Decomposing Response Errors in Food Consumption Measurement: Implications for Survey Design from a Survey Experiment in Tanzania

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

KU LEUVEN

Decomposing Response Errors in Food Consumption Measurement: Implications for Survey Design from a Survey Experiment in Tanzania

Jed Friedman,[a] Kathleen Beegle,[a] Joachim De Weerdt,[b] and 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 that are differentially influenced by contrasting types of survey response error. Although 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. The 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 the lessons for improved consumption survey design. In light of these findings, the relative performance of common survey designs is discussed, and design lessons are drawn to enhance the accuracy of item-specific consumption reporting and, consequently, the measures of total household food consumption.

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 surveys (Fiedler et al. 2008). While levels of hunger and nutrition covary with household resource availability, the role of resources relative to other driving forces is debated (Deaton 1997). The evidence cited in this debate has been influenced by the reliability of measures of food consumption and economic resources (Bouis, Haddad, and Kennedy 1992; Gibson and Kim 2013).

This paper focuses on the measurement of food consumption. It leverages a recent survey experiment to study the performance of commonly used consumption survey modules to shed light on the nature of reporting errors in consumption data. The experiment involved randomly allocating one of eight consumption survey modules to a nationally representative sample of Tanzanian households. An individual diary supervised on a daily basis has been taken as the benchmark, or gold standard, survey approach. This approach was adopted because of the scope of the resources and the care teams devoted to the survey (see below). The accuracy of the other modules is assessed with respect to this benchmark. Previous work associated with the same experiment has explored the relative performance of the eight modules in terms of mean consumption, inequality, poverty, and the prevalence of hunger (Beegle et al. 2012; De Weerdt et al. 2016; Gibson et al. 2015). These studies concentrate on total household-level consumption aggregates and do not consider variations in performance among individual items, as is done here. Moreover, variations in mean consumption by module, which represents up to 27 percent of the total value in these studies, convey the net effect of all possible types of reporting error, including the opposing impacts of recall and telescoping errors, as well as the difficulty of fully capturing individual consumption opportunities outside the home.

This paper extends previous findings through a more careful focus on the nature of survey reporting errors (relative to the benchmark). We accomplish this by decomposing the sum of reported consumption into a product of two vectors: (1) a vector of binary indicators recording whether the household reports any positive value consumed for each food subgroup or individual food item captured by the survey and (2) a real value vector of the subgroup 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. Furthermore, it 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 surveys. The third section describes the Tanzania survey experiment. The fourth section presents the analytic methods we employ, and the fifth discusses the results. The final section summarizes the findings and discusses the consequent implications for improved survey design.

II. Consumption measurement errors: a brief taxonomy

The degree and nature of measurement error in consumption captured by household surveys depend partly on survey design features.[1] These vary along a large number of dimensions, such as the length of the recall period or the level of item-specific detail sought (Fiedler, Carletto, and Dupriez 2012; Smith et al. 2014). Moreover, because these features affect the estimates of household consumption, comparisons across countries, as well as within countries over time, are compromised when questionnaires change.[2]

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

- *Recall error*. A main concern is that respondents might forget the occurrence of a consumption event. This could result in recall error. Lower salience and longer recall periods make forgetfulness more likely among respondents (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 (Grosh et al. 1995; Scott and Amenuvegbe 1991).
- *Telescoping*. The converse of recall error is telescoping whereby a household compresses consumption that occurred over a longer period of time into 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). Particularly for longer recall periods that typically involve more transactions, respondents may cease trying to enumerate each and instead use rules of thumb to estimate them (Arthi et al. 2016; Blair and Burton 1987; de Nicola and Giné 2014; Gibson and Kim 2007). In this case, rule of thumb error depends on transaction frequency and regularity; less frequent items are likely reported with more error. Whereas recall error biases the consumption estimate downward, and telescoping creates upward bias, there is no obvious direction of bias in responses that resort to the rule of thumb instead of enumeration. We may expect this error to be especially pertinent in hypothetical consumption constructs such as questions about consumption during a usual month. Usual month consumption is an explicit attempt to abstract away from seasonal considerations in consumption; however, this type of question may pose additional cognitive demands relative to a definitive recall period in the immediate past.
- *Personal leave out error*. Yet another source of reporting error is the inability to capture individual consumption by household members accurately 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 the activities of these members outside the home (World Bank 2006).
- *Other error types*. While the analysis in this study focuses 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 last is a respondent who understates her consumption to appear poorer because of a belief that these responses may determine the eligibility for some future social program. There may also be intentional misreporting because of respondent fatigue. So, whether

---

[1] A consumption survey is a household survey that collects detailed consumption data. It has a range of labels, such as household budget survey, living standards survey, or household income, consumption, and expenditure survey.

[2] See Beegle et al. (2016) for an extensive discussion of this issue in Sub-Saharan Africa.

the respondent is presented with a long or a short list of consumption items can influence the quality of the responses.[3]

- *Diary versus recall surveys*. 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 because 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. The extent to which diaries are supervised to ensure they are regularly filled is thus a key design feature. Unsupervised diaries may effectively become self-administered recall modules with endogenous recall periods if some types of respondents do not fill them in every day and, hence, render them subject to varying degrees of recall, telescoping, and rule of thumb reporting. Diaries administered among individuals should also prove better at capturing individual consumption outside the household (i.e. reduced personal leave out error), 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 based on different methods of data capture (diary versus recall questionnaires), levels of respondent (individual versus household), recall periods, or degree of commodity detail may not be comparable. We have designed the survey experiment used here in part to assess the extent to which variations 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 errors.

III. The Tanzania survey experiment

The Tanzania survey experiment, conducted to shed light on the implications of survey design variations in food consumption measurement, systematically contrasts various design features. We strategically selected eight survey designs to reflect the most common methods utilized in low-income countries and that are typical of the scope of variation one is likely to find in consumption surveys. We then randomly assigned these eight designs to over 4,000 total households. Given the sample size and the random assignment of survey designs, differences in mean measurement performance may be attributed with a high degree of confidence to the survey design rather than potential confounders.

