

Friedman, Jed; Beegle, Kathleen; De Weerd, Joachim; Gibson, John K.

Working Paper

Decomposing response error in food consumption measurement: Implications for survey design from a randomized survey experiment in Tanzania

LICOS Discussion Paper, No. 375

Provided in Cooperation with:

LICOS Centre for Institutions and Economic Performance, KU Leuven

Suggested Citation: Friedman, Jed; Beegle, Kathleen; De Weerd, Joachim; Gibson, John K. (2016) : Decomposing response error in food consumption measurement: Implications for survey design from a randomized survey experiment in Tanzania, LICOS Discussion Paper, No. 375, Katholieke Universiteit Leuven, LICOS Centre for Institutions and Economic Performance, Leuven

This Version is available at:

<https://hdl.handle.net/10419/172027>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



LICOS Centre for Institutions and Economic Performance

Centre of Excellence



LICOS Discussion Paper Series

Discussion Paper 375/2016

Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania

Jed Friedman, Kathleen Beegle, Joachim De Weerd, John Gibson

KU LEUVEN

Faculty of Economics And Business

LICOS Centre for Institutions and Economic Performance
Waaistraat 6 – mailbox 3511
3000 Leuven
BELGIUM

TEL: +32-(0)16 32 65 98

FAX: +32-(0)16 32 65 99

<http://www.econ.kuleuven.be/licos>



Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania

Jed Friedman^a, Kathleen Beegle^a, Joachim De Weerd^b, John Gibson^c

Abstract: There is wide variation in how consumption is measured in household surveys, both across countries and over time. This variation may confound welfare comparisons in part because these alternative survey designs produce consumption estimates differentially influenced by contrasting types of survey response error. While previous studies have documented the extent of net error in alternative survey designs, little is known about the relative influence of the different response errors that underpin a survey estimate. This study leverages a recent randomized food consumption survey experiment in Tanzania to shed light on the relative influence of these various error types. The observed deviation of measured household consumption from a benchmark is decomposed into item-specific consumption incidence and consumption value so as to investigate effects related to (a) the omission of any consumption and then (b) the error in value reporting conditional on positive consumption. Results show that various survey designs exhibit widely differing error decompositions and hence a simple summary comparison of the total recorded consumption across surveys will obscure specific error patterns and inhibit lessons for improved consumption survey design. In light of these findings, the relative performance of common survey designs are discussed and design lessons are drawn in order to enhance the accuracy of item-specific consumption reporting and, consequently, measures of total household food consumption.

JEL: C81, D12

Keywords: Food consumption, Household surveys, Response error, Recall, Telescoping

Author affiliations: ^a The World Bank, ^b University of Antwerp and KU Leuven, ^c University of Waikato

We wish to thank Francisco Ferreira, Alberto Zezza, two anonymous referees, and seminar participants at the FAO. 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

I. Introduction

Consumption or income, valued at prevailing market prices, is the workhorse metric of human welfare in economic analysis; poverty is almost universally defined in these terms. In low- and middle-income countries, these measures of household resource availability are typically assessed through household surveys. The global diversity in survey approaches is vast, with little rigorous evidence concerning which particular approach, in conjunction with which context, yields the most accurate resource estimate. Many other key dimensions of welfare, such as nutrition intake and hunger, are also widely assessed through household consumption survey (Fiedler et al., 2008). While levels of hunger and nutrition co-vary with household resource availability, the role of resources *vis-à-vis* other driving forces is debated (Deaton, 1997). The resolution to these debates has been influenced by the reliability of measures of both food consumption and economic resources (Bouis et al., 1992; Gibson and Kim, 2013).

This study focuses on food consumption measurement, leveraging a recent survey experiment to study the performance of commonly used consumption survey modules in order to shed light on the nature of reporting error in consumption data. The experiment randomly allocated to a nationally representative sample of Tanzanian households one of eight consumption survey modules. A daily supervised individual diary with dependent household members assigned to a unique adult respondent is taken as the benchmark, or “gold standard”, survey approach given the resources and care survey teams devoted to it (described in detail below). The accuracy of the other modules is assessed with respect to this benchmark. Previous work from the same experiment (Beegle et al., 2012, Gibson et al. 2015, De Weerd et al., 2016) explores relative performance of the eight modules in terms of mean consumption, inequality, poverty, and the prevalence of hunger. These studies concentrate on household-level total aggregates for food, non-food, and total consumption and do not consider variation in performance amongst individual items, as is done here. Moreover, variation in mean consumption by module – which is up to 27 percent of total value in these studies – conveys the net effect of all possible types of reporting error including opposing impacts of recall and telescoping error, as well as the difficulty of fully capturing individual consumption opportunities outside the home.

This paper extends the previous findings with a more careful focus on the nature of survey reporting error (relative to the benchmark). We do so by decomposing the sum of reported consumption into a product of two vectors: a vector of binary indicators recording whether the household reports any positive value consumed for each food sub-group or individual food item captured by the survey, and a real valued vector of the sub-group or item-specific value consumed. This framework – akin to a separate analysis of the extensive and intensive margin of reporting food consumption – allows for an exploration of the relative importance of the different types of reporting error in the seven survey designs. Further, this framework can relate the relative importance of these error types to individual commodity characteristics.

The next section briefly reviews the types of error in food consumption measurement captured by household survey. The third section describes the Tanzania survey experiment. The fourth section presents the analytic methods we employ and the fifth presents the results. The final section summarizes the findings and discusses the consequent implications for improved survey design.

II. Errors in consumption measurement: A brief taxonomy

The degree and nature of measurement error in consumption captured by household surveys depends, in part, on survey design features.¹ These design features vary along a large number of dimensions (Fiedler et al. 2012; Smith et al. 2014) such as the length of recall period or the level of item-specific detail sought. Moreover, as these features affect the estimates of household consumption, comparisons across countries and within countries over time are compromised when questionnaires change (see the extensive discussion of this issue for sub-Saharan Africa in Beegle et al., 2016).

Reporting error occurs when the information relayed by the respondent to the interviewer is not accurate. This error can take various forms including:

Recall error. A main concern is that respondents forget the occurrence of a consumption event resulting in what is termed recall error. Lower salience and longer recall periods make it more likely for respondents to forget (Sudman and Bradburn, 1973). Several studies show that, all else equal, the longer the period of recall, the lower the reported consumption per standardized unit of time (Scott and Amenuvegbe, 1991; Grosh et al, 1991).

Telescoping. The converse of recall error is telescoping, where a household compresses consumption that occurred over a longer period of time in to the reference period and thus reports consumption greater than the actual value.

Rule of thumb error. Respondents may not always recall and count events (Menon, 1993). Especially for longer recall periods that typically have more transactions, respondents may cease trying to enumerate each one and instead use rules of thumb to estimate them (Blair and Burton, 1987; Gibson and Kim, 2007; de Nicola and Gine, 2014 and Arthi et al., 2016). In this case, what is termed rule-of-thumb error depends on the frequency and regularity of the transaction with less frequent items reported with more error. Whereas recall error downwardly bias the consumption estimate, and telescoping creates upward bias, there is no obvious direction of bias for responses that resort to rule of thumb instead of enumeration. We may expect this error to be particularly pertinent for hypothetical consumption constructs such as questions about consumption in a “usual” month. “Usual” month consumption is an explicit attempt to abstract away from seasonal considerations in consumption, however this type of question may present additional cognitive demands compared to a definitive recall period in the immediate past.

Personal leave out error. Yet another source of reporting error is the inability to accurately capture individual consumption by household members if it occurs outside the purview of the survey respondent. This may be more significant for certain types of food, such as snacks or meals taken outside the home, or for personal goods such as mobile telecommunications. The degree of inaccuracy is likely to increase with the number of adult household members and with the diversity of their activities outside the home (World Bank, 2005).

Other error types. While the analysis in this study will focus on the four types of reporting error listed above, misreporting can also arise from other sources such as rounding error, social desirability bias, and strategic responses. An example of the latter is a respondent who exaggerates (or understates)

¹ The term “consumption survey” is used to generically refer to a household survey that collected detailed consumption information. Such surveys follow a range of labels such as Household Budget Survey, Living Standards Survey, and Household Income, Consumption and Expenditure Survey.

her consumption in order to appear poorer (or richer) due to a belief that the responses given may determine eligibility for some future social program. There may also be intentional misreporting in the light of respondent fatigue, so whether the respondent is presented with a long or a short list of consumption items can influence the quality of the responses.²

Errors in diary versus recall survey. The consumption diary is the main alternative to the recall approach to consumption measurement. It is generally expected that diaries suffer less from recall or telescoping errors, since the consumption is intended to be recorded either simultaneously or soon after it occurs. Of course this presumed accuracy is only achieved if the diary is used as intended. Thus the extent to which diaries are supervised to ensure they are regularly filled remains an important design feature. Unsupervised diaries may end up being, effectively, self-administered recall modules with endogenous recall periods if some types of respondents do not fill them in every day and, hence, subject to varying degrees of recall, telescoping, and rule of thumb reporting. Diaries administered at the individual level should also be better at capturing the individual consumption outside the household, leading to a higher level of measured household total consumption (Grootaert, 1986).

As a net result of these various types of reporting error, consumption estimates with different methods of data capture (diary versus recall questionnaires), levels of respondent (individual versus household), recall period, or degree of commodity detail may not be comparable. We designed the survey experiment used here in part to assess the extent to which variation across these dimensions affects item-specific and summary consumption measures in relation to the benchmark measure of the daily-supervised individual diary. We chose this diary design, described in more detail in the next section, to minimize the influence of recall, telescoping, personal leave-out, and rule of thumb error.

III. The Tanzania survey experiment

We conducted the Tanzania survey experiment to shed light on the implications of survey design variation in food consumption measurement by systematically contrasting various design features. Eight survey designs were strategically selected to reflect the most common methods utilized in low-income countries and are typical of the scope of variation one is likely to find in consumption surveys. We then randomly assigned these eight designs to over 4000 total households. Given the sample size and the random assignment of survey designs, differences in mean measurement performance can be attributed to the survey design rather than potential confounders with a high degree of confidence.

The designs differ by method of data capture (diary or recall survey), designated respondent (household head or other householder), length of reference period, number of items in the recall list, and nature of the cognitive task required of the respondent. Table 1 summarizes each of these designs – modules we number 1-5 are recall designs and modules 6-8 are diaries. For the food recall modules, households report the value consumed from three sources – purchases, home production, and gifts or payments. Modules 1 and 2 contain a list of 58 food items; Module 3 has a subset list that consists of the 17 most important food items that constitute, on average, 77 percent of food consumption expenditure in Tanzania based on the national Household Budget Survey 2000/01. To make module 3 comparable, we

² Beegle et al. (2012) find a reduction from 49 to 41 minutes when reducing the list of food for a one-week recall from 58 to 17 items. Interview times increased to 76 minutes for the 58-item list when using the more cognitively demanding ‘typical month’ phrasing.

scale up reported expenditures for that module (by $1/0.77$). Module 4 has a list of 11 food items. It is an aggregated version of the list of 58 food items where, for example, several individual listed vegetables are aggregated into “vegetables” as one item. The 58 individual food items in modules 1 and 2, those that are in the subset in module 3, as well as the aggregation of them for module 4 are shown in Appendix Table 1. The appendix table also lists seven items of a twelfth food group – “Meals outside the home”. Even though this food-outside-the-home group is asked in an identical manner across all modules – as a detailed 7 day-recall – we include it in the decomposition analysis since it is a food category that grows in importance as national incomes rise.

Among the recall modules, module 5 deviates from a reporting of actual consumption over a specified time period. Instead it asks for “usual” consumption, following a recommendation in Deaton and Grosh (2000), where households report the number of months in which the food item is usually consumed by the household and the average monthly value of what is consumed in those months. These questions aim to measure permanent rather than transitory living standards, without interviewing the same households repeatedly throughout the year. Hence, module 5 introduces two key differences from the other recall modules: a longer time frame and a different (and, we propose, more complicated) cognitive task required of respondents.

The three diary modules are of the standard “acquisition type”. Specifically, they add everything that came into the household through harvests, purchases, gifts, and stock reductions and subtract everything that went out of the household through sales, gifts, and stock increases. Modules 6 and 7 are household diaries in which a single diary is used to record all household consumption activities. These two household diaries differ by the frequency of supervision that each received from trained survey staff. Households assigned the infrequent diary received supervisory visits weekly while those with the frequent diary were visited every other day.

