption quantity reports are affected, 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 datasets approximately follow the probability (P):

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

where $d \in \{1, ..., 9\}$ 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.[21]

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 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 Appendix A reports the observed first-digit distributions in our data and compares them to the distribution predicted by Benford's law.[22] 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 units' specific average Euclidean distances for the two groups are statistically different by regressing the mean distance on our binary treatment variable. Table A4 in Appendix A 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 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 Weerdt 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) are used as the dependent variables. For household food consumption, we observe that the bias decreases with household head's education and increases with household size.[23] The former result suggests that 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

---

[21] The Euclidian distance is calculated as the square root of the sum of squared differences between the observed percentage and the percentage predicted by 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.

[22] We calculated these distributions using a user-written Stata routine devised by Jann (2007).

[23] Table 2 reports that the difference in household size between the two household groups is not statistically different from zero.

to the food consumption data. Overall, these heterogenous impacts imply that adjustment factors to account for the bias caused by the phone survey mode cannot be easily developed.

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***	0.073	−0.427***	−0.224
	(0.060)	(0.117)	(0.143)	(0.183)
Phone survey mode × head's education in years	0.015**		0.011	
	(0.007)		(0.014)	
Phone survey mode × household size		−0.041*		−0.027
		(0.022)		(0.027)
Household level controls	Yes	Yes	Yes	Yes
Sub-city fixed effects	Yes	Yes	Yes	Yes
Observations	795	795	795	795
R^2	0.595	0.595	0.227	0.227

Note: OLS 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' construction based on own data.

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.[24] 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 Appendix A). 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 subsamples (see Figure A4 in Appendix A). Third, we show that our results are robust to the controlling for enumerator random effects (Table A5, column 3 in Appendix A) as well as Mundlak (1978)

---

[24] 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).

correlated random effects (Table A5, column 4 in Appendix A).[25] 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 with 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 three 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 that phone and in-person surveys took, phone enumerators were able to conduct about three times as many interviews in a day as in-person enumerators because the survey mode allowed them to make the next call as soon as they were ready, while the in-person survey required enumerators to travel to the next household. However, there were 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 neighbourhoods 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 sizeable gain in statistical power and thus improvement in the precision of consumption and poverty estimates.[26] 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.

---

[25] The random effects estimator controls for enumerator heterogeneity by decomposing the unobserved heterogeneity to variance occurring between 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.

[26] 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 sizeable 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.

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 (Dabalen et al. 2016; Dillon 2012; Hoogeveen et al. 2014). Meanwhile, the 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 Weerdt 2008; Hirvonen et al. 2021).[27] Indeed, there have been only few attempts to measure household consumption to inform how the progress towards meeting the first Sustainable Development Goal of 'no poverty' has been affected by the pandemic. Finally, there remains considerable uncertainty around the implications of the use of the phone survey mode for 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 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 results in a decrease in the reported household per capita consumption expenditures of 23 per cent, 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 reinterpret 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 that the results likely underestimate all categories of food consumption.

The mechanism appears to be linked to survey fatigue, which results in phone survey respondents greatly underestimating consumption quantities but not whether they consumed the item during the recall period. Our heterogeneity analysis suggests that the bias increases when more people eat within the household, possibly because of the increased cognitive burden in remembering larger

---

[27] At the same time, with imperfect and non-random mobile phone access in rural areas, the data may not be representative, as poor people and those 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).

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, which 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.[28] 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 Appendix A). Consequently, the households 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 mobile 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 suggests that shorter modules systematically underestimate consumption levels and thus overestimate poverty headcounts (Beegle et al. 2012; Jolliffe 2001; Pradhan 2009). Therefore, when adjusting consumption module length, survey practitioners need to achieve a 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.[29]

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; Douidich 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

---

[28] However, 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 per cent of rural households have access to a phone, and those that have access 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.

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

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-19 pandemic and the conflict between Russia and Ukraine, an open question remains 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 one, 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 set-up. In this method, respondents are randomly assigned fractions of the full questionnaire and the missing data are then imputed using multiple imputation techniques (Raghunathan 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). There remains an open question about whether split designs could be used to generate low-bias estimates of poverty incidence with phone surveys.

References

Abate, G.T., A. de Brauw, J. Gibson, K. Hirvonen, and A. Wolle (2020). 'Telescoping Causes Overstatement in Recalled Food Consumption: Evidence from a Survey Experiment in Ethiopia'. IFPRI Discussion Paper 01976. Washington, DC: IFPRI. https://doi.org/10.2499/p15738coll2.134161

Abate, G.T., K. Baye, A. de Brauw, K. Hirvonen, and A. Wolle (2021a). 'Video-Based Behavioral Change Communication to Change Consumption Patterns'. IFPRI Discussion Paper 02052. Washington, DC: IFPRI. https://doi.org/10.2499/p15738coll2.134725

Abate, G.T., A. de Brauw, K. Hirvonen, and A. Wolle (2021b). 'Measuring Consumption Over the Phone: Evidence from a Survey Experiment in Urban Ethiopia'. IFPRI Discussion Paper 2087. Washington, DC: International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/p15738coll2.134939

Abay, K.A., G.T. Abate, C.B. Barrett, and T. Bernard (2019). 'Correlated Non-classical Measurement Errors, "Second Best" Policy Inference, and the Inverse Size–Productivity Relationship in Agriculture'. *Journal of Development Economics*, 139: 171–84. https://doi.org/10.1016/j.jdeveco.2019.03.008

Abay, K.A., L.E. Bevis, and C.B. Barrett (2021a). 'Measurement Error Mechanisms Matter: Agricultural Intensification with Farmer Misperceptions and Misreporting'. *American Journal of Agricultural Economics*, 103: 498–522. https://doi.org/10.1111/ajae.12173

