INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

RESEARCH PROGRAM ON Agriculture for Nutrition and Health
CGIAR
Led by IFPRI

IFPRI Discussion Paper 02087

December 2021

Measuring Consumption over the Phone

Evidence from a Survey Experiment in Urban Ethiopia

Gashaw T. Abate

Alan de Brauw

Kalle Hirvonen

Abdulazize Wolle

CGIAR Research Program on Agriculture for Nutrition and Health

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Gashaw T. Abate (g.abate@cgiar.org) is a Research Fellow in the Markets, Trade, and Institutions Division at the International Food Policy Research Institute (IFPRI), Washington, DC.

Alan de Brauw (a.debrauw@cgiar.org) is a Senior Research Fellow in the Markets, Trade, and Institutions Division at IFPRI, Washington, DC.

Kalle Hirvonen (k.hirvonen@cgiar.org) is a Senior Research Fellow in the Development Strategy and Governance Division at IFPRI, Addis Ababa, Ethiopia.

Abdulazize Wolle (abdulazize.wolle@gmail.com) is in the Economics Department at the State University of New York at Albany, Albany, New York.

Contents

Tables

Figure

ABSTRACT

The paucity of reliable and timely household consumption data in many low- and middle-income countries have made it practically impossible to assess how global poverty has evolved during the COVID-19 pandemic. Despite the burst of phone surveys, there has been few attempts to collect household consumption data. To test the feasibility of collecting consumption data over the phone, we conducted a survey experiment in urban Ethiopia, randomly assigning a balanced sample to either a phone or an in-person interview. The average value of per capita consumption is 23 percent lower, and the estimated poverty headcount is twice as high in the phone survey relative to the in-person survey. We see evidence of survey fatigue occurring early on in phone interviews but not in in-person interviews, and the bias is correlated with household characteristics. While the phone survey mode provides lower costs, it cannot replace in-person surveys for household consumption measurement.

Keywords: Survey experiment, Phone survey, Survey fatigue, Food consumption, Household surveys

ACKNOWLEDGMENTS

We thank Kibrom Abay, Joachim De Weerdt, and conference participants in the Methods and Measurement Conference organized by the IPA-Northwestern Research Methods Initiative for useful comments. We are grateful to the households that participated in this study. We acknowledge NEED NUTRITIONAL and their survey staff (Abraha Weldegerima, Alemayehu Deme, Nadiya Kemal, Fikirte Sinkineh, Meskerem Abera, Rediet Dadi, Fitsum Aregawi, Asnakech Yosef, Ekiram Shimelis, Biruktawit Abebe, Yared Tilahun, Tesfaye Eana, Yosef Regasa, Ashenafi Hailemariam, Selamawit Genene, Shambel Asefa, Teshale Hirpesa, Getachew Buko, Habtamu Ayele, Ayinalem Reta, Mohamed Meka, Kibrom Tadesse, and Huluhager Endashaw) for collecting the data. Any and all errors are the sole responsibility of the authors. This work was undertaken as part of, and funded by, the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH). The opinions expressed here belong to the authors, and do not necessarily reflect those of A4NH or CGIAR.

ACRONYMS

CAPI	Computer-assisted personal interview
CATI	Computer-assisted telephone interview
CSA	Central Statistical Agency
DHS	Demographic and Health Surveys
FCS	Food consumption score
HCES	Household Consumption Expenditure Surveys
HDDS	Household diet diversity score
IFPRI	International Food Policy Research Institute
IRB	Institutional review board
LSMS	Living Standards Measurement Survey
OLS	Ordinary least squares
RCT	Randomized control trial
RECOVR	Innovations for Poverty Action
SDG	Sustainable Development Goals
WFP	World Food Programme

I. INTRODUCTION

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

The main pivot by researchers during the early part of the pandemic was to begin conducting phone surveys. There was a veritable explosion of efforts to collect some type of data to monitor situations over the phone, including major coordinated efforts by Innovations for Poverty Action (RECOVR) and the World Bank LSMS team (Gourlay, et al., 2021). These efforts have played an important role in helping us understand some of the socioeconomic consequences of the pandemic. In terms of living standards, these surveys have generally asked questions about job loss and loss of income, and they tend to show substantial, concerning negative effects (Egger, et al., 2021, Josephson, et al., 2021, Miguel and Mobarak, 2021). Yet these findings are all based on crude measures; e.g., asking whether household income was lower, the same, or higher than it had been at the same time of the year 12 months ago.

Although these measures have allowed for some information about how living standards have changed during the pandemic, there remain obvious ways that phone surveys cannot replace in-person surveys. Some variables require physical measurement; for example, it is impossible to study how stunting prevalence is evolving among children under 5 years of age without in-person data collection. And other variables such as poverty incidence require complex measurement, such as the value of household

consumption expenditures[1], so most phone surveys have not attempted to collect them, in trying to minimize the time spent on the phone.

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

The surveys that have tried to collect consumption data over the phone during the pandemic suggest the increases in poverty incidence are not as severe as either the crude income measures or models would suggest. Egger, et al. (2021) report on phone surveys in Kenya and Sierra Leone that collected data on food consumption in both countries and non-food in Kenya, and find that the value of food consumption increased in both countries, offset by a decline in non-food consumption in Kenya.[2] Janssens, et al. (2020) study a sample of households in Kenya who keep 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 had changed.