Module 8 is a personal diary, where each adult member keeps their own diary and consumption of children is captured in the diaries of the adults who knew most about their daily activities. Diary entries are specific to an individual and should leave no scope for double-counting purchases or self-produced goods. It is possible that a “gift” could be given to the household and accidentally recorded by two individuals. However, interviewers were trained to cross-check individual diaries for similar items purchased, produced, or gifted that occur on the same day and to query these during the checks. In many cases, one person will acquire food for the household (such as buying 5 kilograms of rice), which is entered in the diary of the person acquiring the food. Thus the personal diary is not an individual’s record of food consumption. Rather, it records the food acquired for the household by each member even if for several members to consume (as well as food consumed outside the household). Supervision visits occurred every other day for each individual respondent with a diary. This intensive supervision of the personal diary sample would be impractical for most surveys; these investments were made in order to establish a benchmark for analytic comparisons. We view module 8 as close to a 24-hour food-intake approach, not only because of the intensity of supervision but also because of the detailed cross-checks on meals to minimize food in-flows and out-flows that may be otherwise missed. Module 8 arguably provides the most accurate estimate of total household food consumption.

The field work was conducted from September 2007 to August 2008 in both rural and urban areas from seven districts across Tanzania: one district from each of the regions of Dodoma, Pwani, Dar es

Salaam, Manyara, and Shinyanga and two districts in the Kagera Region.³ The districts were purposively selected to capture variations in socio-economic characteristics. In each district, 24 communities were randomly selected from the 2002 Census, with probability-proportional-to-size (PPS). Within communities, a random sub-village (enumeration area, EA) was chosen and all households therein were listed. Twenty-four households per sub-village were randomly selected to participate and three households were randomly assigned to each of the eight modules. Among the original households selected there were 13 replacements due to refusals. Three households that started a diary were dropped because they did not complete their final interview. Another five households were dropped due to missing data on some of their key household characteristics, yielding a final sample size of 4,029 households.⁴

The basic characteristics of the sampled households generally match those from the nationally representative 2006-07 Household Budget Survey. The randomized assignment of households to the eight different questionnaire variants was successful in terms of balance across various characteristics relevant for consumption and consumption measurement (this analysis is presented in Beegle et al. 2012).

In regards reporting error, there are several points to note about the survey experiment. The recall modules 1-5 ask the respondent about consumption and not acquisition of food. These questionnaires record details on meals consumed outside the home by household members as well as meals within the household that were shared with non-household members. The diaries are acquisition diaries that account for food given to animals (e.g. scraps, or left-overs), food used for seed, food taken from stocks and food brought into the household by children (individual diary only). At the end of each week, there is a review of the main meals the household ate each day and additional information is recorded if any components for these meals were not captured in the diaries. This is important as the 2012 State of Food Insecurity report (FAO, 2012) incorporated, for the first time, tentative estimates of food losses, which lead to a significant revision of some of the world hunger numbers. Our diaries explicitly account for any food that has been used for seed, fed to animals or thrown away. The recall modules do this implicitly by asking about food consumed, which would eliminate seeds and animal feed being counted as consumption, but may not account for food scraps and left-overs that are fed to animals.

The survey was administered on paper. To minimize data entry errors, all questionnaires were entered twice and discrepancies were adjudicated. As non-standard units are common in Tanzania, the experiment collected conversion factors during a community price survey conducted by the field supervisors in each sample community. Supervisors used a food weighing scale to obtain a metric value of food-specific non-standard unit combinations. Median district-level metric conversion rates were used to convert non-metric units into kilograms or litres. Where district-level conversion rates were not available, the sample median was used. In a handful of cases where neither was available, measurements at the survey's headquarters were taken after the fieldwork was done. Further details on the experiment are described in Beegle et al. (2012) including the relative costs to field each module.

Table 2 presents the summary results of the consumption survey experiment, reporting the difference in the log per-capita consumption measure of each design relative to the benchmark individual

³ The survey teams were small, extensively trained on all modules, and well supervised. The teams stayed in the field for the entire 12 month study period in part to ensure that well-trained survey teams consistently applied the survey modules across all districts and also to abstract away from seasonal concerns that may interact with specific survey designs.

⁴ There is almost no item non-response in the consumption section of the recall modules, i.e. all respondents answer all questions for all consumption items (including a response of no, or zero, consumption).

diary.⁵ The estimates in Table 2 derive from regressions of the natural logarithm of food, non-food, and total consumption on binary indicators for module assignment (with the benchmark personal diary as the left-out category). Since the survey experiment was randomized, the regressions include no covariate controls except for survey cluster – the village or urban area sampling unit within which households were randomized to the various survey designs. The regressions in Table 2 show that, with the exception of 7-day recall with the long list, the other modules record between 8 and 33 percent less food consumption compared with the personal diary (column 3). The impact on total consumption is of a similar magnitude (column 2). Regarding the diary approach to food consumption, having just one respondent complete the diary for an entire household is associated with significantly lower food consumption of 13-20 percent, most likely because some unobservable personal consumption of other householders is omitted by the respondent maintaining the diary. Differences in frequent non-food consumption are also observed, especially in the diaries, again suggesting the importance of accurately recording personal consumption.⁶

Regarding the recall survey approach, all mean food expenditures are lower than the benchmark. The mean from the 7-day long list lies nearest to the benchmark value while modules with longer recall periods (14-day or “usual” month) or more aggregated consumption categories (the collapsed list) record food consumption that is 17 to 33 percent lower. Even though the 7-day long list comes closest to the mean benchmark food consumption value in this experiment, it is difficult to definitively extrapolate that the 7-day long list will be the most accurate of the recall designs when applied to different settings. As the net deviation of each module from the benchmark is the product of the contrasting influence of various types of reporting error, different settings may present differing magnitudes of underlying error types. The error decomposition analysis below is a first attempt to disentangle the relative influence of these types of reporting errors.

Beegle et al. (2012) also investigate how salient and easily observed household characteristics, those assumed to determine actual consumption levels, may also affect the accuracy of consumption reporting. Investigated characteristics include: (a) household size – where it was determined that recall modules under-report consumption even more as the size of the household increases, (b) urban location – where household diaries significantly under-report consumption in urban areas (but not rural) suggesting the relative prevalence of personal consumption opportunities in urban areas, (c) the education level of the household head – education had little relation to module performance except for the “usual” month approach where inaccuracy was higher among low educated households, and (d) household wealth as

⁵ While the experiment focused on food consumption measurement, each survey also recorded non-food consumption. For less-frequently purchased items, such as durable goods, clothes, and health care, all surveys and diaries employed a one-month or 12-month recall design (where households assigned to diaries were administered a non-food consumption survey at the end of the two week study period). For more frequently purchased non-food items such as soap or transport, consumption of these goods were either asked in recall form for the recall modules 1-5 (with period of recall corresponding to that for food) or recorded as diary entries for households assigned a diary.

⁶ Because the questionnaire wording and structure for the non-frequent non-food consumption section was identical across the 8 modules, it is perhaps surprising to see significantly negative coefficients for modules 1, 4, and 7 compared to the benchmark. Such differences can result from three sources: respondent fatigue, since these recalled items come after lengthy food recall sections in modules 1-5 or after a two-week diary; cognitive framing; and differing ability to capture personal non-frequent non-food consumption outside the purview of the main respondent. Contrary to concerns of respondent fatigue, module 4 with the collapsed food categories and shorter interview time yielded significantly less (14 per cent less) non-frequent non-food consumption. Possibly the lack of follow-up during the diary period made the module 7 respondents less diligent in the non-frequent non-food section of their final interview.

captured by a household asset index – under-reporting in recall modules is even greatest among the poorest households and the mean deviation significantly declines with wealth. It is currently an open question whether these household characteristics, shown to be important mediators for consumption reporting accuracy, are affected to differing degrees by the different types of reporting error. This possibility will also be investigated in the error decomposition framework introduced below.⁷

IV. Reporting error decomposition

The previous analysis of consumption reporting errors focused on a net measure of misreporting. As discussed, this masks two aspects to reporting consumption: that any consumption occurred and, when it did occur, the value of that consumption. Our main analytic approach of this paper is to examine these two aspects of misreporting, again in comparison with the benchmark module, by modeling total food consumption as a product of two vectors where each ordered element in the two vectors corresponds to an individual food good f . The first vector records, through an indicator function, whether the household reports any positive consumption of f . The second vector records the stated consumption value of each element. More formally, total consumption C recorded for household h by survey module m can be written as the following:

$$C_{hm} = \overline{I(C_{fhm} > 0)} * \overline{(C_{fhm} | C_{fhm} > 0)}$$

where the first vector in the above product is the consumption incidence vector and the second vector the consumption value.⁸

This decomposition enables a separate analysis of survey design effects on consumption incidence and consumption value (or quantity). Different survey designs may differentially affect these two sources of error and simple summary cross-module comparisons of total consumption may obscure these error patterns and consequently inhibit lessons for improvement in consumption survey design. Further, different research questions may also not be equally worried about errors in each of these vectors. For example, food diversity indices are often based on incidence, rather than value or quantity.

A straightforward regression framework is used to analyze the relative performance of the seven survey designs in relation to the benchmark module 8. For the specification with respect to consumption incidence we have:

$$I(C_{fhm}) = \beta_{m=8} + \beta_{fm}M_m + \varepsilon_{fhm}$$

⁷ Another important consideration is how the characteristics of the enumerator affect the interview quality and response error. Unfortunately this measurement experiment cannot shed much light on this question. First, the survey modules were equally balanced across enumerators, so any difference in relative module performance cannot be attributed to differential enumerator quality. Second, the characteristic distributions of the enumerators are much more uniform than for the general population – all of the enumerators finished secondary school but none yet joined university, were between 20 and 30 years of age, and from urban areas. This narrow range severely limits an analysis of response heterogeneity by enumerator characteristics. While data quality is a function, in part, of enumerator effort and quality, these characteristics are not easily observable. Future work along these lines might consider pre-fieldwork cognitive testing of enumerators to supplement inquiries of this nature.

⁸ In order for the two vectors to have the same dimension, and thus allow for total consumption to equate the inner-product, the value vector above needs to include the zero consumption values. Therefore the depiction above of the vector as consumption values conditional on positive consumption is purely stylistic to highlight the decomposition analysis to follow.

where M is a vector of indicators for module type. The individual diary, $m=8$, is the excluded category and therefore the constant $\beta_{m=8}$ represents the mean benchmark incidence. Regressions include survey cluster fixed effects and are estimated with OLS.⁹

Earlier work has demonstrated that household characteristics interact with survey design in non-trivial ways to produce error, and so may also interact in differential ways with respect to consumption incidence and value. Understanding the presence of these interaction affects can also be used to inform consumption survey design. Therefore an extended regression framework includes a household characteristic X – either the number of adult or child household members, household location, education level of the household head, or asset wealth – and interacts this characteristic with the survey module indicator, M :

$$I(C_{fhm}) = \beta_{m=8} + \beta_{fm}M_m + \beta_{fx}X_h + \beta_{fmx}M_mX_h + \varepsilon_{fhm}$$

In this specification the coefficient of interest is β_{fmx} , which relates how module effects are mediated by the household characteristics.

The same two regression specifications given above are used to explore the survey design effects on the value of consumption (conditional on a positive value) by replacing the dependent variable with the consumption value and dropping all observations that report zero consumption for that specific food item. Results will explore consumption incidence and then consumption value in turn.

We then extend this analysis by relating module specific reporting error for a particular food good to select characteristics of that good. It is possible for a survey design that minimizes error with respect to certain types of food goods to not be as effective with other food types. Consequently, the analysis will compare the design error as estimated above with respect to item specific features as consumption incidence (common or rare items), consumption value, the share of consumption from self-production, the frequency of market purchase, and the storability or perishability of goods. Such analysis will unveil some of the mechanisms that underlie the misreporting and enhance the relevance of our results outside the specific context of our survey experiment.

V. Results

Survey design and the report of consumption incidence

The consumption decomposition results begin with Table 3, which presents consumption incidence for twelve food groups relative to the benchmark module. The consumption incidence estimated by the benchmark is given by the constant term. Several lessons are immediately apparent, beginning with the relative performance of the 7-day and 14-day recall modules. These recall modules record significantly lower consumption incidence for most food groups. For example, while 67% of benchmark households report the consumption of Tubers, the 7- and 14-day long list recall designs (modules 1 and 2) report a significantly lower consumption incidence of 58% - 59%. The only food group reported at the same frequency as the benchmark is Vegetables; two food groups, Oils/Fats and Beverages, are actually reported at a significantly higher incidence of 5 - 6 percentage points. These two exceptions are true only for the long-list recall modules. The 7-day subset list and collapsed list (modules 3 and 4) under-report

⁹ Results are appreciably similar if binary response models (probit or logit) are used in place of OLS.

consumption for all food groups. Indeed the downward bias in incidence is even larger in magnitude for the 7-day subset and collapsed lists. For example, Tuber consumption incidence is estimated at 52% - 54%. While the consumption incidence of the 7-day shortened list may be expected to be lower than the 7-day long list, as that module design asks about a fewer number of individual food items, there is no prior expectation for the 7-day collapsed list to record a lower consumption incidence. The fact that the collapsed list does indeed record a lower incidence for all food items suggests that important consumption items are excluded due to a lack of cognitive prodding that the longer list explicitly incorporates in its design.

