Respondent Fatigue Reduces Dietary Diversity Scores Reported from Mobile Phone Surveys in Ethiopia during the COVID-19 Pandemic

Kibrom A Abay,[1] Guush Berhane,[2] John Hoddinott,[3] and Kibrom Tafere[4]

[1]International Food Policy Research Institute, Cairo, Egypt; [2]International Food Policy Research Institute, Washington, DC, USA; [3]Division of Nutritional Sciences, Charles H. Dyson School of Applied Economics and Management, Department of Global Development, Cornell University, Ithaca, NY, USA; and [4]Development Research Group, The World Bank, Washington, DC, USA

ABSTRACT

Background: The computer-assisted telephone interview (CATI) has been used extensively during the COVID-19 pandemic, but the effects of respondent fatigue during these interviews on responses to questions about diet are unknown.

Objectives: We designed an experiment that randomized the placement of a survey module on the dietary diversity of rural Ethiopian women and assessed whether responses were altered by placing this module earlier or later in a phone survey.

Methods: Two CATIs were implemented; in the second, women were randomly assigned to answer questions on diet diversity either earlier or later in the interview. Women's Dietary Diversity Scores were the primary outcome. Secondary outcomes were dichotomous measures of consumption from four or more and five or more food groups and consumption of food groups consumed frequently, often, and rarely. Impacts were assessed using a respondent fixed effects model.

Results: Delaying the food consumption module by 15 min in the interview led to an 8%–17% ($P < 0.01$) decrease in reported Dietary Diversity Scores, a 28% ($P < 0.01$) decrease in the number of women who consumed a minimum of four dietary groups, and a 40% ($P < 0.01$) and 11% ($P < 0.01$) decrease in the reporting of consumption of animal source foods and fruits and vegetables, respectively. Moving the food consumption module closer to the beginning of the interview increased the number of reported food groups consumed by older women, women with a below-median education level, and women in larger households.

Conclusions: Our findings suggest that comparisons of descriptive statistics across studies and countries on metrics such as food security and dietary quality may be confounded by where these modules are placed in the interview, thus highlighting trade-offs between volume of information collected and data quality when designing CATI surveys. *J Nutr* 2022;152:2269–2276.

Keywords: COVID-19, response fatigue, phone survey, dietary diversity, Ethiopia

Introduction

The outbreak of pandemics and conflicts make monitoring welfare outcomes such as food security particularly important. However, such events (e.g., COVID-19 and Ebola) create substantial obstacles to using traditional methods that employ in-person (face-to-face) interviews (1, 2). This, with increased penetration of mobile phones and continued improvements in access to the Internet, has spurred interest in remote data collection using tools such as web sites, online polls, text messages, and phone surveys. For example, the World Bank, in collaboration with National Statistical Offices, has transformed the traditional multicountry and multiround face-to-face Living Standards Measurement Study–Integrated Surveys on Agriculture program into high-frequency monthly phone survey following the outbreak of the COVID-19 pandemic (3). Alongside this increased interest and use of surveys that do not rely on face-to-face interviews has been an outpouring of methodological guides and experimentation on aspects of remote data collection (4–6). Although these remote data collection methods continue to fulfill rapidly evolving needs for timely data to inform policy responses, they also suffer from some limitations (7). Focal points of discussion in this literature are respondent access to forms of remote interviewing techniques (e.g., nonrandom differences in ownership of mobile phones or Internet connectivity); levels and differences in response rates across platforms such as the computer-assisted telephone interview (CATI), interactive voice response, and

short message service; the value of providing incentives to respondents to participate in and complete these interviews; and the implications of all these considerations for sample representativeness.

To the best of our knowledge, there has been much less discussion about the impact of the length of time required from respondents to answer questions and the response fatigue associated with these remote methods, beyond noting concerns that long surveys might generate higher rates of nonresponse. As noted by authors such as Dabalen et al. (1), despite some anecdotal pieces of evidence based on recent household surveys, there is no nuanced and systematic evidence on the potential impact of survey length and related designs in remote data collection methods. However, it is not hard to surmise why the duration of remote methods such as a CATI might affect the accuracy of responses being provided. With in-person interviews, enumerators can use visual cues to see whether respondents are beginning to tire. They can suggest taking a short break, perhaps getting a drink or having a brief stretch to allow the respondent and the enumerator to refresh. Such cues are not available when interviews are conducted remotely. Furthermore, when remote methods such as a CATI are used, it is harder to ensure that the interview is taking place in an environment where the respondent is not subject to distractions (e.g., children or elders calling for help), the likelihood of which might increase with longer interviews. In localities where time charges for mobile use are high relative to incomes, respondents might feel that they need to rush their responses as interviews drag on, even if the call time is being paid for. Yet, although work on face-to-face interviews has shown that seemingly small changes in survey methods and designs (e.g., the placement of questions) can distort stylized descriptive statistics as well as statistical analysis and associated conclusions (8–13), this issue appears to have received less attention even when the pandemic has forced many researchers to shift to remote data collection methods.

Understanding whether and how fatigue in phone surveys affects the responses received is the focus of our experimental study. We examine responses to a module aimed at characterizing the diet diversity of mothers. We use this measure for several reasons. Dietary diversity is a common indicator of quality of diets, which has been shown to be correlated with nutritional outcomes (14–17). Dietary diversity is usually operationalized using a simple count of foods or food groups consumed over a given reference period, mostly over the last 24 h, which makes

it well suited for a CATI (14, 18). Because of their simplicity, dietary diversity indicators are sometimes used to assess overall household food security (18). Dietary diversity indicators are also commonly employed to examine nutritional transition, food system transformations, and the effects of shocks on household food security (19, 20). As such, they are of interest to nutritionists, economists, and others who study these topics.