These surveys suggest it might be plausible to conduct phone surveys to measure consumption and therefore poverty incidence, particularly if survey efforts first attempt to develop some rapport with

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

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

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

We can then compute poverty incidence using both the consumption measures generated by our phone sample, versus the in-person sample. Note that it is best to at least initially be agnostic about which sample provides closer to a "true" approximation of the distribution of consumption, and therefore poverty incidence. Indeed, an important challenge in survey experiments such as ours is that we do not observe the "true value" against 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 consumption items appear in the 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. In heterogeneity analysis, we find that bias is attenuated with more educated household heads and is positively related to household size. [3] This finding implies that the measurement error in phone survey mode is non-classical in nature and, as a result, cannot be easily corrected with standard methods used in the literature (Bound, et al., 2001).

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

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 the consumption measures that are 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 to those from an in-person survey. Finally, many researchers have hypothesized that the phone survey mode is likely to be considerably more vulnerable to response fatigue than the in-person mode, leading to the widespread recommendation to keep phone-based interviews short, and to avoid complex questions (Dabalen, et al., 2016, Gourlay, et al., 2021). Our results provide empirical support to this hypothesis. However, in our case, both survey modes result in similar estimates when measuring diet-based food security suggesting that the phone survey mode is appropriate for measuring simpler and cognitively less demanding indicators, as long as the interview time is kept relatively short (Abay, et al., 2021).

II. THE SURVEY EXPERIMENT, DATA AND METHODS

2.1. The survey experiment

We designed a survey experiment to understand the implications of using 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 instruments in both survey modes were identical and had four sections. The interview began with a brief section containing only three questions needed to construct household size and its dependency ratio. In the first main section, respondents were asked to report on the household's food consumption for each item from a list of 118 food items, grouped into eight food groups. We first went through this list of 118 items asking whether the household consumed the item in the past seven days or not. The survey instrument was programmed to carry forward all items that were consumed in the past seven days to the next sub-section that asked about the consumption frequency ('on how many days was the item consumed') and quantity ('amount consumed') within the seven-day period. The second main section of the questionnaire included a short module asking household 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). The full questionnaire is included in the Online Appendix.

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	7 days	7 days
Recall period in the non-food consumption module (*)	1 month or 12 months	1 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: (*) 1 month for non-food expenditures such as toiletries and utilities and 12 months for expenditures such as school fees and health expenses.

2.2. Household sample

The household sample for this survey experiment originates from a randomized control trial (RCT) to assess the impact of video-based behavioral change communication on fruit and vegetable consumption in Addis Ababa, Ethiopia (Abate, et al., 2021). The baseline and endline surveys for the RCT took place in September 2019 and February 2020, respectively.[4] The sample of 930 households was randomly selected from six sub-cities, 20 *woredas* (districts), and 40 *ketenas* (neighborhoods; or clusters of households) within Addis Ababa.[5] Comparison of household characteristics against those reported in other surveys from Addis Ababa suggest that the sample is representative of the households residing in the city (Hirvonen, et al., 2020).

The endline survey was administered just before the COVID-19 pandemic was declared in 2020, a setup that was highly optimal for launching COVID-19 phone surveys. To monitor the food security situation in Addis Ababa during the pandemic, we selected a random sub-sample of 600 households for monthly phone surveys (Hirvonen, et al., 2021). In total, four phone survey rounds were carried out between June

[4] 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.

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

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). The survey experiment contrasting consumption data collected via in-person and phone modes was administered over a 10-day period in September 2021 (i.e., one year after the last COVID-19 phone survey).[6] The sampling frame for this study was based on 895 households that were interviewed during the in-person survey conducted in February 2020, the endline survey of the video RCT. Out of the 895 households, 448 were randomly selected for an in-person interview and 447 for a phone interview.[7] A total of 797 households were interviewed; 421 in the in-person group and 376 in the phone group.[8] Administering the consumption modules over the phone took 41 minutes on average (median) and while the average (median) interview duration was 43 minutes for an in-person visit. The quality of the connection was generally good across the phone interviews, and based on enumerators' assessment, rarely affected the quality of the interview.[9]

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

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

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

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

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

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 the Hawassa University in Ethiopia. Informed oral consent was obtained from all participants at the start of the interview. Enumerators provided respondents a brief overview of the study objectives and informed them that their participation in the study was entirely voluntary.

2.3. Data

Food consumed at home was reported in terms of quantities consumed, which we converted into local currency units (Ethiopian birr) using retail price data collected by the Central Statistical Agency (CSA) of Ethiopia. We used the retail price data for Addis Ababa from February 2020 (the latest month available to us) and then used a food-specific consumer price index for Addis Ababa to express our food consumption data in September 2021 prices. Food consumption outside the home as well as non-food expenditures 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 comes from the HCES collected every five years. The HCES survey is conducted throughout the Ethiopian calendar year to address consumption seasonality and covers nearly 400 food items and more than 850 non-food items. The latest HCES was administered in 2015/16, after which prices for food and non-food items have both been rising annually at a double-digit rate. Considering the high inflation rate and the considerable methodological differences between our survey and the HCES, we do not attempt to update the HCES poverty line for September 2021. Instead, we calibrate our poverty line for the in-person sample to match the 16.8 percent poverty headcount based on the national poverty line and reported for Addis Ababa using the 2015/16 HCES (FDRE, 2018).