By contrast to the other recall modules, the “usual” month approach to recall survey design reports significantly higher consumption incidence for almost all food groups (with the sole exception of cereals – already consumed by 96% of benchmark households). This difference is most likely due to the very different cognitive demand of considering a “usual month”, which prompts respondents to report significantly higher consumption incidence than actually recorded in the benchmark. Finally the two household diary modules (modules 6 and 7) tend to report lower consumption incidence for various food groups such as Fruits (nine percentage points lower) and Meals Outside the Home (seven to nine percentage points lower). While the frequency of the household diary supervision doesn’t appear to influence the accuracy of consumption incidence measurement (as the rates are equal for the weekly and thrice-weekly supervised diaries), the two household diaries systematically record lower incidence compared to the personal diary.

Overall, these results show that an important component of recall error is the omission of any positive value of consumption for particular items. It is possible that a portion of this error is due to personal leave-out error, where the household respondent likely misses some individual consumption. However since the magnitudes of the incidence shortfall are relatively high for all recall modules (except the “usual” month) and occur even for nearly universally consumed items such as cereals, this indicates that a key channel for recall error to manifest is complete forgetting (or deliberate suppression).¹⁰ By contrast, the “usual” month approach prompts households to report a far higher monthly incidence of consumption than the benchmark suggesting a very different pattern of reporting error for this module. As the hypothetical nature of the question excludes telescoping as the cause, it is likely that the rules-of-thumb used by respondents underlie the misreports. We provide further evidence of this below by showing that the overestimates are worse for infrequently purchased items. Finally, the consistent shortfall in incidence of the two household diaries with respect to the personal diary points to the importance of personal leave-out error as this is the main driver of divergence between modules 6 and 7

¹⁰ The fact that the 7-day recall tends to report even lower consumption incidence than the 14-day recall can in principle be due to less diversified actual consumption in a one-week period than in a two-week period. Note however that the 7-day recall reports relatively lower incidence than the 14-day recall for those selected food groups in a non-linear fashion (going against expectation if the lower incidence is solely due to less frequent consumption across weeks). The conclusion that the even lower reported incidence in the 7-day recall is largely driven by recall error of greater magnitude (than the 14-day recall) is supported by a comparison of the 7-day recall module with consumption incidence recorded in the first week of the personal diaries. The shortfall in incidence is largely the same when compared with the first week of the personal diaries as with both weeks. Unfortunately it is impossible to conduct a similar analysis for the usual month as the personal diary was only collected for a two-week period. However combining personal diaries from two households fielded within the same calendar month can simulate consumption incidence over a one-month period. This exercise also reveals higher reported incidence by the usual month than in the personal diaries, suggesting that a significant portion of the higher consumption incidence in the usual month is indeed due to response error.

and the benchmark.

Key household characteristics may exacerbate (or moderate) the module-specific reporting error in consumption incidence. This can be explored, as described in the previous section, by interacting the module indicator with select characteristics. Table 4 summarizes these results by reporting the food groups for which significant interaction effects were estimated.¹¹ The effect of household characteristics on incidence report not only depends on the characteristic but also the particular food sub-group. For recall modules in general, the tendency to under-report incidence is mediated by urban location, household head education, and household wealth, at least for numerous key food groups such as Cereals, Sugars, and Meat and Fish (this is because while the module effect is negative, most of the interaction terms are positive). That is, under-reporting of any consumption in these food groups is highest for rural, low education, and low wealth households. For some food groups the number of household members, especially the number of children, also tends to exacerbate under-reporting. This implies that more disadvantaged households, those that are rural, have low education, have more children, and fewer assets are more likely to omit consumption during the survey experience. This will further exaggerate their monetary poverty status. These households may especially benefit from increased enumerator attention and explicit prompting for consumption incidence on a good by good basis.

In contrast, the household diaries seldom have significant interactions with household characteristics, suggesting that the downward bias in consumption incidence from these modules is fairly constant across all households. Exceptions to this include Fruit, Pulses, and Nuts/Seeds consumption incidence recorded for urban households, where the reported incidence of these groups is significantly lower. This implies that the individual consumption of select food groups is more likely to be missed in urban households with diaries than rural ones – perhaps these items are even more commonly eaten outside the home in urban areas than rural.

Survey design and the value of consumption

The same analytic framework used for the analysis of relative consumption incidence is applied to reports of consumption value (conditional on positive consumption) in Tanzania Shilling amounts.¹² Table 5 summarizes the module design effects for the reported consumption values, all converted to monthly equivalents. Differential reporting behavior by module type is clear. The 7-day recall records significantly higher consumption values for most food groups; the four exceptions are Tubers, Vegetables, Meat and Fish and Oils/Fats where the quantities reported are not different than the benchmark. As these goods are typically more perishable than other types and consequently purchased more frequently, perhaps the tendency to over-report consumption value is mitigated by these characteristics. In contrast, the 14-day recall values are all lower than the 7-day recall values and generally exhibit negative value errors. Only for Cereals, Dairy, and Meals Out do the 14-day recall values exceed the benchmark values, for all others it is lower and often significantly so.

The 7-day subset list (Module 3) tends to report even greater positive valued errors than the 7-day long list, which must be due to the module's focus on only the most commonly consumed items as that is

¹¹ The specific interaction terms are presented in Appendix Table 2. The main effects estimated by equation (3) are suppressed for ease of exposition.

¹² The monetary value results can also be interpreted as the effect on reporting quantities (kilos or litres).

the only design feature that distinguishes the two modules. In contrast the collapsed list exhibits both over-reporting and under-reporting.

It is not clear what causes these value error patterns. Over-reporting could occur when the value of salient episodes of consumption gets telescoped in to the recall period – presumably salient episodes are constituted by larger consumption values. Equally plausible is that respondents do not value each and every individual consumption event, but use a rule of thumb to do so. Over-reporting could then also occur if larger (and therefore more salient) episodes are used as the rule-of-thumb. The fact that reporting errors for some key food groups (such as Tubers, Sugars, Vegetables, and Fruits) switch from positive to negative as the recall period shifts from 7 to 14 days suggests one of two possibilities depending on what underlies the reporting behavior: (i) the negative influence of recall error outweighs the positive influence of telescoping as the recall period extends in length from 7 to 14 days whereas telescoping dominated the shorter period, (ii) alternatively, if rule-of-thumb reporting is utilized for both periods, rule-of-thumb tends to over-report to a greater degree with a shorter recall period.

A very different reporting pattern is seen for the “usual” month. While this module recorded a higher consumption incidence for most food groups (significantly higher than the benchmark in all but two cases), the values recorded by this module design are all significantly lower, with the sole exception of Cereals where the value reported is not significantly different than the benchmark. It appears that the particular cognitive challenges of the “usual” month approach, at least in a highly seasonal setting such as rural Tanzania where consumption patterns can vary widely throughout the year, result in overestimation of consumption incidence for most foods and an underestimation of value.

The two household diaries exhibit a distinct pattern of value reporting error. As the expected main driver of reporting error between either household diary and the personal diary would be the inability to fully capture personal consumption outside the home, the values reported should most diverge for such food goods. This is indeed what we find – the household diaries record 36 - 39% lower value of Fruit consumption, 37 - 42% lower value of Beverages, and 35 - 36% lower value of Meals Outside the Home. In contrast, Cereals, Tubers, and other basic foodstuffs are closer in value to the benchmark, although significantly lower at times for the frequently supervised household diary. The higher consumption values for the infrequently supervised diary vis-à-vis the frequently supervised one may suggest the partial presence of telescoping or rule-of-thumb errors (similar to what we observe with the 7-day recall) if indeed the infrequently supervised diary defaults to a short-period recall survey as a result of less frequent updating of the diary.

Table 6 summarizes how consumption value reporting error co-varies with particular household characteristics by listing the precisely estimated interaction terms as given by equation (3). (Appendix Table 3 relays the magnitude of all the estimated interaction terms.) Characteristics of households do mediate the degree of error in the value reported, although the patterns are less clear than for consumption incidence. One fairly clear result concerns the number of adults in the household; the more adults the greater the negative error in consumption valuation for goods likely consumed outside the home such as Fruits and Meals Out. This appears true for both recall and diary formats and indicates additional attention is warranted for households with many adults but only one survey respondent. Recall modules administered to households with many children (but not diaries) tend to differ with respect to the benchmark for various basic foodstuffs such as Vegetables, Fruits, and Meat/Fish. In general, the recall modules undervalue consumption for these item, so this undervaluation increases for households with many children – again indicating the importance of increased attention when larger households are

administered recall surveys.

Urban households administered recall modules also tend to underreport select food groups, especially so for the subset and collapsed list recall surveys (modules 3 and 4), but also exhibit over reporting for other food groups. There are some other specific results for various food groups and either the education or wealth level of the household, but no systematic pattern emerges, making it hard to generalize survey design lessons with respect to value reported and household measures of economic status. This is in contrast to the results in Table 3 that contained more clear patterns with consumption incidence error and select household characteristics, where more disadvantaged households are more likely to entirely omit consumption of key food groups.

Item characteristics and reporting error

This section explores the relation between select commodity characteristics and reporting error. Investigated commodity characteristics include (a) either commonly or uncommonly consumed items (in terms of the consumption incidence as reported in the benchmark individual diary), (b) the monthly value of the consumed item, (c) the frequency of purchase/consumption, (d) the share of item consumed out of own-production, and (e) the storability or perishability of the item.

Since most modules record consumption information on up to 58 individual food items, we construct the item-specific incidence and value error for 54 items relative to the incidence and value recorded in the personal diary benchmark.¹³ This analysis is not possible for the 7-day collapsed list module (module 4) as it lacks sufficient disaggregated information. In addition, the 7-day subset list (module 3) only contains item specific information for the 17 individual items listed and so is also removed from subsequent analysis.

For the remaining 5 modules, Figure 1 depicts the consumption incidence error (with respect to the benchmark) as a function of the commonality of the item (as measured by the item incidence recorded in the benchmark) for each of the 54 food items. A fitted linear regression line is also included in the module specific plots. Differential patterns in incidence error across modules are readily apparent. The incidence error of modules 1 and 2, while greater in magnitude for less commonly consumed items (especially for those items with a benchmark frequency less than .2), shows no relation with the commonality of the consumed item. The fitted regression line with a slope very close to zero confirms this lack of a linear relation. In contrast the “usual” month Module 5 reveals a decidedly negative relationship between consumption incidence and incidence error, with the least commonly consumed items in the benchmark reporting a very high positive incidence error. Clearly the “usual” month exercise leads to a gross overstatement of consumption frequency for these rare items, which include, for example, Macaroni and Spaghetti or Pork. In contrast the more frequently consumed items have a much lower incidence error (albeit still positive). Finally for the diaries, the pattern of error is reversed – the less frequently consumed items are actually reported with even less frequency in the household diaries, most likely reflecting, again, the influence of personal leave out error for these items while the more commonly consumed items are captured to a relative degree of accuracy by the household respondent.

¹³ The long list contained 58 food items. Due to possible confusion in the diary entries between (a) paddy rice and husked rice, (b) maize grain, maize cob, and maize flour, and (c) millet grain and millet flour, these categories are combined for this item-specific analysis, resulting in a total of 54 commodities available for analysis.

Another striking feature of Figure 1 relates to the spread of the reporting error. In all five figures there is a clear tendency for the incidence error to be much further away from zero (in either positive or negative direction) for less commonly consumed items and for the spread to reduce substantially with higher consumption incidence. This patterns fits well with some of the social and cognitive psychology literature, discussed in Section 2, which highlights salience and regularity as key dimensions in frequency reporting, in particular when respondents do not ‘count and recall’ but instead use rate-based estimation techniques.

Now that we have established incidence reports are more accurate for high-incidence items (in terms of less spread in the reporting error), we can ask a parallel question concerning the value reported. Here we may hypothesize that the salience of high-value items in recall surveys leads to lower errors in their value reporting. Figure 2 plots value error as a function of benchmark consumption value. The first thing to note is, in general, no convergence to the benchmark value for higher value items, as was the case for incidence reporting of commonly consumed items. Recall that Table 5 revealed a mix of over- or under-reporting for module 1 (14-day long list), depending on the food group. The relevant scatter-plot in Figure 2 further suggests that the net over-reporting error is greater for consumption items of relatively low value. As the benchmark value increases, the net error approaches zero and then even turns negative at high value items signifying that items largest in value suffer from under-reporting. The same directional pattern is apparent for module 2 (7-day long list) where each food group in Table 5 displays net over-reporting. Figure 2 suggests that particularly high relative errors occur for foods at below median consumption value.

The “usual” month also shows the same error gradient as the other two recall modules, although here the net error for most food items is negative (as was true for the food groups listed in Table 5). The magnitude of this error increases with the benchmark consumption value suggesting that the highest value consumption items are associated with the greatest degree of underreporting. This likely indicates that either recall error dominates value reporting for this module or a rule-of-thumb value reporting for the “usual month” induces negative error that grows in magnitude with consumption value. Finally for the diary modules, and in contrast to the three recall modules analyzed, there is little systematic reporting error with respect to value. Clearly the bigger challenge with the household diaries is ensuring the recording of any consumption for some food groups (the incidence), rather than recording accurate consumption values conditional on incidence.