The designs differ by method of data capture (diary or recall survey), designated respondent (household head or other household member), length of reference period, number of items in the recall list, and nature of the cognitive task required of the respondents. Table 1 summarizes each of these designs. The modules we number 1–5 are recall designs, and modules 6–8 are diaries. For the food recall modules, households report the value of items consumed from three sources: purchases, home production, and gifts or payments. Modules 1 and 2 contain a list of 58 food items. Module 3 is associated with a subset list that consists of the 17 most important food items, which 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 scale up reported expenditures for that module (by 1/0.77). Module 4 is associated with a list of 11 food items. It is an aggregated version of the list of 58 food items whereby, for example, several

---

[3] Beegle et al. (2012) find a drop from 49 to 41 minutes in interview times if the food list is cut from 58 to 17 items in a one-week recall. Times for a 58-item list rise to 76 minutes if the typical, more cognitively demanding "usual" month recall is used.

listed vegetables are aggregated into one item, vegetables. The specific 58 individual food items in modules 1 and 2, those that are in the subset in module 3, and the aggregation for module 4 are shown in appendix table 1. The appendix table also lists seven items of a 12fth food group, meals outside the home. Although this food-outside-the-home group is collected in an identical manner across all recall modules (as a detailed 7-day recall), we include it in the decomposition analysis because it is a food category that grows in importance as national incomes rise.

Among the recall modules, module 5 deviates from the reporting of actual consumption over a specified period. Instead, it asks for usual consumption following a recommendation in Deaton and Grosh (2000) whereby households report the number of months in which the food item is usually consumed and the average monthly value of what is consumed during 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 relative to the other recall modules: a longer time frame and a distinct 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, whereby each adult member keeps their own diary, and the consumption of children is captured in the diaries of the adults who know most about the daily activities of the children. 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 the 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 a not an individual's record of food consumption. Rather, it records the food acquired for the household by each member even if the food is for the consumption of several members (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 in most surveys. The investments were made 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 inflows and outflows that may be otherwise missed. Module 8 arguably provides the most accurate estimate of total household food consumption.

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

and Shinyanga and two districts in the Kagera Region.[4] The districts were purposively selected to capture variations in socioeconomic characteristics. In each district, 24 communities were randomly chosen from the 2002 census based on probability-proportional-to-size criteria. Within communities, a random subvillage (enumeration area) was chosen, and all households therein were listed. Per subvillage, 24 households 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 because of refusals. Three households that started a diary were dropped because they did not complete their final interview. Another five households were dropped because of missing data on some of the key household characteristics, yielding a final sample size of 4,029 households.[5]

The basic characteristics of the sampled households generally match those from the nationally representative national Household Budget Survey 2007. 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.[6]

In regard to reporting error, there are several points to note about the survey experiment. The recall modules 1–5 ask the respondent about consumption, but not food acquisition. The 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 (for example, scraps or leftovers), 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 of these meals were not captured in the diaries. This is important because the 2012 State of Food Insecurity report incorporated, for the first time, tentative estimates of food losses, which led to a significant revision of some of the world hunger numbers (FAO, WFP, and IFAD 2012). 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 the food consumed, which eliminates the counting of seeds and animal feed as consumption, but may not eliminate food scraps and leftovers 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. Because nonstandard 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 nonstandard unit combinations. Median district-level metric conversion rates were used to convert nonmetric units into kilograms or liters. If 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 implementation, including the relative costs to field each module, are described in Beegle et al. (2012).

---

[4] The survey teams were small, extensively trained on all modules, and well supervised. They stayed in the field for the entire 12-month study period to ensure that well-trained survey teams consistently applied the modules across all districts and also to abstract away from seasonal concerns that might have interacted with specific survey designs.
[5] There is almost no item nonresponse in the consumption section of the recall modules, that is, all respondents answered virtually all questions on all consumption items, including a response of no, or zero, consumption.
[6] This analysis is presented in Beegle et al. (2012).

Table 2 presents the summary results of the consumption survey experiment. It reports the difference in the log per capita consumption measure of each design relative to the benchmark individual diary.[7] The estimates in table 2 derive from regressions of the natural logarithm of food, nonfood, and total consumption on binary indicators for module assignment (whereby the benchmark personal diary is the left out category). Because the survey experiment was randomized, the regressions include no covariate controls except for the 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 modules record between 8 percent and 33 percent less food consumption compared with the personal diary (column 3). The impact on total consumption is at a similar magnitude (column 2). In the diary approach to food consumption, the use of only one respondent to complete the diary for an entire household is associated with significantly lower food consumption, by 13–20 percent, most likely because some share of unobservable personal consumption of the other household members is omitted (not captured) by the respondent maintaining the diary. Differences in frequent nonfood consumption are also observed, especially in the diaries, again suggesting the importance of accurately recording personal consumption.[8]

Regarding the recall survey approach, all mean food expenditures are lower than the benchmark. The mean of the 7-day long list lies nearest to the benchmark value, while modules with longer recall periods (14 days or the usual month) or more aggregated consumption categories (the collapsed list) record food consumption that is 17 percent 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 extrapolate definitively that the 7-day long list will be the most accurate of the recall designs if it is applied in different settings. Because 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 the possible effect of salient and easily observed household characteristics—those assumed to determine actual consumption levels—on the accuracy of consumption reporting. The characteristics investigated include the following: (1) household size: it was determined that recall modules underreport consumption even more as the size of the household increases; (2) urban

---

[7] While the experiment focused on food consumption measurement, each survey also recorded nonfood 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 (whereby households assigned to diaries were administered a nonfood consumption survey at the end of a two-week study period). For more frequently purchased nonfood items such as soap or transport, the consumption was either asked in recall form in the recall modules 1–5 (in which the period of recall corresponded to that for food) or recorded as diary entries for households assigned a diary.