Abay, K.A., G. Berhane, J.F. Hoddinott, and K. Tafere (2021b). 'Assessing Response Fatigue in Phone Surveys: Experimental Evidence on Dietary Diversity in Ethiopia'. IFPRI Discussion Paper 02017. Washington DC: IFPRI. https://doi.org/10.2499/p15738coll2.134381

Ambel, A., K. McGee, and A. Tsegay (2021). 'Reducing Bias in Phone Survey Samples: Effectiveness of Reweighting Techniques Using Face-to-Face Surveys as Frames in Four African Countries'. Policy Research Working Paper 9676. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-9676

Ambler, K., S. Herskowitz, and M.K. Maredia (2021). 'Are We Done Yet? Response Fatigue and Rural Livelihoods'. *Journal of Development Economics*, 153: 102736. https://doi.org/10.1016/j.jdeveco.2021.102736

Ameye, H., J. De Weerdt, and J. Gibson (2021). 'Measuring Macro- and Micronutrient Intake in Multi-Purpose Surveys: Evidence from a Survey Experiment in Tanzania'. *Food Policy*, 102: 102042. https://doi.org/10.1016/j.foodpol.2021.102042

Backiny-Yetna, P., D. Steele, and I.Y. Djima (2017). 'The Impact of Household Food Consumption Data Collection Methods on Poverty and Inequality Measures in Niger'. *Food Policy*, 72: 7–19. https://doi.org/10.1016/j.foodpol.2017.08.008

Baird, S., J. Hamory, and E. Miguel (2008). 'Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)'. Berkeley, CA: Center for International and Development Economics Research, University of California, Berkeley.

Beaman, L., and A. Dillon (2012). 'Do Household Definitions Matter in Survey Design? Results from a Randomized Survey Experiment in Mali'. *Journal of Development Economics*, 98: 124–35. https://doi.org/10.1016/j.jdeveco.2011.06.005

Beegle, K., J. De Weerdt, J. Friedman, and J. Gibson (2012). 'Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania'. *Journal of Development Economics*, 98: 3–18. https://doi.org/10.1016/j.jdeveco.2011.11.001

Bell, A., and K. Jones (2015). 'Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data'. *Political Science Research and Methods*, 3: 133–53. https://doi.org/10.1017/psrm.2014.7

Benford, F. (1938). 'The Law of Anomalous Numbers'. *Proceedings Of the American Philosophical Society*, 78(4): 551–72.

Bound, J., C. Brown, and N. Mathiowetz (2001). 'Measurement Error in Survey Data.' In J.J. Heckman and E. Learner (eds), *Handbook of Econometrics*, volume 5. Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4412(01)05012-7

Brubaker, J., T. Kilic, and P. Wollburg (2021). 'Representativeness of Individual-Level Data in COVID-19 Phone Surveys'. Policy Research Working Paper 9660. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-9660

Caeyers, B., N. Chalmers, and J. De Weerdt (2012). 'Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment'. *Journal of Development Economics*, 98: 19–33. https://doi.org/10.1016/j.jdeveco.2011.12.001

Carletto, C., S. Savastano, and A. Zezza (2013). 'Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship'. *Journal of Development Economics*, 103: 254–61. https://doi.org/10.1016/j.jdeveco.2013.03.004

Christiaensen, L., P. Lanjouw, J. Luoto, and D. Stifel (2012). 'Small Area Estimation-Based Prediction Methods to Track Poverty: Validation and Applications'. *The Journal of Economic Inequality*, 10: 267–97. https://doi.org/10.1007/s10888-011-9209-9

Dabalen, A., A. Etang, J. Hoogeveen, E. Mushi, Y. Schipper, and J. von Engelhardt (2016). *Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-0904-0

Dang, H.-A.H., T. Kilic, C. Carletto, and K. Abanokova (2021) 'Poverty Imputation in Contexts without Consumption Data'. Policy Research Working Paper 9838. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-9838

De Weerdt, J. (2008). 'Field Notes on Administering Shock Modules'. *Journal of International Development*, 20: 398–402. https://doi.org/10.1002/jid.1435

De Weerdt, J., K. Beegle, J. Friedman, and J. Gibson (2016). 'The Challenge of Measuring Hunger through Survey'. *Economic Development and Cultural Change*, 64: 727–58. https://doi.org/10.1086/686669

De Weerdt, J., J. Gibson, and K. Beegle (2020). 'What Can We Learn from Experimenting with Survey Methods?' *Annual Review of Resource Economics*, 12: 431–47. https://doi.org/10.1146/annurev-resource-103019-105958

Deaton, A., and M. Grosh (2000). 'Consumption'. In M. Grosh and P. Glewwe (eds), *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of Living Standards Measurement Study*. Washington, DC: World Bank.

Deaton, A., and S. Zaidi (2002). *Guidelines for Constructing Consumption Aggregates for Welfare Analysis*. Washington, DC: World Bank.