Besides consumption incidence and value, we investigate three other dimensions of the individual food goods – the frequency of market purchase, the share of consumption out of self-production, and whether the food is storable or perishable. Table 7 lists the mean proportional errors relative to the benchmark for these three characteristics by each of the five modules. Often any difference in mean error across each characteristic is not statistically significant, in part because the power of the test is relatively low – there are only 54 observations informing the two means. For this reason the standard errors are not shown. Nevertheless Table 7 contains some fairly suggestive results relating food characteristics to the degree of reporting error.

The frequency of purchase of a particular food appears to convey a salience reflected in the accuracy of report. For frequently purchased items (defined as those above the median purchase incidence of 1.4 times in a two week period), the incidence error is lower for all five modules investigated, especially for the “usual” month where the error magnitude is less than half as much as the over-reporting for infrequently purchased foods. The value error is also far lower for frequently purchased goods among

the recall modules. Combined, this suggests that additional attention must be paid to infrequently consumed goods to improve accuracy of both the reported incidence and the reported value conditional on incidence. For the diaries, the frequency of purchase is not substantially related to the value of the good recorded as we might expect in settings if transactions are recorded soon after they occur, regardless of purchase frequency.

Somewhat surprisingly, there is no apparent relation of the share of self-production to incidence reporting error – across all five modules the incidence error for goods more commonly consumed from self-production and those seldom consumed are very close. The same holds true for consumption value error except for the 14- and 7-day recall. Under 14-day recall, value error is positive for foods consumed with a high share of self-production, while negative for foods with a low share. On the other hand, the 7-day module exhibits a much higher degree of over-reporting for foods with a low share of consumption from self-production.

In terms of storability, 26 individual food items are defined as storable (i.e. dry goods such as maize flour) and 28 as perishable (i.e. the various types of fresh vegetables). There are some apparent differences in errors between the two characteristics. The incidence error is always greater in absolute magnitude for perishable goods – the “usual” month overstates the incidence of perishable goods while the other four understate the incidence relative to the benchmark. With respect to consumption value, perishable goods also tend to demonstrate reporting errors of greater magnitude, i.e. the over-reporting in the 7-day module is greater for perishables, as is the magnitude of under-reporting for the “usual” month and the diaries. The 14-day recall reveals directionally different error for the two types of goods with over-reporting for storable goods and a slight under-reporting error for perishables. There is no equivalent pattern for the two diary modules.

VI. Concluding lessons for food consumption survey design

This paper utilizes data from a Tanzania consumption experiment to describe the nature of reporting error in food consumption surveys. The goal of this exercise is not merely to document error patterns but to generate hypotheses for an improved consumption survey design. Of course any such suggestions must be regarded as preliminary as the conclusions are derived from one setting (albeit national in scope) that is largely rural with diets based on a particular staple crop. Further work is needed to establish the generalizability of any findings reported here. However we hope that the focus on underlying mechanisms of misreporting error will increase the external validity of the findings reported here.

We use a simple analytic framework of error decomposition, which compares consumption incidence and consumption value to a benchmark of intensively supervised individual diaries. This basic decomposition immediately points to clear patterns in module divergence from the benchmark.

The omission of any consumption is a major cause of bias in the standard recall modules (except the “usual” month). While the 7-day long list module comes closest in its estimate of mean household food consumption (as shown in Table 2), the incidence of consumption is significantly lower than the benchmark for most food groups. The same finding holds true for the other 7-day recall designs (modules 3 and 4) and also, to a lesser degree, for the 14-day long list.

While consumption incidence is largely downward biased for the 7-day recall, the reported values, conditional on positive consumption, exhibit a large degree of upward bias for most food groups, whether commonly consumed (Cereals) or not (Dairy). The relative equality in mean consumption of the 7-day recall to the benchmark individual diary may therefore be due to the happenstance of off-setting errors: negative errors in incidence multiplied by positive errors in value. This approximately equal off-setting error magnitude may not translate to other settings, thus raising questions about a characterization of the 7-day recall as the most accurate recall design. The 7-day subset list recall module (module 3) also exhibits an extensive degree of over-reporting of value. We conclude that value over-reporting (due to telescoping or rule-of-thumb) is most pronounced in the short 7-day window. On the other hand, the 14-day recall module also exhibits positive value errors for some food-groups but to a lesser degree than the 7-day period. The 14-day recall also exhibits net negative error for other, mostly perishable food groups. Clearly the recall period that yields the greatest accuracy will vary with the nature of the good in question – an issue we return to below.

This simple decomposition exercise, which raises some very stark patterns in differential response by recall module, suggests that methods to improve the accuracy of recall surveys should consider a dual track approach: (a) efforts related to the prompting of households to report any positive consumption, as recall modules under-report consumption incidence for almost any food group, and (b) efforts aimed to improve the accuracy of consumption value reported, conditional on the household reporting any consumption.

Regarding efforts to improve the accuracy of consumption incidence, note that the absolute magnitude of recall error is particularly high for foods that are not commonly consumed. Therefore further prompting in the interview and perhaps the use of locally salient images to aid the memory of survey respondents may be avenues to explore, especially for less common items.

Regarding efforts to improve the accuracy of consumption value reports, note that over-reporting of value is a particularly prominent error for short recall periods such as 7 days. Unlike the incidence error, we do not have a strong lead on what causes the misreports. If telescoping of high-value consumption items is the driver, then one approach sporadically discussed in the literature is to bracket the recall period with a surveyor visit at the start of the relevant period as a reference to the respondent for the visit when the recall is reported. The initial visit from a survey team may help delineate the exact period of recall and improve accuracy, however this approach would need to be validated before being taken up more broadly. It also has consequences for field work structure and cost (in cases where mobile teams rather than resident enumerators are used).

We extended the basic analytic framework with respect to how food reporting errors across modules interact with key socio-economic characteristics of the households and with food characteristics such as perishability or prevalence of home production.

With respect to interactions with household characteristics, the bias in consumption incidence is most pronounced for disadvantaged households such as those with low education or household wealth, larger numbers of household members, and those residing in rural areas. Thus the difficulty of recall to capture consumption is exacerbated for the more disadvantaged households. The degree of error in value is also attenuated, albeit to a far less degree, by education, wealth, urban location, and small household size. Survey methods to improve accuracy may thus need to be particularly sensitive to poorer and less educated households, perhaps with more considered and attentive prompting and an enhanced use of

images and easily understood commonly used quantity measures.

With respect to item characteristics, recall modules also tend to report lower consumption incidence relative to the benchmark, as well as report higher magnitudes of consumption value error, for less frequently purchased items. A separate strategy may need to be employed for the relatively rare purchased items. Similarly, recall models tend to have incidence error of greater absolute value for perishable items than durables (although both types of goods are reported at lower frequency than in the benchmark). 7-day recall also exhibits greater value errors for perishables than durables. For both incidence and value reports these findings suggest scope to tailor the recall period to the individual good, although the current knowledge base is not yet sufficient to inform the optimal recall period. Regarding the subset and collapsed recall modules (Modules 3 and 4), Beegle et al. (2012) report that the time savings from these modules is not especially significant: 41 and 42 minutes to complete these modules, respectively, versus 49 minutes for the long list 7-day module. Given the relative gain in accuracy from the longer-list version, only the most time-constrained surveys (where a saving of seven interview minutes has a real benefit on data quality) should consider these shorter modules.

Apart from the recall modules rooted in the very recent past (7 or 14 days), the “usual” month approach seeks the respondent to abstract away from a concrete time period and report on consumption for an idealized construct. We have demonstrated that this approach results in a very different error pattern, with consumption incidence significantly higher than the benchmark. Where “forgetting any consumption” appears to be a problem with standard recall modules, the “usual” month exercise yields a “crowding-in” of consumption to a large degree, especially for food items that are not commonly consumed in the benchmark. While consumption incidence is biased upward in the “usual” month, the consumption value appears to be significantly under-reported for all food groups except Cereals. Compounding these problems, the error magnitude in both incidence and value is substantially worse for disadvantaged households. Finally, this module took 76 minutes to complete on average – far longer than the 49 minutes for the 7-day recall. The longer length to completion likely derives from the large cognitive burden placed on respondents by the “usual” month approach. As this large cognitive burden also results in a stark divergence from the benchmark in a variety of dimensions, the “usual-month” is not a survey design we can recommend.

The error patterns in the household diaries are somewhat more straightforward to interpret as the point of divergence with the benchmark is the reliance of one householder to record consumption for all members, and hence the influence of personal leave-out likely takes prominence (along with differences in the frequency of supervision for the less frequently supervised module). The household diaries do record 7% - 10% lower consumption incidence for food goods such as Fruits, Beverages, and Outside Meals, which are more likely to be consumed by individual members when alone or outside the household. The reported values are also lower for many of the same commodities. Interestingly for some food groups, the negative error in consumption value for the infrequently supervised household diary is smaller in magnitude than for the frequently supervised one. In one example, the frequently supervised diary yields a consumption value for Nuts and Seeds 17% less than the benchmark while the infrequently supervised one is only 7% less. Since, as described earlier, the infrequently supervised diary can transform into recall surveys over relatively brief periods between supervisory visits if the household does not fully adhere to the diary format, the larger consumption values for the infrequently supervised diary may point to a similar positive reporting error phenomenon as seen in the 7-day recall survey. This error,

in turn, partly compensates for the leave-out error thus bringing the value from the infrequent diaries actually closer to the benchmark.

Since household diaries are substantially less resource intensive than the benchmark personal diary, they will remain a commonly utilized tool. As suggested by these results, some steps can be taken to improve accuracy: even if one householder will record all consumption, prompting explicit consultation with all adult household members around the consumption of items commonly consumed alone or outside the house may serve to reduce the degree of personal leave-out error. The fieldwork of diary supervisors may also be structured so that more frequent supervision is possible (by optimizing work routes with a constraint on repeat visit). Where this isn't possible then bounding notices such as stickers for dates of past and future visits can call explicit attention to relevant time periods for the consumption records.

Our analysis has focused on the influence of errors generated by forgetting, telescoping, the use of rules-of-thumb, and personal leave-out error, and has suggested changes in survey design that may reduce the influence of such error types. We conclude with an additional brief discussion of small design changes, unlikely to impact survey costs, which may reduce the influence of other types of response error not investigated here. Two examples are given below and we refer interested readers to a more comprehensive treatment of this issue in Smith et al. (2014).

First, intentional error could also stem from interviewers subtly guiding respondents to give answers that minimize interview length, or who rush to complete the questionnaire. We can assume that such errors become more likely as questionnaires get longer and if supervision is limited. This type of error has also been observed in high frequency panels where follow-up survey rounds are likely less accurate than earlier rounds (Halpern-Manners and Warren, 2012). Extensive enumerator training and active field supervision that emphasizes adherence to study design can minimize these errors.

Second, throughout most of the developing world households do not typically purchase, harvest, or consume their food in standard units (kilograms or liters). Some surveys force reporting in standardized units and there are doubts about the accuracy of these reports when made by people who rarely transact in metric units. Alternatively, consumption surveys can allow the respondent to report in local units, such as bunches, heaps, tins, buckets, or bundles. However, in order to aggregate food consumption, these local units must be converted into standard units. If common foods are more likely to be reported in non-standard units (i.e. pieces of cassava, bunches of bananas) and conversion factors are inadequate, then unit conversion error could significantly distort the resulting value estimates. A final relatively low-cost addition to existing consumption surveys would be to ensure that such conversion factors, which are often geographically specific, are locally relevant, well specified, and systematically collected by survey teams.