In light of these issues and the limited knowledge about them, we designed and implemented an experiment that randomized the placement of a survey module on women's dietary diversity as part of a longitudinal study that tracks the impact of the COVID-19 pandemic on households' food security in rural Ethiopia. Our objective was to assess whether responses were altered by placing this module earlier or later in a phone survey.

Methods

Study design and participants

We use data from two rounds of CATIs collected in June 2020 and December 2020. These build on previous in-person surveys conducted in March and August 2019 to understand the impact of the nutrition-sensitive components of Ethiopia's Productive Safety Nets Programme (PSNP) in four regions (Tigray, Amhara, Oromia, and SNNP). A stratified random sampling procedure had been used to select sample households in the 2019 surveys. From a list of *woredas* (districts) in each region where the nutrition-sensitive PSNP was operational, 22 were randomly selected, using probability proportional to size of population and program coverage. Within each woreda, three rural *kebeles* (subdistricts) were randomly selected, and within these, one enumeration area (EA) was randomly chosen. For each selected EA, a household list was constructed using the following selection criteria: having a child aged 0 to 23 mo (called the index child) and being a PSNP beneficiary or, if not, being considered poor based on a subjective ranking scheme applied to PSNP and non-PSNP households. From this list of eligible households, five PSNP and five non-PSNP households were randomly selected for interview. A total of 2640 households from 264 EAs and 88 woredas were surveyed in March 2019, of which 2551 were then reinterviewed in August 2019 (21).

In the first phone survey, conducted June 2020, we contacted and interviewed 1497 women (about 59%) of the 2551 who had been surveyed in August 2019 (22). Prior to the commencement of the second phone survey, war broke out in the Tigray region; because of the near-total blackout in telecommunications, we were unable to contact the 378 women in that region. For this reason, we limited the second phone survey, carried out December 2020, to women living in Amhara, Oromia, and SNNPR who had participated in the June 2020 round, successfully contacting 1109 of the 1,1119 women in those three regions. In all rounds, the primary respondent was the mother or caregiver who had provided responses to questions on the nutrition-sensitive components of the PSNP administered during the 2019 in-person interviews.

The long in-person questionnaires administered in March and August 2019 were considerably shortened when we shifted to phone surveys. We retained only core modules that focused on household food security, maternal food consumption, child-feeding practices, and access to nutrition and health services. These modules were shortened to minimize respondent burden while preserving the framing and comparability of questions across rounds. Questions were simplified to fit interview protocols using phones. Doing so reduced the time taken to administer the survey from the ≥ 2 h needed for the in-person survey to a median 26 min for interviews conducted by CATI. In the June 2020 survey, we kept mothers' and children's dietary diversity modules around the middle of the survey instrument for all respondents. However, feedback from our enumerators indicated that respondents were tiring toward the end of the phone survey, particularly when asked about long lists of items, such as the 17 yes/no questions about food groups that mothers had consumed the previous day.
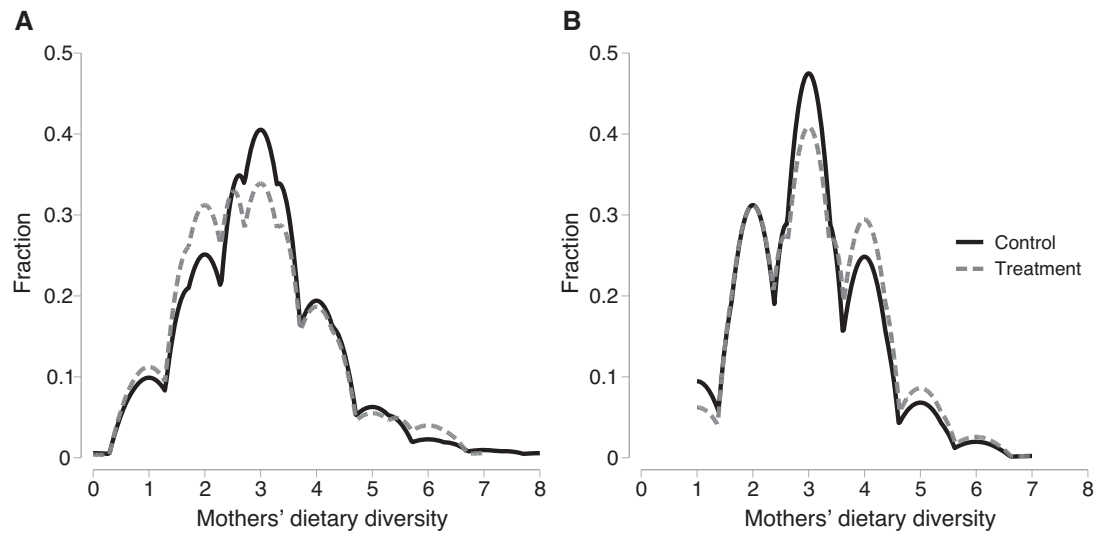
FIGURE 1 Kernel density of mothers' dietary diversity for the control and treatment groups at baseline and posttreatment. The distribution of mothers' Dietary Diversity Score for treatment and control households (A) at baseline and (B) after treatment.