[8] Because the questionnaire wording and structure for the nonfrequent nonfood consumption section were identical across the eight modules, it is perhaps surprising to see significantly negative coefficients for modules 1, 4, and 7 relative to the benchmark. Such differences can result from three sources: respondent fatigue as the recalled items in these modules come after the lengthy food recall sections in modules 1–5 or after a two-week diary; cognitive framing; and variations in the ability to capture personal nonfrequent nonfood 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 (by 14 percent) nonfrequent nonfood consumption. Possibly the lack of follow-up during the diary period made the module 7 respondents less diligent in the nonfrequent nonfood section of the final interview.

location: household diaries significantly underreport consumption in urban areas (but not rural areas) suggesting the relative prevalence of personal consumption opportunities in urban areas; (3) the educational attainment of the household head: education had little relation to module performance except in the usual month approach, wherein inaccuracy was greater among less well educated households; and (4) household wealth as captured by a household asset index: the underreporting in recall modules is greatest among the poorest households and the 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 various types of reporting error. This possibility is investigated in the error decomposition framework introduced in the next section.[9]

IV. Reporting error decomposition

Earlier analyses of consumption reporting errors has focused on a net measure of total misreporting. This masks two aspects of consumption reporting: whether any consumption occurred and, if it did occur, the value of the consumption. Our main analytic approach in this paper is to examine these two aspects of misreporting in comparison with the benchmark module by modeling total food consumption as a product of two vectors whereby each ordered element of 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:

$$C_{hm} = \overrightarrow{I(C_{fhm} > 0)} * \overrightarrow{(C_{fhm}|C_{fhm} > 0)} \qquad (1)$$

where the first vector in the product is the consumption incidence vector, and the second vector is the consumption value.[10]

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 can obscure these error patterns and, consequently, inhibit the lessons for improvement in consumption survey design. Furthermore, different research questions may not be as equally concerned about the errors in each of these vectors. For example, food diversity indicators are often based solely on incidence, rather than value or quantity.

---

[9] Another important consideration is the effect of the characteristics of the enumerator on the interview quality and response error. Unfortunately, the 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 are much more uniform across the enumerators than across the general population; all the enumerators had completed secondary school, but none had yet entered university, were between 20 and 30 years of age, and were 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 prefieldwork cognitive testing of enumerators to supplement the inquiries of this nature.

[10] If the two vectors are to have the same dimension and thus allow total consumption to equate to the inner-product, the consumption value vector needs to include the zero consumption values. Therefore, the depiction of the vector as consumption values conditional on positive consumption is purely stylistic to highlight the decomposition analysis to follow.

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 the following:

$$I(C_{fhm}) = \beta_{m=8} + \beta_{fm}M_m + \varepsilon_{fhm} \qquad (2)$$

where M is a vector of indicators for module type. The individual diary, $m = 8$, is the excluded category, and the constant $\beta_{m=8}$ therefore represents the mean benchmark incidence. Regressions include survey cluster fixed effects and are estimated with ordinary least squares.[11]

Earlier work has demonstrated that household characteristics interact with survey design in nontrivial ways to produce error and, so, may also interact in differential ways with respect to consumption incidence and value. An understanding of the presence of these interaction affects can also inform consumption survey design. An extended regression framework thus includes a household characteristic X – for example the number of adult or child household members, household location, the educational attainment of the household head, or asset wealth – and interacts this characteristic with the survey module indicator, M:

$$I(C_{fhm}) = \beta_{m=8} + \beta_{fm}M_m + \beta_{fx}X_h + \beta_{fmx}M_mX_h + \varepsilon_{fhm} \qquad (3)$$

In this specification, the coefficient of interest is β_{fmx}, which relates how module effects on incidence reporting 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.

The next section first explores consumption incidence and then consumption value. 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 be less effective with other food types. Consequently, the analysis compares the design error as estimated above with respect to item-specific features such as consumption incidence (i.e. common or rare items), consumption value, the share of consumption from home production, the frequency of market purchase of a food item, and the storability or perishability of the good. This analysis unveils some of the mechanisms underlying the misreporting and enhances the relevance of our results beyond 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 illustrates the consumption incidence for 12 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 among most food groups. For example, while 67 percent of benchmark households

---

[11] The results are appreciably similar if binary response models (probit or logit) are used in place of ordinary least squares.

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 percent. 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) underreport the consumption of 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 percent–54 percent. While the consumption incidence of the 7-day short list may be expected to be lower than the 7-day long list because the former module design asks about a fewer number of individual food items, there is no prior expectation that the 7-day collapsed list will record lower consumption incidence. The fact that the collapsed list does record a lower incidence for all food items suggests that important consumption items are excluded because of a lack of cognitive prodding that the longer list explicitly incorporates in the design.

By contrast with the other recall modules, the usual month approach to recall survey design reports significantly higher consumption incidence among almost all food groups, with the sole exception of Cereals, which are consumed by 96 percent of the benchmark households. This difference most likely derives from the different cognitive demand of considering a usual month, which apparently prompts respondents to report significantly higher consumption incidence relative to the actual consumption recorded in the benchmark. Finally, the two household diary modules (modules 6 and 7) tend to report lower consumption incidence among various food groups such as Fruits (9 percent lower) and Meals Outside the Home (7 percent–9 percent lower). While the frequency of the household diary supervision does not appear to influence the accuracy of consumption incidence measurement because the rates are equal for the weekly and thrice-weekly supervised diaries, the two household diaries systematically record lower incidence relative 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 arises because of personal leave out error, whereby the household respondent likely misses some individual consumption. However, because the magnitudes of the incidence shortfall are relatively high for all recall modules (except the usual month) and occur even in the case of nearly universally consumed items such as cereals, this indicates that a key channel for recall error is complete forgetting (or deliberate suppression).[12] By contrast, the usual month approach prompts households to report a far higher monthly incidence of consumption than the benchmark, suggesting a different pattern of reporting error in this module. Given that the hypothetical nature of the