We also use our food consumption data to study how the phone survey mode affects household dietary diversity, an indicator of household food security (Hoddinott and Yohannes, 2002). First, we computed the HDDS of Swindale and Bilinsky (2006) by grouping the 118 food items in our consumption module into 12 food groups: cereals; roots and tubers; vegetables; fruits; meat, poultry and offal; eggs; fish and seafood; pulses, legumes and nuts; milk and milk products; oil and fats; sugar and honey; and miscellaneous foods. The HDDS is a sum of all food groups from which the household consumed food items during the seven-day recall period, with a minimum of one and maximum of 12. Second, we constructed the 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[10] and allocating more weight to protein-rich foods. The weighted FCS index ranges between zero and 112, with higher scores indicating a better food security situation.

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

[10] 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).

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.

2.4. Estimation methods

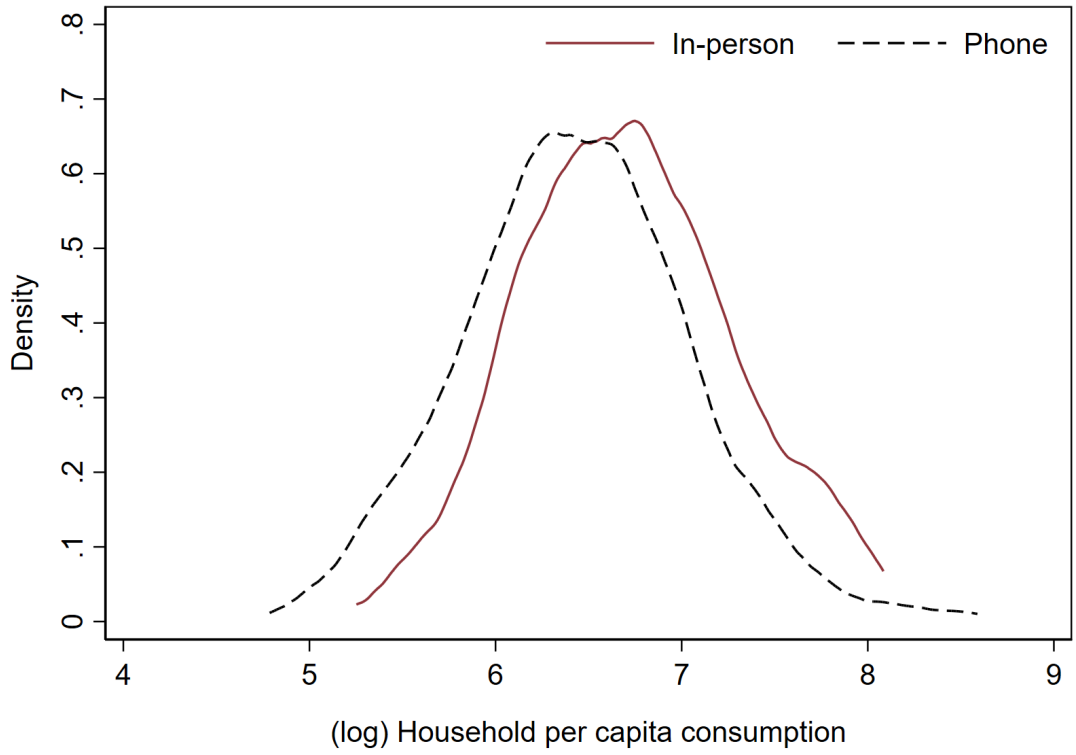
We quantify the difference in reported household per capita consumption values across the two groups using ordinary least squares (OLS). In our most basic model, we regress both the per capita consumption value and its logarithm on a binary treatment variable valued one if the household was randomly selected into the phone group, and zero if into the in-person group. In subsequent models, we control for differences in basic household characteristics (household size, and household head's sex and level of education in years) as well as sub-city fixed effects. Finally, when we discuss percentage differences derived from the coefficients in semi-log regressions they are based on the approximate unbiased variance estimator of 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. The standard errors in all household level regressions are clustered at the enumeration area (*ketena*) level.

III. 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 on the left of the distribution estimated for the in-person group, indicating that the phone survey resulted in smaller reported consumption values across the board.

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



Note: N=795 households

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 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.

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

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

3.2. Components of consumption

Food consumed at home represents 50.3 percent of the total consumption in the in-person group and 55.8 percent in the phone survey group.[11] The regression estimates reported in Column 1 of Table 5 indicate that the reported per capita food consumption values are 13 percent lower on average when the phone

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

survey mode is used (95-% CI: -5.5; -20.7). However, we do not find strong evidence to suggest that some food groups were more affected than others. We re-estimated the main regression using the value of food consumption for each of seven categories of food as the dependent variable; in Figure A2 in the Appendix, we observe that all the coefficient estimates are negative and suggest 10 to 25 percent lower consumption, with overlapping confidence intervals.