Desiere, S., and D. Jolliffe (2018). 'Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?' *Journal of Development Economics*, 130: 84–98. https://doi.org/10.1016/j.jdeveco.2017.10.002

Di Maio, M., and N. Fiala (2020). 'Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect'. *The World Bank Economic Review*, 34: 654–69. https://doi.org/10.1093/wber/lhy024

Dillon, B. (2012). 'Using Mobile Phones to Collect Panel Data in Developing Countries'. *Journal of International Development*, 24: 518–27. https://doi.org/10.1002/jid.1771

Douidich, M., A. Ezzrari, R. Van der Weide, and P. Verme (2016). 'Estimating Quarterly Poverty Rates using Labor Force Surveys: A Primer'. *The World Bank Economic Review*, 30: 475–500. https://doi.org/10.1093/wber/lhv062

Eckman, S., F. Kreuter, A. Kirchner, A. Jäckle, R. Tourangeau, and S. Presser (2014). 'Assessing the Mechanisms of Misreporting to Filter Questions in Surveys'. *Public Opinion Quarterly*, 78: 721–33. https://doi.org/10.1093/poq/nfu030

Egger, D., E. Miguel, S.S. Warren, A. Shenoy, E. Collins, D. Karlan, D. Parkerson, A.M. Mobarak, G. Fink, and C. Udry (2021). 'Falling Living Standards during the COVID-19 Crisis: Quantitative Evidence from Nine Developing Countries'. *Science Advances*, 7: eabe0997. https://doi.org/10.1126/sciadv.abe0997

FDRE (2018). 'Poverty and Economic Growth in Ethiopia 1995/96–2015/16'. Addis Ababa: Planning and Development Commission, Federal Democratic Republic of Ethiopia (FDRE).

Fiedler, J.L., and D.M. Mwangi (2016) 'Improving Household Consumption and Expenditure Surveys' Food Consumption Metrics: Developing a Strategic Approach to the Unfinished Agenda'. IFPRI Discussion Paper 1570. Washington, DC: IFPRI.

Friedman, J., K. Beegle, J. De Weerdt, and J. Gibson (2017). 'Decomposing Response Error in Food Consumption Measurement: Implications for Survey Design from a Randomized Survey Experiment in Tanzania'. *Food Policy*, 72: 94–111. https://doi.org/10.1016/j.foodpol.2017.08.016

Garlick, R., K. Orkin, and S. Quinn (2020). 'Call Me Maybe: Experimental Evidence on Frequency and Medium Effects in Microenterprise Surveys'. *The World Bank Economic Review*, 34: 418–43. https://doi.org/10.1093/wber/lhz021

Gibson, J., and B. Kim (2007). 'Measurement Error in Recall Surveys and the Relationship between Household Size and Food Demand'. *American Journal of Agricultural Economics*, 89: 473–89. https://doi.org/10.1111/j.1467-8276.2007.00978.x

Gibson, J., and B. Kim (2010). 'Non-classical Measurement Error in Long-Term Retrospective Recall Surveys'. *Oxford Bulletin of Economics and Statistics*, 72: 687–95. https://doi.org/10.1111/j.1468-0084.2010.00599.x

Gibson, J., K. Beegle, J. De Weerdt, and J. Friedman (2015). 'What Does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?' *Oxford Bulletin of Economics and Statistics*, 77: 466–74. https://doi.org/10.1111/obes.12066

Glazerman, S., M. Rosenbaum, R. Sandino, and L. Shaughnessy (2020). *Remote Surveying in a Pandemic: handbook*. Newark, DE: Innovation for Poverty Action (IPA).

Gourlay, S., T. Kilic, and D.B. Lobell (2019). 'A New Spin on an Old Debate: Errors in Farmer-Reported Production and Their Implications for Inverse Scale–Productivity Relationship in Uganda'. *Journal of Development Economics*, 141: 102376. https://doi.org/10.1016/j.jdeveco.2019.102376

Gourlay, S., T. Kilic, A. Martuscelli, P. Wollburg, and A. Zezza (2021). 'High-Frequency Phone Surveys on COVID-19: Good Practices, Open Questions'. *Food Policy*, 105: 102153. https://doi.org/10.1016/j.foodpol.2021.102153

Hirvonen, K., G.T. Abate, and A. de Brauw (2020) 'Food and Nutrition Security in Addis Ababa, Ethiopia during COVID-19 Pandemic: May 2020 Report'. IFPRI-ESSP working paper 143. Washington, DC, Ethiopia Strategy Support Program (ESSP), IFPRI. https://doi.org/10.2499/p15738coll2.133731

Hirvonen, K., A. de Brauw, and G.T. Abate (2021). 'Food Consumption and Food Security during the COVID-19 Pandemic in Addis Ababa'. *American Journal of Agricultural Economics*, 103: 772–89. https://doi.org/10.1111/ajae.12206

Hoddinott, J., and Y. Yohannes (2002). 'Dietary Diversity as a Food Security Indicator'. IFPRI-FCND discussion paper 136. Washington, DC: Food Consumption and Nutrition Division (FCND), IFPRI.

Hoogeveen, J., K. Croke, A. Dabalen, G. Demombynes, and M. Giugale (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. https://doi.org/10.1080/02255189.2014.876390

Hughes, S., and K. Velyvis (2020). 'Tips to Quickly Switch from Face-to-Face to Home-Based Telephone Interviewing'. Mathematica, 1 April. Available at: www.mathematica.org/blogs/tips-to-quickly-switch-from-face-to-face-to-home-based-telephone-interviewing (accessed 11 December 2021).

Jann, B. (2007). 'DIGDIS: Stata Module to Analyze the Distribution of Digits'. Statistical Software Components S456853. Boston: Boston College Department of Economics, revised 18 January 2021.

