

Poverty Measurement in the Era  
of Food Away from Home  
Testing Alternative Approaches in Vietnam

*Gabriela Farfan*  
*Kevin Robert McGee*  
*Julie Perng*  
*Renos Vakis*



**WORLD BANK GROUP**

Poverty and Equity Global Practice

&

Development Data Group

January 2019

## Abstract

Food consumed outside the home in restaurants or other food establishments is a growing segment of consumption in many developing countries. However, the survey methods that are utilized to collect data on expenditures on food away from home are often simplistic and could potentially result in inaccurate reporting. This study addresses the potential inaccuracy of commonly used methods and tests potentially superior methods to inform best practices when collecting data on consumption of food away from home. A household survey experiment was implemented in Hanoi, Vietnam, to test these different methods. Using a food away from home consumption diary as a benchmark, the study finds that many of the alternative methods considered—including asking about consumption in one line (the existing practice in Vietnam) or asking each individual

about their food away from home—lead to underreporting (33 and 22 percent underestimates, respectively). Surprisingly, using one respondent and helping them with recall with a simple worksheet as well as bounding (two-visits) results in food away from home estimates that are indistinguishable from those reported in the benchmark diary. This finding implies that there is a more cost-effective way to collect accurate data on food away from home than an intensive daily diary. Furthermore, it highlights the inaccuracy associated with collecting data on consumption of food away from home from a single question in a survey. Although limited analysis can be conducted on the implications for poverty, the study finds that the profiles of the poorest households differ across different methods of collecting information on food consumed away from home.

---

This paper is a product of the Poverty and Equity Global Practice and the Development Data Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at [gfarfan@worldbank.org](mailto:gfarfan@worldbank.org), [kmcgee@worldbank.org](mailto:kmcgee@worldbank.org), [jperng@worldbank.org](mailto:jperng@worldbank.org), and [rvakis@worldbank.org](mailto:rvakis@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# Poverty Measurement in the Era of Food Away from Home: Testing Alternative Approaches in Vietnam<sup>1</sup>

*Gabriela Farfan, Kevin Robert McGee, Julie Perng, Renos Vakis*

*The World Bank*

Keywords: survey methodology, food consumption, welfare measurement, poverty, data collection.

JEL codes: O1, C81, C93, I31, I32.

---

<sup>1</sup> This was joint work with the World Bank's Living Standards Measurement Study (LSMS) and in response to the UN Statistical Commission's call for improved guidelines on data collection. We are grateful to colleagues at the General Statistics Office of Vietnam, especially Lo Thi Duc, for their continuous support during design, implementation, and analysis of this study. We appreciate the support from Michael Lokshin, Zurab Sajaia, and Michael Wild who assisted with the retrieval and analysis of the paradata; as well as Tu Chi Nguyen, Linh Hoang Vu and Ha Thi Ngoc who helped us throughout the implementation of the project. The study also benefited from comments from Gero Carletto, Alberto Zezza, Gabriel Demombynes, Talip Killic, Obert Pimhidzai; as well as participants of various workshops at the WB. Financial support for this research has been received from the World Bank's Knowledge for Change window as well as the Food and Agriculture Organization's funding through the "Global Strategy for Improving Agricultural and Rural Statistics" initiative. The views expressed in this paper are those of the authors and do not necessarily reflect those of the World Bank or any of its affiliated organizations. All errors and omissions are our own.

## 1. Introduction

Accurate measurement of food consumption is key to the assessment and monitoring of multiple welfare dimensions, including food security, nutrition, health, and poverty. These household welfare measures are, in turn, central to the design of many development projects and policies aimed at improving well-being. It is therefore critically important for effective policy design and targeting to ensure that food consumption is collected with minimal measurement error.

With economies developing, and consumption patterns rapidly changing across the developing world, some traditional survey modules are becoming less informative over time. One recent trend is the growth in the popularity of meals prepared, packaged, and consumed outside the home (in commercial establishments and locations such as schools, work, restaurants or on the street) in contrast to the more traditional meals prepared and consumed within the home. As a result, food consumed outside the house is taking up an ever-growing share of households' food budget. For example, the percentage of households reporting consuming meals outside the home increased from 20 to 46 percent between 1981 and 1998 in the Arab Republic of Egypt, from 23 to 39 percent between 1994 and 2010 in India, and from 42.3 to 72.4 percent in Colombia (Smith, 2015; Smith et al., 2014; DANE). In China, household per-capita expenditure in FAFH rose at an average annual rate 9.5 percent from 2002 to 2011, and the share of FAFH in total food expenditure rose from 18.2 to 21.5 percent (You, 2014).

However, food consumption survey modules that were designed with more traditional eating patterns in mind may be failing to appropriately measure this growing segment of food consumption, with important consequences for welfare measurement. In India, Smith (2015) find that the increase in undernourishment at the time of falling poverty rates can be partly explained by lack of measurement of calorie consumption from FAFH. Borlizzi and Cafiero (2017) show that failure to account for children's school meal consumption in Brazil results in an overestimation of inequality in calorie consumption. Farfan, Genoni, and Vakis (2017) find that total poverty rates are 16 percent lower (or almost 6 percentage points) while extreme poverty rates are 18 percent higher (or 1 percentage point) when FAFH is taken into account. The impacts on the poverty gap and severity of poverty indicators are even larger. Furthermore, the direction of the change cannot be established ex-ante. Despite the importance of FAFH for welfare measurement, no study has rigorously tested the accuracy of different methods for collecting FAFH consumption information. This lack of evidence is highlighted in the methodological guidelines on food consumption recently endorsed by the UN Statistical Commission, which call for more research on the collection of FAFH (FAO and WB, 2018). Our study is the first to undertake this effort.

Collecting information on FAFH through household surveys raises several methodological challenges, and scalable best practices have not been yet defined. A recent study that looks at the most recent nationally representative survey across many developing countries finds great variation in practices and quality of information collected (Smith et al., 2014). For example, 10 percent of the surveys analyzed do not have any reference to FAFH. Among those that do, 24 percent dedicate only one line to FAFH, and only 35 percent account for snacks. The aim of this study is to build systematic evidence by testing alternative modules and protocols, and ultimately identifying best practices for the collection of FAFH consumption in household surveys.

To test different methods of collecting this consumption, we conducted a randomized experiment in Hanoi, Vietnam, within the framework of the Vietnam Household Living Standards Survey (VHLSS). Important considerations for the design of the experiment were not only potential improvements in accuracy, but also cost and scalability. The randomized control trial (RCT) consisted of five different experimental arms. Each arm tested alternative modules and protocols addressing methodological issues not only specific to the collection of food away from home (such as *what* to report and *who* should report), but also from traditional sources of measurement error (such as lack of knowledge or memory). The treatment arms draw from the survey methodology and behavioral sciences literatures to encourage more accurate reporting.

Drawing on the comparison across treatment arms the following lessons emerge. First, asking about household-level FAFH in a one-line question within a larger module significantly underestimates households' FAFH consumption (by about 33 percent). Second, the introduction of behaviorally-informed changes to survey protocols can significantly improve FAFH measurement, to the point of making it indistinguishable from our first-best benchmark measure (i.e. a heavily supervised individual-level diary) at a significantly reduced cost. In particular, we introduced an initial visit prior to the interview to mark the beginning of the recall period and make the recall period more salient. Furthermore, we introduced a tool to help respondents track household FAFH consumption in order to make FAFH more salient and thereby improve FAFH reporting. Third, we find that a household-level FAFH recall outperforms an individual-level recall when the household recall is accompanied by these changes to survey protocols (a bounding visit and FAFH recording tool).

The structure of the paper is as follows: Section 2 summarizes background literature on food and FAFH measurement; Section 3 describes the experimental design and sample; Section 4 presents the main results; Section 5 follows with robustness checks, unintended consequences of the experimental variations, and exploration of results beyond the means; Section 6 provides a discussion on trade-offs across treatment arms, including a cost-benefit analysis; and Section 7 concludes.

## 2. Background

Despite the growing importance of FAFH in household expenditures, most national household surveys in developing countries largely focus on measuring consumption of food consumed at home. Furthermore, even when surveys attempt to better capture FAFH, little is known about how this information should be captured. This lack of attention and knowledge is the result of the difficulties in including FAFH within current food consumption modules, as well as the lack of rigorous evidence around best practices for collecting information on FAFH. With household surveys not well-designed to adequately capture these different types of consumption, FAFH consumption is often poorly measured.

A number of methodological challenges and various sources of measurement error arise when including FAFH. To begin with, simply distinguishing between ‘food consumed at home’ and ‘food consumed outside’ is not enough. Under this classification it is not clear where take-away meals belong (i.e. those ready-to-eat meals that are produced outside, for example at a restaurant, but brought to eat at home). Careful definition, classification of types of foods, and an organized framework is therefore needed to prevent double counting or omission.

In addition to clear classification of consumption types, there are myriad other potential sources of measurement error when collecting FAFH consumption. We focus in this study on three sources: omission, telescoping, and the level of disaggregation. Measurement of FAFH is particularly vulnerable to these issues, and as such, the need to address them formed the basis for our design of the different treatment arms.

Omission of consumption can occur due to accidental omission, purposeful deception, noncompliance, or lack of knowledge. Omission of consumption is likely to happen in any recall survey, but is expected to be more common when a single household member provides information for the entire household as a result of purposeful and unintentional misinformation as well as separate spending accounts for different members of the household (Dillman and House, 2012).<sup>2</sup> Reporting on food consumed away from home poses additional challenges since it is quite possible that the informant will not be fully aware of the outside consumption of all other household members. While he or she can often directly observe food consumed at home, the same level of observation is much less likely for FAFH. Therefore, the informant is often unable to accurately report the FAFH consumption of other household members.

---

<sup>2</sup> While asking every member of the household for data (versus one representative respondent) could lead to higher noncompliance, individual-level (rather than household-level) diaries often lead to higher total reporting than those using one proxy respondent’s diary, implying that single representative diaries (or surveys) can miss out on expenditures (Browning et al., 2014).

Lack of knowledge may also occur due to memory and cognitive bandwidth, or lack of frequency or salience of purchases. Safir and Goldenberg (2008), testing the use of visual booklets intended to help in respondents' answers, found that recall aids were a bigger predictor than mode effects (e.g. telephone versus in-person interviews) of data quality. Forgotten consumption may lead to an underestimation of consumption (Crossley & Winter, 2014; Silberstein & Jacob, 1989; Sudman, Bradburn, & Schwarz, 1997).

The tendency to “telescope” when reporting consumption is another common source of measurement error. Telescoping occurs when a respondent reports consumption from outside of the reference period. Because respondents cannot always clearly define the starting point of the reference period, they mistakenly include expenditures incurred outside of the reference period and thereby over-report their expenditures (Neter and Waksberg, 1964). The issue of telescoping is more likely to occur with infrequent purchases (Brennan et al., 1996; Beegle et al., 2012; Browning et al., 2014). Although FAFH consumption is becoming more commonplace, for many households it can still be a relatively infrequent purchase and thus the issue of telescoping is a real concern when collecting FAFH information.

One method to reduce telescoping is to make the reference period more salient for the respondent through a technique called bounding. With bounding, the interviewers start with an initial interview which defines, or landmarks, the beginning of the reference period. After the initial interview, a second interview is scheduled, and data are collected on respondents *since the last interview* rather than the more typical “since ‘x’ period of time ago” (Fowler, 1995; Sudman et al., 1984). This can allow a point in one’s memory which helps pinpoint which events occurred before or after, and possibly allow a more accurate “dating” of the starting point (Gaskell et al., 2000).

However, evidence about the efficacy of bounding is mixed. Silberstein (1990) found that estimates from an initial, unbounded interview were higher than reported consumption from later waves (for which the initial interview was the bound). This could indicate that the bounding reduced the over-reporting of consumption created by telescoping. However, Elkin (2013) offers an alternative interpretation arguing that the difference was tied to other factors, and further recommends that the costs of bounding interviews do not outweigh the advantages of lowering measurement error due to telescoping.

Lastly, the level of disaggregation that the respondent is asked to report their consumption in can also have a significant impact on their reporting accuracy. While asking about consumption for a long list of distinct items or categories can often lead to higher (and presumably more accurate) expenditures, this can also result in higher levels of refusal, under-reporting, or non-completion due to a higher response burden

(Rolstad et al., 2011; Zezza et al., 2017).<sup>3</sup> Aggregating items also yields cost savings and reduces the length of the survey.