Given this feedback, we modified our December 2020 survey instrument to assess the effect of response fatigue on a measure of women's food security: diet diversity. Specifically, we introduced a randomized assignment of respondents to one of two questionnaire types that differed only in the placement of the module on women's diet. Half of our respondents were randomly assigned to the treatment group, moving the instrument on women's dietary diversity to a position approximately 15 min earlier in the interview. Following some background questions at the start of the interview, mothers assigned to the treatment group were asked to respond to a list of the food items that they had consumed in the last 24 h (yes/no format). Mothers assigned to the control group were asked the same set of questions in the middle of the interview. The placement of these questions for the control group was the same as the June 2020 survey. To maintain balance by PSNP status—an important source of heterogeneity in our sample—and within administrative regions, randomization was stratified by PSNP beneficiary status and region.

In the CATI surveys, we collected detailed information about the phone calls, including interview date and time, number of call attempts made, interviewer identifiers, interview duration, and other features of the interview. We use these characteristics as control variables in our regression analysis. The August 2019 in-person survey, collected before the start of the COVID-19 pandemic, contains detailed background information about the sample households that were later interviewed in the phone surveys (CATI). We use these data when we disaggregate our results by prepandemic household characteristics. We also use these data, instead of the June 2020 survey, as an alternative preintervention sample.

Ethical approval for the study was obtained from the institutional review board at the International Food Policy Research Institute (USA; FWA 00,005,121). As the study does not evaluate a health-related biomedical or behavioral outcome, it was not registered. For this reason, these analyses should be considered exploratory.

Outcome variables

We used the information provided by mothers on their consumption of different foods to create variables reflecting consumption of the following food groups during the previous day: all starchy staple foods, beans and peas, nuts and seeds, dairy, flesh foods, eggs, vitamin A–rich dark greens leafy vegetables, other vitamin A–rich vegetables and fruits, other vegetables, and other fruits. Affirmative responses were summed to create a variable, Dietary Diversity Score, that ranges in value from 0 to 10. This is our primary outcome. Following the FAO and FHI 360 (23), we constructed a secondary 0/1 variable that equals 1 if women met minimum dietary diversity dummy requirements:

consumption of foods from five or more categories on the previous day. However, because much of the distribution mass of these scores lies to the left of the five-group cutoff (Figure 1), we constructed a second dichotomous measure: whether the mother had consumed four food groups or more. To assess whether the placement of questions on diet affected answers to foods consumed frequently, often, and rarely, we aggregated the 10 food groups into three categories and constructed three dichotomous variables equaling 1 if, during the previous day, the woman reported consuming the following: staples, beans, and nuts (foods consumed frequently); vegetables and fruits (foods consumed often); and animal source foods (foods consumed rarely). These are also secondary outcomes.

Our survey instrument included a module on child diet diversity. Mothers were asked whether the index child had consumed from each of the following food groups during the previous day: grains, roots and tubers, legumes and nuts, dairy products, flesh foods, eggs, vitamin A–rich fruits and vegetables, and other fruits and vegetables. Affirmative responses were summed to create a variable, Child Dietary Diversity Score, that ranges in value from 0 to 7. In addition, following the WHO, we constructed a 0/1 variable that equals 1 if the index child met minimum dietary diversity dummy requirements: consumption of foods from four or more categories on the previous day (24).

Statistical analysis

We calculated means and standard deviations for all outcome variables, for variables used to assess whether the treatment and control groups are balanced, and for variables included as controls in our regressions. We tested whether differences in means between the treatment and control groups are statistically significant, and we note whether these differences are statistically significant at the 10%, 5%, and 1% levels. We also conducted joint significance tests by regressing the treatment dummy on baseline characteristics.

We estimated a respondent fixed effects model where the outcome variables were respondent-reported measures of diet and dietary diversity: the continuous measure (Dietary Diversity Scores) and the dichotomous measures just described. We estimated the following equation:

$$Y_{mt} = \alpha_m + \beta_0 Round_t + \beta_1 Treatment_{mt} + \gamma X_{mt} + \varepsilon_{mt} \qquad (1)$$

where Y_{mt} refers to measures of food consumption of mother m in round t. *Round* is the survey round indicator that takes a value of 1 for the December 2020 survey and 0 for the June 2020 round. This captures seasonal factors, such as the availability of foods that might affect diet diversity. *Treatment$_{mt}$* is a dummy variable equal to 1 for mothers receiving the dietary diversity module early in the

TABLE 1 Means and mean differences in observable characteristics across the control and treatment group households[1]