---

[12] That the 7-day recall tends to report lower consumption incidence than the 14-day recall may, in principle, arise because of the less diversified actual consumption in a one-week period relative to a two-week period. However, the 7-day recall reports relatively lower incidence than the 14-day recall for those selected food groups in a nonlinear fashion (going against expectation if the lower incidence derives solely from less frequent consumption across weeks). The conclusion that the lower reported incidence in the 7-day recall is largely driven by recall error of greater magnitude (relative to the 14-day recall) is supported by a comparison of the 7-day recall module with the consumption incidence recorded in the first week of the personal diaries. The shortfall in incidence is largely the same relative to the first week of the personal diaries as with both weeks. It is impossible to conduct a similar analysis for the usual month as the personal diary was only collected for a two-week period. Nonetheless, 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 arises because of response error.

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 in the two household diaries with respect to the personal diary points to the importance of personal leave out error because this is the main driver of divergence between modules 6 and 7 and the benchmark.

Key household characteristics may exacerbate (or moderate) the module-specific reporting error in consumption incidence. This can be explored by interacting the module indicator with the select characteristics mentioned earlier Table 4 summarizes these results by reporting the food groups on which significant interaction effects have been estimated.[13] The effect of household characteristics on the incidence reported not only depends on the characteristics, but also on the particular food subgroup. For recall modules in general, the tendency to underreport incidence is mediated by urban location, the education of the household head, 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.) Thus, the underreporting of any consumption in these food groups is greatest among rural, less well educated, and low-wealth households. For some food groups, the number of household members, especially the number of children, also tends to exacerbate underreporting. This implies that more disadvantaged households, those that are rural, have less education, have more children, and have fewer assets are more likely to omit consumption during the survey experience. This will 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 in these modules is fairly constant across all households. Exceptions to this include the consumption incidence of Fruit, Pulses, and Nuts/Seeds recorded among 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 in rural areas.

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 shillings.[14] Table 5 summarizes the module design effects of 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, for which the quantities reported are not different than the benchmark. Because these goods are typically more perishable than other types and, consequently, purchased more frequently, perhaps the tendency to overreport 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 in the

---

[13] The specific interaction terms are presented in appendix table 2. The main effects estimated in equation (3) are suppressed for ease of exposition.

[14] The monetary value results can also be interpreted as the effect on reporting quantities (kilograms or liters).

case of Cereals, Dairy, and Meals Out do the 14-day recall values exceed the benchmark values; for all others, they are lower, often significantly so.

The 7-day subset list (module 3) tends to report greater positive value errors than the 7-day long list, which must derive from the module's focus on only the most commonly consumed items because that is the only design feature that distinguishes the two modules. In contrast, the collapsed list exhibits both overreporting and underreporting.

It is not clear what causes these value error patterns. Overreporting could occur if the value of salient episodes of consumption is telescoped into 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. Overreporting could then also occur if larger (and therefore more salient) episodes are used as the rule of thumb. The fact that reporting errors on 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: (1) 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 dominates in the shorter period; (2) alternatively, if rule of thumb reporting is utilized for both periods, rule of thumb tends to overreport to a greater degree during a shorter recall period.

A different reporting pattern is evident for the usual month. While this module recorded a higher consumption incidence on most food groups (significantly higher than the benchmark in all but two cases), the values recorded in this module design are significantly lower, with the sole exception of Cereals, for which 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, consistently result in overestimation of consumption incidence on most foods and in an underestimation of value.

The two household diaries exhibit a distinct pattern of value reporting error. Because the expected main driver of reporting error between the household diary and the personal diary would be the inability to capture personal consumption outside the home fully, the values reported should diverge most for the relevant food goods. This is indeed what we find. The household diaries record 36 percent – 39 percent lower value in Fruit consumption, 37 percent–42 percent lower value in Beverages, and 35 percent – 36 percent lower value in 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 relative to the frequently supervised diary 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 covaries with particular household characteristics by listing the precisely estimated interaction terms as given by equation (3).[15] The characteristics of households do mediate the degree of error in the value reported, although the patterns are less clear than in the case of consumption incidence. One fairly clear result concerns the number of adults

---

[15] Appendix table 3 relays the magnitude of all estimated interaction terms.

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 recall and for 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 items, and this undervaluation increases for households with many children, indicating once more the importance of greater attention if recall surveys are administered to larger households.

Urban households administered recall modules also tend to underreport select food groups; this is especially so for the subset and collapsed list recall surveys (modules 3 and 4). However, such households also exhibit overreporting for other food groups. There are some other specific results for various food groups and either the educational attainment or wealth level of the household, but no systematic pattern emerges, making survey design lessons difficult to generalize with respect to value reported and household measures of economic status. This is in contrast to the results shown in table 3, which contains more clear patterns in consumption incidence error and selected household characteristics, whereby more disadvantaged households are more likely to omit the consumption of key food groups entirely.

Item characteristics and reporting error

This section explores the relation between commodity characteristics and reporting error. Investigated commodity characteristics include (1) either commonly or uncommonly consumed items in terms of the consumption incidence as reported in the benchmark individual diary; (2) the monthly value of the consumed item; (3) the frequency of purchase/consumption; (4) the share of the item consumed from home production; and (5) the storability or perishability of the item.

Because 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.[16] This analysis is not possible for the 7-day collapsed list module (module 4) because that module 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 consequently is also removed from subsequent analysis.