While the appropriate level of disaggregation for FAFH is unclear, asking on a single line to account for all FAFH consumption for the entire household may result in high levels of underreported consumption, as well as an increase in measurement error (Fiedler & Yadav, 2017). One-line questions often give lower overall estimates of consumption and can be difficult for respondents to answer (Browning & Crossley, 2001; Gray et al., 2008). Still, around 24% of nationally representative household surveys collect FAFH data with just a single aggregated line (Smith et al., 2014). This extreme level of aggregation for a growing segment of household expenditures could likely lead to further measurement error.

The purpose of the current study is not to test the impact of each of these dimensions separately, but to identify scalable and cost-effective solutions that improve the accuracy of measures coming from modern surveys and data collection methods. The design of the experiment draws from pre-existing evidence and provides further insight on the measurement effects resulting from the level of reporting (household or individual), bounding methods, and the level of aggregation in order to identify the alternatives that work best when collecting FAFH consumption, and why they work best.

Our benchmark measure of ‘true FAFH consumption’ is that captured through an individual-level daily diary rather than asking about consumption in a given period (which is the recall method). With this approach, respondents are asked to keep a diary which allows them to record all FAFH consumption as it happens (or at least on a daily basis). Personal diaries have often been used as a benchmark for collecting information on food consumption (Beegle et al., 2012). Despite the “gold standard” status of diary methods, discrepancies can occur due to compliance or supervision issues, diary fatigue,<sup>4</sup> education, and even changes in spending behaviors<sup>5</sup> (Beegle et al., 2012; Browning et al., 2014; Burke et al., 2011). These drawbacks are significant enough that there is some disagreement about the accuracy of diary measurements.<sup>6</sup> Because of this, diary methods need strict protocols to increase compliance and reduce response burden to yield accurate consumption data (Browning et al., 2014). Greater interviewer

---

<sup>3</sup> In addition, the impacts of disaggregation on measures of inequality and welfare are still unclear (Beegle et al., 2012; Browning et al., 2014; Hurd & Rohwedder, 2013).

<sup>4</sup> For example, Ahmed et al. (2006) find a drop off in diary records of around 10 percent between the first and second week of the diary. In this and other studies, respondents stopped measuring their consumption as time went on due to the burden.

<sup>5</sup> In other contexts (weight loss, financial and health habits and instrument take-up), surveys have been found to have impacts on behaviors (Burke et al., 2011; Crossley et al., 2017; Zwane et al., 2011).

<sup>6</sup> In national budget surveys, recall and interview survey measurements can sometimes work as well as or better than diary measurements (Bee et al., 2014; Browning et al., 2014).



involvement in ensuring diary completion, including oral questioning and enumerator assistance, can blur the line between a diary and recall survey, but allow the method to act as a benchmark of reliable consumption (Beegle et al., 2012). Given these challenges and the importance of a highly accurate benchmark, we undertook strict protocols for implementation of the FAFH diary.

### 3. Experimental design and context

To test different food expenditure data collection methods, we implemented a randomized survey experiment in Vietnam across approximately 2,400 households in urban Hanoi. In collaboration with the General Statistics Office (GSO) of Vietnam, we took the biennial VHLSS as a starting point for design of the experiment and implemented four additional treatments varying the recall period and FAFH collection method. The survey, fielded between August and October 2016, was conducted entirely using computer assisted personal interview (CAPI) methods utilizing the World Bank’s CAPI software Survey Solutions.

#### 3.1 Experimental design

Households across 40 enumeration areas (EAs) in urban Hanoi were randomly assigned to one of five treatment groups. The randomization was performed within the EA so that in each EA there were 12 households assigned to each treatment group. The five treatment arms had several variations in terms of the recall period, the respondent answering the FAFH section, the number of times the household is visited, as well as the methods used in inducing accurate recall (see Figure 1). For the purpose of this paper – which focuses on variations in data collection methods to capture FAFH consumption rather than on the recall period - we omit analysis of the data collected in the “status quo” arm, T0, which was unrelated to FAFH.

The first arm, T1, followed the standard VHLSS methodology and asked for an aggregate of all FAFH for only three categories: (1) total FAFH of present household members, (2) total FAFH of temporarily absent household members, and (3) other FAFH. We expect that these household-level aggregated categories will lead to mismeasurement of FAFH and this arm to perform least well compared to the benchmark. Although FAFH was technically collected for three different categories or lines, we shall still refer to T1 as the “one-line” treatment arm since the majority of FAFH consumption is reported on the first line (FAFH of present members).

The next treatment arm, T2, collected FAFH using a daily individual-level diary, and is the benchmark against which all other arms will be compared. Finally, the next two arms, T3 and T4, tested possible alternatives which are less intensive than a diary (T2), but more intensive than collecting from three highly aggregated categories (T1). T3 was collected at the individual-level directly from all adult members of the household. By contrast, T4 (like T1) was collected from a single individual informant for the entire

household. However, additional elements were incorporated in T4 to potentially improve the accuracy of the household-level reporting of FAFH. For all new treatment arms (T2, T3, T4) a separate FAFH module was designed and implemented whereas FAFH in T1 was collected as additional items inside the overall VHLSS consumption module.

Before providing details on the FAFH component of each treatment arm, we describe now how this component of food consumption fits into the broader survey. Total food consumption can be split into (1) at-home, (2) FAFH, and (3) take-away meals. For this study, FAFH is defined as food purchased *and* consumed outside the home, such as at restaurants, work, or a friend's house. **Take-away** meals are meals that are produced outside of the home but brought back and consumed at home. Any food that was prepared at home regardless of where it is consumed is considered to be **at-home** consumption. The at-home portion of the food consumption module implemented in the study was exactly the same across all treatment arms. Take-away meals consumption was asked in a separate subsection within the household-level food consumption module only in treatment arms T2, T3 and T4. In T1, which followed the original VHLSS protocol, take-away meals were not separately identified. It was therefore very important that interviewers described in detail what each of these components capture in order to avoid double-counting or omitted consumption.

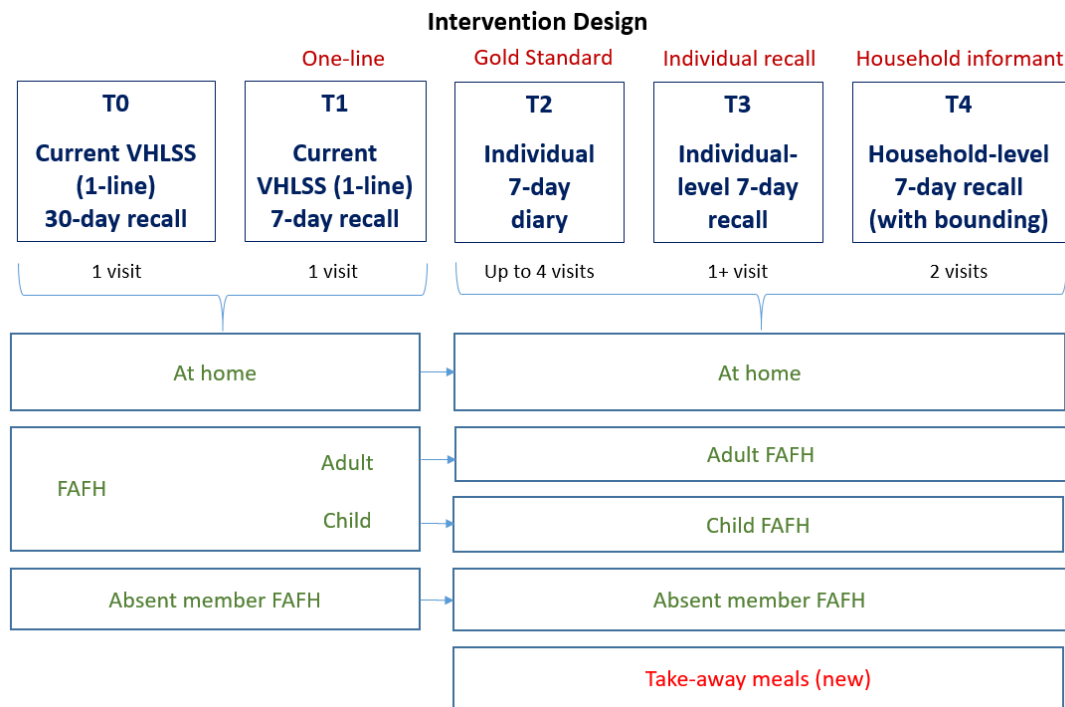
FAFH can be further split into adult FAFH, child FAFH, and FAFH of absent members.<sup>7</sup> The one-line recall arm (T1) collected FAFH on a single line for all members of the household (adults and children) present in the household. However, this aggregation could potentially lead to accidental omission of some FAFH, especially for children. Therefore, we differentiated between adult and child FAFH consumption in the other three arms (T2, T3, and T4). Adult FAFH was collected in a separate module (described in detail below) which varied across treatment arms. Child FAFH consumption continued to be asked at the household level but was disaggregated into two categories: (1) lunch at school and (2) any other child FAFH. The consumption of absent members is likely to be largely composed of FAFH since the member has not been present in the home. This category is critically important for accurate measurement of FAFH and thus the consumption of these members is explicitly captured separate from present members. Following the standard VHLSS methodology, consumption of absent members is included as another line item in the consumption module in T1. However, in T2, T3, and T4, consumption of absent members was

---

<sup>7</sup> Absent household members are defined as those persons who would not be present at the household during the entire day of the interview. Typically, these were individuals who were temporarily traveling, boarding at school, or in the hospital.

captured in a separate question with a prompt for the respondent describing what constitutes an absent member.

Figure 1 Five-arm randomized control trial testing various methods of measuring food consumption.



Below, we describe in greater detail the differences between treatment arms, with a focus on how we measure FAFH for adults in the household. With the exception of T1, adult FAFH was collected in an entirely separate module specifically designed for each of the experimental arms. Adult FAFH was collected for nine separate meal events, differentiating main meals (breakfast, lunch and dinner), snacks (morning, afternoon, and evening), and drinks (bottled water, alcohol, other drinks). This structure was meant to assist the respondent in thinking in more detail about FAFH, and to ensure they did not omit any FAFH consumption. The explicit mention of snacks, for example, is particularly important as most of the snacking is likely to take place out of the home. The module also asked for the most common place of consumption, and collected the value of all food consumed, whether purchased or received for free. Questions were designed so that there was no double-counting of expenditures (e.g. for family meals outside the home paid for by one member). More information on each of the experimental treatment arms can be found in Appendix 9.1.

Unlike the newly designed modules, treatment arm T1, the 7-day “one-line” recall, followed the current and standard methodology of the VHLSS, where households were asked to recall their food consumption in one question: “How much has your household consumed over the past 7 days of outdoor meals and drinks

(breakfast, lunch, dinner)?” As mentioned above, it then allowed to separately identify consumption among present household members, temporarily absent household members, and others. As a result, this structure not only aggregates across all household members, but also across all possible meal events, and snacks are not explicitly mentioned. The purpose of the one-line recall in this study was then to quantify the magnitude of the measurement error when asking about FAFH in one-line.<sup>8</sup>

The second arm (T2, the individual diary) acted as the benchmark or comparison group. It consisted of a personal food diary provided to each adult member of the household which helped them track, item-by-item and day-by-day, consumption of all FAFH over a period of seven days.<sup>9</sup> Enumerators provided instructions to the respondent (and other household members who were in the house at the time) and left a booklet with the same instructions. To reduce non-compliance, fatigue, cognition, and motivation issues, enumerators were instructed to contact all adults in the household within two to three days of the first visit to verify that all members were filling out the diaries and to answer any questions from the respondents. Furthermore, interviewers asked all household members to leave their booklets at a pre-specified location so that the interviewer could review them in person in up to three more household visits throughout the week. Adherence to protocols was necessary to help ensure that this arm did indeed collect the most accurate data and could be trusted as the benchmark measure. This adherence was partially verified by paradata automatically collected by Survey Solutions (the CAPI software). The paradata recorded the date and time of interviewer log-ins, which is correlated with each visit or contact the enumerator made with the household. The analysis reveals that diary module households had the largest number of days of interviewer log-ins (around 5.11), which provides an indication that enumerators conducted the requisite number of visits. For more discussion on the implementation of the diary, see Appendix 9.2.

Like the individual diary module, the third arm (T3, individual-level recall) asked every single individual in the household separately about their FAFH consumption rather than relying on a single household informant to report FAFH consumption for all members. This arm, however, used a less intensive method than the diary to survey each member. The interviewer asked each adult respondent to recall his/her FAFH consumption over the last seven days. Those who were not at home while the interview was being conducted and would not be available while the interviewer was present in the EA were interviewed by

---

<sup>8</sup> A second purpose of this arm is to analyze the impact of shortening the recall period from 30 days (current VHLSS design) to 7 days (international guidelines) on total consumption, by comparing T0 and T1. A separate paper analyzes this dimension.