	Control		Treatment		
	n	Mean	*n*	Mean	Mean difference (95% CI)
Outcome variables measured in Jun 2020					
Mothers'					
Dietary Diversity Score	555	2.989	554	2.897	0.092 (−0.06, 0.24)
Minimum dietary diversity, ≥5 groups	555	0.095	554	0.096	0.00 (−0.04, 0.04)
Minimum dietary diversity, ≥4 groups	555	0.279	554	0.273	0.01 (−0.05, 0.06)
Mother consumed					
Staples	555	0.964	554	0.977	−0.01 (−0.03, 0.01)
Animal source food	555	0.222	554	0.202	0.02 (−0.03, 0.07)
Vegetable fruits	555	0.771	554	0.706	0.07** (0.01, 0.12)
Household and respondent characteristics measured in Aug 2019					
Male-headed household, dummy	555	0.93	554	0.93	0.00 (−0.03, 0.03)
Age of household head, dummy	555	37.27	554	37.83	−0.54 (−1.83, 0.75)
Education of household head, y	555	3.62	554	3.59	0.02 (−0.41, 0.44)
Age of the mother, y	551	29.22	553	28.43	0.787** (0.06, 1.52)
Education of mother, y	555	3.12	554	3.26	−0.14 (−0.66, 0.38)
Fasting mother, dummy	555	0.14	554	0.13	0.01 (−0.03, 0.05)
Age of the child, mo	555	30.77	554	31.19	−0.42 (−1.39, 0.55)
Household size	555	5.78	554	5.68	0.10 (−0.13, 0.33)
Livestock assets, TLU	555	3.30	554	3.42	−0.12 (−0.62, 0.39)
Corrugated iron roof, dummy	555	0.55	554	0.55	0.00 (−0.06, 0.06)
Access to electricity, dummy	555	0.43	554	0.40	0.02 (−0.03, 0.08)
Farm size, ha	555	0.90	554	0.96	−0.06 (−0.17, 0.05)
Poor housing condition, dummy	555	0.19	554	0.18	0.01 (−0.04, 0.05)
Food gap, mo	555	2.49	554	2.56	−0.07 (−0.30, 0.15)
Food insecure household	555	0.77	554	0.76	0.02 (−0.03, 0.07)
Amhara region	555	0.34	554	0.34	0.00 (−0.05, 0.06)
Oromia region	555	0.34	554	0.34	0.00 (−0.05, 0.06)
SNNP region	555	0.32	554	0.32	0.00 (−0.06, 0.05)

[1]Southern Nations, Nationalities, and Peoples, Tropical Livestock Units.
**$P < 0.05$.

interview and 0 for those receiving the same module later. This variable takes the value 0 for all respondents in the baseline round. X_{mt} is a vector of time-variant observable mother characteristics and interview features. These are mothers' fasting status and the following temporal factors: interview day and time, interview duration, number of call attempts made, and enumerator fixed effects. The coefficient associated with the treatment indicator in Equation 1, β_1, captures the impact of being asked questions about diet earlier in the interview.

We disaggregated our sample by four baseline characteristics: respondents with ages above and below the median; education attainments above and below the median; household size above and below the median; and household wealth—as measured by an index constructed from a principal component analysis based on prepandemic holdings of durable goods—above and below the median. For each group, we estimated Equation 1 to ascertain heterogeneity in treatment effects. We also conducted a falsification test. The randomization of the placement of the questions on the mothers' diet had no effect on the placement of children's food consumption module; this was kept in the same place in the June and December 2020 survey rounds. Thus, our treatment should not affect reported dietary diversity outcomes for children. According to standard practice (23, 24), we constructed a continuous Dietary Diversity Score and indicator variable for the minimum Dietary Diversity Score of children: a variable that assumes a value of 1 if the number of food groups consumed in the 24 h preceding the interview was ≥4; 0, otherwise. The falsification model includes the same regressors listed in Equation 1.

We clustered standard errors at the EA level, the lowest sampling unit (25, 26). Statistical significance is reported at the 10%, 5%,

and 1% levels. All statistical analysis was conducted using Stata version 16.1.

Results

Table 1 presents the baseline balance test between the treatment and control groups. We did not reject the null hypothesis of equality of means between the treatment and control groups for each outcome variable the we consider. With respect to baseline characteristics, apart from mothers' age, we again did not reject the null hypothesis of equality of means between the treatment and control groups. When we regressed the treatment dummy on the baseline characteristics listed in Table 1, we found that the joint significance F test is 0.83 ($P > F = 0.72$), indicating that we could not reject the null hypothesis that all coefficients associated with these regressors were jointly zero (**Supplemental Table 1**).

Table 2 presents regression results focusing on mothers' dietary diversity and minimum dietary diversity. Odd-numbered columns provide results from a sparse specification that includes only dummy variables for treatment and survey round. Even-numbered columns include treatment status, survey round, a dummy variable indicating whether the mother was fasting, duration of interview, time of interview, a dummy variable if interview was conducted in the afternoon, and the number of call attempts, as well as controls for interview day of the week and enumerator dummy variables.

We focus our discussion on results obtained from regressions with the full set of controls, noting that there are no important differences between the results in columns 1 and 2 and columns 3 and 4. Column 2 shows that mothers who were asked the diet diversity module early

TABLE 2 Impact of early placement of module on maternal Dietary Diversity Score: respondent fixed effects estimates[1]

	Dietary Diversity Score		Minimum diet diversity dummy ($\geq$5 food groups)		Minimum diet diversity dummy ($\geq$4 food groups)	
	1	2	3	4	5	6
Treatment: early placement	0.23***	0.25***	0.02	0.03	0.07**	0.08**
95% CI	(0.06, 0.39)	(0.09, 0.42)	(−0.02, 0.07)	(−0.02, 0.07)	(0.00, 0.14)	(0.01, 0.15)
Round	−0.03	−0.13	−0.02	−0.03	−0.00	−0.05
95% CI	(−0.16, 0.10)	(−0.29, 0.03)	(−0.06, 0.01)	(−0.07, 0.01)	(−0.06, 0.05)	(−0.12, 0.01)
Controls[2]	No	Included	No	Included	No	Included
Interview day[3]	No	Included	No	Included	No	Included
Enumerator fixed effect[3]	No	Included	No	Included	No	Included
Mean of dependent	2.985	2.985	0.090	0.090	0.292	0.292
R^2	0.01	0.08	0.00	0.06	0.01	0.06
N	2234	2234	2234	2234	2234	2234

[1] For column numbers, see Results section.
[2] Controls include a dummy variable indicating whether the mother was fasting, the duration of interview, the time of interview, a dummy variable if the interview was conducted in the afternoon, and the number of call attempts.
[3] Dummy variable.
** $P < 0.05$.
*** $P < 0.01$.

in the interview (treatment group) reported consumption of 0.25 more food groups compared with those who were asked the module later in the interview. At the mean Dietary Diversity Score, these results are equivalent to an 8.4% reduction in maternal diet diversity for respondents whose food consumption module was delayed by 15 min.