About 60 percent of the households in our sample report to have consumed food items outside of their home in the past seven days. This reporting incidence varies by survey mode with households in the phone survey group being 13 percentage points less likely to report to have consumed foods outside their home (Table 4, column 2). A regression based on a non-logged outcome variable shows that the food expenditures outside of the home are 40.2 percent lower in the phone group relative to the in-person group (Table 4, column 3).[12]

All the households in our sample report positive (non-zero) non-food consumption values. Column 4 in Table 4 shows the impact of the phone survey mode when the dependent variable is logged weekly per capita non-food consumption. On average, the phone survey mode lowers the reported non-food consumption by 30.1 percent (95-% CI: -15.5; -42.1).

[12] Considering that the mean value in the in-person group is 53.91 birr, the difference of 21.66 birr estimated with OLS translates to 40.2 percent (21.66/53.91).

Table 4: Impact of phone survey mode on components of household consumption

Dependent variable:	(1) (ln) Household per capita food consumption at home	(2) Household consumed food outside home (0/1)	(3) Household per capita food consumption outside home	(4) (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: Ordinary least squares regression. Unit of observation is household. Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with * p < 0.10, ** p < 0.05, *** p < 0.01.

3.3. Poverty estimates

Next, we estimate the impact of using phone survey mode on poverty estimates. Since poverty is defined at the individual level, we need to convert our data from household to individual level. To do so, we use a weighted least square regression method where the weights are frequency weights based on household size. Using our calibrated poverty line, in Table 5 we estimate that poverty rate is 17 percentage points higher when phone survey mode is used compared to when consumption data are collected through in-person visits (95%-CI: 9.99; 24.1). Since the poverty rate in the in-person sample is calibrated at 16.82 percent, using the phone survey mode effectively doubles the poverty rate in this context.

Table 5: Impact of phone survey mode on poverty rate

	(1)	(2)
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 is poverty rate. Household level controls include household size (number of members), indicator variable for male-headed households, and household head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with * p < 0.10, ** p < 0.05, *** p < 0.01.

3.4. Measures of food security

In Table 6, we report the impacts of using the phone survey mode on two widely used diet-based food security measures, HDDS and FCS. Both can be computed from the food consumption survey data. All four reported impact estimates are relatively small in magnitude and not statistically different from zero. The HDDS and FCS do not require respondents to estimate quantities consumed, only whether the food item was consumed in the past 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 when measuring diet-based food security to in-person surveys but leads to much lower estimates of the value of household food or non-food consumption.

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

IV. EXPLORING MECHANISMS

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, we would conclude that the poverty headcount is twice as high using the phone survey data as the data collected in-person. Here, we study whether survey fatigue can help explain differences between results of the two survey modes.

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

Evidence from in-person surveys suggests that survey fatigue can lead to under reporting and overall deterioration of data quality in some settings (Ambler, et al., 2021, Roberts, et al., 2010, Schündeln, 2018), but not always (Laajaj and Macours, 2021). In a recent phone survey conducted in rural Ethiopia, Abay, et al. (2021a) estimate that delaying the timing of a dietary diversity module by 15 minutes

increased the likelihood that the respondents reported not to have consumed from certain food groups, resulting in an 8 percent decline in the mothers' dietary diversity score.[13]

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.[14] Specifically, we implemented two versions of this food consumption module, ordering the food groups differently (see Table A1 in the Appendix). For example, in version 1, mango appeared as the 5th item while in version 2, it appeared as the 73rd item. Similarly, in version 1, rice was the 52nd item on the list while in version 2, it was the 11th item on the list. Exploiting this variation, we use the food item-level data to construct a variable that takes on the value of 1 when each food appears later in the questionnaire relative to the other version, and 0 otherwise. Using the example above, this variable would be 1 when mangoes appear as the 73rd item, and when rice appears as the 52nd item. Using our food item level data, we then regressed the weekly household per capita consumption of the food item on this binary variable capturing the item's relative position in the questionnaire, and the indicator variable for the phone survey mode. To assess whether the impact of delaying when the item is asked in the module differs between phone and in-person survey modes, we interact the two variables and include the interaction term in the regression. In these regressions we control for food item fixed effects, meaning that our estimates are identified from variation in the survey mode or relative position in the questionnaire for the same food items. As additional controls, we include household size, an indicator variable for male-headed households, the head's years of education, and sub-city fixed effects.

Table 7 provides the results. In Column 1, we estimate the model without the interaction term. Moving the item later in the questionnaire results in a report that is, on average, 5.8 percent lower for the item than

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

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

if it takes on its earlier position.[15] Using the phone survey mode, the average report suggests the value of consumption is 15.5 percent lower than found with the in-person survey mode.[16] In Column 2, we estimate the model with the interaction term. The basic variable now captures the effect of placing the item later in the questionnaire in the in-person survey; this coefficient is close to zero and not statistically significant. The CI is relatively tight around zero (95-% CI: -0.0167; 0.0016) indicating that survey fatigue does not play a role in the in-person survey mode, at least in this relatively early part of the questionnaire. In contrast, the coefficient on the interacted variable is negative, relatively large in magnitude, and statistically different from zero; it suggests that delaying an item in the phone survey mode leads to a report that is 11.9 percent lower on average than an item occurring later in the in-person survey.[17] 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)
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: Ordinary least squares regression. Unit of observation is food item consumed (or not) in each household. Number of food items is 118 and number of households is 795 resulting in 93,810 observations. Dependent variable is weekly 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.