<sup>9</sup> The diary was also structured around the 9 meal events mentioned before, and then respondents had to list every item consumed under each meal event.

phone. While this data collection method was likely to impact the accuracy of the information reported, it helped prevent low response rates or high proxy response rates, as it was logistically difficult to interview all members of a household in person.<sup>10</sup> This protocol was pilot-tested, and from field observations it was concluded that the FAFH module was short and simple enough to make the phone interview a cost-effective option.

Lastly, the final arm (T4 or the “bounding/salience” arm) tested the potential of relying on the traditional single-informant practice for the collection of FAFH. The selection of the household informant followed the same protocols that VHLSS had for implementing the whole consumption module, which asks the interviewer to identify the most knowledgeable person about household expenses. In order to correct for the potentially higher measurement error when collecting from a single household informant, this arm aimed to reduce information asymmetries and improve the memory and reporting of the informant by drawing lessons from survey methodology, psychology and behavioral sciences. More specifically, this arm used salience and bounding to mitigate response error by implementing the following:

- First, the interview was implemented in two visits, with households being administered a “dry-run” of the FAFH module in the first visit by asking about FAFH consumption in the past 24 hours. Informants were also informed that they would be revisited in 7 days’ time and an appointment was agreed when they would be available for the second interview. The households were visited seven days later where the full FAFH information “since the last visit” was collected. The dry-run in the first visit was intended to have two effects: (1) exposing the household informant to the type of information the enumerators were looking for (and thereby making the information salient), and (2) helping to provide a clear starting point for the reference period over which consumption is asked in the second visit (bounding) and thereby combat telescoping.
- Second, the household informant was given a worksheet to help him/her remember to keep track of all household members’ FAFH consumption throughout the week. This sheet had very little information, and its goal was to remind the informant to ask others about their consumption outside the home and help the informant keep track of what he/she and everyone else was eating throughout the week (see Appendix 9.1.3 for more information).

---

<sup>10</sup> Answering a survey over the phone, which can make reporting harder, is likely to solicit lower-quality responses (and higher item non-response rates) than answering in person (Browning et al, 2014; Safir and Goldenberg, 2008; Tourangeau et al., 2000).

### 3.2 Randomization

The GSO conducted the random assignment of 480 households to each arm. Households within an EA were randomly assigned to the interviewers working in the EA. Each interviewer was assigned an equal number of households from each treatment arm within the EA. Randomization at the interviewer-level was undertaken to reduce the potential for interviewer error that could bias FAFH and other responses. Randomization within the EA also reduces the potential for any bias stemming from variation across EAs. An additional sample of replacement households were likewise randomly assigned to each arm, and only surveyed after households from the main sample could not be reached or refused to participate.

After dropping households with missing information in critical areas for this study (mainly, FAFH consumption), there are over 475 households in each treatment arm, except for the 7-day individual diary, which has 473 households. About 20 percent of the final sample in each treatment arm consists of “replacement” households, with only T3 having a higher share (25.8%) than the other arms (see Table 1). In Section 5.1, we test for the effects of non-response by running the models only with the original sample and find that the results hold.

*Table 1 Samples of original and replacement households for each treatment arm.*

	(1) T1: 7-day one-line recall	(2) T2: 7-day individual diary	(3) T3: 7-day individual recall	(4) T4: 7-day HH informant
Original	381	378	354	378
Replacement	97	95	123	98
Total	478	473	477	476

By design, random assignment should result in groups that are statistically equivalent in both observable and unobservable characteristics, and therefore allow for differences across groups to be attributed to the experiment. In practice, however, some deviations may occur. Table 2 shows a balance table comparing differences in means of various household and survey characteristics across treatment groups. We focused on the demographic and socioeconomic characteristics of households.

In our sample, the average household head<sup>11</sup> is slightly under 54 years old and is male in 63 percent of households. He/she has attained university-level education in 26 percent of the households. The average

---

<sup>11</sup> The respondent ID was unfortunately not collected and thus we cannot test differences across respondent characteristics.

household size is slightly over four members, and the average number of children living in the household is somewhat lower than one.

Column 11 shows results of a joint orthogonality test of treatment arms and indicates that there is overall balance in household characteristics across all treatment arms except for the number of adults in the household, the incidence of a university-educated household-head, and the incidence that the household owned a computer. When looking at pairwise comparisons relative to T2 (which is the treatment arm against which results are going to be interpreted), the table shows that there is no difference between T2 and T4 (column 9), and only two of the three differences mentioned above remain between T1 and T2 (column 7). Results are, however, somewhat worse when comparing the sample under T3 (column 8). While the difference in the household-head's education level disappears, there are more differences related to the composition of T3 households (fewer number of adults, fewer number of in-household members, and fewer number of members not living in the household). Because the performance of the arms may be tied to these differences, we ran several robustness checks and find that results hold (see Section 5).

While the experiment was designed to capture differences in FAFH expenditure, expenditure on other items (non-food purchases, education, health, etc.) are expected to be the same across all treatment arms. As shown in Table 2, total annual spending on non-food items and festival spending<sup>12</sup> were balanced across arms.

Another consideration regarding the balance across groups is the timing of the interview. Complicated scheduling and logistics (particularly for the multiple-visit arms) forced the experiment to be implemented such that the day of week that the first visit was conducted was not balanced across arms. Some treatment arms had to be conducted more often on certain days of the week (Table 7). All treatments have at least some interviews starting in each day of the week. Further discussion of these details can be found in Appendix 9.2. In Section 5, we perform robustness checks and show that results hold after accounting for this lack of balance in day of the week.

---

<sup>12</sup> Festival spending is expenditure on food specifically for festivals or holidays. It is exclusive of recurrent food consumption.

### 3.3 Descriptive statistics

Table 3 provides further summary statistics on average spending observed across treatment arms. The total per capita annual spending<sup>13</sup> is around 65 million VND<sup>14</sup> (over \$2,900 USD in 2016 PPP terms); total food spending constitutes a little over 43 percent of this amount. Of total food spending, FAFH accounts for around 25 percent of the expenditures (though, as we show in the main results in Section 4, this varies across treatment arms). The table also provides some preliminary results on reported FAFH expenditures. T2 has the highest FAFH expenditure, followed by T4, T3, and T1. We explore these results in more detail in the next section.

---

<sup>13</sup> Development of the total spending aggregate followed the standard VHLSS methodology. Note that we do not include complete housing spending, as we could not impute rent values for households that did not rent. Thus, we do not include household rent for any of our sample.

<sup>14</sup> The average total per capita expenditures, measured by the VHLSS in 2014 and adjusted for inflation and spatial price adjustments, was around 51.7 million VND in the Red River Delta urban region. This was around 1.3 million VND lower than the respondents in the 30-day one-line recall (summary statistics provided upon request); this control arm is the one that matches the prior VHLSS surveys. The 2016 exchange rate was 22360.37 VND = \$1 USD.



Table 2 Balance table of characteristics across arms. In Columns (1)-(4), standard errors are listed in parentheses below the means. Columns (5)-(7) show p-values from t-tests of the differences of the means of each variable in each group. Column (8) shows the p-value from a joint orthogonality test of treatment arms.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	T1: 7-day one-line recall	T2: 7-day individual diary	T3: 7-day individual recall	T4: 7-day HH bounding/ salience	T1 vs. T2, p-value	T2 vs. T3, p-value	T2 vs. T4, p-value	p-value from joint orth. test of treatment arms
HH head is male	0.63 (0.02)	0.61 (0.02)	0.65 (0.02)	0.62 (0.02)	0.60	0.24	0.89	0.65
Age of HH head	53.43 (0.67)	54.28 (0.68)	54.41 (0.64)	53.12 (0.63)	0.38	0.89	0.21	0.42
HH head is university-level	0.23 (0.02)	0.28 (0.02)	0.25 (0.02)	0.29 (0.02)	0.05*	0.19	0.85	0.09*
Age of spouse	49.36 (0.76)	50.67 (0.73)	49.82 (0.71)	49.20 (0.70)	0.21	0.40	0.15	0.47
Spouse is university-level	0.27 (0.02)	0.27 (0.02)	0.27 (0.02)	0.26 (0.02)	0.97	0.96	0.80	0.99
Number of HH members	4.05 (0.09)	4.22 (0.08)	3.94 (0.08)	4.15 (0.08)	0.16	0.02**	0.55	0.10
<i>Number of in-HH members</i>	3.89 (0.08)	4.06 (0.08)	3.84 (0.07)	4.03 (0.08)	0.14	0.04**	0.81	0.13
<i>Number of adults in HH</i>	2.91 (0.05)	3.07 (0.06)	2.91 (0.05)	3.03 (0.06)	0.04**	0.04**	0.63	0.08*
<i>Number of others not living in HH</i>	0.16 (0.02)	0.16 (0.02)	0.10 (0.02)	0.12 (0.02)	0.91	0.06*	0.17	0.18
<i>Number of children in HH</i>	0.98 (0.05)	0.99 (0.05)	0.93 (0.04)	1.00 (0.05)	0.88	0.34	0.87	0.69
Owns computer	0.50 (0.02)	0.55 (0.02)	0.48 (0.02)	0.56 (0.02)	0.18	0.05*	0.68	0.06*
HH owns the accommodation where they live	0.90 (0.01)	0.90 (0.01)	0.92 (0.01)	0.90 (0.01)	0.97	0.41	0.86	0.76
HH accommodation separate house	0.86 (0.02)	0.86 (0.02)	0.84 (0.02)	0.85 (0.02)	0.82	0.39	0.67	0.85
Total annual spending on non-food	26808 (1334.45)	25862.66 (1138.89)	26147 (1412.11)	27528 (1154.58)	0.59	0.88	0.30	0.79
Total festival food spending	1654 (61.41)	1586.12 (56.73)	1689 (66.07)	1659 (63.01)	0.41	0.23	0.39	0.68
Percent of sample which was replacement	0.20 (0.02)	0.20 (0.02)	0.26 (0.02)	0.21 (0.02)	0.94	0.04**	0.85	0.10 0.65
N	478	473	477	476				

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 3 Summary statistics across treatment groups. Housing values do not include data on rent. All values are in 1000 VND and are annual per capita values.

	(1)	(2)	(3)	(4)
	T1: 7-day one-line recall	T2: 7-day individual diary	T3: 7-day individual recall	T4: 7-day HH bounding/ salience
<b>Total spending</b>				
Total spending	63,824	64,747	64,908	66,623
Total food spending	26,971	28,799	28,766	28,575
<b>Food spending: Breakdown</b>				
At-home spending (inc. takeaway)	22,262	20,272	22,086	20,931
FAFH food spending	5,709	8,527	6,680	7,644
<b>FAFH per capita: Breakdown</b>				
FAFH food spending (in HH members)	4,051	8,255	6,374	7,339
FAFH food spending (in HH adults)	.	7,555	5,689	6,614
<b>AH per capita: Breakdown</b>				
Takeaway food spending	.	631	605	635
<b>Miscellaneous spending</b>				
Total education spending	5,774	5,650	5,146	6,160
Total health spending	3,166	3,299	3,467	3,237
Total housing spending	3,460	3,262	3,373	3,349
Total durable spending	8,581	9,052	8,500	9,345
<b>Observations</b>	478	473	477	476

## 4. Main results

Our basic estimating equation is as follows:

$$y_h = \alpha + \sum_{i=1, i \neq 2}^4 \beta_i D_{ih} + \varepsilon_h \quad (1)$$

where  $D_{ih}$  is a dummy indicating treatment group  $i \in \{1,3,4\}$ . The diary treatment T2 is excluded ( $i \neq 2$ ) and serves as the benchmark. Household  $h$  is assigned to one of the four treatment arms and has outcome

$y_h$  (the consumption measure of interest) and idiosyncratic error term  $\varepsilon_h$ . The  $\beta_i$  captures the difference each module makes on the consumption measurement relative to the diary arm.

Table 4 shows the main results. The initial outcome variable we focus on is annual levels of per capita food away from home spending (column 1 and shown visually in Figure 2). Next, in order to see whether results are driven by differences in the incidence of FAFH, we estimate on the probability that the household reported having zero food away from home spending (column 2). Lastly, we present the difference in the value of FAFH conditioned on reporting positive FAFH consumption (column 3). All regression coefficients should be interpreted as deviations away from our benchmark individual diary measure (T2). Under this framework, a positive (negative) coefficient implies that the food measurement in an arm is likely over (under) estimated.