Columns 3 and 4 of Table 2 show results for the minimum dietary diversity dummy of mothers based on the cutoff of consumption of foods from five or more categories on the previous day. The randomized placement of questions had no effect on this outcome. That said, as the mass of the distribution lies to the left of the five–food group cutoff for the Minimum Diet Diversity–Women indicator (Figure 1), we assessed a second dichotomous measure: consumption of four food groups or more (columns 5 and 6). This showed that mothers in the treatment group were 8.1 percentage points more likely to meet this threshold.

Table 3 presents results for aggregations of food groups consumed frequently (staples, beans, and nuts), often (fruits and vegetables), and rarely (animal source foods). The results in columns 1, 3, and 5 are based on a model that includes just treatment and survey round dummies. Columns 2, 4, and 6 include a wider set of controls. Table 3 shows that treatment households are 8.6 percentage points more likely to report consuming animal source foods and vegetables and fruits. Delaying the food consumption module by 15 min led to a 40% decrease relative to

the control group in the number of mothers who reported consuming animal source foods (mean: 22%) and an 11% decrease in mothers who reported consuming vegetables and fruits (mean: 76%). There was no impact on the foods considered frequently consumed: staples, beans, and nuts.

Table 4 presents results disaggregated by respondent characteristics using the full set of controls; results excluding these covariates were similar and are available on request. The first two columns provide results disaggregated by mother's median age (29 y). Older mothers, but not younger mothers, reported 0.5 more food groups when asked about these food groups earlier in the interview, a 17% decrease for women whose food consumption module was delayed by approximately 15 min. Next, we split the sample by median level of education (3 y of schooling) in columns 3 and 4. Moving the food consumption module closer to the beginning of the interview increased (relative to the control group) the number of food groups by mothers with below-median education level by 0.21 groups. Results from splitting the sample by median household size (five members) are presented in columns 5 and 6 of Table 4. Moving the module earlier resulted in a higher number of food groups reported by women in larger households by 0.46 groups, a 15% fall relative to the control group for respondents who answered the diet diversity questions later, but had no effect on reporting by

TABLE 3 Fatigue effects on dietary diversity of mothers by food groups: respondent fixed effects estimates[1]

	Staples, beans, and nuts		Animal source foods		Vegetables and fruits	
	1	2	3	4	5	6
Treatment: early placement	−0.01	−0.02	0.08**	0.09***	0.09***	0.09***
95% CI	(−0.04, 0.01)	(−0.04, 0.00)	(0.02, 0.14)	(0.03, 0.15)	(0.03, 0.15)	(0.03, 0.14)
Round	0.03***	0.03***	−0.03	−0.04	−0.01	−0.05*
95% CI	(0.01, 0.05)	(0.01, 0.05)	(−0.08, 0.01)	(−0.10, 0.02)	(−0.06, 0.04)	(−0.11, 0.00)
Controls[2]	No	Included	No	Included	No	Included
Interview day[3]	No	Included	No	Included	No	Included
Enumerator fixed effect[3]	No	Included	No	Included	No	Included
Mean of dependent variable	0.981	0.981	0.216	0.216	0.755	0.755
R^2	0.01	0.06	0.01	0.06	0.01	0.07
N	2234	2234	2234	2234	2234	2234

[1] For column numbers, see Results section.
[2] Controls include a dummy variable indicating whether the mother was fasting, the duration of interview, the time of interview, a dummy variable if the interview was conducted in the afternoon, and the number of call attempts.
[3] Dummy variable.
* $P < 0.10$.
** $P < 0.05$.
*** $P < 0.01$.

TABLE 4 Heterogeneous effects of treatment on mothers' diet diversity: respondent fixed effects estimates[1]

	Maternal age		Maternal education		Household size		Household wealth	
	< Median 1	≥ Median 2	< Median 3	≥ Median 4	< Median 5	≥ Median 6	< Median 7	≥ Median 8
Treatment: early placement	0.06	0.51***	0.21*	0.13	0.06	0.46***	0.173	0.378***
95% CI	(−0.17, 0.29)	(0.25, 0.78)	(−0.01, 0.44)	(−0.11, 0.38)	(−0.17, 0.30)	(0.21, 0.71)	(−0.04, 0.38)	(0.12, 0.64)
Round	−0.05	−0.24**	−0.19*	0.05	−0.01	−0.25**	−0.118	−0.183
95% CI	(−0.27, 0.16)	(−0.46, −0.03)	(−0.40, 0.01)	(−0.19, 0.29)	(−0.24, 0.22)	(−0.45, −0.04)	(−0.34, 0.10)	(−0.43, 0.06)
Controls[2]	Included	Included	Included	Included	Included	Included	Included	Included
Interview day[3]	Included	Included	Included	Included	Included	Included	Included	Included
Enumerator fixed effect[3]	Included	Included	Included	Included	Included	Included	Included	Included
Mean of dependent	2.979	2.991	2.832	3.167	2.945	3.023	2.786	3.194
R^2	0.11	0.10	0.07	0.13	0.09	0.11	0.08	0.11
n	1170	1050	1223	1001	1098	1136	1146	1088