Janssens, W., M. Pradhan, R. de Groot, E. Sidze, H.P.P. Donfouet, and A. Abajobir (2020). 'The Short-Term Economic Effects of COVID-19 on Low-Income Households in Rural Kenya: An Analysis using Weekly Financial Household Data'. *World Development*, 138: 105280. https://doi.org/10.1016/j.worlddev.2020.105280

Jolliffe, D. (2001). 'Measuring Absolute and Relative Poverty: The Sensitivity of Estimated Household Consumption to Survey Design'. *Journal of Economic and Social Measurement*, 27: 1–23. https://doi.org/10.3233/JEM-2003-0192

Jones, L., and J. von Engelhardt (2020). 'Insights into the Utility and Management of Mobile Phone Panel Surveys: Evidence from Surveys of Household Resilience in Myanmar'. https://doi.org/10.2139/ssrn.3732529

Josephson, A., T. Kilic, and J.D. Michler (2021). 'Socioeconomic Impacts of COVID-19 in Low-Income Countries'. *Nature Human Behaviour* 5:557–565. https://doi.org/10.1038/s41562-021-01096-7

Kaiser, M. (2019). 'Benford's Law as an Indicator of Survey Reliability: Can We Trust Our Data?' *Journal of Economic Surveys*, 33: 1602–18. https://doi.org/10.1111/joes.12338

Kilic, T., and T.P. Sohnesen (2019). 'Same Question But Different Answer: Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies'. *Review of Income and Wealth*, 65: 144–65. https://doi.org/10.1111/roiw.12343

Kopper, S., and A. Sautmann (2020). 'Best Practices for Conducting Phone Surveys'. Abdul Latif Jameel Poverty Action Lab (J-PAL), 20 March. Available at: www.povertyactionlab.org/blog/3-20-20/best-practices-conducting-phone-surveys (accessed 14 December 2021).

Laajaj, R., and K. Macours (2021). 'Measuring Skills in Developing Countries'. *Journal of Human Resources*, 56: 1254–95. https://doi.org/10.3368/jhr.56.4.1018-9805R1

Laborde, D., W. Martin, and R. Vos (2021). 'Impacts of COVID-19 on Global Poverty, Food Security, and Diets: Insights from Global Model Scenario Analysis'. *Agricultural Economics*, 52(3): 375–90.

Lakner, C., N. Yonzan, D.G. Mahler, R.A.C. Aguilar, and H. Wu (2021). 'Updated Estimates of the Impact of COVID-19 on Global Poverty: Looking Back at 2020 and the Outlook for 2021'. World Bank, 11 January. Available at: https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty-looking-back-2020-and-outlook-2021 (accessed 11 December 2021).

Mancini, G., and G. Vecchi (2021). 'On the Construction of a Consumption Aggregate for Inequality and Poverty Analysis'. Washington, D: International Bank for Reconstruction and Development and World Bank.

McKenzie, D., and M. Rosenzweig (2012). 'Preface for Symposium on Measurement and Survey Design'. *Journal of Development Economics*, 1: 1–2. https://doi.org/10.1016/j.jdeveco.2012.02.002

Melesse, M.B., M. van den Berg, A. de Brauw, and G.T. Abate (2019). 'Understanding Urban Consumers' Food Choice Behavior in Ethiopia: Promoting Demand for Healthy Foods'. IFPRI-ESSP Working Paper 131. Washington, DC: IFPRI.

Miguel, E., and A.M. Mobarak (2021). 'The Economics of the COVID-19 Pandemic in Poor Countries'. NBER Working Paper 29339. Cambridge, MA, National Bureau of Economic Research (NBER). https://doi.org/10.3386/w29339

Mundlak, Y. (1978). 'On the Pooling of Time Series and Cross Section Data'. *Econometrica: Journal of the Econometric Society*, 46(1): 69–85. https://doi.org/10.2307/1913646

Pape, U. (2021). 'Measuring Poverty Rapidly Using Within-Survey Imputations'. Policy Research Working Paper 9530. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-9530

Pape, U., and J. Mistiaen (2015). 'Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu High Frequency Survey'. Policy Research Working Paper 8430. Washington, DC: World Bank.

Pradhan, M. (2009). 'Welfare Analysis with a Proxy Consumption Measure: Evidence from a Repeated Experiment in Indonesia'. *Fiscal Studies*, 30: 391–417. https://doi.org/10.1111/j.1475-5890.2009.00101.x

Raghunathan, T.E., and J.E. Grizzle. (1995). 'A Split Questionnaire Survey Design'. *Journal of the American Statistical Association*, 90: 54–63. https://doi.org/10.1080/01621459.1995.10476488

Roberts, C., G. Eva, N. Allum, and P. Lynn (2010). 'Data Quality in Telephone Surveys and the Effect of Questionnaire Length: A Cross-National Experiment'. ISER Working Paper 2010–36. Colchester: Institute for Social and Economic Research (ISER), University of Essex.

Sánchez-Páramo, C., R. Hill, D.G. Mahler, A. Narayan, and N. Yonzan (2021). 'COVID-19 Leaves a Legacy of Rising Poverty and Widening Inequality'. The World Bank, 7 October. Available at: https://blogs.worldbank.org/developmenttalk/covid-19-leaves-legacy-rising-poverty-and-widening-inequality (accessed 11 December 2021).

Schündeln, M. (2018). 'Multiple Visits and Data Quality in Household Surveys'. *Oxford Bulletin of Economics and Statistics*, 80: 380–405. https://doi.org/10.1111/obes.12196

Stifel, D., and L. Christiaensen (2007). 'Tracking Poverty Over Time in the Absence of Comparable Consumption Data'. *The World Bank Economic Review*, 21: 317–41. https://doi.org/10.1093/wber/lhm010

Sumner, A., C. Hoy, and E. Ortiz-Juarez (2020). 'Estimates of the Impact of COVID-19 on Global Poverty'. WIDER Working Paper 2020/43. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/800-9

Swindale, A., and P. Bilinsky (2006). 'Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide'. Washington, DC: Food and Nutrition Technical Assistance Project, Academy for Educational Development.