For the remaining five 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 at less than 0.2), shows no directed relation with the commonality of the consumed item. The fitted regression line with a slope 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; the least commonly consumed items in the benchmark report a high positive incidence error. Clearly, the usual month exercise leads to a gross overstatement of

---

[16] The long list contains 58 food times. Because of possible confusion in the diary entries among (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.

consumption frequency for these rare items, which include, for example, Macaroni and Spaghetti or Pork. In contrast, the more frequently consumed items are associated with a much lower incidence error (albeit still positive). Finally, for the diaries, the pattern of error is reversed. The less frequently consumed items are reported at even less frequency in the household diaries, most likely reflecting 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.

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

Now that we have established that incidence reports are more accurate for high-incidence items (in terms of less spread in the reporting error), we 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 the value reporting. Figure 2 plots value error as a function of benchmark consumption value. The first thing to note is that, in general, there is no reduction in variance for higher value items, contrary to the case of the incidence reporting of commonly consumed items. Recall that table 5 reveals a mix of over- or underreporting for module 1 (14-day long list), depending on the food group. The relevant scatterplot in figure 2 suggests that the net overreporting error is greater for consumption items of relatively low value. As the benchmark value increases, the net error approaches zero and then turns negative at high-value items, signifying that the items largest in value suffer from underreporting. The same directional pattern is apparent for module 2 (7-day long list), on which each food group in table 5 displays net overreporting. Figure 2 suggests that particularly high relative errors are associated with foods with a consumption value below the median.

The usual month shows the same error gradient as the other two recall modules, although, here, the net error for most food items is negative (as is true of the food groups listed in table 5). The magnitude of this error rises 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 in the value reporting in this module, or rule of thumb value reporting on the usual month induces negative errors that grow 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 of home production, and whether the food is storable or perishable. Table 7 lists the mean proportional errors relative to the benchmark for these three characteristics for 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

13

only 54 observations informing the two means. For this reason, standard errors are not shown. Nonetheless, table 7 shows suggestive results that relate food characteristics and the degree of reporting error.

The frequency of purchase of a particular food appears to convey a salience reflected in the accuracy of the reporting. 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 in all five modules investigated, but especially in the usual month, where the error magnitude is less than half as much as the overreporting 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 purchased goods to improve the accuracy in recall surveys 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.

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

In terms of storability, we define 26 individual food items as storable, e.g. dry goods such as maize flour, and 28 as perishable, e.g. 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: thus, 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, that is, the overreporting in the 7-day module is greater for perishables, as is the magnitude of underreporting for the usual month and the diaries. The 14-day recall reveals directionally different errors for the two types of goods: overreporting for storable goods and a slight underreporting error for perishables – this is the one module where the absolute value error is greater for storable goods.

VI. Concluding lessons for food consumption survey design

This paper relies on data from a consumption experiment in Tanzania to describe the nature of reporting errors in food consumption surveys. The goal is not merely to document error patterns, but to generate hypotheses for 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 and in which diets are based on a particular staple crop (maize, by caloric intake). More work is needed to establish the generalizability of any findings reported here. However, we hope that our focus on the underlying mechanisms of misreporting error will increase the external validity of our findings.

We use a simple analytic framework of error decomposition that 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 the estimate of mean household food consumption (see table 2), the incidence of consumption is significantly lower than the benchmark for most food groups. The same finding holds 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 in 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 (e.g. Cereals) or not (e.g. Dairy). The relative equality in mean consumption of the 7-day recall and the benchmark individual diary may therefore derive from the happenstance of offsetting errors: negative errors in incidence multiplied by positive errors in value. This approximately equal offsetting 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 overreporting of value. We conclude that value overreporting (because of 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 to which we return below.

This simple decomposition exercise, which raises some stark patterns in differential response according to recall module, suggests that methods to improve the accuracy of recall surveys should consider a dual-track approach: (1) efforts related to the prompting of households to report any positive consumption because recall modules underreport the consumption incidence associated with almost any food group and (2) efforts aimed at improving the accuracy of consumption value reporting, conditional on the household reporting any consumption.

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

Regarding efforts to improve the accuracy of consumption value reports, note that overreporting of value is a particularly prominent error for short recall periods such as seven 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 for the respondent during the second 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 fieldwork structure and cost (in cases where mobile teams rather than resident enumerators are used).

We have extended the basic analytic framework with respect to how food reporting errors across modules interact with the key socioeconomic characteristics of households and with food characteristics such as perishability or the prevalence of home production. With respect to interactions with household characteristics, the bias in consumption incidence is most pronounced among disadvantaged households

such as those with low educational attainment, low household wealth, larger numbers of household members, and those located in rural areas. The difficulty of recall to accurately capture consumption is thus exacerbated in more disadvantaged households. The degree of error in value is also attenuated, albeit to a far less degree, by greater education, greater wealth, urban location, and small household size. Survey methods to improve accuracy may thus need to be particularly sensitive to poorer and less well educated households, perhaps with more closely considered and attentive prompting and an enhanced use of images and easily understandable quantity measures.

With respect to item characteristics, recall modules also tend to report lower consumption incidence relative to the benchmark, as well as higher magnitudes of consumption value error, for less frequently purchased items. A separate strategy may need to be employed for the relatively more rarely purchased items. Similarly, recall modules tend to contain incidence errors of greater absolute value for perishable items than for durables (although both types of goods are reported at lower frequency than in the benchmark). For perishables relative to durables, 7-day recall also exhibits greater value errors. For incidence and value reports, these findings suggest there is 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) find 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 in the most time-constrained surveys (where a savings of seven interview minutes has a real benefit in data quality) should these shorter modules be considered.