[1] For column numbers, see Results section.
[2] Controls include a dummy variable indicating whether the mother was fasting, the duration of interview, the time of interview, a dummy variable if the interview was conducted in the afternoon, and the number of call attempts.
[3] Dummy variable.
* $P < 0.10$.
** $P < 0.05$.
*** $P < 0.01$.

women in smaller households. Columns 7 and 8 show that moving the food consumption module earlier in the questionnaire had no impact on reporting by households with below-median wealth but led to lower reported consumption by women in less poor households whose module was delayed: a decrease in 0.37 food groups for a 12% reduction.

Our falsification test showed that the randomized placement of the mothers' food consumption module had no effect on the reported dietary diversity of children (**Table 5**). These results were also robust to inclusion of controls for respondent characteristics, temporal features of the survey, and enumerator fixed effects.

As a final robustness check, we replaced the June 2020 baseline phone survey with the in-person survey conducted in August 2019. Results in **Supplemental Table 2** show that placement of the mothers' food consumption module earlier in the interview was associated with higher dietary diversity. **Supplemental Table 3** shows that this was driven by underreporting of infrequently consumed food items, mainly animal source foods, whereas **Supplemental Table 4** replicates our falsification test, showing that the reported dietary diversity of children was unaffected.

Discussion

We designed and implemented an experiment that randomized the placement of a survey module on women's dietary diversity in a CATI. Delaying the timing of mothers' food consumption module by 15 min led to lower reported Dietary Diversity Scores by −17% and to a 28% decrease relative to the control group in the number of mothers who met the minimum dietary diversity (defined at four food groups or higher). This result was driven by lower reporting of infrequently consumed food groups, including animal source foods and fruits and vegetables. A 15 minute delay in the timing of mothers' food consumption module led to fewer mothers reporting consumption of animal source foods and fruits and vegetables (40% and 11%, respectively). We found no impact on the consumption of more frequently consumed foods such as staples, beans, and nuts.

Our study has strengths. We used a randomized design to assess whether the placement of questions on diet diversity within the CATI survey affected responses, giving us confidence that we were identifying the effect of this placement and not other confounding factors. Our use of multiple survey rounds, with the availability of baseline characteristics from an earlier survey, allowed us to show that the random assignment was balanced and that the results were robust to the inclusion or not of additional control variables. Furthermore, these temporal features permitted us to capture additional contexts of the interview that might have affected our results. For instance, we controlled for interviewer fixed effects, thus allowing us to address time-invariant interviewer fatigue, which could have interacted with response fatigue in ways that could affect the estimates on the outcome of interest (27). These data also let us ascertain whether the impact of the placement of these questions varied by respondent age, education, household size, and household wealth. When we conducted a falsification test using an outcome, Child Dietary Diversity Score, that was not affected by the randomization, we did not find any statistically significant impact on that outcome.

Our study has weaknesses. We did not directly measure respondent fatigue. It is possible that our results could reflect fatigue on the part of our survey enumerators. However, we controlled for this partly through our inclusion of enumerator fixed effects. Furthermore, enumerator fatigue would not explain why we observed different effects by women and household characteristics. Our study population is rural, relatively poor, and one where ownership of mobile phones is limited. Responses from less well-educated women appeared to be more sensitive to the placement of the questions on diet. For these reasons, caution should be used when extrapolating these results to other settings.

The COVID-19 pandemic has spurred interest in the use of remote data collection techniques, including phone surveys (CATI), in developing country contexts. This interest has sparked new methodological work focusing on the advantages and disadvantages of different forms of remote data collection, the use of incentives to increase response rates,

TABLE 5 Falsification test: effect of treatment on child diet diversity and respondent fixed effects estimates[1]

	Dietary Diversity Score		Minimum diet diversity dummy	
	1	2	3	4
Treatment: early placement	0.08	0.09	0.05	0.05
95% CI	(−0.11, 0.27)	(−0.11, 0.28)	(−0.03, 0.13)	(−0.04, 0.13)
Round	0.08	0.12	0.02	0.01
95% CI	(−0.06, 0.22)	(−0.06, 0.30)	(−0.04, 0.08)	(−0.07, 0.09)
Controls[2]	No	Included	No	Included
Interview day[3]	No	Included	No	Included
Enumerator fixed effect[3]	No	Included	No	Included
Mean of dependent variable	2.690	2.690	0.213	0.213
R^2	0.01	0.06	0.01	0.05
N	1763	1763	1763	1763

[1] For column numbers, see Results section.
[2] Controls include a dummy variable indicating whether the mother was fasting, the duration of interview, the time of interview, a dummy variable if the interview was conducted in the afternoon, and the number of call attempts.
[3] Dummy variable.

and how to address sample representativeness (27, 28). By contrast, to the best of our knowledge, attention given to associated response fatigue and its implications is limited. We found large impacts on reported food consumption; these were especially notable given that the temporal difference in the placement of the food consumption module between the treatment and control groups was short, 15 min. The large effect on the percentage of women meeting the four–food group minimum—a 28% reduction relative to the control group—suggests that the impact of question placement on outcomes where continuous variables have been converted to dichotomous outcomes will depend on two factors: 1) the magnitude of the response to question placement and 2) where the distribution mass of the underlying continuous variable lies relative to the threshold.