VII. References

- Arthi V, Beegle K, De Weerd J and Palacios-Lopez A. 2016. "Measuring household labor on Tanzanian farms." Mimeo.
- Beegle K, De Weerd J, Friedman J, Gibson J. 2012. "Methods of household consumption measurement through surveys: Experimental results from Tanzania." *Journal of Development Economics*, 98(1): 3–18.
- Beegle K, Christiaensen L, Dabalen A, Gaddis I. 2016. Poverty in a Rising Africa. Africa Poverty Report. Washington, DC: World Bank.
- Blair E, Burton S. 1987. "Cognitive processes used by survey respondents to answer behavioral frequency questions." *Journal of Consumer Research*, 15(3): 280-288.
- Bouis H, Haddad L, Kennedy E. 1992. "Does it matter how we survey demand for food? Evidence from Kenya and the Philippines." *Food Policy*, 17(5): 349-360.
- de Nicola F, Giné X. 2014. "How accurate are recall data? Evidence from coastal India." *Journal of Development Economics*, 106(1): 52-65.
- Deaton A. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. The World Bank: Washington DC.
- Deaton A, Grosh M. 2000. "Consumption", in *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*, M Grosh and P Glewwe, eds. New York: Oxford University Press (for World Bank).
- De Weerd J, Beegle K, Friedman J, Gibson J. 2016. "The challenge of measuring hunger of Measuring Hunger through Survey." *Economic Development and Cultural Change* (forthcoming).
- FAO. 2012. *The State of Food Insecurity in the World: Economic Growth is Necessary but Not Sufficient to Accelerate Reduction of Hunger and Malnutrition*. Rome: Food and Agriculture Organization of the United Nations.
- Fiedler J, Smits M, Dupriez O, Friedman J. 2008. "Household income and expenditure surveys: a tool for accelerating the development of evidence-based fortification programs." *Food and Nutrition Bulletin*, 29(4): 306-319.
- Fiedler J, Carletto C, Dupriez O. 2012. "Still waiting for Godot? Improving household consumption and expenditures surveys (HCES) to enable more evidence-based nutrition policies." *Food and Nutrition Bulletin*, 33(3): S242-S251.
- Gibson J, Beegle K, De Weerd J, Friedman J. 2015. "What does variation in survey design reveal about the nature of measurement errors in household consumption?" *Oxford Bulletin of Economics and Statistics*, 77(3): 466-474.
- Gibson J, Kim B. 2013. "How reliable are household expenditures as a proxy for permanent income? Implications for the income–nutrition relationship." *Economics Letters*, 118(1): 23-25.
- Gibson, J, Kim, B. 2007. "Measurement error in recall surveys and the relationship between household size and food demand." *American Journal of Agricultural Economics*, 89(2): 473-489.
- Grootaert, C. 1986. "The use of multiple diaries in a household expenditure survey in Hong Kong." *Journal of the American Statistical Association*, 81(396): 938-944.

- Grosh M, Zhao Q, Jeancard H. 1995. "The sensitivity of consumption aggregates to questionnaire formulation: Some preliminary evidence from the Jamaican and Ghanaian LSMS surveys." Improving the Policy Relevance of the Living Standards Measurement Study Surveys, Research Paper Series Paper No. 6. The World Bank: Washington DC
- Halpern-Manners A, Warren J. 2012. "Panel conditioning in longitudinal studies: Evidence from labor force items in the Current Population Survey." *Demography*, 49(4): 1499-1519.
- Menon G. 1993. "The effects of accessibility of information in memory on judgments of behavioral frequencies." *Journal of Consumer Research* 20(3): 431-440.
- Scott C, Amenuvegbe B. 1991. "Recall loss and recall duration: An experimental study in Ghana." *Inter-Stat*, 4(1): 31-55.
- Smith L, Dupriez O, Troubat N. 2014. "Assessment of the reliability and relevance of the food data collected in national Household Consumption and Expenditure Surveys." IHSN Working Paper No. 008.
- Sudman S, Bradburn N. 1973. "Effects of time and memory factors on response in surveys." *Journal of the American Statistical Association*, 68(344): 805-815.
- World Bank. 2005. "Russian Federation: Reducing poverty through growth and social policy reform." Report 28923-RU. The World Bank: Washington DC.

Table 1. Survey experiment consumption modules

Module	Consumption measurement design	Recall/survey period	Number of households
1	Long list of 58 food items; 14 day recall	14 day	503
2	Long list of 58 food items; 7 day recall	7 day	504
3	Subset list of 17 food items; 7 day recall	7 day	504
4	Collapsed list of 11 broad food categories	7 day	504
5	Long list of 58 food items; usual 12 month recall	Usual 12 months	504
6	14 day household diary with frequent visits ^a	14 days	503
7	14 day household diary with infrequent visits ^b	14 days	503
8	14 day individual diaries with frequent visits ^a	14 days	503
			4,029

Note: (a) Frequent visits entailed daily visits by the local assistant and visits every other day by the survey enumerator for the duration of the 2-week diary. (b) Infrequent visits entail 3 visits: to deliver the diary (day 1), to pick up week 1 diary and drop off week 2 diary (day 8), and to pick up week 2 diary (day 15). Households assigned to the infrequent diary but who had no literate members (about 18 percent of the 503 households) were visited every other day by the local assistant and the enumerator.

Table 2. Proportional deviation of log PCE from benchmark consumption module

Module type	Ln total PCE	Ln food PCE	Ln non-food, frequent (recall or diary)	Ln non-food, non-frequent (all recall)
1. Long list, 14 day recall	-0.161*** -0.037	-0.167*** -0.037	-0.104 -0.067	-0.105* -0.06
2. Long list, 7 day recall	-0.039 -0.037	-0.017 -0.037	-0.134** -0.067	-0.096 -0.06
3. Subset list, 7 day recall	-0.071* -0.037	-0.079** -0.037	-0.112* -0.067	-0.09 -0.06
4. Collapsed list, 7 day recall	-0.283*** -0.037	-0.332*** -0.037	-0.104 -0.067	-0.138** -0.06
5. Long list, usual month	-0.207*** -0.037	-0.268*** -0.037	0.023 -0.067	-0.013 -0.06
6. HH diary, frequent	-0.173*** -0.037	-0.196*** -0.037	-0.279*** -0.067	-0.046 -0.06
7. HH diary, infrequent	-0.136*** -0.037	-0.129*** -0.037	-0.244*** -0.067	-0.105* -0.06
Number of households	4,025	4,025	3,942	4,016

Note: *** p<0.01, ** p<0.05, * p<0.1. Omitted category is personal diary. Standard errors in parentheses. Columns 3 and 4 have smaller sample sizes due to some households with zero non-food expenditures. 83 households have no frequent non-food expenditures (6, 6, 1, 9, 7, 17, 17, and 20 households for modules 1-8, respectively). 9 households have no non-frequent non-food expenditures (3, 3, 1, and 2 households for modules 5, 6, 7 and 8, respectively).

Source: Beegle et al. (2012)

Table 3. Consumption incidence reported for each food group, relative to benchmark

Module	Cereals	Tubers	Sugars	Pulses	Nuts/seeds	Vegetables
1. Long list, 14 day recall	-0.048*** (0.015)	-0.088*** (0.019)	-0.141*** (0.024)	-0.055** (0.025)	-0.067*** (0.024)	-0.006 (0.012)
2. Long list, 7 day recall	-0.067*** (0.015)	-0.084*** (0.019)	-0.158*** (0.024)	-0.111*** (0.025)	-0.150*** (0.024)	-0.008 (0.012)
3. Subset list, 7 day recall	-0.093*** (0.015)	-0.134*** (0.019)	-0.232*** (0.024)	-0.166*** (0.025)	-- (0.024)	-0.022* (0.012)
4. Collapsed list, 7 day recall	-0.091*** (0.015)	-0.150*** (0.019)	-0.228*** (0.024)	-0.136*** (0.025)	-0.414*** (0.024)	-0.085*** (0.012)
5. Long list, usual month	0.022 (0.015)	0.110*** (0.019)	0.109*** (0.024)	0.201*** (0.025)	0.217*** (0.024)	0.020* (0.012)
6. HH diary, frequent	-0.028* (0.015)	-0.031* (0.019)	-0.095*** (0.024)	-0.026 (0.025)	-0.070*** (0.024)	-0.010 (0.012)
7. HH diary, infrequent	-0.006 (0.015)	-0.002 (0.019)	-0.074*** (0.024)	-0.000 (0.025)	-0.097*** (0.024)	-0.008 (0.012)
Constant (omitted category mean)	0.956*** (0.049)	0.672*** (0.063)	0.436*** (0.080)	0.828*** (0.084)	0.739*** (0.079)	0.890*** (0.040)

Module	Fruits	Meat/fish	Dairy	Oils/fats	Beverages	Meals out
1. Long list, 14 day recall	-0.073*** (0.023)	-0.055*** (0.018)	-0.076*** (0.027)	0.058*** (0.018)	0.048** (0.023)	-0.149*** (0.026)
2. Long list, 7 day recall	-0.124*** (0.023)	-0.095*** (0.018)	-0.106*** (0.027)	0.060*** (0.018)	0.056** (0.023)	-0.195*** (0.026)
3. Subset list, 7 day recall	-- (0.018)	-0.111*** (0.018)	-0.158*** (0.027)	-0.206*** (0.018)	-- (0.018)	-0.139*** (0.026)
4. Collapsed list, 7 day recall	-0.218*** (0.023)	-0.190*** (0.018)	-0.134*** (0.027)	-0.184*** (0.018)	-0.295*** (0.023)	-0.173*** (0.026)
5. Long list, usual month	0.205*** (0.023)	0.038** (0.018)	0.237*** (0.027)	0.074*** (0.018)	0.280*** (0.023)	-0.111*** (0.026)
6. HH diary, frequent	-0.092*** (0.023)	-0.004 (0.018)	-0.062** (0.027)	-0.000 (0.018)	-0.079*** (0.023)	-0.070*** (0.026)
7. HH diary, infrequent	-0.090*** (0.023)	-0.006 (0.018)	-0.022 (0.027)	0.004 (0.018)	-0.074*** (0.023)	-0.088*** (0.026)
Constant (omitted category mean)	0.563*** (0.075)	0.761*** (0.060)	0.415*** (0.090)	0.816*** (0.059)	0.467*** (0.076)	0.574*** (0.084)

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated with a linear probability model (OLS) that regresses a binary indicator for positive consumption on indicators for module type as well as community fixed effects.

Table 4. Summary of interaction effects between consumption incidence and select household characteristics, by module type and food group

Household characteristic:	Number of adults		Number of children		Urban		Education of head		Household wealth	
Module	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1
1. Long list, 14 day recall	Dairy		Sugars, Nuts/seeds, Dairy		Meals out	Sugars	Oils/fats	Tubers, Sugars, Pulses, Meat/fish, Meals out		Sugars
2. Long list, 7 day recall	Tubers, Pulses, Fruits, Meals out		Sugars	Cereals		Cereals, Sugars, Nuts/seeds	Oils/fats	Tubers, Sugars, Pulses, Nuts/seeds, Meat/fish, Meals out		Cereals, Tubers, Sugars, Pulses, Nuts/seeds, Meat/fish, Meat/fish
3. Subset list, 7 day recall			Dairy, Oils/fats		Tubers	Cereals, Sugars, Oils/fats	Vegetables	Tubers, Sugars, Pulses, Meat/fish, Oils/fats, Meals out		Cereals, Tubers, Sugars, Pulses, Meat/fish, Oils/fats
4. Collapsed list, 7 day recall		Cereals	Sugars, Fruits, Meats/fish		Tubers, Nuts/seeds, Beverages	Cereals, Sugars, Vegetables, Oils/Fats		Tubers, Sugars, Pulses, Meat/fish, Oils/fats, Meals out	Nuts/seeds	Cereals, Sugars, Pulses, Fruits, Meat/fish, Oils/fats
5. Long list, usual month	Dairy, Meals out		Dairy, Oils/fats		Pulses, Nuts/seeds, Fruits, Beverages		Pulses, Nuts/seeds, Fruits, Beverages	Dairy, Meals out	Pulses, Nuts/seeds, Fruits, Beverages	Dairy
6. HH diary, frequent	Pulses		Nuts/seeds		Fruits		Pulses, Fruits		Dairy	
7. HH diary, infrequent	Beverages				Pulses, Nuts/seeds, Fruits			Beverages, Meals out		Sugars

Note: Interactions estimated with a linear probability model (OLS) that regresses a binary indicator for positive consumption on indicators for module type, a household characteristic, and an interaction term between module type and the household characteristic, as well as community fixed effects. Only food groups with interaction coefficients estimated at $p < .10$ are reported.