*Table 4 Regression of annual consumption on treatment status. The dummy variable for treatment by T2, the diary arm, is the omitted category. Column (2) provides marginal effects reporting no food away from home on treatment groups, and Column (3) shows estimated food away from home spending among households that report positive FAFH. The mean reporting 0 FAFH in the diary (omitted) arm is 8.678%. All values are in 1000 VND.*

	(1)	(2)	(3)
	FAFH	Reported 0 FAFH	FAFH (Conditional)
T1: 7-day one-line recall	-2,818***	0.19***	-1,649***
	(548)	(0.03)	(617)
T3: 7-day individual recall	-1,847***	0.07**	-1,543***
	(549)	(0.03)	(594)
T4: 7-day HH bounding/salience	-883	0.03	-713
	(549)	(0.03)	(590)
Constant	8,527***		9,315***
	(389)		(415)
Observations	1,904	1,904	1,622
R-squared	0.015	0.038	0.006
T1=T3	0.08	0.00	0.86
T1=T4	0.00	0.00	0.13
T3=T4	0.08	0.18	0.17

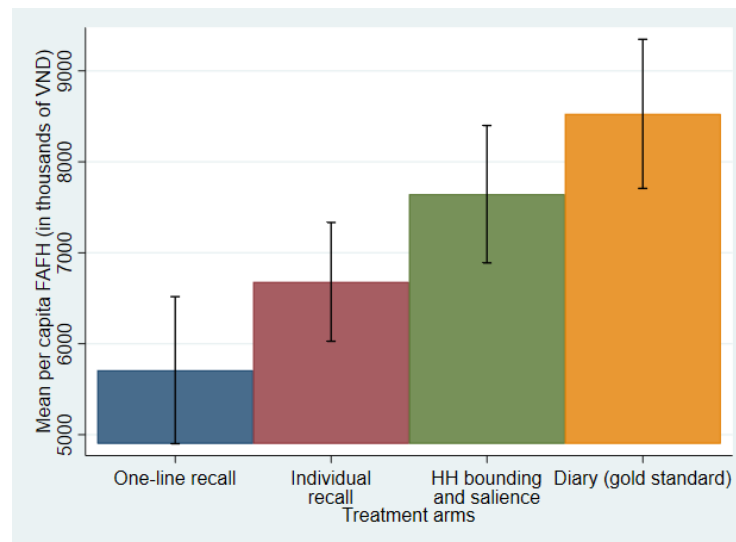
Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 4.1 One-line recall

Results in the first row of Table 4 show that having a consumption module with a single line asking about food away from home substantially underestimates this consumption. The results in Column 1 indicate that annual per capita FAFH expenditure was 2.8 million VND lower than in the diary treatment on average; Figure 2 shows that respondents who were asked about FAFH in one line reported on average significantly lower consumption (around 33 percent) than diary households. The overall lower FAFH expenditures are driven in part by the fact that these respondents were 19 percentage points more likely (than the diary method respondents) to report no food away from home (Column 2 in Table 4). This is a substantial difference, as the occurrence of zero FAFH in the diary arm is less than 9 percent.

Figure 2. Average levels of food away from home per capita of each treatment arm, with individual diary on the right (from left to right: T1: one-line recall, T3: individual recall, T4: bounding/saliency, and T2: diary).



In addition to being significantly lower than the diary treatment (the omitted group), the one-line recall results are also significantly different from the other treatment groups which used more than a line. The exception is the results of FAFH conditional on reporting (Column 3), in which this arm is significantly different from the diary but not the others.

Many factors can explain the difference between a single-line asked to one person and a daily diary filled by each member in the household. First, the diary asked for more detailed information. As discussed earlier, surveys with finer disaggregated items have been found to more accurately capture consumption levels (Pradhan, 2009). Similarly, providing an organized framework with more detailed and clear FAFH questions can also break down the cognitive bandwidth needed to report on this consumption (Fielder & Yadav, 2017). Second, the diary allowed individuals to answer for him or herself (rather than relying on the knowledge of one proxy person). Finally, diaries collected the information on a daily basis. Therefore,

if implemented well to reduce issues such as compliance, fatigue, and illiteracy, the diary is expected to result in more accurate information.

## 4.2 Individual recall

Respondents in T3, the individual recall arm, report 22 percent lower food away from home consumption on average than the diary method, but their reports come significantly closer to the diary than the one-line recall treatment. Column 1 indicates that FAFH expenditure was on average 1.8 million VND lower than in the diary treatment, while Column 3 shows that the households' conditional FAFH is 1.5 million VND lower. Therefore, the difference is driven in part by the extensive margin: these households are seven percentage points more likely to report no food away from home (Column 2) – down from the 19 percentage points observed in T1.

Based on these results, the individual recall arm greatly outperforms the one-line recall. This improvement makes sense as this module, like the diary, is more disaggregated and collects information from the most knowledgeable person – each individual. On the other hand, the information was collected in a less intensive way, and without much personal support from either the enumerator or from something physical such as a diary.

## 4.3 Bounding and salience

The third row of Table 3 presents the results from the bounding/salience arm (T4). The point estimate (Column 1) is not statistically significant, suggesting that there is no significant difference between the bounding/salience and diary modules. Therefore, this alternative worked significantly better than both the one-line and the individual recall arms. The similarities with the diary are driven in part by the equal incidence of zero FAFH reported (the extensive margin): the share of households reporting zero FAFH expenditures in the bounding/salience arm is indistinguishable from the diary arm (Column 2).

The comparatively good performance of the bounding/salience variant can be partially attributed to the introduction of innovative features to its design. It follows the traditional survey protocol of asking one proxy informant about the whole household's consumption of FAFH but addresses the measurement error introduced by this practice with tools used over the course of two visits. More discussion around the effects from bounding and salience is present in Section 6.

Taken together, our results indicate the following: (a) one line is not enough to accurately capture FAFH expenditure (FAFH is underreported by 33 percent); (b) an individual-level recall reduces under-reporting of FAFH, but reported FAFH is still underreported by 22 percent; and (c) the bounding/salience arm performs best and is not significantly different from the benchmark.

## 5. Additional results

### 5.1 Robustness checks

In order to assess the strength of our results under different approaches or assumptions, we ran a series of robustness checks. In all cases the underlying story remains consistent: the bounding/saliency arm is always the best performing arm, almost always indistinguishable from the diary and often statistically closer than the individual recall arm; the individual-level recall underestimates FAFH consumption between 16 and 25%; and the one-line recall arm consistently underestimates FAFH consumption by about 33%.

Table 5 shows key robustness tests in which we explore three key avenues: (1) restricting analysis to *adult* FAFH – as this is the focus of the experiment and the only module that changes across all treatment arms, (2) treatment of potential extreme values, and (3) issues with field implementation. For reference, column 1 shows the original results.

In column 2 we test whether limiting FAFH to only adults present in the household (i.e. excluding the FAFH of children and absent members) has any effect on our results. In column 3, we test sensitivity of outliers by winsorizing per-capita food away from home expenditures at the 1st and 99th percentiles. Columns 4 through 7 test potential implementation issues. In column 4, we add additional control variables for household characteristics that were not balanced across the treatment groups (see Table 2). In column 5, we add fixed effects for the day of the week the first visit was implemented. Additionally, the model includes date (month and day) and enumeration area fixed effects in order to control for community-wide shocks and time effects. In column 6, we address the fact that some members' FAFH in the diary and individual recall arms were missing and run the regressions without these 64 partially incomplete households. Finally, in column 7, we only use data from households which were in the original sample (i.e. exclude replacements).

Overall, the results do not substantially change for any of the robustness checks. FAFH in the bounding/saliency arm becomes marginally significantly different (i.e. at 10 percent level) than the diary arm in columns 2, 3 and 4, but the point estimates remain very similar in magnitude. Similarly, the results for the one-line and individual recall arms remain consistent both in significance and magnitude across the different specifications.

One additional consideration is whether different protocols could cause consumption *habits* to change, and therefore the results confound this with changes in reporting *accuracy*. The diary and bounding/saliency (T2 and T4) arms could potentially cause a change in behavior since these modules asked about FAFH

consumption after the first visit, and because the enumerator provided a high level of interest on understanding FAFH measurements, thus making FAFH consumption more salient throughout the week.

We aim to rule out the theory of behavioral change by looking at patterns of consumption throughout the week from the data reported day-by-day in the diary arm. A decline or increase in consumption over time may indicate that the household member was updating his or her personal consumption of FAFH over time. One hypothesis would be, for example, that the salience generated over FAFH consumption triggers an internal budgeting exercise of the costs of eating out and therefore members substitute for more meals at home. When examining this issue, we found that reported FAFH consumption does not consistently change over time for diary respondents. Furthermore, the number of missing or blank (zero FAFH consumption) diary entries is not significantly different across the 7 days covered by the diary.<sup>15</sup> This indicates that there is no consistent pattern of change in consumption resulting from multiple visits. While this analysis is merely descriptive, it helps support our assertion that the overall observed differences in food consumption are due to changes in reporting behaviors and not a result of changes in consumption patterns.

---

<sup>15</sup> Results available upon request.

Table 5 Regression of annual consumption on treatment status. The dummy variable for treatment by T2, the diary arm, is the omitted category. All results (except for column 2) are for per capita FAFH. Column 1 shows original results without robustness checks, column 2 shows a regression of per capita FAFH spending (for adults in the household only) on treatment groups, column 3 shows a regression of winsorized (at 1st and 99th percentiles) FAFH spending on treatment groups, column 4 includes various additional covariates, column 5 includes controls for day of the week of the first visit, column 6 drops “incomplete” households, and column 7 excludes replacement households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Original results	Adults in HH	Winsorized	Including covariates	With day of week FE	Without missing	Excl. replacement households
T1: 7-day one-line recall	-2,818*** (548)		-2,848*** (482)	-2,606*** (535)	-2,754*** (553)	-2,686*** (556)	-2,683*** (634)
T3: 7-day individual recall	-1,847*** (549)	-1,866*** (522)	-1,681*** (483)	-1,710*** (536)	-1,665*** (581)	-1,377** (568)	-1,917*** (646)
T4: 7-day HH bounding/salience	-883 (549)	-941* (522)	-799* (483)	-976* (535)	-710 (569)	-751 (557)	-459 (635)
HH head is male				-346 (399)			
HH head is university-level				3,122*** (451)			
Number of HH members				-1,429*** (400)			
Number of in-HH members				956** (459)			
Number of children in HH				346 (281)			
Owens computer				2,389*** (400)			
Constant	8,527*** (389)	7,555*** (370)	8,302*** (342)	8,360*** (673)	8,422*** (398)	8,395*** (400)	8,436*** (449)
Observations	1,904	1,426	1,904	1,888	1,904	1,838	1,491
R-squared	0.015	0.009	0.020	0.081	0.017	0.014	0.015
T1=T3	0.08		0.01	0.09	0.06	0.02	0.23
T1=T4	0.00		0.00	0.00	0.00	0.00	0.00
T3=T4	0.08	0.08	0.07	0.17	0.10	0.26	0.02

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



## 5.2 Beyond FAFH: Unintended impacts

Although the FAFH experiment focused on a single module within a larger consumption survey, the changes in protocols could potentially affect the responses in other sections of the questionnaire. We can directly test such knock-on effects for at-home consumption because that module and its protocols were the same across all treatment arms. Unintended changes in at-home reporting are also of substantial interest since it is also a critical component for many measures of household welfare (Kilic & Sohnesen, 2017).

Columns 2-4 of Table 6 show results from an estimation of Equation 1 using at-home food spending, total food spending, and total consumption expenditure as the dependent variables.<sup>16</sup> An evident – and puzzling – message from column 2 is that the report of food consumed at home differs between the individual recall and the food diary.<sup>17</sup> Reports across all other arms are, as expected, statistically equivalent.

A closer look at the distribution of AH consumption across treatment arms reveals that part of the difference is driven by the large number of T3 households with high at-home consumption. Fifty percent of the top 1 percent of at-home consumption households come from treatment arm 3. When we trim the top 1 percent of the at-home consumption distribution based on the full sample, the difference disappears. However, this results in significantly higher at-home food expenditures for the one-line recall arm relative to the diary.

---

<sup>16</sup> The log results are not shown, but they mirror the results in Table 6.

<sup>17</sup> This result is robust to many alternative specifications, including the inclusion of household-level covariates. However, household characteristics do have some impact as this result does not hold when we use alternative methods to individualize household consumption (instead of using per-capita consumption).