In Table A3 in the Appendix we replicate this analysis, but now only considering the responses to the Yes/No questions if the household consumed the item or not in the seven-day period. Interestingly, all coefficients in the interacted model appear insignificant, implying that only consumption quantity reports

[15] Calculated as -0.230/3.97 using the estimates reported in column 1 of Table 7.
[16] Calculated as -0.615/3.97 using the estimates reported in column 1 of Table 7.
[17] Calculated as [-0.014+(-0.458)]/3.97 using the estimates reported in column 2 of Table 7.

are affected, but not responses on whether the household consumed the item or not. This finding is in line with our earlier result according to which diet-based food security measures do not seem to be affected by variation in survey mode.

4.2. Heterogeneity

The results show that using the phone survey mode leads to substantial under-estimation 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 (or 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) is used as the dependent variable. For food household consumption, we see that the bias decreases with household head's education and increases with household size.[18] The former result suggests that more educated households can better overcome survey fatigue in phone surveys. In contrast, the cognitive burden increases with the number of consumption events within the recall period (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 coefficient estimates are of the same sign and magnitude as the food consumption coefficients, but not statistically different from zero, possibly because of the larger variation in the data relative 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.

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

Table 8. Regression results from interaction models

Dependent variable:	(1) (ln) weekly food consumption per capita	(2) (ln) weekly food consumption per capita	(3) (ln) weekly non-food consumption per capita	(4) (ln) weekly 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: Ordinary least squares regression. Unit of observation is household. Household level controls include household size (number of members), indicator variable for male-headed households, and head's education in years. Standard errors are clustered at the enumeration area level and reported in parentheses. Statistical significance denoted with * p < 0.10, ** p < 0.05, *** p < 0.01.

V. 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.[19] 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 A4 in the appendix). Second, to explore whether one poorly performing enumerator in the phone survey group could explain our results, we assess the sensitivity of our result to omitting one enumerator at a time from the sample. Results are remarkably robust to running the main regression across these 21 sub-samples (see Figure A3 in the appendix). Third, we show that our results are robust to the controlling for enumerator random effects (see Column 3 in Table A4 in the appendix) as well as Mundlak (1978) correlated random effects (see Column 4 in Table A4 in the appendix); unfortunately, we cannot control for enumerator fixed effects due to collinearity between treatment status and enumerator assignment.[20] Still, in examining Figure A3 we cannot conclude that enumerator effects had much influence on the overall estimate of the average difference between reports.

In combination with the result in Table 7, this result provides further evidence that survey fatigue among the phone survey respondents dampened their report of food consumption relative to the in-person survey. The mechanism appears to be by reducing the amount of food consumed within categories, instead of

[19] 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).

[20] The random effects estimator controls for enumerator heterogeneity by decomposing the unobserved heterogeneity to variance occurring between enumerators and within enumerators (i.e., across different interviews conducted by the same enumerator). The key assumption of the random effect estimator is that the correlation between the treatment status and the random effects is zero, or in the correlated random effects model, that it takes on a specific parameter. We acknowledge that, in our application, this assumption may not hold. However, simulation studies suggest that the 'heterogeneity bias' stemming from the violation of this assumption is relatively small (see Bell and Jones, 2015). Considering this 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.

omitting entire categories, though reports of food away from home and non-food consumption toward the end of the survey are lower as well among phone survey participants. As a result, phone surveys would still be good for measuring food security measures such as the HDDS and FCS but would not be as useful for measuring the poverty headcount or other measures related to the amount of consumption expenditures.

VI. 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, WFP (2017) had built 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).[21] Indeed, there have been only few attempts to measure household consumption to inform how the progress toward meeting the first Sustainable Development Goal of 'No Poverty' has been affected by the pandemic. Finally, there remains considerable uncertainty on the implications on the use of the phone survey mode on data quality, particularly in low- and middle-income country contexts where the pre-pandemic rollout of phone survey technology and testing had been relatively slow (Gourlay, et al., 2021).

Our research began 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

[21] At the same time, with imperfect and non-random mobile phone access in rural areas, the data may not be representative as the poor and people in more remote areas may have less access to phones or be in coverage areas when phone surveys are fielded, respectively (Ambel, et al., 2021, Brubaker, et al., 2021).

in Addis Ababa, Ethiopia, randomly assigning a balanced and representative sample either to a phone or an in-person interview mode. We find the phone survey mode leads to a statistically significant and large underestimation of household consumption. Relative to the in-person survey mode, the phone survey mode decreases the reported household per capita consumption expenditures by 23 percent, on average. Consequently, the estimated poverty rate is twice as high when the phone survey mode is used. The phone survey bias appears related to household size, suggesting that the bias is worse when there are more people eating within the household.

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

The mechanism appears to be linked to survey fatigue that results phone survey respondents to greatly under-estimate consumption quantities, but not whether they consumed the item during the recall period. The bias is attenuated by education; when the household head is more educated, the bias is reduced, suggesting that more educated individuals can overcome issues of attention. Nonetheless, the overall estimates suggest that while phone surveys can provide large cost savings (Gourlay, et al., 2021), they cannot replace in-person surveys for household consumption and poverty measurement. 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 it has been used by the WFP (2017) in their Vulnerability Analysis and Mapping surveys.