Our finding that differences in responses to questions about diet vary by food groups is consistent with the notion that routinely consumed food groups are easy to remember and less taxing cognitively (10, 12, 29). By contrast, recalling foods consumed less frequently is more taxing and thus more vulnerable to respondent fatigue. Complementing this idea is another explanation—namely, that these patterns could reflect respondent incentives and responses to lengthy interviews. For instance, if each additional question involving a *yes* response brings (or is perceived to bring) follow-up questions, this may encourage time-constrained and fatigued respondents to respond *no* and likely more so for less frequent items (9). Different respondents may exhibit varying vulnerability to response fatigue for several reasons (7, 30). First, recalling diets consumed in the last 24 h entails varying cognitive burden on respondents with different age and education profiles. For example, older and less educated mothers may be more vulnerable to response fatigue than younger and more educated mothers because the latter are likely to be mobile savvy and familiar with long-running telephone conversations. Second, the breadth of consumption items that households usually consume varies across households, which implies varying cognitive burden to recall consumed food groups in the last 24 h. For instance, maternal diets for poorer households may be monotonous (21), which require less cognitive burden to recall. Third, mothers with more familial responsibilities may be more vulnerable to fatigue because larger families may involve more complex intrahousehold diet distribution as well as greater demands that may distract attention from a more accurate recollection of events. Our findings are consistent with these observations.

Our findings have some important implications for inference and survey designs. First, they suggest that comparisons of descriptive statistics across studies and countries on key welfare metrics such as food security and dietary quality may be confounded by the placement of these modules in the survey instruments. Second, variations in vulnerability to response fatigue by respondent characteristics suggest that some of the errors in phone survey–based analysis cannot be treated as classical measurement error (31, 32). These nonclassical errors are likely to render inferential biases on relationships and impacts involving these fatigue-prone phone survey data, such as underestimating income elasticities for diet quality. Third, our findings highlight the important trade-offs between volume of information collected (length of surveys) and data quality, which need to be taken into account when designing CATI surveys. The results suggest that questions that involve significant cognitive burden or those that assume strategic importance to a specific research agenda should be asked toward the start of the interview to reduce potential biases due to fatigue.

Although our findings come from a specific women's sample and phone surveys in rural Ethiopia, the implication and relevance of our findings are likely to extend to other settings involving CATI surveys. Whereas the impact of fatigue in choice experiments and diaries is well documented (33–38), the effect of fatigue in the usual multimodule and long rural household survey remains understudied. Two recent exceptions (30, 39) formally tested for the effect of fatigue on measuring skills and number of rural activities. However, we are not aware of other studies examining the impact and implication of response fatigue in CATI surveys. Additional work on this topic, particularly as it relates to diet quality, will be of value.

Data Availability

Individual participant data that underlie the results reported in this article and that were collected during the trial will be available immediately after publication, with no end date after deidentification. Data will be available to those who wish to access the data, and the study protocol, statistical analysis plan, and analytic code will be shared. Data are available indefinitely at: https://doi.org/10.7910/DVN/UT9NN3.

References

1. Dabalen A, Etang A, Hoogeveen J, Mushi E, Schipper Y, von Engelhardt J. Mobile phone panel surveys in developing countries: a practical guide for microdata collection. Washington (DC): World Bank; 2016. (Directions in development). doi:10.1596/978-1-4648-0904-0

2. Etang A, Himelein K. Monitoring the Ebola crisis using mobile phone surveys. In: Hoogeveen J, Pape U, editors. Data collection in fragile states. Cham (Switzerland): Palgrave Macmillan; 2022, 15–-31 doi:10.1007/978-3-030-25120-8_2

3. World Bank. L SMS-supported high-frequency phone surveys on COVID-19. [2022 Mar 9]. Available from[Internet]: https://www.worldbank.org/en/programs/lsms/brief/lsms-launches-high-frequency-phone-surveys-on-covid-19

4. Dillon B. Using mobile phones to collect panel data in developing countries. J Int Dev 2012;24(4):518–27.

5. Angrist N, Bergman P, Evans DK, Hares S, Jukes MCH, Letsomo M. Practical lessons for phone-based assessments of learning. BMJ Glob Health 2020;5(7):e003030.

6. Lau CQ, Cronberg A, Marks L, Amaya A. In search of the optimal mode for mobile phone surveys in developing countries: a comparison of IVR, SMS, and CATI in Nigeria. Surv Res Methods 2019;13(3):305–18. doi:10.18148/srm/2019.v13i3.7375

7. Gourlay S, Kilic T, Martuscelli A, Wollburg P, Zezza A. High-frequency phone surveys on COVID-19: good practices, open questions. Food Policy 2021;105:102153.

8. Kilic T, Sohnesen T. Same question but different answer: experimental evidence on questionnaire design's impact on poverty measured by proxies. Rev Income Wealth 2019;65(1):144–65.

9. De Weerdt J, Gibson J, Beegle K. What can we learn from experimenting with survey methods? Annual Review of Resource Economics 2020;12(1):431–447.

10. Beegle K, De Weerdt J, Friedman J, Gibson J. Methods of household consumption measurement through surveys: experimental results from Tanzania. J Dev Econ 2012;99(1):3–18.