Apart from the recall modules rooted in the recent past (7 or 14 days), the usual month approach seeks to encourage respondents 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 different error pattern whereby consumption incidence is significantly higher than the benchmark. Whereas forgetting any consumption appears to be a problem in 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 approach, the consumption value appears to be significantly underreported for all food groups except Cereals. Compounding these problems, the error magnitudes in both incidence and value are substantially worse among disadvantaged households. Finally, this module took an average of 76 minutes to complete, far longer than the 49 minutes for the 7-day recall. The longer span to completion likely derives from the large cognitive burden placed on respondents by the usual month approach. Because 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 interpretation of the error patterns in the household diaries is somewhat more straightforward because the point of divergence from the benchmark is the reliance on one household member to record the consumption of all members, and, hence, the influence of personal leave out likely becomes more prominent (along with differences in the frequency of supervision in the less frequently supervised module). The household diaries do record 7 percent – 10 percent lower consumption incidence for food goods such as Fruits, Beverages, and Outside Meals, which are more likely to be consumed by individual members alone or outside the household. The reported values are also lower for many of the same commodities. For some food groups, the negative error in consumption value is smaller in magnitude in the infrequently supervised

household diary than in the frequently supervised one. In one example, the frequently supervised diary yielded a consumption value for Nuts and Seeds 17 percent less than the benchmark, while the corresponding yield of the infrequently supervised diary was only 7 percent less. Because the infrequently supervised diary can be transformed into a recall survey 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 total consumption value from the infrequent diaries closer to the benchmark.

Because household diaries are substantially less resource intensive than the benchmark personal diary, they will remain a commonly used tool. As suggested by these results, steps can be taken to improve accuracy. Thus, even if one household member records all consumption, prompting explicit consultation with all adult household members around the consumption of items commonly consumed alone or outside the home may serve to reduce personal leave out errors. The fieldwork of diary supervisors may also be structured so that more frequent supervision is possible (by optimizing work routes exhibiting a constraint on repeat visits). Where this is not possible, bounding notices such as stickers showing the dates of the past and future visits can call attention to the relevant time periods for the consumption records.

Our analysis focuses on the influence of errors generated by forgetting, telescoping, the use of rules of thumb, and personal leave out errors and proposes changes in survey design that may reduce this influence. We conclude with an additional brief discussion of small design changes unlikely to impact survey costs that may reduce the influence of other types of response error not investigated here. Two examples are given below. We refer interested readers to a more comprehensive treatment of this issue in Smith, Dupriez, and Troubat (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, in 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 if they are 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, to aggregate food consumption these local units must be converted into standard units. If common foods are more likely to be reported in nonstandard units, that is, pieces of cassava and bunches of bananas, and conversion factors are inadequate, 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.

References

Arthi, Vellore, Kathleen Beegle, Joachim De Weerdt, and Amparo Palacios-Lopez. 2016. "Measuring Household Labor on Tanzanian Farms." Unpublished working paper, World Bank, Washington, DC.

Beegle, Kathleen, Luc Christiaensen, Andrew Dabalen, and Isis Gaddis. 2016. *Poverty in a Rising Africa*. Africa Poverty Report. Washington, DC: World Bank.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson. 2012. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics* 98 (1): 3–18.

Blair, Edward, and Scot Burton. 1987. "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions." *Journal of Consumer Research* 14 (2): 280–88.

Bouis, Howarth E., Lawrence J. Haddad, and Eileen Kennedy. 1992. "Does It Matter How We Survey Demand for Food? Evidence from Kenya and the Philippines." *Food Policy* 17 (5): 349–60.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Washington, DC: World Bank; Baltimore: Johns Hopkins University Press.

Deaton, Angus, and Margaret E. Grosh. 2000. "Consumption." In *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*, vol. 1, edited by Margaret E. Grosh and Paul Glewwe, 91–134. Washington, DC: World Bank.

de Nicola, Francesca, and Xavier Giné. 2014. "How accurate are recall data? Evidence from coastal India." *Journal of Development Economics* 106 (1): 52–65.

De Weerdt, Joachim, Kathleen Beegle, Jed Friedman, and John Gibson. 2016. "The Challenge of Measuring Hunger." *Economic Development and Cultural Change* (forthcoming).

FAO (Food and Agriculture Organization of the United Nations), WFP (World Food Programme), and IFAD (International Fund for Agricultural Development). 2012. "The State of Food Insecurity in the World 2012: Economic Growth Is Necessary but Not Sufficient to Accelerate Reduction of Hunger and Malnutrition." FAO, Rome.

Fiedler, John L., Calogero Carletto, and Olivier Dupriez. 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 Supplement): S242–S251.

Fiedler, John L., Marc-Francois Smitz, Olivier Dupriez, and Jed Friedman. 2008. "Household Income and Expenditure Surveys: A Tool for Accelerating the Development of Evidence-Based Fortification Programs." *Food and Nutrition Bulletin* 29 (4): 306–19.

Gibson, John, Kathleen Beegle, Joachim De Weerdt, and Jed Friedman. 2015. "What Does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?" *Oxford Bulletin of Economics and Statistics* 77 (3): 466–74.

Gibson, John, and Bonggeun Kim. 2007. "Measurement Error in Recall Surveys and the Relationship between Household Size and Food Demand." *American Journal of Agricultural Economics* 89 (2): 473–89.

———. 2013. "How Reliable Are Household Expenditures as a Proxy for Permanent Income? Implications for the Income–Nutrition Relationship." *Economics Letters* 118 (1): 23–25.

Grootaert, Christiaan. 1986. "The Use of Multiple Diaries in a Household Expenditure Survey in Hong Kong." *Journal of the American Statistical Association* 81 (396): 938–44.

Grosh, Margaret E., Qing-hua Zhao, and Henri-Pierre Jeancard. 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 6, World Bank, Washington, DC.

Halpern-Manners, Andrew, and John Robert Warren. 2012. "Panel Conditioning in Longitudinal Studies: Evidence from Labor Force Items in the Current Population Survey." *Demography* 49 (4): 1499–1519.

Menon, Gaeta. 1993. "The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies." *Journal of Consumer Research* 20 (3): 431–40.