REFERENCES

Abate, G.T., K. Baye, A. de Brauw, K. Hirvonen, and A. Wolle. 2021. "Video-Based Behavioral Change Communication to Change Consumption Patterns." IFPRI Discussion Paper 02052. Washington D.C., International Food Policy Research Institute (IFPRI).

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 D.C., International Food Policy Research Institute (IFPRI).

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* 139:171-184.

Abay, K.A., G. Berhane, J.F. Hoddinott, and K. Tafere. 2021a. "Assessing response fatigue in phone surveys: Experimental evidence on dietary diversity in Ethiopia."2017. Washington D.C., International Food Policy Research Institute (IFPRI).

Abay, K.A., L.E. Bevis, and C.B. Barrett. 2021b. "Measurement Error Mechanisms Matter: Agricultural intensification with farmer misperceptions and misreporting." *American Journal of Agricultural Economics* 103:498-522.

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 D.C., The World Bank.

Ambler, K., S. Herskowitz, and M.K. Maredia. 2021. "Are we done yet? Response fatigue and rural livelihoods." *Journal of Development Economics* 153:102736.

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.

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.

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* 98:124-135.

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.

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* 3:133-153.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement error in survey data." In *Handbook of econometrics.* Elsevier, pp. 3705-3843.

Brubaker, J., T. Kilic, and P. Wollburg (2021) "Representativeness of Individual-Level Data in COVID-19 Phone Surveys." Policy Research working paper 9660. Washington D.C., The World Bank.

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.

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* 103:254-261.

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 D.C.: The World Bank.

De Weerdt, J. 2008. "Field notes on administering shock modules." *Journal of International Development* 20:398-402.

De Weerdt, J., K. Beegle, J. Friedman, and J. Gibson. 2016. "The challenge of measuring hunger through survey." *Economic Development and Cultural Change* 64:727-758.

De Weerdt, J., J. Gibson, and K. Beegle. 2020. "What can we learn from experimenting with survey methods?" *Annual Review of Resource Economics* 12:431-447.

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.

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* 34:654-669.

Dillon, B. 2012. "Using mobile phones to collect panel data in developing countries." *Journal of International Development* 24:518-527.

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.

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

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.

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* 34:418-443.

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 - 474.

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* 89:473-489.

---. 2010. "Non-classical measurement error in long-term retrospective recall surveys." *Oxford Bulletin of Economics and Statistics* 72:687-695.

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.

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*: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 D.C., Ethiopia Strategy Support Program (ESSP) of the International Food Policy Research Institute (IFPRI).

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* 103:772-789.

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.

Hughes, S., and K. Velyvis. "Tips to Quickly Switch from Face-to-Face to Home-Based Telephone Interviewing." Mathematica, Accessed: 11 December 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.

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.

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." Available at https://ssrn.com/abstract=3732529. Accessed: 14 December 2021.

Josephson, A., T. Kilic, and J.D. Michler. 2021. "Socioeconomic impacts of COVID-19 in low-income countries." *Nature Human Behaviour* 5:557-565.

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* 65:144-165.

Kopper, S., and A. Sautmann. 2020. "Best practices for conducting phone surveys." Abdul Latif Jameel Poverty Action Lab (J-PAL), Accessed: 14 December 2021.

Laajaj, R., and K. Macours. 2021. "Measuring skills in developing countries." *Journal of Human Resources* 56:1254-1295.

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*.

Lakner, C., N. Yonzan, D.G. Mahler, R.A.C. Aguilar, and H. Wu. "Updated estimates of the impact of COVID-19 on global poverty: Looking back at 2020 and the outlook for 2021." The World Bank, Accessed: 11 December 2021.

McKenzie, D., and M. Rosenzweig. 2012. "Preface for symposium on measurement and survey design." *Journal of Development Economics* 1:1-2.

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 D.C., International Food Policy Research Institute (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).

Mundlak, Y. 1978. "On the pooling of time series and cross section data." *Econometrica: journal of the Econometric Society*:69-85.

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 Series.

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, Accessed: 11 December 2021.

Schündeln, M. 2018. "Multiple visits and data quality in household surveys." *Oxford Bulletin of Economics and Statistics* 80:380-405.

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.

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. Washington D.C., FANTA FHI 360.

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* 72:132-145.

van Garderen, J.K., and C. Shah. 2002. "Exact interpretation of dummy variables in semilogarithmic equations." *The Econometrics Journal* 5:149-159.