Table 6 Regression of annual consumption on treatment status. The dummy variable for treatment by T2, the diary arm, is the omitted category. Column 1 (2) shows a regression of (natural logged) per-capita at-home spending on treatment groups, column 3 (4) regresses (natural logged) per capita total food spending on treatment groups, and column 5 (6) regresses (natural logged) per capita total spending (including takeaway, health, housing and education spending) on treatment groups. All values are in 1000 VND.

	(1)	(2)	(3)	(4)	(5)
	FAFH	At-home	Total food	Total consumption	Percent at-home items answered
T1: 7-day one-line recall	-2,818*** (548)	991 (746)	-1,828* (1,038)	-923 (2,762)	-0.00 (0.01)
T3: 7-day individual recall	-1,847*** (549)	1,815** (746)	-33 (1,039)	161 (2,764)	-0.00 (0.01)
T4: 7-day HH bounding/salience	-883 (549)	660 (746)	-224 (1,039)	1,876 (2,765)	-0.01 (0.01)
Constant	8,527*** (389)	20,272*** (529)	28,799*** (736)	64,747*** (1,958)	0.41*** (0.00)
Observations	1,904	1,904	1,904	1,904	1,903
R-squared	0.015	0.003	0.002	0.001	0.001
T1=T3	0.08	0.27	0.08	0.69	0.63
T1=T4	0.00	0.66	0.12	0.31	0.20
T3=T4	0.08	0.12	0.85	0.53	0.42

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

In an attempt to explain this puzzle, we explored potential explanations linked to field implementation. In the one-line (T1) and individual-level (T3) recall arms, at-home consumption was collected in the first and only visit to the household, before asking about FAFH. In the diary (T2) and bounding/salience (T4) arms, at-home consumption was collected on the last visit.<sup>18</sup> We thus might expect that the length of the survey, the order of modules, or other components of this setup to have an impact on the at-home responses (Kilic & Sohnesen, 2017). However, the difference in reporting only arises in one of the treatment arms

<sup>18</sup> The suggested protocol was to implement the whole consumption module in the last visit, so that the reference periods for the different components coincide (i.e. at-home and FAFH consumption correspond to the same week). Everything but the consumption module should have been implemented in the first visit. In practice, however, some interviewers collected some consumption information in previous visits.

that are implemented in one visit, and therefore the source of the discrepancies should originate somewhere else.

Thus, we focus next on potential implementation differences between the individual-level recall arm (T3) and the one-line recall arm (T1) that could explain the results. In particular, we look into changes in protocols that were introduced to T3 to improve FAFH collection but that could have unexpectedly impacted the at-home module's reported consumption. Taking advantage of the paradata that were collected through the implementation of CAPI, we explore two hypotheses: interruptions to the interview and "group interviewing". Both hypotheses stem from the fact that enumerators were expected to collect data from all adult members in the household when implementing T3 but not when implementing T1.

Interruptions to the interview in general - or the at-home module in particular - may have occurred more often in the individual-level recall arm because enumerators may have tried to get adult members' information as they were leaving or coming into the house as a way to maximize response rates. This practice, in turn, would have extended the interview time. Results do not show, however, support for this hypothesis. The recorded time between the start of the survey and the start of the AH module, as well as the time it took to complete the AH module, are statistically the same in T1 and T3 interviews.

The hypothesis of "group interviewing" also builds on the idea that interviewers would try to get any adult member as they stop by. In this case, however, instead of interrupting the ongoing interview, the interviewer would ask them to 'stand by' for a bit. The potential implication of this practice is that extra people may have participated in answering the AH section. We explore this hypothesis by looking into two indicators. First, we test whether the share of at-home items with positive incidence is higher among T3 households, based on the rationale that the presence of more individuals should have reduced omission. Second, we analyze the correlation between the value of AH consumption and the number of days over which FAFH was collected. The more days FAFH collection is spread over, the less likely group interviewing occurred (and thus the lower the chance that that it resulted in higher AH consumption). Neither of the two patterns were observed in the data.

In this section we explored alternative hypotheses that could explain the higher at-home consumption reported in the individual recall arm. Unfortunately, we fail to find strong evidence to substantiate the proposed explanations, and therefore the reason behind this finding remains an open question. The results do open, however, areas of future research that would further examine unintended effects in survey design. Regarding the implications for the current study, we remain conservative in our conclusions regarding the individual recall arm due to this as yet unexplained effect.

### 5.3 Beyond the means

Up to now we have focused on the impact that changing the survey instrument (and protocols) has on FAFH consumption at the mean. However, a few additional questions can be explored to shed light on how FAFH measurement might impact policy and the targeting of the poor. We look at the following questions: a) Does FAFH have an impact on who is identified to be at the bottom of the distribution,<sup>19</sup> and does it generate a re-ordering of households? b) does FAFH accuracy vary by household characteristics? and c) does FAFH accuracy vary by level of consumption?

In order to address the first question, we analyze differences in the profiles of the poorest quintile of households across treatment arms. If there is significant re-ranking of households based on household-per capita consumption, the characteristics of the bottom quintile of total consumption should change as well. The results (available upon request) comparing these household characteristics across arms show some differences. For example, the poorest households identified in the one-line recall (versus the diary) are more likely to have a nonmarried household head and with a lower education level. In addition, the poorest households identified in the individual recall arm have an older household head (and spouse) who is less likely to be at a university-level education than the poor in the bounding/salience arm. Additionally, the poorest households in the individual recall have more females working for a wage or salary than either the diary or bounding/salience arms. All together, these differences suggest, for example, that once under-reporting is accounted for, households with less educated heads or with more females working are “no longer poor”. Having more females in the labor force is consistent with higher FAFH. Finally, the household heads of the poorest households in the bounding/salience arm are less likely than those in the diary arm to be self-employed in agriculture. However, there are no differences in household size or household composition across arms.

Next, we explore whether the performances of different treatment arms vary by household characteristics (Beegle et al., 2012). If characteristics typically associated with poverty also result in misreporting of FAFH, then this could suggest significant implications for poverty measurement. To examine this, we investigate whether characteristics of the household head (age, gender, and education), household size, number of adults, and whether the household owns a house affect reported FAFH by treatment arm (results available upon request). The results fail to identify characteristics that are consistently associated with mismeasurement of FAFH. The only characteristic that seems to partially

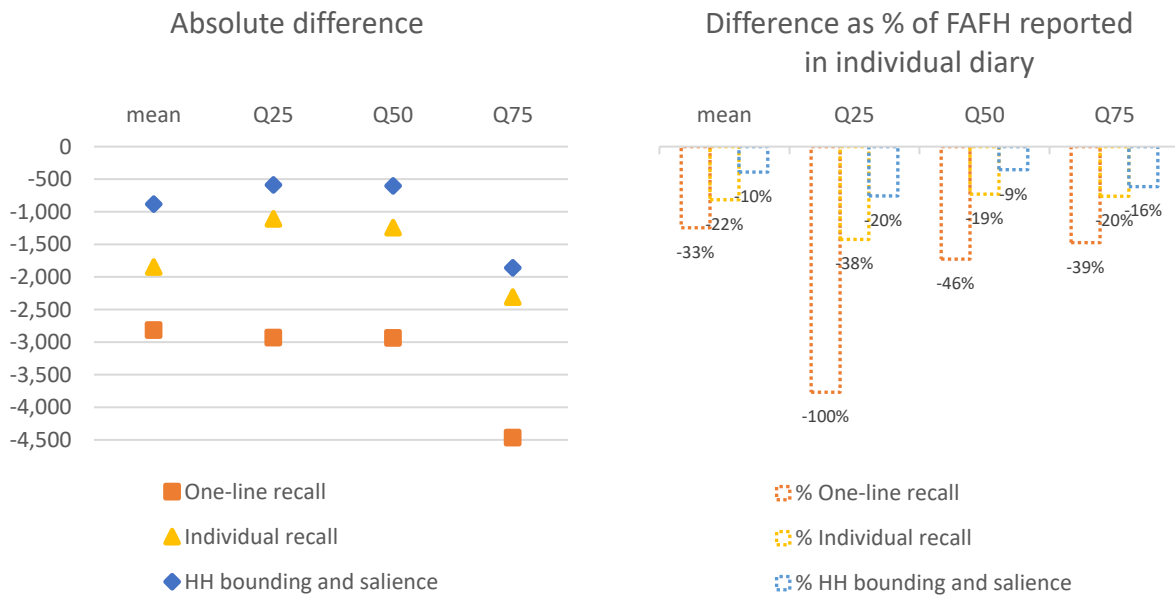
---

<sup>19</sup> We do not focus directly on the impact of FAFH measurement on poverty because poverty rates are extremely low in our setting. Estimates from 2010 put the poverty rate in the Red River Delta region (which includes urban Hanoi) under 2 percent.

explain differences in reported FAFH consumption is the number of adults in the one-line arm: the higher the number of adults, the lower the level of reported FAFH, and therefore the higher the level of under-reporting relative to the diary.

Finally, we examine whether the main results vary across the distribution of expenditures. Figure 3 below shows the results of a quantile regression of food away from home across treatment arms. Three results emerge from the analysis: (1) the ranking of performance across treatment arms holds at all points of the distribution; (2) the absolute difference between each arm and the diary is larger the higher the level of FAFH consumption, particularly at the 75<sup>th</sup> percentile where even the bounding/salience arm becomes statistically different; and (3) while in absolute terms the distance to the diary is smaller at lower quantiles of the distribution, in relative terms (as percentage of FAFH as reported in the diary), this difference is largest at the bottom of the distribution.

Figure 3: Difference in reported FAFH at different points of the distribution (quantile regression results)



In sum, although we cannot test directly the impact that improving the reporting of FAFH can have on poverty rates, we do find some indications that the different FAFH measurement methods do have implications for poverty measurement and targeting. First, the profile of households at the bottom of the distribution varied across some of the treatment arms, suggesting that miss-measurement of FAFH consumption has implications for the identification of these households. Additionally, while the absolute difference in the value of FAFH reported across treatment arms was highest among richer households – who tend to consume larger quantities of FAFH, the difference as a share of total FAFH consumption was

highest among poorer households. This suggests that efforts to improve the accuracy in reported consumption of FAFH even among those households who do not consume much in absolute terms is critically important.

## 6. Discussion

Three conclusions can be drawn from the results presented thus far: (1) One-line is not enough to collect information on FAFH; (2) The bounding/salience arm performed the best at collection of FAFH consumption; and (3) despite the fact that the individual recall arm was less accurate than the bounding/salience arm, it did perform significantly better than the one-line option. The different methods implemented present some significant trade-offs in terms of accuracy and cost of implementation. We explore further both of these aspects.

### 6.1 Accuracy

From our analysis, we have a good understanding of the relative accuracy of each treatment arm. The next important question to answer is why certain variations outperformed others. Was it due to the respondent (individual vs informant), partial collection of FAFH information (particularly for the individual-level recall arm), bounding, or salience? While we cannot empirically determine which of the components made the largest contribution to accuracy, we explore the trade-offs of each arm as well as examine some qualitative information collected from enumerators to give some indications of successful components. An evaluation of the trade-offs (in terms of accuracy and cost) and qualitative information combined with our empirical findings will allow us to provide guidance to survey practitioners on the most effective design of an FAFH consumption module.

In terms of accuracy, the individual-level arm had the advantage of eliciting information from the best informant possible: the individual. However, it came at the cost of requiring interviewers to reach all adults in the household, which was only partially addressed by allowing phone interviews. Additionally, the arm was subject to the sources of measurement error affecting all recall surveys: informants were asked to make the cognitive effort of recalling the consumption over the last seven days, without any help in identifying the starting of the recall period or the tracking of consumption throughout the reference period.

In contrast, the bounding/salience arm had the advantage of following traditional practices of interviewing only one person, but at the cost of eliciting the information from a less well-informed individual. To minimize the measurement error associated with relying on the report of one informant, a second visit was introduced. This allowed: a) giving the informant the opportunity to know what information was going to be asked and to collect the necessary information from other household members,

b) bounding the recall period through the implementation of a 24-hour recall in the first visit, and c) salience of FAFH through the introduction of a worksheet.

The better performance of the bounding/salience arm suggests that the elements introduced through the first visit were not only enough to compensate for the original knowledge gap of the household informant, but also helped address other sources of miss-reporting. A potential factor that could additionally explain the differences is the fact that some interviews in T3 were done by phone, potentially lowering data quality of the individual recall.