Table 5. Log consumption value (conditional on positive consumption) relative to benchmark, by food group

Module	Cereals	Tubers	Sugars	Pulses	Nuts/seeds	Vegetables
1. Long list, 14 day recall	0.036 (0.052)	-0.208*** (0.061)	-0.026 (0.076)	-0.026 (0.070)	-0.001 (0.088)	-0.182*** (0.054)
2. Long list, 7 day recall	0.244*** (0.052)	0.013 (0.061)	0.244*** (0.076)	0.233*** (0.071)	0.291*** (0.092)	-0.081 (0.054)
3. Subset list, 7 day recall	0.193*** (0.053)	0.136** (0.062)	0.511*** (0.079)	0.574*** (0.073)	-- (0.073)	0.227*** (0.054)
4. Collapsed list, 7 day recall	-0.119** (0.052)	-0.163*** (0.062)	0.286*** (0.079)	0.210*** (0.072)	0.100 (0.119)	-0.368*** (0.055)
5. Long list, usual month	0.024 (0.051)	-0.539*** (0.057)	-0.758*** (0.070)	-0.513*** (0.065)	-0.459*** (0.081)	-0.409*** (0.053)
6. HH diary, frequent	-0.156*** (0.052)	-0.074 (0.059)	-0.083 (0.074)	-0.171** (0.069)	-0.173* (0.089)	-0.163*** (0.054)
7. HH diary, infrequent	-0.060 (0.051)	-0.006 (0.059)	0.063 (0.074)	-0.009 (0.068)	-0.074 (0.090)	-0.043 (0.054)
Constant (omitted category mean)	11.085*** (0.174)	9.357*** (0.228)	8.901*** (0.365)	9.380*** (0.222)	9.689*** (0.300)	9.611*** (0.186)

Module	Fruits	Meat/fish	Dairy	Oils/fats	Beverages	Meals out
1. Long list, 14 day recall	-0.126 (0.084)	-0.247*** (0.068)	0.440*** (0.142)	-0.246*** (0.058)	0.090 (0.103)	0.413*** (0.097)
2. Long list, 7 day recall	0.188** (0.086)	-0.022 (0.069)	0.746*** (0.147)	0.019 (0.058)	0.389*** (0.103)	0.364*** (0.098)
3. Subset list, 7 day recall	-- (0.069)	-0.120* (0.069)	0.828*** (0.156)	0.222*** (0.063)	-- (0.063)	0.604*** (0.096)
4. Collapsed list, 7 day recall	-0.024 (0.090)	-0.396*** (0.071)	0.673*** (0.152)	0.055 (0.062)	0.745*** (0.125)	0.362*** (0.098)
5. Long list, usual month	-0.790*** (0.078)	-0.448*** (0.066)	-1.064*** (0.122)	-0.408*** (0.057)	-0.296*** (0.097)	0.352*** (0.095)
6. HH diary, frequent	-0.394*** (0.084)	-0.196*** (0.067)	0.024 (0.142)	-0.040 (0.059)	-0.374*** (0.108)	-0.348*** (0.094)
7. HH diary, infrequent	-0.360*** (0.084)	-0.104 (0.067)	0.152 (0.137)	0.005 (0.059)	-0.424*** (0.108)	-0.364*** (0.095)
Constant (omitted category mean)	8.703*** (0.363)	9.588*** (0.255)	8.429*** (0.448)	9.069*** (0.208)	8.297*** (0.453)	10.663*** (0.413)

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated with a regression for the log value of consumption on indicators for module type as well as community fixed effects.

Table 6. Summary of interaction effects between log consumption value and select household characteristics, by module type and food group

Household characteristic:	Number of adults		Number of children		Urban		Education of head		Household wealth	
Module	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1	Negative interaction, p<0.1	Positive interaction, p<0.1
1. Long list, 14 day recall	Fruits, Meals out		Vegetables, Meals out	Nuts/seeds		Sugars, Meals out	Pulses, Meat/fish	Tubers, Sugars, Oils/fats, Meals out		Sugars
2. Long list, 7 day recall	Vegetables, Meals out		Vegetables, Fruits, Meals out		Nuts/seeds	Cereals, Sugars	Pulses, Meals out	Tubers, Sugars, Nuts/seeds, Meat/fish, Oils/fats	Pulses, Fruits	Cereals, Tubers, Sugars, Nuts/seeds, Meat/fish
3. Subset list, 7 day recall	Meals out		Vegetables, Meat/fish, Oils/fat, Meals out		Cereals, Sugars, Oils/fats	Tubers	Pulses, Meat/fish, Oils/fats, Meals out	Tubers, Sugars, Vegetables	Cereals, Sugars, Pulses, Oils/fats	Tubers, Meat/fish
4. Collapsed list, 7 day recall	Fruits, Meat/fish, Meals out		Meat/fish		Cereals, Sugars, Nuts/seeds, Vegetables	Tubers, Oils/fats, Beverages	Pulses, Meat/fish, Oils/fats, Meals out	Tubers, Sugars	Cereals, Sugars, Pulses, Nuts/seeds, Meat/fish, Oils/fats	Fruits
5. Long list, usual month	Vegetables, Fruits, Meals out		Sugar, Vegetables, Fruits, Meat/fish		Pulses, Beverages	Nuts/seeds, Fruits, Dairy		Pulses, Nuts/seeds, Fruits, Dairy, Oils/fats, Beverages, Meals out		Pulses, Nuts/seeds, Fruits, Dairy, Beverages
6. HH diary, frequent	Beverages, Meals out			Tubers, Nuts/seeds, Fruits	Fruits		Fruits	Pulses		Dairy
7. HH diary, infrequent	Fruits, Meals out	Dairy		Tubers, Pulses, Meat/fish, Oils/fats	Pulses, Nuts/seeds, Fruits		Meals out		Sugars	

Note: Interactions estimated with a regression of log consumption value on indicators for module type, a household characteristic, and an interaction term between module type and the household characteristic, as well as community fixed effects. Only food groups with interaction coefficients estimated at $p < .10$ are reported.

Table 7. Proportional mean deviation from benchmark consumption incidence or value, by frequency of purchase, own production, and storability/perishability

Frequency of purchase	1. Long list, 14 day recall	2. Long list, 7 day recall	5. Long list, usual month	6. HH diary, frequent	7. HH diary, infrequent
<i>Proportional mean deviation from benchmark consumption incidence</i>					
Infrequently purchased	-0.052	-0.200	1.391	-0.160	-0.166
Frequently purchased	-0.049	-0.127	0.463	-0.113	-0.100
<i>Proportional mean deviation from benchmark consumption value</i>					
Infrequently purchased	0.443	0.481	-0.624	-0.059	0.096
Frequently purchased	-0.134	0.206	-0.469	-0.087	-0.020
<i>Consumption out of own production</i>					
	1. Long list, 14 day recall	2. Long list, 7 day recall	5. Long list, usual month	6. HH diary, frequent	7. HH diary, infrequent
<i>Proportional mean deviation from benchmark consumption incidence</i>					
High share from own production	-0.072	-0.168	0.735	-0.121	-0.115
Seldom from own production	-0.024	-0.120	0.646	-0.127	-0.118
<i>Proportional mean deviation from benchmark consumption value</i>					
High share from own production	0.082	0.163	-0.525	-0.098	0.009
Seldom from own production	-0.072	0.400	-0.488	-0.060	0.008
<i>Storability</i>					
	1. Long list, 14 day recall	2. Long list, 7 day recall	5. Long list, usual month	6. HH diary, frequent	7. HH diary, infrequent
<i>Proportional mean deviation from benchmark consumption incidence</i>					
Storable foods	-0.030	-0.104	0.556	-0.119	-0.109
Perishable foods	-0.066	-0.180	0.811	-0.128	-0.123
<i>Proportional mean deviation from benchmark consumption value</i>					
Storable foods	0.084	0.229	-0.468	-0.073	0.005
Perishable foods	-0.057	0.313	-0.542	-0.086	0.011

Note: Proportional mean deviations reported for the 5 modules that record item consumption at sufficient levels of disaggregation - Module 3 (shortened list) and Module 4 (collapsed list) excluded. Reported numbers are the mean proportional deviations for all items in category relative to that reported in Module 8 (benchmark individual diaries). Infrequently purchased items are defined as those purchased less than the median purchase incidence in the benchmark module. This median purchase incidence is 1.39 times in a two week period. Common consumption out of own production is defined as at least 10% of total value of consumption derived from own production (there are 29 such food items). There are 26 food items classified as storable and 28 as perishable. All means weighted by consumption incidence as measured in benchmark module.

Figure 1. Proportional mean deviation in individual item consumption incidence relative to benchmark consumption incidence, by survey module, with fitted regression line

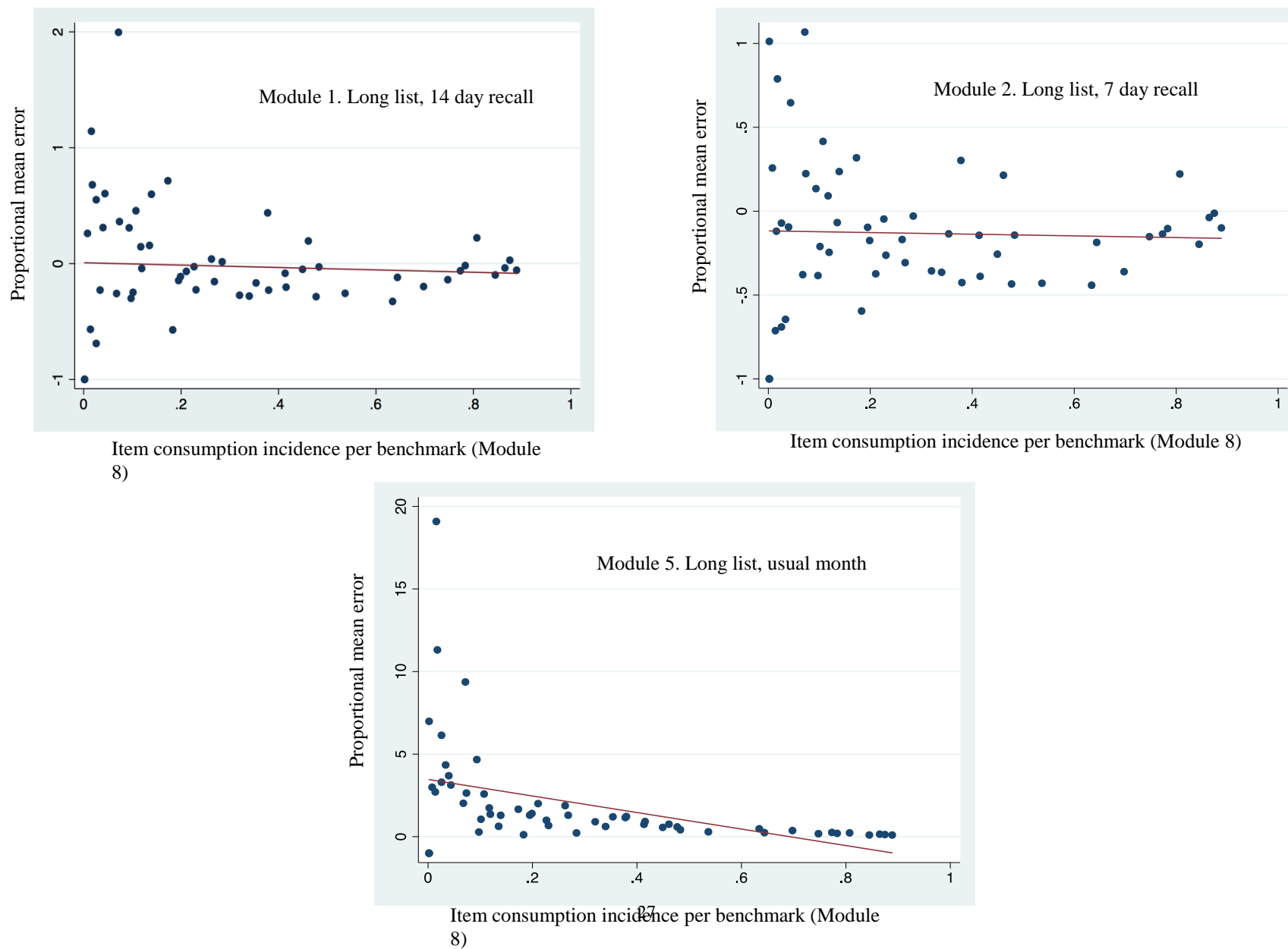


Figure 1 (cont.). Proportional mean deviation in individual item consumption incidence relative to benchmark consumption incidence, by survey module, with fitted regression line

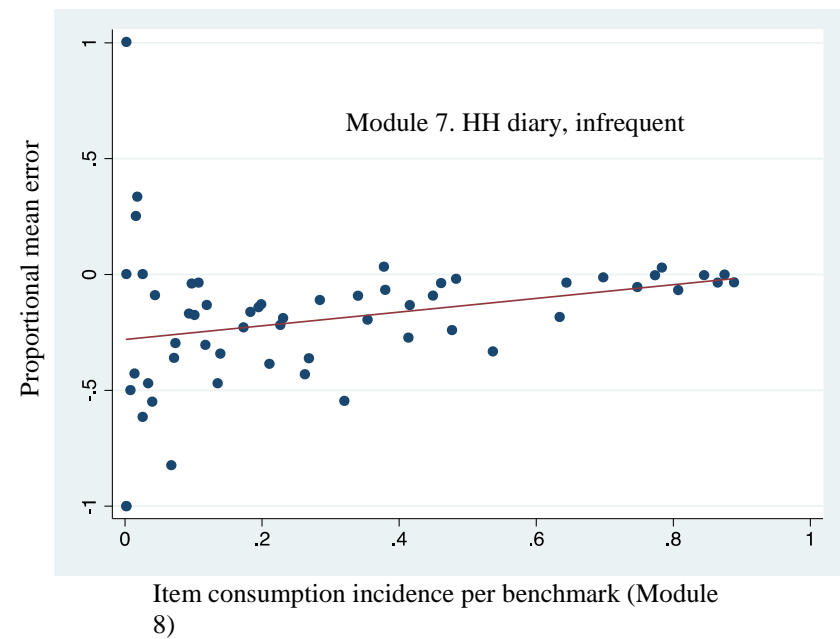
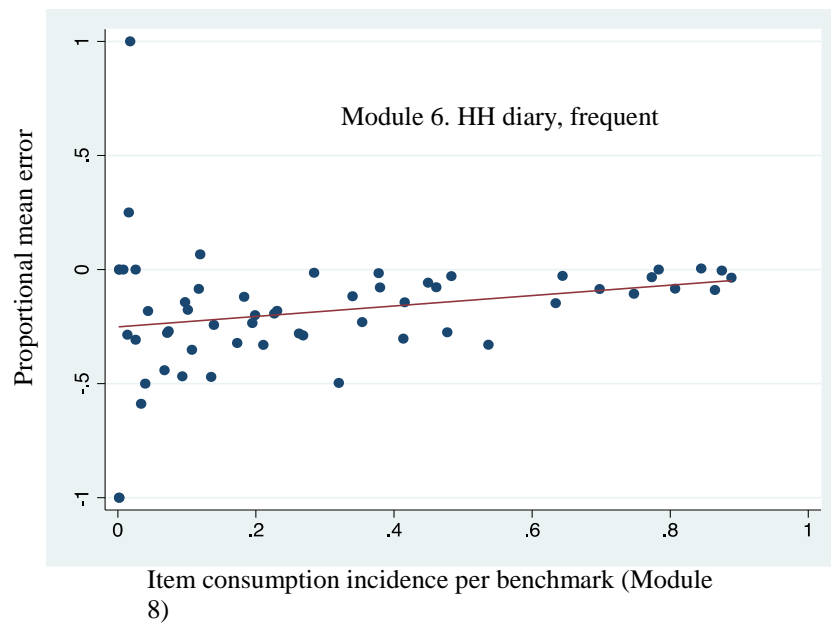