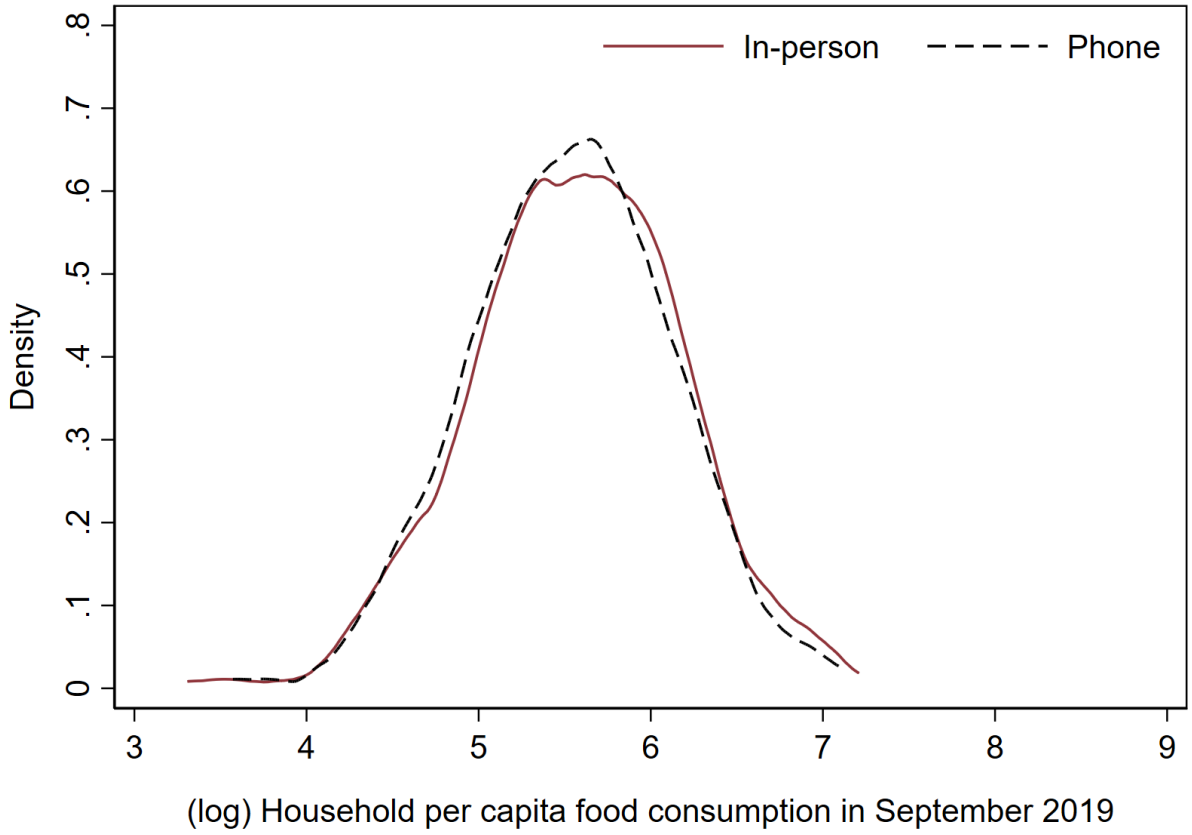
WFP. 2017. *Conducting Mobile Surveys Responsibly - A Field Book for WFP Staff*. Rome: World Food Programme (WFP), Vulnerability Analysis & Mapping (VAM).

--- (2008) "Food consumption analysis: Calculation and use of the food consumption score in food security analysis." Rome, World Food Programme (WFP), Vulnerability Analysis and Mapping Branch (ODAV).

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.

APPENDIX: ADDITIONAL RESULTS

Figure A1. Distribution of (ln) weekly food consumption per capita (in birr) in September-2019, by survey mode in August-2021



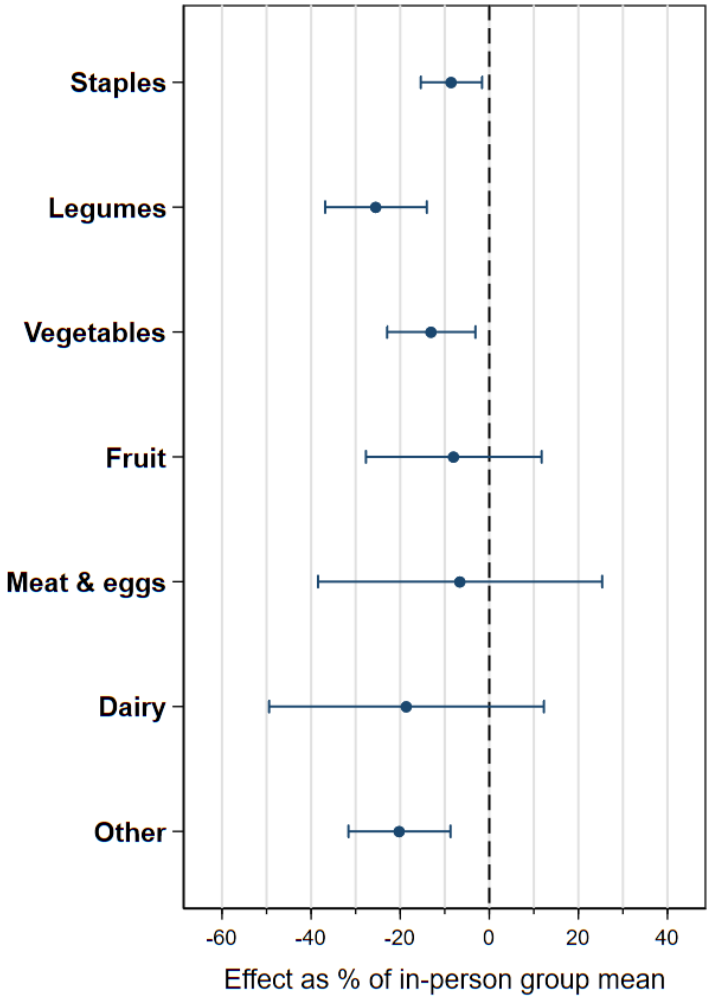
Note: N=795 households.