Qualitative feedback from enumerators further confirmed that these additional elements help. They reported that the FAFH worksheet provided to household informants was often very or fully complete when they returned in the second visit, suggesting it was a useful tool to capture FAFH. When asked which of the two components of the bounding/salience arm (bounding visit with 24-hour FAFH recall or the FAFH worksheet) was more useful to help respondents remember their consumption, enumerators slightly favored the 24-hour FAFH recall (bounding) over the worksheet (salience).

## 6.2 Implementation and cost

In order to design a survey instrument with the right components, researchers and policy makers must balance not only accuracy, but feasibility and cost. Feasibility is dependent on the setting and survey logistical constraints – for example, can phone calls be used to interview individuals not present in the household? Can field operations accommodate two visits to the household 7 days apart? Regarding cost, treatment arms vary widely in terms of time spent with respondents, and, separately, time spent making the actual visit to the household. While it is very hard to accurately cost each arm, we present in Table 6 a few rough estimates of the relative costs among arms. First, we price each arm based on GSO's budget and their estimations regarding the relative costs of each arm (rows 1 and 2). Second, we estimate interview time based on data entry time calculated from the paradata and instruction time estimated from field observations. Estimates do not include interviewers' travel time, which also varies by treatment arm.

The first row of Table 6 presents the estimated cost per interview when these relative costs are applied to the overall budget, and the second row presents the estimated field cost (i.e. excluding fixed costs such as training, software, and cleaning/data processing). The field costs of the one-line recall (T1), diary (T2), individual-recall (T3), and bounding/salience (T4) variants are \$57.93 USD, \$115.87 USD, \$62.02 USD, and \$77.25 USD per household interviewed, respectively. Going beyond one line comes at a cost: the cheapest alternative (T3, the individual-level recall) is around 7% more expensive. The bounding/salience arm, in comparison, is around 33% more costly, but allows results which are indistinguishable from the gold standard of the diary arm (which is around 100% more expensive than the one-line arm).

One important consideration regarding the cost of each arm is the cost of interviewer travel. The travel cost for any survey will vary significantly depending on the local context, dispersion of the sample, and fieldwork organization. For this survey, the travel costs were relatively low since the sample was restricted to households within urban Hanoi and thus travel time to and between survey areas was low. For national surveys that cover a wider area, travel costs associated with the multiple visits required for T2 and T4 can be substantially higher. The fieldwork organization of the survey also has implications for the cost of each arm. Travel costs will be high if mobile teams of interviewers are moving throughout a wider area to conduct interviews, while travel costs will be minimal if interviewers remain for an extended period in their assigned survey area (i.e. resident interviewers). These details can have a dramatic effect on the cost and feasibility of the different arms examined in this study. For national surveys with mobile survey teams, the cost of T2 and T4 could be prohibitive and the less preferred but still improved T3 would be more practical. However, for a survey with resident interviewers or covering a very small area, the cost of T4 would be reasonable.

When looking at the length of the interview, time is calculated by adding up the time spent on each of the questions in the FAFH section which is automatically calculated by the CAPI software (row 3). Additionally, a row on instruction time is added based on field observations during the training as it took more time to explain and implement arms T2 and T4 relative to the other two. The instruction time spent on diaries is estimated to be 15 minutes, while that on the sheet and scheduling the following visit is estimated to be 7.5 minutes. Finally, the last row presents an alternative interview time estimate, where the time spent on the 24-hour recall is not accounted for. The reason for such estimate is to estimate how much of the difference between T3 and T4 is driven by this additional module.

As expected, the diary module took by far the longest time. It takes on average about 42 minutes (including instruction time) to implement the diary. Furthermore, as shown in Figure 4, this is not driven by extreme outliers: the median time is 36.9 minutes. In contrast, interview time takes slightly over one minute when applying the one-line recall. With regards to the two preferred arms (T3 and T4), time spent on each correlates with their level of accuracy. That is, the improvements achieved through the bounding/saliency arm come at a cost. Not only is the mean about twice as high, but the whole distribution is shifted to the right. When the 24-hour recall portion is eliminated, however, time spent comes very close to the individual recall arm, at about 8 minutes on average.



Table 6 Cost and length of interview for each arm; Note: The entry time corresponds highly with the response time, though if proxying for response time, would be a lower bound for both the one-line recall and diary arms. The FAFH module entry time includes the bounding portion, the 24-hour recall module, in the second row.

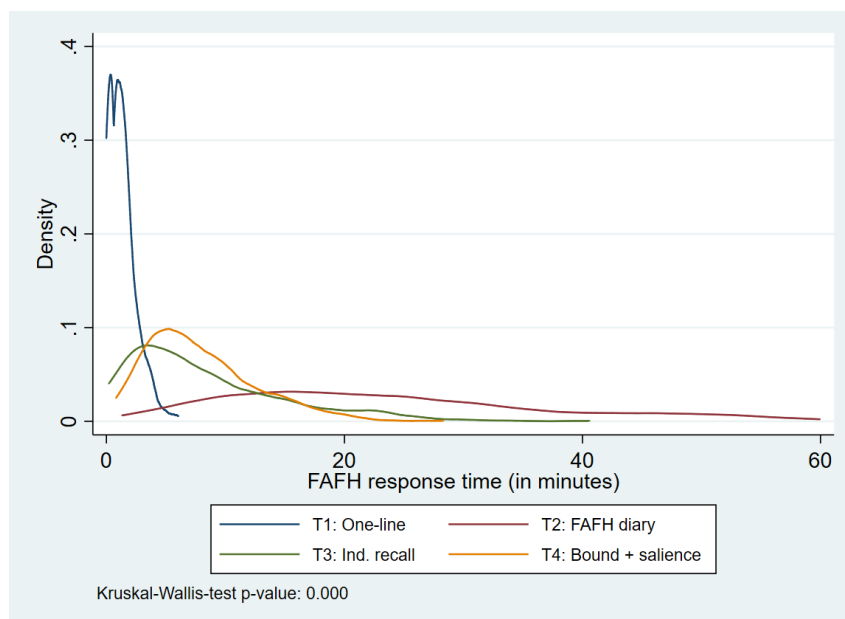
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	T1: 7-day one-line recall	T2: 7-day individual diary	T3: 7-day individual recall	T4: 7-day HH bounding/ salience	T1 vs. T2*	T2 vs. T3*	T2 vs. T4*	T3 vs. T4*	p-value from joint orthogonality test of treatment arms
Total cost per survey (upper bound)	\$102.10	\$160.04	\$106.19	\$121.42					
Field cost per survey (excl. fixed costs)	\$57.93	\$115.87	\$62.02	\$77.25					
FAFH entry time (in minutes)	1.30 (0.05)	27.00 (0.92)	8.27 (0.31)	7.92 (0.21)	0.00	0.00	0.00	0.34	0.00
Instruction time	0	15	0	7.5					
FAFH entry time (including instruction time)	1.30 (0.05)	42.00 (0.92)	8.27 (0.31)	15.42 (0.21)	0.00	0.00	0.00	0.00	0.00
FAFH entry time (including instruction time) without 24-hour bounding in T4 (in minutes)	1.30 (0.05)	42.00 (0.92)	8.27 (0.31)	8.37 (0.05)	0.00	0.00	0.00	0.74	0.00
N	478	473	477	476					

\*p-values

In light of our results on both the measurements and the cost of each method, we are left with topics for future exploration to further refine the best methodology for FAFH collection. An analysis of which component of the bounding/salience arm worked best could help us understand further how to help improve measurements with minimal cost. For example, the bounding component contributed greatly to both the time and actual costs since it required two separate visits to each household. Is the 24-hour recall section necessary, or would the first visit alone be enough to mark the recall period? Elkin (2002) conducts a similar cost analysis for a U.S. survey, recommending elimination of the bounding interview (to which he attributes up to 12 percent of costs and time). Additionally, is it necessary to have a visit with tracking sheet, or can

that first contact be done by phone? This then leads to the question of whether or not the follow-up can be done by phone; if so, costs could be further reduced.

Figure 4. Kernel density of time to enter the FAFH module responses (without instruction time) over treatment groups; the plot is truncated at 60 minutes.



## 7. Conclusion

Food consumed outside the home is becoming increasingly important in developing countries, but survey methodologies currently implemented have not been adapted to be able to capture these changing food patterns, often leading to inaccurate reporting. For example, 10 percent of consumption surveys do not make any reference to FAFH, almost a quarter of those that do only collect household-level FAFH through a one line-item in the entire module, and only 35 percent mention snacks explicitly - when snacking is most likely to take place outside the home.<sup>20</sup>

While the implications of poor-quality data for the design and monitoring of public policy are increasingly evident (see for example, Smith 2015; Borlizzi and Cafiero 2017; Farfan et al. 2017), very little is known about how to improve data collection practices in a cost-effective way. Accurate collection of FAFH consumption within traditional household surveys comes with methodological and implementation challenges that are not straightforward to address, and methodological guidelines are lacking.

---

<sup>20</sup> Smith et al., 2014.

This paper marks the first effort to fill this information gap. The objective of the study was to test alternative approaches for the collection of FAFH with the ultimate goal of informing the identification of scalable best practices. To that end, we designed a survey experiment in urban Hanoi, Vietnam, in collaboration with the General Statistical Office. The performance of three different alternatives was tested against a ‘gold standard’ – a heavily supervised individual-level diary. The first arm was the status-quo in their current living standards survey (VHLSS), where one household member is asked to report total household FAFH consumption in only one line within the food consumption module. The design of the other two arms was informed by the survey methodology and behavioral sciences literatures and included not only the application of a separate (more detailed) module for the collection of FAFH but also followed different implementation protocols.

First, we tested the performance of an ‘individual-level recall’ arm, which asked every adult household member to report their own consumption of food away from home, under the rationale that each person is the best informant for their own consumption. Second, we tested a variant of a household-informant recall arm (‘bounding and salience’), with the objective of looking for an option that would still follow the most common field practice (i.e. interview one household informant) while minimizing the expected measurement error that comes from relying on the report of one person. More specifically, the distinctive elements introduced in this arm include: the implementation of the interview in two visits, with a 7-day period between visits; the administration of a 24-hour recall FAFH module in the first visit; and the hand-out of a simple worksheet. These changes were intended to: (a) reduce the information asymmetries associated with the fact that one person does not know everyone else’s consumption, (b) facilitate the tracking of consumption throughout the week, and (c) make the recall period more salient and therefore minimize the report of consumption that took place outside the reference period.

The main results of the study are (1) One line is not enough to collect accurate information on FAFH (it underestimates FAFH consumption by 33 percent); (2) individual recall is more reliable than the status-quo, but it also led to substantial underreporting (it underestimates FAFH consumption 22 percent); and (3) the bounding and salience arm dramatically improved the accuracy of the report to the point that made it statistically indistinguishable from the individual diary. Additionally, under-reporting of FAFH is not only relevant at the mean. While we are limited in the analysis we can do on poverty, we do find that the profile of households at the bottom of the distribution does change across survey arms. Moreover, while the consumption of FAFH grows with household income, the size of under-reporting as a percent of FAFH is higher among households with lower FAFH consumption.

In terms of implications for future surveys, two main take-aways can be drawn from the analysis: First, moving beyond one line should become a priority in countries where that is still the practice; second, there

is a more cost-effective way than a diary to collect FAFH information. However, the specificities of the best option should still be based on the context. On the one hand, moving beyond one line comes at a cost: the least costly alternative – individual recall – is around 7% more expensive, while the best performing arm is around 33% more costly. On the other, each of the two improved treatment arms tested in this experiment is associated with different field and implementation logistics that can be more or less appropriate depending on the setting (i.e. geographic distributions of teams, transportation costs, time spent in each location, number of visits usually made to the household, prevalence of cell-phones, etc.). Finally, there is still much to be learned. For example, it remains an open question whether some tweaks to the individual recall arm can improve its accuracy, or whether eliminating or changing some of the components of the salience and bounding arm can lower its costs without sacrificing accuracy. Additionally, the external validity of the results is unknown, and the scope of the research question was limited for better capturing the *value* of FAFH. Much less is known still about how to measure *quantities* consumed, and its implications for the measurement of the caloric and nutritional value of food consumption.

In sum, this work highlights the inaccuracy associated with collecting FAFH consumption from a single question in a survey, and the high potential of using lessons from survey methodology and behavioral sciences to identify cost-effective alternatives that can better inform policy making over time. When considering approaches to measuring food away from home, policy makers should consider their own context, costs, and needs to decide on how to improve methods of data collection. This study offers a few lessons and provides useful insights into ways to go about it.