Tourangeau, R., L.J. Rips, and K. Rasinski (2000). *The Psychology of Survey Response*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Troubat, N., and K. Grünberger (2017). 'Impact of Survey Design in the Estimation of Habitual Food Consumption: A Study Based on Urban Households of Mongolia'. *Food Policy*, 72: 132–45. https://doi.org/10.1016/j.foodpol.2017.08.019

Van Garderen, J.K., and C. Shah (2002). 'Exact Interpretation of Dummy Variables in Semilogarithmic Equations'. *The Econometrics Journal*, 5: 149–59. https://doi.org/10.1111/1368-423X.00078

WFP (2008). 'Food Consumption Analysis: Calculation and Use of the Food Consumption Score in Food Security Analysis'. Rome: Vulnerability Analysis and Mapping Branch, WFP.

WFP (2017). *Conducting Mobile Surveys Responsibly: A Field Book for WFP Staff*. Rome: Vulnerability Analysis and Mapping Branch, WFP.

Wieser, C., A.A. Ambel, T. Bundervoet, and A.H. Tsegay (2020). 'Monitoring COVID-19 Impacts on Households in Ethiopia: Results from a High-Frequency Phone Survey of Households. Report #1'. Washington, DC: World Bank. https://doi.org/10.1596/33824

Zezza, A., C. Carletto, J.L. Fiedler, P. Gennari, and D. Jolliffe (2017). 'Food Counts: Measuring Food Consumption and Expenditures in Household Consumption and Expenditure Surveys (HCES). Introduction to the Special Issue'. *Food Policy*, 72: 1–6. https://doi.org/10.1016/j.foodpol.2017.08.007

Appendix A

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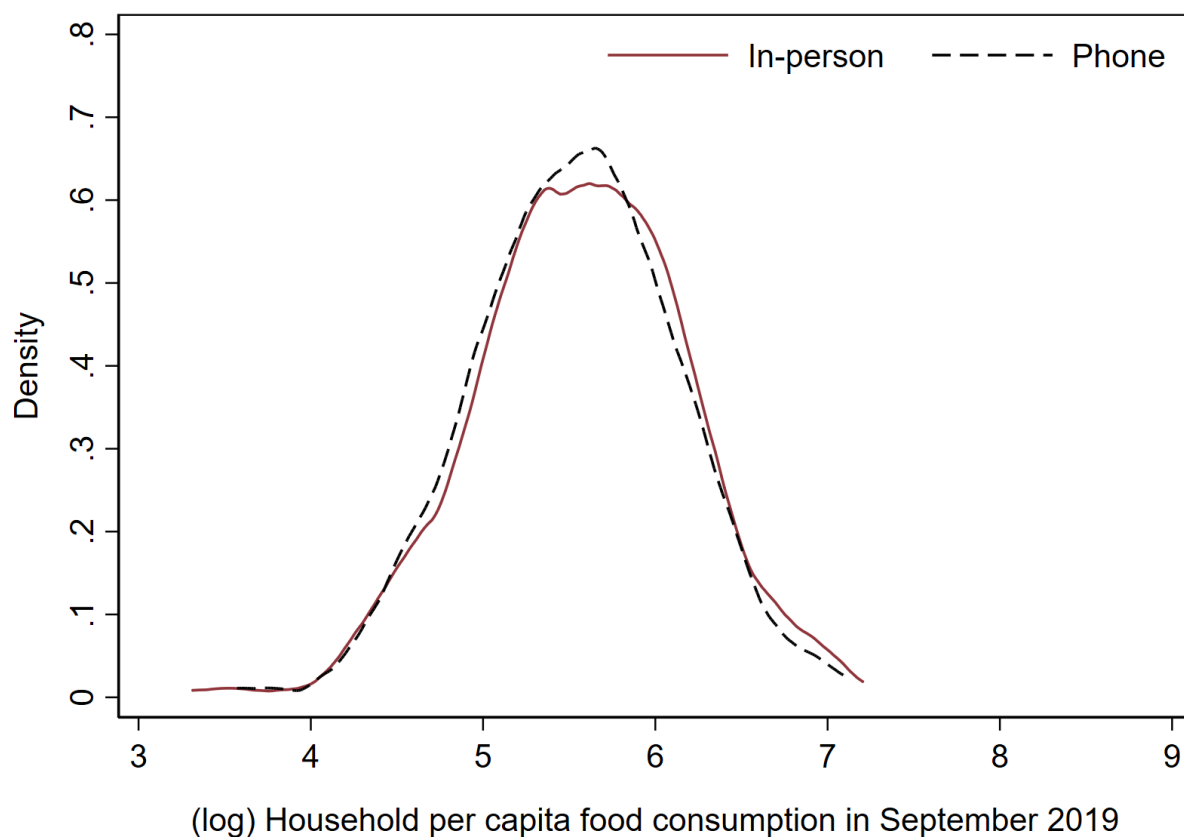
Note: N = 795 households.

Source: authors' illustration based on own data.

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 and in-person survey	August 2021	800	Food and non-food consumption modules

Source: authors' construction based on own data.

Table A2: Order of 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 food groups appearing in different order in the questionnaire; this table shows the order of food groups in both questionnaire types.

Source: authors' construction based on own data.

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***	−0.004
	(0.003)	(0.004)
Phone survey mode	−0.008*	−0.004
	(0.005)	(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: OLS 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 of 1 if the household reported having consumed the item in the past week, 0 otherwise; 0/1 = binary variable; standard errors are clustered at the food item level and reported in parentheses; statistical significance denoted with * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: authors' construction based on own data.