Figure 2. Proportional mean deviation in individual item consumption value relative to benchmark consumption value, by survey module, with fitted regression line

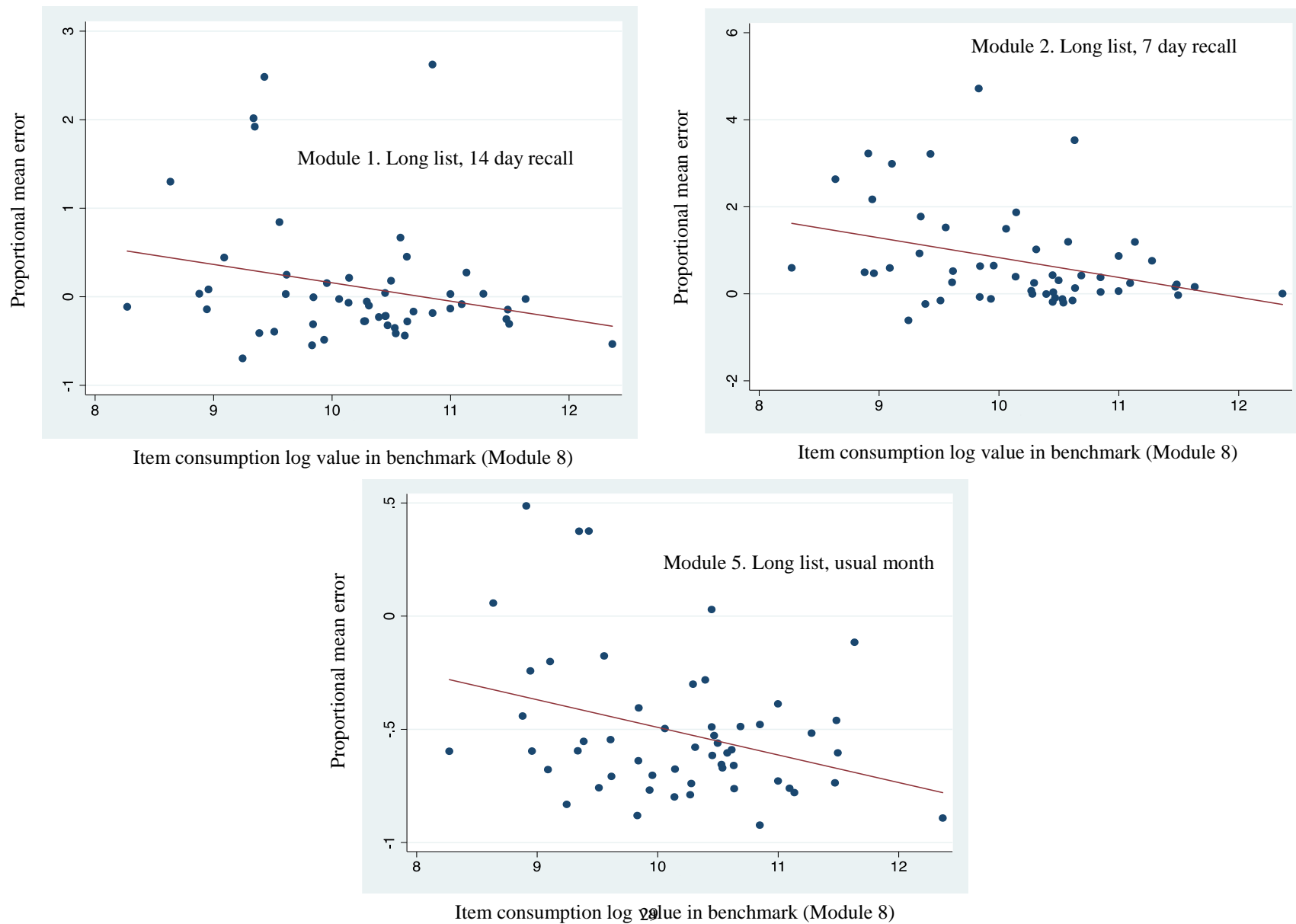
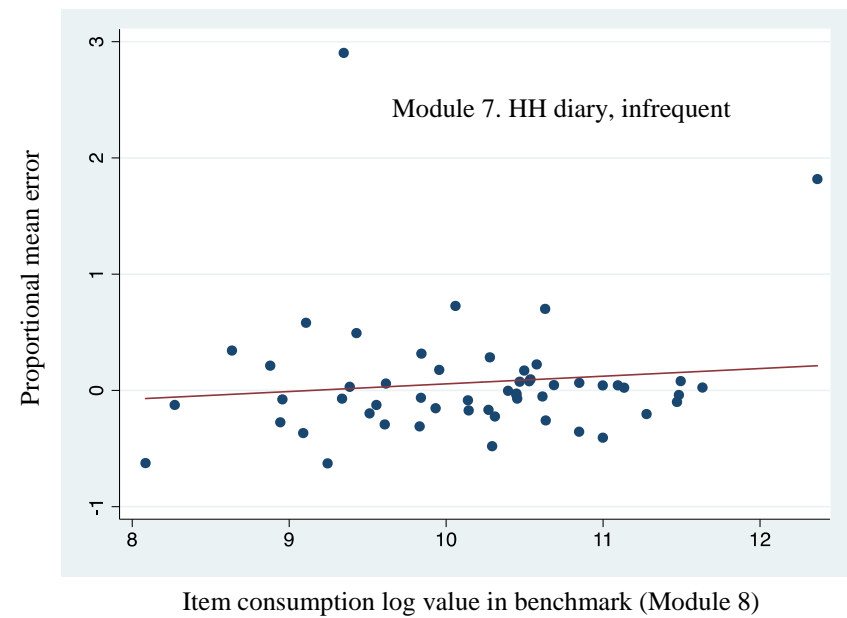
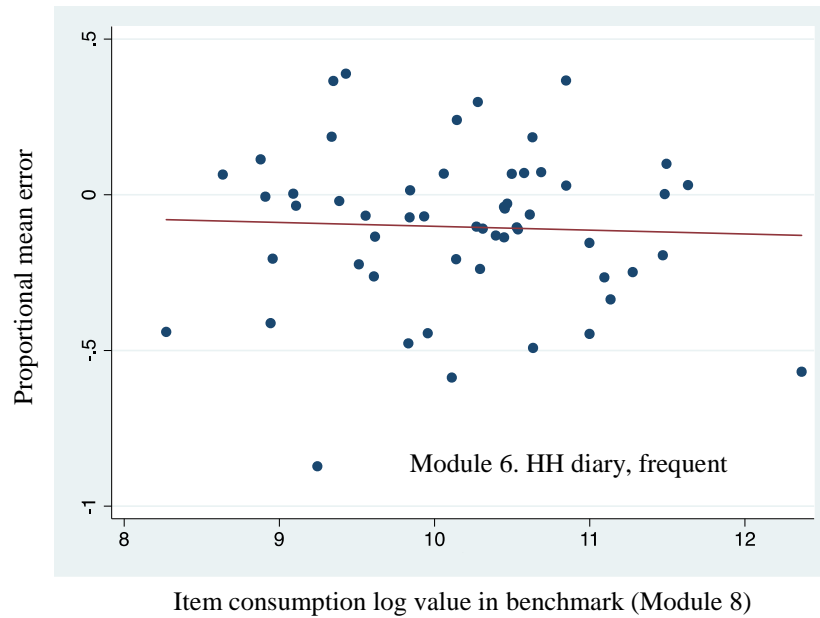


Figure 2 (cont.). Proportional mean deviation in individual item consumption value relative to benchmark consumption value, by survey module, with fitted regression line



Appendix Table 1. List of individual food items and subgroups used by the various survey designs

Food group	Item	Item included in module 3 (subset list)
Cereals and cereal products	Rice (paddy)	
	Rice (husked)	✓
	Maize (green, cob)	
	Maize (grain)	✓
	Maize (flour)	✓
	Millet and sorghum (grain)	
	Millet and sorghum (flour)	✓
	Wheat, barley grain, and other cereals	
	Cost of grinding	
	Bread	
	Buns, cakes, and biscuits	
	Macaroni, spaghetti	
	Other cereal products	
Tubers and plantains	Cassava fresh	✓
	Cassava dry/flour	✓
	Sweet potatoes	✓
	Yams/ cocoyams	
	Irish potatoes	
	Cooking bananas, plantains	✓
	Other starches	
Sugar and sweets	Sugar	✓
	Honey, syrups, jams, marmalade, jellies, canned fruits	
	Sweets	
Pulses, dry	Peas, beans, lentils, and other pulses	✓
Nuts and seeds	Groundnuts in shell/shelled	
	Coconuts (mature/immature)	
	Cashews, almonds, and other nuts	
	Seeds and products from nuts/seeds (excl. cooking oil)	
Vegetables	Onions, tomatoes, carrots, and green pepper, other vlungo	✓
	Spinach, cabbage, and other green vegetables	✓
	Canned, dried, and wild vegetables	
Fruits	Ripe bananas	
	Citrus fruits (oranges, lemons, tangerines, etc.)	
	Mangoes, avocados, and other fruits	
	Sugarcane	
Meat, meat products, and fish	Goat meat	
	Beef including minced, sausage	✓
	Pork including sausages and bacon	
	Chicken and other poultry	
	Wild birds and edible insects	
	Other domestic/wild meat products	
	Eggs	
Milk and milk products	Fresh fish and seafood (including dagaa)	✓
	Dried/salted/canned fish and seafood (incl. dagaa)	
	Fresh milk	✓
	Milk products (like cream, cheese, yoghurt etc., but excl. fats)	
Oil, fats, spices	Canned milk/milk powder	
	Cooking oil	✓
	Butter, margarine, ghee, and other fat products	
	Salt	
Beverages	Other spices	
	Tea (leaves or prepared)	
	Coffee and cocoa (beans, ground, instant, prepared)	
	Other raw materials for drinks	
	Bottled/canned soft drinks (soda, juice, water)	
	Bottled beer	
	Local brews	
Meals outside the household	Wine and spirits	
	Full meals (breakfast, lunch, dinner)	
	Barbecued meat, chips, roast bananas, and other snacks	
	Kibuku and other local brews	
	Wine, commercial beers, and spirits	
	Sodas and other non-alcoholic drinks	
	Sweets and ice-cream	
	Snacks	

Note: Table lists all food groups and individual food items recorded by the five recall survey designs. Module 3 (subset list) only records food items indicated in column three. Module 4 (collapsed list) records the first 11 food groups (all but meals outside the household) as 11 distinct aggregates. For all 5 survey designs, meals outside the household recorded in identical manner as seven day recall for each of the seven items in that food group.