Scott, Christopher, and Ben Amenuvegbe. 1991. "Recall Loss and Recall Duration: An Experimental Study

in Ghana." *Inter-Stat* 4 (1): 31–55.

Smith Lisa C., Olivier Dupriez, and Nathalie Troubat. 2014. "Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys." IHSN Working Paper 008, International Household Survey Network, Food and Agriculture Organization of the United Nations, and World Bank, Washington, DC.

Sudman, Seymour, and Norman N. Bradburn. 1973. "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association* 68 (344): 805–15.

World Bank. 2006. *Reducing Poverty through Growth and Social Policy Reform in Russia*. Report 35519. Directions in Development Series. Washington, DC: World Bank.

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***	-0.167***	-0.104	-0.105*
	-0.037	-0.037	-0.067	-0.06
2. Long list, 7 day recall	-0.039	-0.017	-0.134**	-0.096
	-0.037	-0.037	-0.067	-0.06
3. Subset list, 7 day recall	-0.071*	-0.079**	-0.112*	-0.09
	-0.037	-0.037	-0.067	-0.06
4. Collapsed list, 7 day recall	-0.283***	-0.332***	-0.104	-0.138**
	-0.037	-0.037	-0.067	-0.06
5. Long list, usual month	-0.207***	-0.268***	0.023	-0.013
	-0.037	-0.037	-0.067	-0.06
6. HH diary, frequent	-0.173***	-0.196***	-0.279***	-0.046
	-0.037	-0.037	-0.067	-0.06
7. HH diary, infrequent	-0.136***	-0.129***	-0.244***	-0.105*
	-0.037	-0.037	-0.067	-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***	-0.088***	-0.141***	-0.055**	-0.067***	-0.006
	(0.015)	(0.019)	(0.024)	(0.025)	(0.024)	(0.012)
2. Long list, 7 day recall	-0.067***	-0.084***	-0.158***	-0.111***	-0.150***	-0.008
	(0.015)	(0.019)	(0.024)	(0.025)	(0.024)	(0.012)
3. Subset list, 7 day recall	-0.093***	-0.134***	-0.232***	-0.166***	--	-0.022*
	(0.015)	(0.019)	(0.024)	(0.025)	--	(0.012)
4. Collapsed list, 7 day recall	-0.091***	-0.150***	-0.228***	-0.136***	-0.414***	-0.085***
	(0.015)	(0.019)	(0.024)	(0.025)	(0.024)	(0.012)
5. Long list, usual month	0.022	0.110***	0.109***	0.201***	0.217***	0.020*
	(0.015)	(0.019)	(0.024)	(0.025)	(0.024)	(0.012)
6. HH diary, frequent	-0.028*	-0.031*	-0.095***	-0.026	-0.070***	-0.010
	(0.015)	(0.019)	(0.024)	(0.025)	(0.024)	(0.012)
7. HH diary, infrequent	-0.006	-0.002	-0.074***	-0.000	-0.097***	-0.008
	(0.015)	(0.019)	(0.024)	(0.025)	(0.024)	(0.012)
Constant (omitted category mean)	0.956***	0.672***	0.436***	0.828***	0.739***	0.890***
	(0.049)	(0.063)	(0.080)	(0.084)	(0.079)	(0.040)

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

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated with a linear probability model (OLS) that regresses a binary indicatory 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, Fruits, 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 indicatory 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	-0.208***	-0.026	-0.026	-0.001	-0.182***
	(0.052)	(0.061)	(0.076)	(0.070)	(0.088)	(0.054)
2. Long list, 7 day recall	0.244***	0.013	0.244***	0.233***	0.291***	-0.081
	(0.052)	(0.061)	(0.076)	(0.071)	(0.092)	(0.054)
3. Subset list, 7 day recall	0.193***	0.136**	0.511***	0.574***	--	0.227***
	(0.053)	(0.062)	(0.079)	(0.073)	--	(0.054)
4. Collapsed list, 7 day recall	-0.119**	-0.163***	0.286***	0.210***	0.100	-0.368***
	(0.052)	(0.062)	(0.079)	(0.072)	(0.119)	(0.055)
5. Long list, usual month	0.024	-0.539***	-0.758***	-0.513***	-0.459***	-0.409***
	(0.051)	(0.057)	(0.070)	(0.065)	(0.081)	(0.053)
6. HH diary, frequent	-0.156***	-0.074	-0.083	-0.171**	-0.173*	-0.163***
	(0.052)	(0.059)	(0.074)	(0.069)	(0.089)	(0.054)
7. HH diary, infrequent	-0.060	-0.006	0.063	-0.009	-0.074	-0.043
	(0.051)	(0.059)	(0.074)	(0.068)	(0.090)	(0.054)
Constant (omitted category mean)	11.085***	9.357***	8.901***	9.380***	9.689***	9.611***
	(0.174)	(0.228)	(0.365)	(0.222)	(0.300)	(0.186)

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

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
Proportional mean deviation from benchmark consumption incidence					
Infrequently purchased	-0.052	-0.200	1.391	-0.160	-0.166
Frequently purchased	-0.049	-0.127	0.463	-0.113	-0.100
Proportional mean deviation from benchmark consumption value					
Infrequently purchased	0.443	0.481	-0.624	-0.059	0.096
Frequently purchased	-0.134	0.206	-0.469	-0.087	-0.020

Consumption out of own production

	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
Proportional mean deviation from benchmark consumption incidence					
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
Proportional mean deviation from benchmark consumption value					
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