## 8. Bibliography

- Backiny-Yetna, P., Steele, D., and Djima, I. (2014). “The impact of household food consumption data collection methods on poverty and inequality measures in Niger”. World Bank Group Policy Research Working Paper 7090.
- Bee, Adam, Bruce D. Meyer, and James X. Sullivan (2012). “The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?” No. w18308. National Bureau of Economic Research.
- Beegle, Kathleen, et al. (2012). "Methods of household consumption measurement through surveys: Experimental results from Tanzania." *Journal of Development Economics* 98.1: 3-18.
- Borlizzi, Andrea, Del Grossi, Mauro Eduardo, Carlo, Cafiero (2017). “National food security assessment through the analysis of food consumption data from household budget and expenditure surveys: the case of Brazil’s Pesquisa de Orçamento Familiares 2008/09”. *Food Policy*.

- Brennan, Mike, et al. (1996) "Improving the accuracy of recall data: A test of two procedures." *Marketing Bulletin – Department of Marketing Massey University* 7: 20-29.
- Browning, Martin, Thomas F. Crossley, and Joachim Winter (2014). "The measurement of household consumption expenditures." *Annu. Rev. Econ.* 6.1: 475-501.
- Burke, Lora E., Jing Wang, and Mary Ann Sevick (2011). "Self-monitoring in weight loss: a systematic review of the literature." *Journal of the American Dietetic Association* 111.1: 92-102.
- Crossley, Thomas F., et al. (2017) "Can Survey Participation Alter Household Saving behaviour?" *The Economic Journal* .
- Crossley, T. F., & Winter, J. K. (2014). "Asking households about expenditures: what have we learned?" In *Improving the Measurement of Consumer Expenditures* (pp. 23-50). University of Chicago Press.
- DANE, Departamento Administrativo Nacional de Estadística
- Deaton, Angus, and Margaret Grosh (2000). "Consumption." *Designing household survey questionnaires for developing countries: lessons from ten years of LSMS experience.* 15: 91-133.
- Deaton, Angus, and Valerie Kozel (2005). "Data and dogma: the great Indian poverty debate." *The World Bank Research Observer* 20.2: 177-199.
- Dillman, D., and C. House (2012). "Measuring what we spend: toward a new Consumer Expenditure Survey." *Panel on Redesigning the BLS Consumer Expenditure Surveys.*
- Elkin, I. (2013). Recommendation regarding the use of a CE bounding interview. *Bounding interview project unpublished paper (US Bureau of Labor Statistics, 2012).*
- Farfan, Gabriela, Maria Eugenia Genoni, and Renos Vakis (2017). "You are what (and where) you eat: capturing food away from home in welfare measures." *Food Policy* 72:146-156.
- FAO and The World Bank. (2018). *Food data collection in Household Consumption and Expenditure Surveys. Guidelines for low- and middle-income countries.* Rome.
- Fiedler, J. L., & Yadav, S. (2017). "How can we better capture food away from Home? Lessons from India's linking person-level meal and household-level food data". *Food policy*, 72: 81-93.
- Fowler, F. J. (1995). *Improving survey questions: Design and evaluation* (Vol. 38). Sage.
- Gaskell, G. D., Wright, D. B., & O'Muircheartaigh, C. A. (2000). "Telescoping of landmark events: Implications for survey research." *The Public Opinion Quarterly*, 64(1), 77-89.
- Gibson, Rosalind S (2005). *Principles of Nutritional Assessment.* Oxford university press, USA.

- Gieseeman, Raymond (1987). "The Consumer Expenditure Survey: quality control by comparative analysis." *Monthly Lab. Rev.* 110: 8.
- Goldenberg, Adam and Safir, Karen L. (2008). "Mode Effects in a Survey of Consumer Expenditures October 2008."
- Hurd, Michael D., and Susann Rohwedder (2013). "Expectations and Household Spending."
- Kilic, T. and Sohnesen, T. P. (2017). "Same Question But Different Answer: Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies". *Review of Income and Wealth*, doi:10.1111/roiw.12343
- Neter, J., & Waksberg, J. (1964). "A study of response errors in expenditures data from household interviews." *Journal of the American Statistical Association*, 59(305), 18-55.
- Pradhan, Menno (2009). "Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia." *Fiscal Studies* 30.3-4: 391-417.
- Rolstad, S., Adler, J., & Rydén, A. (2011). "Response burden and questionnaire length: is shorter better? A review and meta-analysis." *Value in Health*, 14(8), 1101-1108.
- Silberstein, Adriana (1990). "First wave effects in the U.S. Consumer Expenditure Interview Survey." *Survey Methodology* 16: 293-304.
- Silberstein, A. R., and C. A. Jacob. (1989). "Symptoms of Repeated Interview Effects in the Consumer Expenditure Survey." In *Panel Surveys*, ed. D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh, 289–303. Hoboken, NJ: Wiley.
- Smith, L. C., Dupriez, O., & Troubat, N. (2014). "Assessment of the reliability and relevance of the food data collected in national household consumption and expenditure surveys." *International Household Survey Network*.
- Smith, L. C. (2015). "The great Indian calorie debate: Explaining rising undernourishment during India's rapid economic growth." *Food Policy*, 50, 53-67.
- Sudman, S., Bradburn, N. M., Schwarz, N., & Gullickson, T. (1997). "Thinking about answers: The application of cognitive processes to survey methodology." *Psychocritiques*, 42(7), 652.
- Sudman, Seymour, Adam Finn, and Linda Lannom. "The use of bounded recall procedures in single interviews." *Public Opinion Quarterly* 48.2 (1984): 520-524.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000). *The psychology of survey response*. Cambridge University Press, 2000.

You, J. (2014). "Dietary change, nutrient transition and food security in fast-growing China." in R. Jha, R. Gaiha and A.B. Deolalika, eds., *Handbook on Food*, pp.204-245. Cheltenham, UK, Edward Elgar Publishing.

Zeza, Alberto, Carletto, Calogero, Fiedler, John L., Gennari, Pietro, Jolliffe, Dean. (2017)

"Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES)." Introduction to the special issue, *Food Policy*, Volume 72: 1-6.

Zwane, Alix Peterson, et al. (2011). "Being surveyed can change later behavior and related parameter estimates." *Proceedings of the National Academy of Sciences* 108.5: 1821-1826.

## 9. Appendix

### 9.1 Materials used in the treatment arms

#### 9.1.1 T2: Individual diary

Figure 5. Example page from the Diary Instruction Booklet, a document left for the household informant and which included general and line-by-line instructions.

#### Example

Member's Name: Hung Nguyen Day 1

Monday Tuesday  Wednesday Thursday Friday Saturday Sunday

Date: \_06\_ / 21\_ / 16\_

Did you consume food outside the home?  YES → fill forms below  NO

FOR INTERVIEWER  
 Result:  COMPLETE  INCOMPLETE  REFUSED  MISSING  
 Filled by:  RESPONDENT  INTERVIEWER

LINE EVENT	1		2		3		4		5		6	
	MEAL	EVENT	Please describe the meal/items you consumed outside the home		Check the place where you consumed the meal/item		Check if you paid for the meal/item or if you received it for free		If you PAID for the meal/item: How much did you spend on your meal/item? VND		If you received the meal/item for FREE: How much would you have spent if you had to pay for meal/item received for free? VND	
<b>Main meals</b>												
1	breakfast	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	Sticky rice		<input checked="" type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input checked="" type="checkbox"/> PAID <input type="checkbox"/> FREE		10			
2	lunch	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	Chicken with rice, orange juice, and cake		<input checked="" type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input checked="" type="checkbox"/> PAID <input type="checkbox"/> FREE		15			
3	dinner	<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO			<input type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input type="checkbox"/> PAID <input type="checkbox"/> FREE					
<b>Snacks or drinks outside of main meals</b>												
4	morning snack	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	yogurt		<input type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input checked="" type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input checked="" type="checkbox"/> PAID <input type="checkbox"/> FREE		2.5			
5	afternoon snack	<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO			<input type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input type="checkbox"/> PAID <input type="checkbox"/> FREE					
6	evening snack	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	Sweet soup		<input checked="" type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input checked="" type="checkbox"/> PAID <input type="checkbox"/> FREE		5			
7	bottled water	<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO			<input type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input type="checkbox"/> PAID <input type="checkbox"/> FREE					
8	alcoholic drinks	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	beer		<input checked="" type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input checked="" type="checkbox"/> PAID <input type="checkbox"/> FREE		6			
9	other drinks	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	tea		<input type="checkbox"/> RESTAURANT <input type="checkbox"/> SMALL FOOD SHOP <input type="checkbox"/> STREET VENDOR <input checked="" type="checkbox"/> SCHOOL / WORK <input type="checkbox"/> FAMILY / FRIEND / OTHERS HOUSE <input type="checkbox"/> OTHER		<input type="checkbox"/> PAID <input checked="" type="checkbox"/> FREE				1.5	



## 9.1.2 T3: Individual recall

Figure 6. FAFH module for T3.

### T3 - individual level recall

#### 5A3.3 ADULT CONSUMPTION AWAY FROM HOME: Consumption expenditures on food and drinks away from home in past 7 days (Adults 15 years old and above)

In this section I'm going to ask you about the food you consumed outside your home in the last 7 days. We want you to consider all the food that was '*prepared and consumed*' outside the home, such as food consumed at restaurants, at work, or at a relative or friend's house.

You should not include food you prepared at home and took to consume at work or somewhere else, or snacks that you took from your pantry (purchased for your household's consumption) and consumed outside the home.

Food item code	Food Item / Food occasion	2 During the past 7 days, did you consume [ITEM] away from home?  DO NOT INCLUDE ITEMS TAKEN FROM YOUR HOME  YES...1 ▶ 3 NO...2 ▶ next item	3 How many days did you consume [ITEM] away from home in the past 7 days?  ▶ 2 next item AFTER GOING OVER ALL ITEMS ▶ 4	4 During the past 7 days, where did you most frequently consume [ITEM] away from home?  RESTAURANT.....1 FOOD SHOP...2 STREET VENDOR.....3 SCHOOL/WORK.....4 FAMILY/FRIEND/OTHER'S HOUSE.....5 OTHER (SPECIFY)....6	5 Did you pay for all or part of the [ITEM] you consumed away from home in the past 7 days?  [IF FOOD WAS PURCHASED BY OTHER HOUSEHOLD MEMBER, MARK 3]  Yes: all...1 ▶ 6 Yes: part...2 ▶ 6 No.....3 ▶ 7	6 During the past 7 days, how much did you spend in total on [ITEM] consumed away from home?	6a 6b 6c COMPONENTS			Interviewer check: If 5=1 ▶ 9 If 5=2 ▶ 7	7 Who paid for the [ITEM] that you received as a gift or from other sources in the last 7 days?  Household member...1 ▶ 9 Non-household member...2 ▶ 8 Both...3 ▶ 8	8 How much would you have spent if you had to pay for [ITEM] that you received as a gift from non-household members in the past 7 days?  VND	9 What did you consume for [ITEM] the last time in the past 7 days?  (Specify the name of dish or content of the meal you consumed)	9a What was the value of the meal/item you consumed?  [If received for free, provide estimated value]  VND
							Amount spent on the food you consumed (same as 6 if didn't share consumption with other individuals)	Amount spent on the consumption of other adult household members (15 years and older)	Amount spent on the consumption of other child household members (under 15 years old)					
<b>Main meals</b>														
	breakfast meal						VND	VND	VND	VND				
	lunch/midday meal													
	dinner/evening meal													
<b>Snacks or drinks outside main meals</b>														
	morning snack													
	afternoon snack													
	evening snack													
	bottled water													
	alcoholic drinks													
	other drinks													

9.1.3 T4: Household-level recall with bounding

Figure 7. 24-hour recall module for T4.

**T4 - 24hr recall (to be applied to household informant in first visit)**

5A3.3A

**5A3.4A ADULT CONSUMPTION AWAY FROM HOME: Consumption expenditures on food and drinks away from home in the previous day (Adults 15 years old and above)**

I'm going to ask you a few questions about yesterday's consumption of "food away from home" by all adults in your household. By FAFH we mean "any meal, snack, or drink that was "produced and consumed" outside the home. Think of all adults (15 years or above) who are currently in your household, and consider all possible places of consumption (restaurant, pub, work, school, friend's house, etc).

. Do not include food that was prepared at home or snacks that were taken from home (for example, from the household's pantry or fridge) to consume at work or somewhere else.