11. Caeyers B, De Weerdt J, Chalmers N. Improving consumption measurement and other survey data through CAPI: evidence from a randomized experiment. J Dev Econ 2012;98(1):19–33.

12. Gibson J, Beegle K, De Weerdt J, Friedman J. What does variation in household survey methods reveal about the nature of measurement errors in consumption estimates? Oxf Bull Econ Stat 2015;77(3):466–74.

13. Ameye H, De Weerdt J, Gibson J. Measuring macro-and micronutrient intake in multi-purpose surveys: evidence from a survey experiment in Tanzania. Food Policy 2021;102:102042.

14. Ruel MT. Operationalizing dietary diversity: a review of measurement issues and research priorities. J Nutr 2003;133(11):3911S–26S.

15. Steyn N, Nel J, Nantel G, Kennedy G, Labadarios G. Food variety and dietary diversity scores in children: are they good indicators of dietary adequacy? Public Health Nutr 2006;9(5):644–50.

16. Kennedy GL, Pedro MR, Seghieri C, Nantel G, Brouwer I. Dietary diversity score is a useful indicator of micronutrient intake in non-breast-feeding Filipino children. J Nutr 2007;137(2):472–7.

17. Moursi MM, Arimond M, Dewey KW, Trèche S, Ruel MT, Delpeuch F. Dietary diversity is a good predictor of the micronutrient density of the diet of 6-to 23-month-old children in Madagascar. J Nutr 2008;138(12):2448–53.

18. Hoddinott J, Yohannes Y. Dietary diversity as a food security indicator. Washington (DC): International Food Policy Research Institute; 2002. IFPRI discussion paper No.: 136.

19. Popkin BM. The nutrition transition in the developing world. Development Policy Review 2003;21(5-6):581–97.

20. Pingali P, Sunder N. Transitioning toward nutrition-sensitive food systems in developing countries. Annual Review of Resource Economics 2017;9(1):439–59.

21. Berhane G, Golan J, Hirvonen K, Hoddinott J, Kim S, Taffesse AS, et al. Evaluation of the nutrition-sensitive features of the fourth phase of Ethiopia's Productive Safety Net Program. Washington, DC: International Food Policy Research Institute; 2020. ESSP working paper No.: 140. doi:10.2499/p15738coll2.133685.

22. Abay KA, Berhane G, Hoddinott J, Tafere K. COVID-19 and food security in Ethiopia: do social protection programs protect? Washington, (DC): World Bank; 2020. Policy research working paper No.: 9475.

23. FAO, FHI 360. Minimum dietary diversity for women: a guide to measurement. Rome (Italy): FAO of the United Nations and FHI 360; 2016.

24. WHO. Indicators for assessing infant and young child feeding practices: part 2 measurement. Geneva (Switzerland); WHO; 2010.

25. Rogers WH. Regression standard errors in clustered samples. Stata Technical Bulletin 1993;13:19–23. Reprinted in Stata Technical Bulletin Reprints 1994;3:88–94.

26. Froot KA. Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. Journal of Financial and Quantitative Analysis 1989;24:333–55.

27. Di Maio M, Fiala F. Be wary of those who ask: a randomized experiment on the size and determinants of the enumerator effect. World Bank Econ Rev 2020;34:654–69.

28. Singer E, Robert MB. Incentives for survey participation: when are they "coercive"? Am J Prev Med 2006;31(5):411–8.

29. Abate GT, De Brauw A, Gibson J, Hirvonen K, Wolle A. Telescoping causes overstatement in recalled food consumption: evidence from a survey experiment in Ethiopia. Washington (DC): International Food Policy Research Institute; 2020. IFPRI discussion paper No.: 01976

30. Ambler K, Herskowitz S, Maredia M. Are we done yet? Response fatigue and rural livelihoods. J Dev Econ 2021;153:102736.

31. Bound J, Brown C, Mathiowetz N. Measurement error in survey data. In: Heckman J, Leamer E , editors. Handbook of econometrics. Amsterdam (Netherlands): Elsevier; 2001. Vol. 5, p. 3705–843.

32. Abay K, Abate G, Barrett C, Bernard T. Correlated non-classical measurement errors, "second best" policy inference, and the inverse size-productivity relationship in agriculture. J Dev Econ 2019;139(1):171–84.

33. Bradley M, Daly A. Use of the logit scaling approach to test for rank order and fatigue effects in stated preference data. Transportation 1994;21(1):167–84.

34. Savage SJ, Waldman DM. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. J Appl Econ 2008;23(3):351–71.

35. Silberstein AR, Scott S. Expenditure diary surveys and their associated errors. In: Measurement errors in surveys. Hoboken (NJ): Wiley-Blackwell; 2011. p. 303–26.

36. Schündeln M. Multiple visits and data quality in household surveys. Oxf Bull Econ Stat 2018;80(2):380–405.

37. Hess S, Hensher DA, Daly A. Not bored yet—revisiting respondent fatigue in stated choice experiments. Transp Res Part A Policy Pract 2012;46:626–44.

38. Battistin E, De Nadai M, Krishnan N. The insights and illusions of consumption measurement: evidence from a large-scale randomization. World Bank; Washington DC; 2020. Policy research working paper No.: 9255.

39. Lajaaj R, Macours K. Measuring skills in developing countries. J Hum Resour 2021;56(4):1254–95.