Figure A2: Impact of phone survey mode on household consumption of different food groups
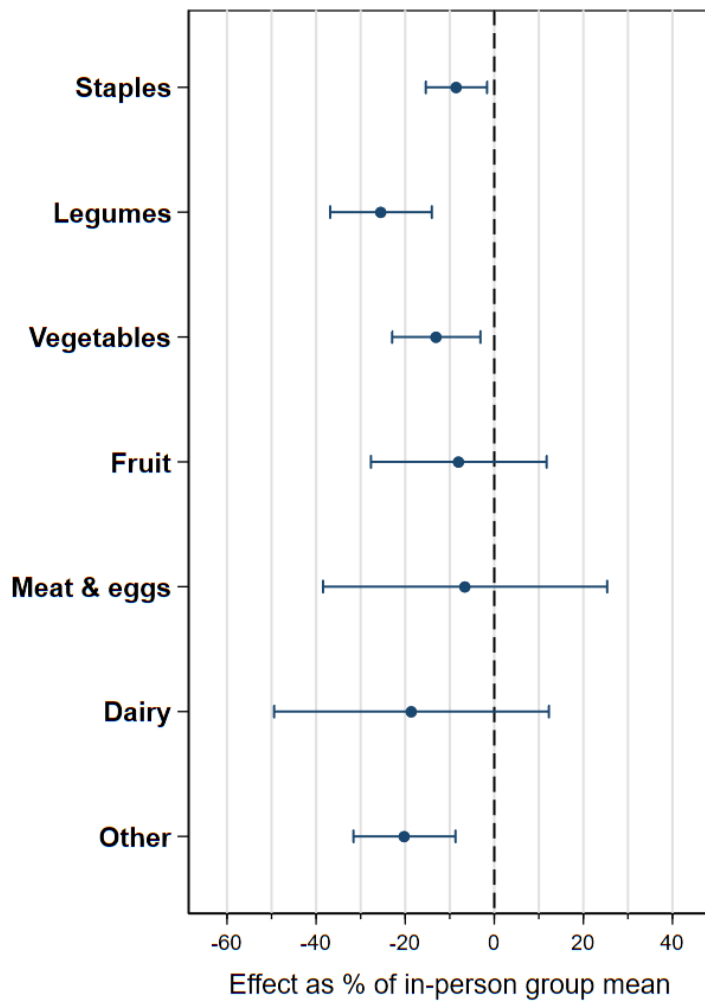
Note: based on OLS regression; unit of observation is household; N = 795; all regressions included household-level controls (household size, indicator variable for male-headed households, and head's education in years) and sub-city fixed effects; dots quantify the difference in household per capita consumption 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 percentage of mean household per capita consumption expenditure value reported in the in-person group; capped bars are 95 per cent confidence intervals, calculated from standard errors clustered at the enumeration area level.

Source: authors' illustration based on own data.

Figure A3: Predicted and observed first-digit distributions, by survey mode

Note: N = 10,526 for the in-person group and 9,042 for the phone group.

Source: authors' illustration based on own data.

Table A4: Testing differences in Euclidean distance to the distribution predicted by Benford's law

	(1)	(2)
Phone survey mode	0.328**	0.328**
	(0.156)	(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 reported in parentheses; statistical significance denoted with * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: authors' construction based on own data.

Table A5: Robustness to controlling for enumerator characteristics

	(1)	(2)	(3)	(4)
Dependent variable			(ln) Household per capita consumption	
Phone survey mode	−0.262***	−0.269***	−0.263***	−0.263***
	(0.054)	(0.054)	(0.059)	(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^2	0.288	0.290	n/a	n/a
R^2 within	n/a	n/a	0.224	0.224
R^2 between	n/a	n/a	0.600	0.652
R^2 overall	n/a	n/a	0.286	0.290

Note: OLS 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 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
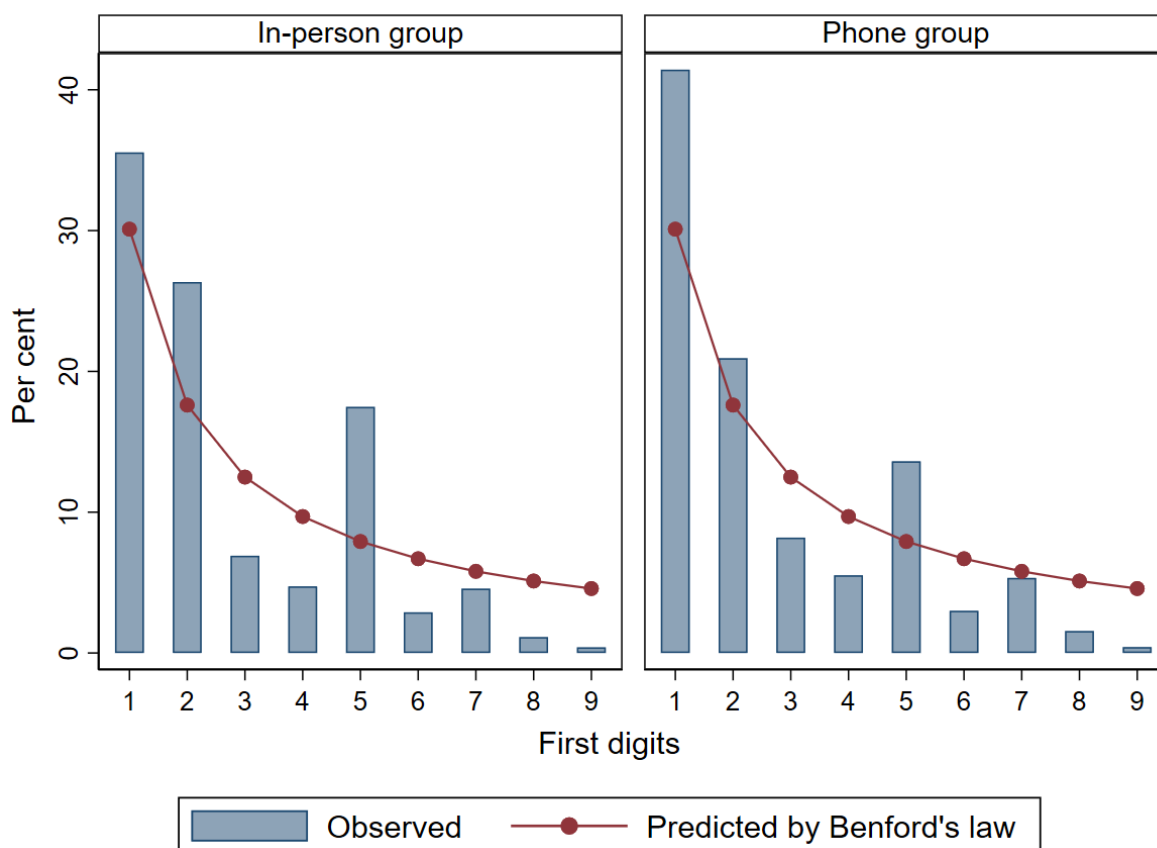
Source: authors' construction based on own data.