Food item / Food occasion	2	4	5	6	6a		6b		7
					Amount spent on food consumed by adult household members (15 years and older)	Amount spent on the consumption of child household members (under 15 years old)	Interviewer check:	How much would your household have spent if it had to pay for the [ITEM] received yesterday as a gift?	
	<p>Yesterday, did you or any member of your household consume [ITEM] away from home?</p> <p><b>DO NOT INCLUDE ITEMS THAT WERE PREPARED AT YOUR HOME</b></p> <p>▶ next item</p>	<p>Where was [ITEM] consumed away from home?</p> <p>(If more than one place applies, report the most common)</p> <p>RESTAURANT.....1            FOOD SHOP ...2            STREET VENDOR.....3            SCHOOL/WORK.....4            FAMILY/FRIEND/            OTHER'S HOUSE.....5            OTHER (SPECIFY)....6</p>	<p>Did your household pay for all or part of the [ITEM] consumed yesterday?</p> <p>Yes: all...1 ▶ 6            Yes: part..2 ▶ 6            No.....3 ▶ 7</p>	<p>How much did your household spend in total on [ITEM] consumed away from home?</p>	VND	VND	VND		VND
<b>Main meals</b>									
breakfast meal									
lunch/midday meal									
dinner/evening meal									
<b>Snacks or drinks outside main meals</b>									
solid snacks									
bottled water									
alcoholic drinks									
other drinks									

Figure 8. Worksheet distributed in the first visit for T4.

Helping you track your household's consumption of  
 "Food Outside the Home"

. For the next seven days including today, use this sheet to help you keep track the food that your household consumes outside the home.

. By "food outside the home" we mean all food produced and consumed away from home, such as a meal at a restaurant, food from street vendors, meals or snacks at work or school, or simply food at some friend's or neighbor's house.

. Remember to ask all adult members in your household about their consumption during this week.

. In 7 days, we'll come back and ask you about it!!

**Next week's appointment:**  
 August

Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3
4	5	6	7	8	9	10

at \_\_\_\_\_ : \_\_\_\_\_

Name the adults in your household: .....

Day of the week:	Did any <u>adult</u> household member had breakfast, lunch, dinner, snacks or drinks <u>outside the home</u> ?	How much did <u>your household</u> <u>paid</u> for that consumption?  VND	If some food was recieved for free, how much would your household have had to pay for that food?  VND
Day 1.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....
Day 2.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....
Day 3.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....
Day 4.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....
Day 5.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....
Day 6.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....
Day 7.....	Yes <input type="checkbox"/> No <input type="checkbox"/>	.....	.....

In designing this worksheet, the research team aimed to balance comprehensiveness with simplicity, to make it user-friendly. The purpose was limited to assist the household respondent in keeping track of everyone's FAFH. Its use was not enforced by interviewers, and the information was not used directly in the second visit. Interviewers would take the worksheet from the informant and then implement the survey module.

## 9.2 Implementation

In this section, we discuss the implementation protocols used for the survey.

### 9.2.1 Proper implementation of the diary

There are a few indications that implementation of the diary went as planned. First, only 6 percent of the daily entries were missing or not completed in our sample and this did not differ significantly over days (i.e. the first through seventh day of the diary) or between individual days. Second, the amount of food away from home reported for the first day is higher than the other days, but it is not significantly higher

than the fourth or the seventh days (and none of the other days are significantly different from each other).<sup>21</sup> In other words, the reported consumption did not drop off in a worrisome pattern over time. Finally, while enumerators filled out diaries for household members in certain circumstances (due to illiteracy and lack of time), we do not believe that this happened often. In a follow-up discussion with enumerators, they generally agreed that the diaries were close to full completion in the second, third, and fourth visits.

### 9.2.2 Fieldwork implementation and coordination

Extensive coordination among teams and treatment arms was necessary to be able to collect the data in the planned timeline. The one-line recall surveys required one visit, the 7-day individual recall (T3) arm required one or more visits (or calls), the bounding/salience arm required two visits 7-days apart, and the diary required 4 separate visits across a period of 8 days<sup>22</sup> (see Figure 9 and Figure 10 for the field schedules for the diary and bounding and salience arms).

---

<sup>21</sup> We did this analysis using reported consumption over days. For some households, certain days were set as missing. When we set these entries to zero rather than to missing, we see the same patterns of consumption over the course of the week as described, but the p-value from a joint orthogonality test of treatment arms is higher than 0.10.

<sup>22</sup> More information on the implementation can be found in Section 3 Experimental design and context.

Figure 9 Diary field schedule.

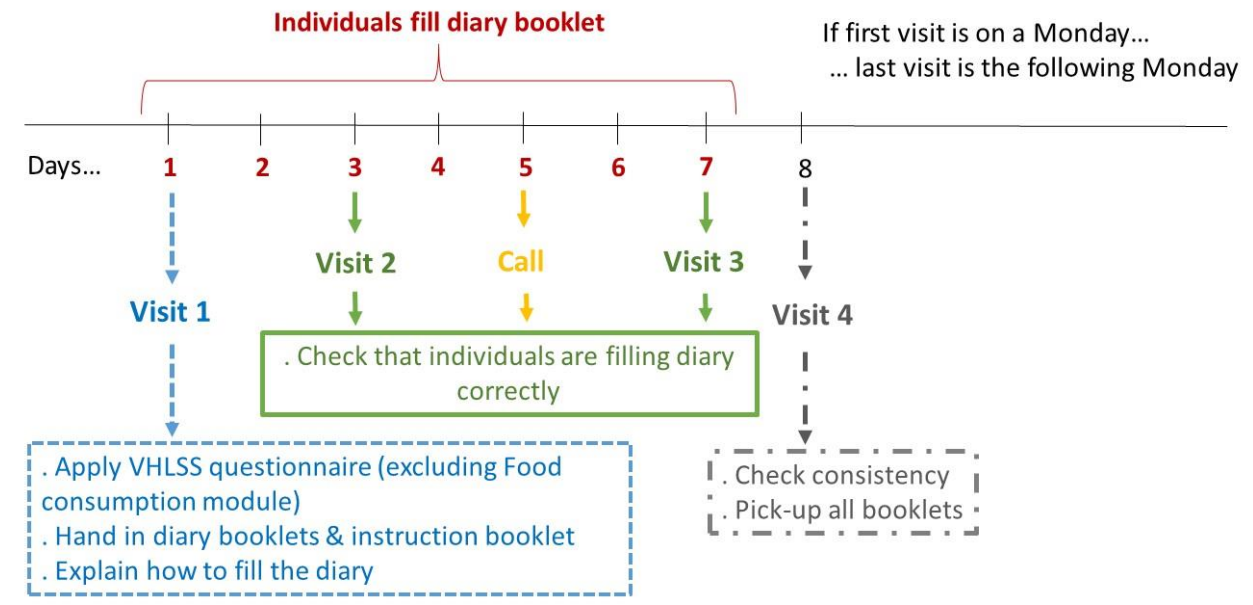
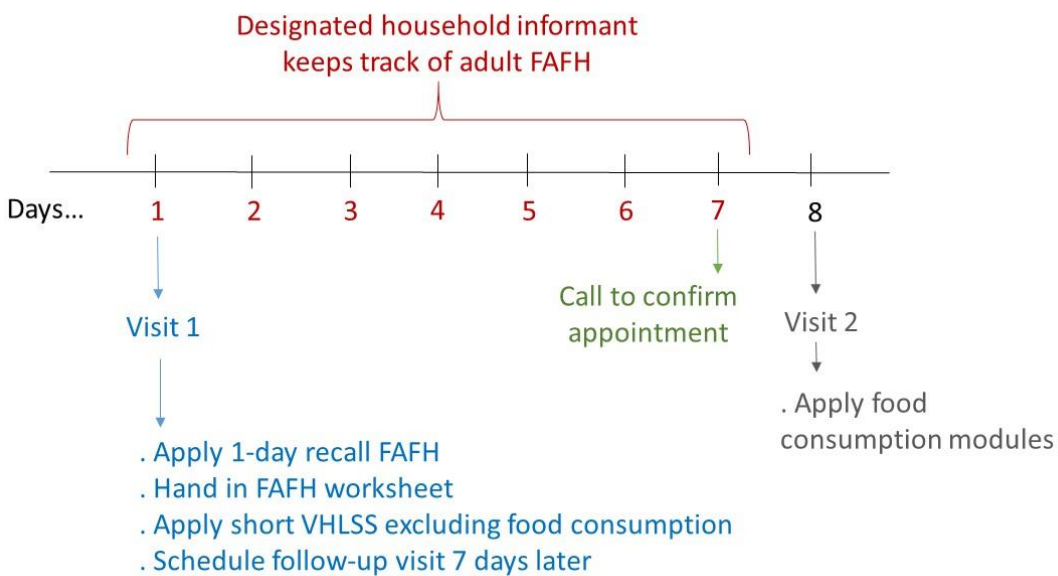


Figure 10 Bounding and salience arm field schedule.



Special attention was paid to the fieldwork schedule to ensure that enumeration areas were completed in the most efficient way possible and also maintain variation across teams. Because of the logistics involved in the visits, all teams were instructed to follow roughly the same bi-weekly schedule. This schedule specified the type of households (i.e. from which arm) that would be interviewed on a given day,

with the cycle repeating after 14 days. As designed, there was little within-team variation in terms of the schedule. Given the complexity of the survey and sensitivity of the timing for the multiple visit arms, it was decided to maintain this fixed schedule within the survey teams and avoid the risk of confusing interviewers and supervisors and thereby compromising the survey.

This created another problem whereby if all teams started fieldwork on the same day, then the day of the week when each treatment group was interviewed would be the same across teams. For example, if individual recall households were interviewed on days 3 and 10 of the 14-day schedule and all teams started on a Monday, then for the entire period of the survey individual-recall households would only be interviewed on Wednesdays. To ensure that there was variation in the days of the week when each treatment was interviewed, the start dates for the field teams was staggered with each team starting the 14-day schedule on a different day of the week.

This logistical characteristic of the start date is not balanced across treatment arms (as can be seen in Column (11) of Table 7). This is a concern because the day of the week households are surveyed can impact their responses. Quite often, FAFH consumption was highest on the weekend. Households that are interviewed on a Monday would likely have a clearer picture of their weekend FAFH consumption since they only must think back two days whereas a household interviewed on a Friday will have to recall their weekend FAFH from 6 and 7 days ago.

To check the robustness of results, the baseline regressions are run with fixed effects controlling for the day of the week (see Table 4). In the consumption variables of interest, the main story of regressions of food away from home on treatment groups does not change. Controlling for whether or not the first interview was conducted on a weekday similarly does not impact results, and the coefficient on the dummy is not significant in the regression (results provided upon request).

Table 7 Balance table of the days in which the first visit was implemented across arms. In Columns (1)-(4), standard errors are listed in parentheses below the means. Columns (5)-(7) show p-values from t-tests of the differences of the means of each variable in each group. Column (8) shows the p-value from a joint orthogonality test of treatment arms.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	T1: 7-day one-line recall	T2: 7-day individual diary	T3: 7-day individual recall	T4: 7-day HH bounding/salience	T1 vs. T2, p-value	T2 vs. T3, p-value	T2 vs. T4, p-value	p-value from joint orthogonality test of treatment arms
Conducted interview on Sunday	0.17 (0.02)	0.20 (0.02)	0.11 (0.01)	0.17 (0.02)	0.28	0.00	0.22	0.00
Conducted interview on Monday	0.10 (0.01)	0.05 (0.01)	0.16 (0.02)	0.22 (0.02)	0.01	0.00	0.00	0.00
Conducted interview on Tuesday	0.15 (0.02)	0.19 (0.02)	0.22 (0.02)	0.05 (0.01)	0.10	0.19	0.00	0.00
Conducted interview on Wednesday	0.17 (0.02)	0.23 (0.02)	0.06 (0.01)	0.19 (0.02)	0.03	0.00	0.21	0.00
Conducted interview on Thursday	0.17 (0.02)	0.08 (0.01)	0.16 (0.02)	0.21 (0.02)	0.00	0.00	0.00	0.00
Conducted interview on Friday	0.04 (0.01)	0.05 (0.01)	0.23 (0.02)	0.05 (0.01)	0.42	0.00	0.79	0.00
Conducted interview on Saturday	0.20 (0.02)	0.21 (0.02)	0.06 (0.01)	0.10 (0.01)	0.87	0.00	0.00	0.00
Conducted interview on a weekday	0.63 (0.02)	0.59 (0.02)	0.83 (0.02)	0.73 (0.02)	0.32	0.00	0.00	0.00
N	478	473	477	476				