Appendix Table 2. Interaction effects of consumption incidence with household characteristics, by module type and food group

Module-characteristic interaction	Cereals	Tubers	Sugars	Pulses	Nuts/seeds	Vegetables	Fruits	Meat/fish	Dairy	Oils/fats	Beverages	Meals out
<i>Number of adults, 15 years or older (sample mean = 2.78 people)</i>												
1. Long list, 14 day recall	0.012	-0.021	-0.011	0.003	-0.014	-0.004	-0.003	0.007	-0.045**	-0.007	-0.024	0.011
2. Long list, 7 day recall	0.006	-0.023*	-0.021	-0.044**	-0.002	-0.006	-0.029*	-0.005	0.005	-0.013	-0.019	-0.031*
3. Subset list, 7 day recall	-0.004	-0.021	0.012	-0.009	--	0.006	--	0.017	-0.016	0.000	--	-0.025
4. Collapsed list, 7 day recall	0.019*	-0.011	-0.010	-0.012	-0.005	0.012	-0.019	0.010	-0.004	0.001	0.000	-0.022
5. Long list, usual month	-0.008	-0.000	-0.003	-0.020	-0.014	-0.011	-0.018	-0.006	-0.052***	-0.012	-0.007	-0.031*
6. HH diary, frequent	0.007	-0.011	0.025	-0.031*	-0.018	0.007	-0.013	0.001	-0.000	-0.004	0.005	-0.029
7. HH diary, infrequent	-0.009	-0.003	0.005	-0.021	-0.008	-0.006	-0.020	-0.001	-0.015	-0.005	-0.026*	-0.018
<i>Number of children, 14 years or younger (sample mean = 2.50 people)</i>												
1. Long list, 14 day recall	0.004	-0.014	-0.032***	-0.002	-0.033***	-0.002	-0.001	-0.001	-0.031**	-0.009	-0.013	0.018
2. Long list, 7 day recall	0.019**	-0.006	-0.026**	-0.010	0.001	0.004	0.011	0.015	-0.017	-0.012	-0.012	-0.006
3. Subset list, 7 day recall	-0.001	-0.012	-0.018	-0.017	--	-0.003	--	0.008	-0.030**	-0.016*	--	-0.020
4. Collapsed list, 7 day recall	-0.003	0.003	-0.040***	-0.003	-0.000	0.010	-0.020*	-0.023**	-0.001	-0.001	0.005	-0.010
5. Long list, usual month	-0.006	0.005	-0.008	0.000	0.014	-0.009	0.015	-0.007	-0.043***	-0.020**	0.014	-0.009
6. HH diary, frequent	0.003	-0.009	-0.007	0.009	-0.022*	0.004	0.001	-0.008	-0.001	-0.007	0.015	-0.005
7. HH diary, infrequent	-0.008	0.008	0.004	0.001	0.000	0.003	0.004	-0.004	-0.005	-0.009	-0.005	-0.004
<i>Urban (proportion in sample = 0.34)</i>												
1. Long list, 14 day recall	0.020	-0.027	0.172***	-0.063	-0.028	-0.008	-0.070	0.015	0.051	-0.043	0.016	-0.089*
2. Long list, 7 day recall	0.077**	0.002	0.217***	-0.013	0.108**	0.004	-0.009	0.023	0.088	-0.028	0.012	0.042
3. Subset list, 7 day recall	0.064**	-0.089**	0.294***	0.010	--	-0.010	--	0.056	0.044	0.211***	--	-0.069
4. Collapsed list, 7 day recall	0.078**	-0.118***	0.288***	0.035	-0.261***	0.078***	0.028	0.011	0.008	0.213***	-0.083*	0.018
5. Long list, usual month	-0.024	0.003	-0.025	-0.217***	-0.189***	0.014	-0.135***	-0.022	0.108*	0.003	-0.146***	-0.085
6. HH diary, frequent	0.042	-0.065	0.032	-0.038	0.011	-0.021	-0.079*	-0.020	-0.090	-0.009	-0.049	-0.086
7. HH diary, infrequent	0.018	-0.059	0.033	-0.115**	-0.083*	-0.006	-0.137***	-0.009	0.038	0.011	-0.020	-0.052
<i>Education of household head (sample mean = 4.73 years)</i>												
1. Long list, 14 day recall	0.001	0.010**	0.013**	0.012*	-0.002	-0.000	0.000	0.010**	0.008	-0.010**	0.001	0.015**
2. Long list, 7 day recall	0.004	0.015***	0.023***	0.019***	0.013**	-0.000	0.006	0.011**	0.010	-0.008*	0.004	0.013*
3. Subset list, 7 day recall	0.002	0.013**	0.026***	0.012*	--	-0.006*	--	0.015***	0.001	0.023***	--	0.015**
4. Collapsed list, 7 day recall	0.001	0.011**	0.016**	0.013**	-0.002	-0.005	0.006	0.014***	-0.000	0.013***	0.007	0.017***
5. Long list, usual month	-0.004	-0.002	-0.010	-0.013*	-0.015**	-0.001	-0.025***	-0.006	0.015**	-0.009**	-0.021***	0.014**
6. HH diary, frequent	-0.001	-0.001	0.005	-0.012*	-0.002	-0.002	-0.013**	-0.004	0.009	-0.007	0.001	0.010
7. HH diary, infrequent	0.001	-0.000	0.002	-0.005	0.001	-0.001	-0.005	-0.001	0.006	-0.006	0.010*	0.018***
<i>Household wealth index (normalized population mean = 0.001)</i>												
1. Long list, 14 day recall	0.007	0.013	0.088***	0.038	0.015	-0.004	0.013	0.019	0.028	-0.025	0.007	0.012
2. Long list, 7 day recall	0.027*	0.047**	0.133***	0.071***	0.076***	0.002	0.048**	0.038**	0.024	-0.020	0.009	0.030
3. Subset list, 7 day recall	0.025*	0.035*	0.165***	0.098***	--	-0.001	--	0.047**	0.015	0.137***	--	0.005
4. Collapsed list, 7 day recall	0.026*	0.010	0.145***	0.072***	-0.072***	0.003	0.063***	0.055***	-0.009	0.088***	-0.010	0.026
5. Long list, usual month	-0.007	0.018	-0.026	-0.049*	-0.061***	0.011	-0.068***	-0.003	0.046*	-0.005	-0.113***	0.005
6. HH diary, frequent	0.003	-0.006	0.036	-0.004	0.026	-0.004	-0.026	0.007	-0.045*	-0.008	0.017	0.016
7. HH diary, infrequent	0.007	-0.007	0.052**	-0.020	-0.025	-0.003	0.001	0.000	0.000	0.001	0.010	0.038

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated with a linear probability model (OLS) that regresses a binary indicator for positive consumption on indicators for module type, a household characteristic, and an interaction term between module type and the household characteristic, as well as community fixed effects. Only coefficients on the interaction term are reported.

Appendix Table 3. Interaction effects of consumption value (conditional on positive consumption) with household characteristics, by module type and food group

Module-characteristic interaction	Cereals	Tubers	Sugars	Pulses	Nuts/seeds	Vegetables	Fruits	Meat/fish	Dairy	Oils/fats	Beverages	Meals out
<i>Number of adults, 15 years or older (sample mean = 2.78 people)</i>												
1. Long list, 14 day recall	-0.060	-0.023	0.031	-0.064	0.008	-0.054	-0.108*	-0.045	-0.053	-0.020	-0.082	-0.221***
2. Long list, 7 day recall	0.012	-0.009	-0.024	0.039	0.040	-0.074**	-0.087	-0.046	0.020	-0.007	-0.002	-0.188***
3. Subset list, 7 day recall	0.045	0.019	-0.073	0.038	--	-0.061	--	-0.044	-0.043	-0.007	--	-0.196***
4. Collapsed list, 7 day recall	-0.021	0.020	-0.079	-0.021	0.032	-0.045	-0.129**	-0.104**	-0.028	-0.059	-0.050	-0.196***
5. Long list, usual month	-0.047	0.003	-0.053	0.013	0.066	-0.102***	-0.102*	-0.047	0.070	0.020	-0.099	-0.153**
6. HH diary, frequent	0.037	0.003	0.047	0.035	0.029	-0.014	-0.096	-0.019	0.052	-0.022	-0.130*	-0.215***
7. HH diary, infrequent	-0.008	0.051	-0.022	0.051	0.079	-0.010	-0.096*	0.025	0.146*	0.021	-0.001	-0.130**
<i>Number of children, 14 years or younger (sample mean = 2.50 people)</i>												
1. Long list, 14 day recall	0.001	-0.032	-0.042	0.018	0.077*	-0.053**	-0.015	-0.015	-0.018	-0.010	-0.064	-0.102**
2. Long list, 7 day recall	-0.004	-0.034	-0.029	0.029	0.077	-0.088***	-0.085*	-0.058	-0.065	-0.016	-0.025	-0.160***
3. Subset list, 7 day recall	0.021	0.005	0.002	0.056	--	-0.089***	--	-0.114***	-0.113	-0.082***	--	-0.099**
4. Collapsed list, 7 day recall	0.013	0.002	-0.043	0.008	0.068	-0.040	-0.049	-0.078**	0.011	-0.037	-0.053	-0.024
5. Long list, usual month	-0.015	0.006	-0.148***	0.008	0.050	-0.080***	-0.080**	-0.054*	0.015	-0.019	-0.075	-0.071
6. HH diary, frequent	-0.020	0.067**	0.048	0.053	0.119***	0.037	0.074*	0.023	0.099	0.036	0.048	-0.049
7. HH diary, infrequent	0.015	0.055*	0.033	0.066**	0.061	-0.002	0.033	0.067**	0.091	0.058**	-0.015	-0.024
<i>Urban (proportion in sample = 0.34)</i>												
1. Long list, 14 day recall	0.039	0.469	0.111***	-0.244	-0.332	-0.002	0.181	0.046	0.279	0.037	-0.292	0.138*
2. Long list, 7 day recall	0.118**	0.492	0.004***	-0.311	-0.247**	0.180	0.014	0.018	0.437	0.201	-0.434	-0.043
3. Subset list, 7 day recall	-0.055**	0.274**	-0.182***	-0.253	--	0.177	--	0.266	0.414	-0.085***	--	-0.031
4. Collapsed list, 7 day recall	-0.229**	0.493***	-0.242***	-0.286	-0.403***	-0.341***	0.140	0.048	0.127	0.043***	0.216*	-0.148
5. Long list, usual month	0.100	0.409	0.708	-0.038***	0.248***	0.419	0.277***	0.271	0.524*	0.196	-0.015***	0.110
6. HH diary, frequent	0.130	-0.055	0.191	0.048	-0.373	-0.089	-0.138*	-0.139	-0.007	-0.087	-0.613	-0.300
7. HH diary, infrequent	-0.178	-0.260	-0.162	-0.198**	-0.491*	-0.127	-0.132***	-0.167	0.036	-0.135	-0.562	-0.308
<i>Education of household head (sample mean = 4.73 years)</i>												
1. Long list, 14 day recall	-0.012	0.024**	0.025**	-0.032*	0.005	-0.017	-0.036	-0.016**	-0.029	0.000**	-0.036	0.019**
2. Long list, 7 day recall	0.008	0.015***	0.038***	-0.010***	0.008**	0.004	-0.042	0.002**	0.017	0.013*	-0.054	-0.002*
3. Subset list, 7 day recall	-0.027	0.011**	0.011***	-0.022*	--	0.008*	--	-0.013***	-0.009	-0.023***	--	-0.002**
4. Collapsed list, 7 day recall	-0.000	0.039**	0.002**	-0.010**	0.013	-0.057	-0.032	-0.030***	-0.023	-0.030***	-0.022	-0.004***
5. Long list, usual month	0.020	0.059	0.104	0.048*	0.066**	0.034	0.018***	0.007	0.122**	0.007**	0.047***	0.047**
6. HH diary, frequent	-0.007	0.022	0.054	0.014*	-0.022	-0.006	-0.036**	-0.011	0.026	0.005	-0.028	-0.014
7. HH diary, infrequent	0.009	0.019	-0.003	0.002	-0.041	-0.001	-0.008	-0.006	0.002	0.005	0.006*	-0.015***
<i>Household wealth index (normalized population mean = 0.001)</i>												
1. Long list, 14 day recall	-0.010	0.282	0.052***	0.041	0.038	0.016	-0.010	-0.004	0.007	0.031	-0.235	0.083
2. Long list, 7 day recall	0.011*	0.240**	0.021***	-0.010***	0.076***	0.090	-0.016**	0.045**	0.195	0.084	-0.236	0.040
3. Subset list, 7 day recall	-0.071*	0.151*	-0.092***	-0.030***	--	0.066	--	0.110**	0.173	-0.076***	--	-0.066
4. Collapsed list, 7 day recall	-0.064*	0.342	-0.126***	-0.005***	-0.075***	-0.270	0.017***	-0.016***	0.141	-0.027***	-0.122	-0.029
5. Long list, usual month	0.062	0.343	0.430	0.143*	0.283***	0.176	0.209***	0.120	0.524*	0.106	0.095***	0.055
6. HH diary, frequent	0.020	0.005	0.108	0.084	-0.064	-0.095	-0.153	-0.086	0.070*	-0.082	-0.216	-0.151
7. HH diary, infrequent	-0.060	0.009	-0.100**	-0.013	-0.099	-0.088	-0.103	-0.113	0.045	-0.094	-0.165	-0.168

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated with a linear probability model (OLS) that regresses a the log value of consumption on indicators for module type, a household characteristic, and an interaction term between module type and the household characteristic, as well as community fixed effects. Only coefficients on the interaction term are reported.