Figure A4: Robustness to leaving one enumerator out of the dataset at a time
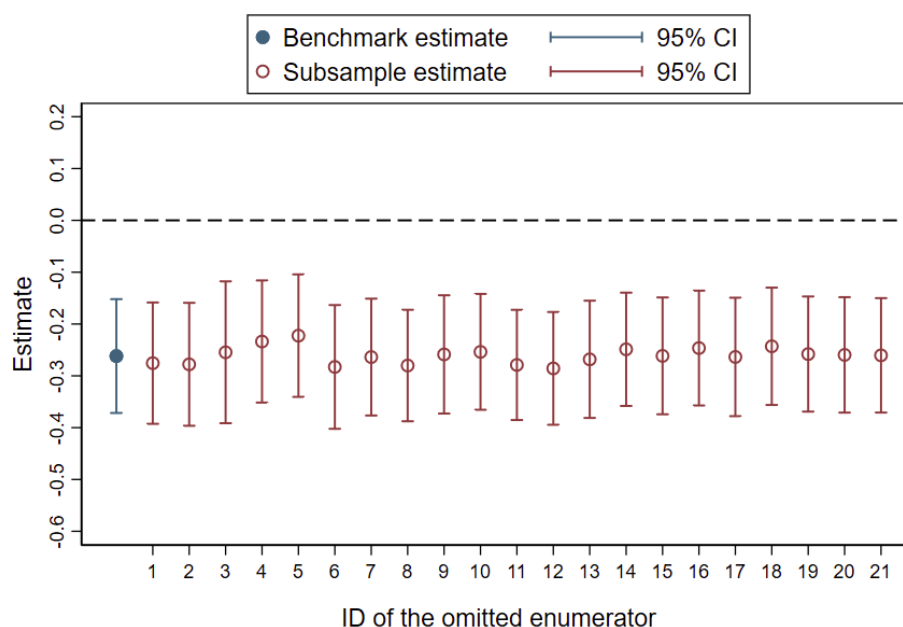
Note: blue solid dot represents the benchmark OLS estimate for full sample reported in column 2 of Table 3; maroon hollow dots are equivalent OLS estimates for 21 different subsamples when one enumerator is dropped from the dataset; capped vertical lines represent the corresponding 95 per cent confidence intervals.

Source: authors' illustration based on own data.

Appendix B: questionnaire

The questionnaire administered in the survey was in electronic format and translated into Amharic.

Section 0. Pre-interview information		
Location Information	Name	Code
Region	[prefilled]	[prefilled]
Sub-city/zone	[prefilled]	[prefilled]
Woreda	[prefilled]	[prefilled]
Kebele	[prefilled]	[prefilled]
Information for household ID number		
Household unique ID	[prefilled]	
Household (Assigned Code)	[prefilled]	
House No (Official if any)	[prefilled]	
Name of respondent	[prefilled]	[update]
Gender of the respondent: 1. Female, 2. Male	[prefilled]	[update]
Mobile phone number (1)	[prefilled]	[update]
Mobile phone number (2)	[prefilled]	[update]
Age of the respondent	(In completed years)	
Enumerator:	Supervisor:	
Date of interview:	Date of check:	

Section 1. Household composition
Including you, how many people live in your household? (people or household members who live and eat together most of the time in your home, that is more than 6 months of the year or more than 3.5 days of the week, on average)
How many household members are less than 15 years old?
How many household members are 65 years or older?

Section 2. Food consumption in the last 7 days (version 1)					
I would like to ask you about all the different foods that your household members have eaten in the last 7 days. Could you please tell me how many days in the past week and how much your household has eaten the following foods?					
ENUMERATOR: First ask whether the household consumed a given food item over the past 7 days and then about the number of DAYS eaten for each food item. Only after that ask about the quantities that the household reported to have consumed.					
Code	Food item	Did your household consume [item] over the last 7 days?	If Yes (consumed), Number of DAYS eaten in past 7 days (1–7 days)	If Yes, how much was consumed:	
		1. Yes 2. No		Quantity	Unit
f1	Lemon				
f2	Banana				
f3	Avocado				
f4	Orange/mandarin				
f5	Mango				
f6	Papaya				
f7	Dates (zebib)				
f8	Pineapple				