Table A1. Order of the food groups in the two versions of the 'food consumed at home' module

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

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

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



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

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

	(1)	(2)
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

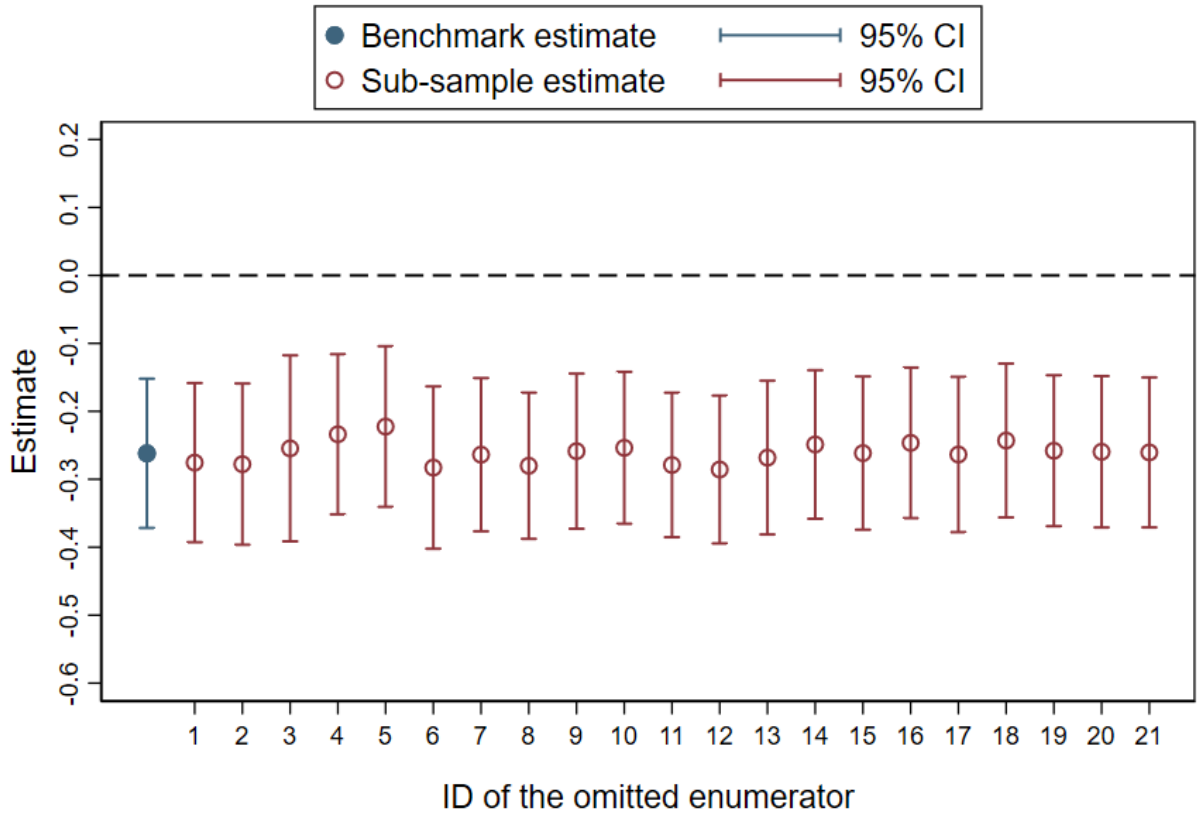
Note: Ordinary least squares regression. Unit of observation is food item consumed (or not) in each household. Number of food items is 118 and number of households is 795 resulting in 93,810 observations. Dependent variable obtains a value 1 if the household reported to have consumed the item in the past week, zero otherwise. 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.

Table A4. Robustness to controlling for enumerator characteristics

	(1)	(2)	(3)	(4)
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: Ordinary least squares regression. Unit of observation is household. Dependent variable is (ln) total per capita consumption (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.

Figure A3. Robustness of leaving one enumerator out of the dataset at a time



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

ONLINE APPENDIX: QUESTIONNAIRE

Note: the questionnaire administered in the survey was in electronic format and translated to 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					
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 consumed [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 & switzcharge				
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, boilled)				
c11	Rice				
c12	Sorghum, flour				
c13	Enjera				
c14	Bread (wheat)				
c15	Spaghetti				
c16	Macaroni				
c17	Biscuits				
c18	Bread (other cereals)				
c19	Beso (milled barely)				
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 & spiced				
ob4	Butter, unspiced but refined				
ob5	Butter, unspiced & 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., Enjera 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	Nonalcoholic 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 (inc. 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 (inc. books, money send to students, etc.)		
14	Health expenses (inc. clinic/hospital fees, medicine, etc.)		
15	Clothes/shoes/fabric for MEN (18 years and older)		

Note: Estimate the cash val

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org