f9	Pome (apple)				
f10	Watermelon				
f11	Guava				
f12	Beles				
f13	Pears				
f14	Strawberries				
f15	Graps				
f16	Cherries				
f17	Kok (peach)				
v1	Onions				
v2	Tomato				
v3	Garlic				
v4	Pepper green				
v5	Ethiopian kale				
v6	Carrot				
v7	Head cabbage				
v8	Lettuce				
v9	Beet root				
v10	Spinach				
v11	Green beans				
v12	Pumpkin				
v13	Celery (savory, fennel, leaves)				
v14	Edible wild vegetables & weeds				
v15	Leek				
v16	Cauliflower and Swiss chard				
v17	Shiferaw/aleko				
v18	Mushroom				
v19	Broccoli				
v20	Eggplant				
v21	Zucchini				
v22	Red pepper				
v23	Cucumber				
v24	Potato/sweet potato				
c1	Teff black, flour				
c2	Teff mixed, flour				
c3	Teff white, flour				
c4	Wheat white, flour				
c5	Wheat emmur, flour				
c6	Wheat, whole grain				
c7	Barley, flour				
c8	Barley, whole grain				
c9	Maize, flour				
c10	Maize (fresh, roasted, boiled)				
c11	Rice				
c12	Sorghum, flour				
c13	Enjera				
c14	Bread (wheat)				
c15	Spaghetti				
c16	Macaroni				
c17	Biscuits				
c18	Bread (other cereals)				
c19	Beso (milled barley)				
c20	Bula flour				
c21	Cerefam, baby food				
c22	Mitin flour, home made				
c23	bulla				
c24	Kocho				
p1	Field peas, flour (shiro)				
p2	Field peas, split (kik)				
p3	Lentils, split (kik)				
p4	Lentils, whole grain				
p5	Faba beans, flour (shiro)				

p6	Faba beans, split (kik)				
p7	Faba bean, whole grain/roasted				
p8	Chick peas, flour (shiro)				
p9	Chick peas, split (kik)				
p10	Chick peas, whole grain/roasted				
p11	Grass peas, flour (shiro)				
p12	Mixed pulses, flour (shiro)				
m1	Beef				
m2	Sheep meat				
m3	Goat meat				
m4	Chicken				
m5	Fish (fresh, frozen, flate)				
ed1	Eggs, indigenous				
ed2	Eggs, non-indigenous (hybrid)				
ed3	Cow milk, raw				
ed4	Cottage cheese				
ed5	Powdered milk				
ob1	Edible oil, imported				
ob2	Edible oil, local				
ob3	Butter, refined and spiced				
ob4	Butter, unspiced but refined				
ob5	Butter, unspiced and non-refined				
ob6	Edible vegetable butter				
s1	Tea leaves				
s2	Coffee beans				
s3	Chaat (qat)				
s4	Sugar				
s5	Salt				
s6	soft drink (leslasa)				
s7	packed fruit juice				

Section 3. Food consumption outside the house in the last 7 days				
Code			In the past 7 days, did members of this household consume any of the following meals or drinks away from home? (from restaurants, relatives or friends)	How much did you or other household members pay, in total in the last 7 days for [MEAL/DRINK]? If free, please estimate what it would have cost if you had to pay.
			1. Yes 2. No	Birr
1	Full meals (e.g., injera made of teff/millet/barley with any type of stew, kocho/kocho with meat, rice with sauce, etc.)	Breakfast		
2		Lunch		
3		Dinner		
4	Snacks such as kolo, bread, biscuits, cakes, etc.			
5	Dairy-based beverages such as milk, yoghurt, etc.			
6	Vegetables and roasted or boiled items such as (carrot, potatoes, boiled/roasted corn, sugarcane, etc.)			
7	Non-alcoholic drinks (coffee, tea, fruit juice, soda, etc.)			
8	Alcoholic drinks			
Section 4. Non-food expenditure				
Last ONE MONTH				
Code	Over the past one month, did your household purchase or pay for any [ITEM]?			How much did your household pay in total?

1	Matches	1. Yes 2. No	Birr
2	Batteries		
3	Candles (tua'af)		
4	Laundry: soap/OMO/endod/besana leaves (local and imported)		
5	Hand/body soap		
6	Other personal care goods and services (incl. sendal, matent, insecticides, salon, barber, toilet paper)		
7	Electricity, charcoal, firewood, kerosene		
8	Water		
9	Telephone, TV/cable subscriptions, postage		
10	Cigarettes, tobacco, suret, gaya		
11	Transport (fare and/or cost of diesel, petrol per month)		
12	Salaries for servants, private guards, community guards, babysitters		
13a	Own house? (if yes, ask them to estimate how much it would have cost if rented)		
13b	Rented house (if yes, ask the rent per month)		
14	Shoe shining, car wash, etc.		
15	Maintenance expenses (tailors, repairs, plumbers, etc.)		
16	Toys, sports, events, other entertainment		
17	Milling services		

Last 12 MONTHS			
Code	Over the past 12 months, did your household purchase or pay for any [ITEM]?		How much did your household pay in total?
1	Clothes/shoes/fabric for MEN (18 years and older)	1. Yes 2. No	Birr
2	Clothes/shoes/fabric for WOMEN (18 years and older)		
3	Clothes/shoes/fabric for BOYS (less than 18 years)		
4	Clothes/shoes/fabric for GIRLS (less than 18 years)		
5	Kitchen equipment (dishes, cooking pots, pans, etc.)		
6	Linens (sheets, towels, blankets) and mattress, etc.		
7	Furniture		
8	Lamp/torch/solar power		
9	Ceremonial expenses		
10	Contributions to informal social security institutions (incl. *Iddir, mahiber*)		
11	Donations to religious institutions (incl. churches and mosques)		
12	Contributions to community development activities (road, school, health, water, etc. developments)		
13a	Contributions to social and political activities (Red Cross, sport, political parties)		
13b	School fees (incl. books, money send to students, etc.)		
14	Health expenses (incl. clinic/hospital fees, medicine, etc.)		
15	Clothes/shoes/fabric for MEN (18 years and older)		

Note: estimate the cash value if payment is made in kind.

Source: authors' construction.