Storability

	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
Proportional mean deviation from benchmark consumption incidence					
Storable foods	-0.030	-0.104	0.556	-0.119	-0.109
Perishable foods	-0.066	-0.180	0.811	-0.128	-0.123
Proportional mean deviation from benchmark consumption value					
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) exlcuded. 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
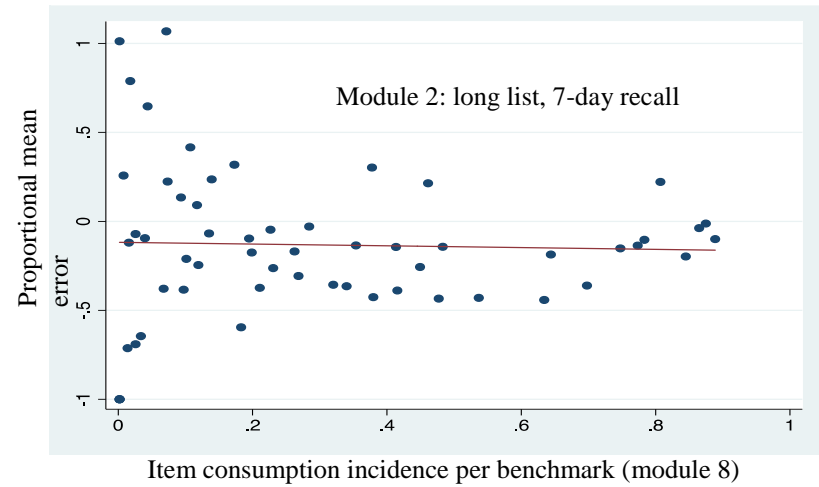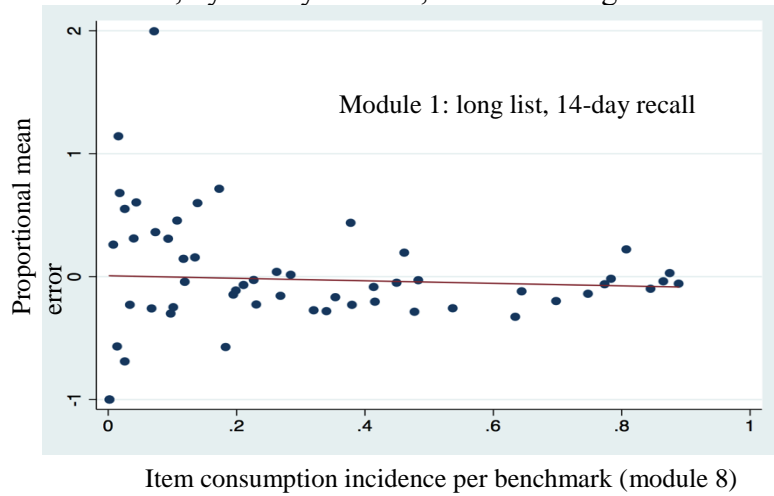
Figure 1 (cont.). Proportional mean deviation in individual item consumption incidence relative to benchmark consumption incidence, by survey module, with fitted regression line
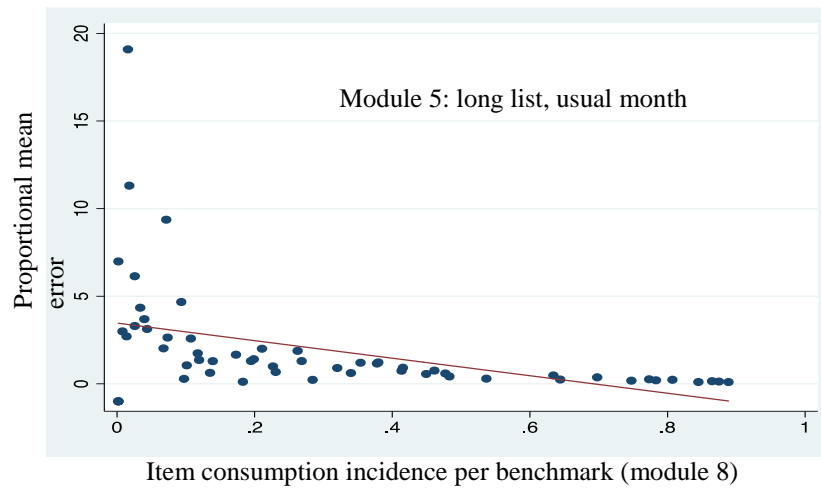
Figure 2. Proportional mean deviation in individual item consumption value relative to benchmark consumption value, by survey module, with fitted regression line

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

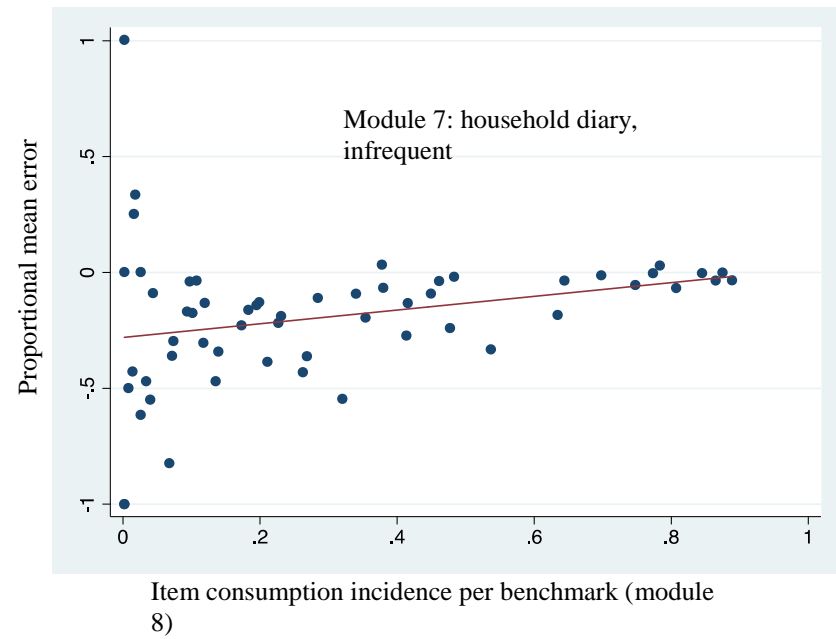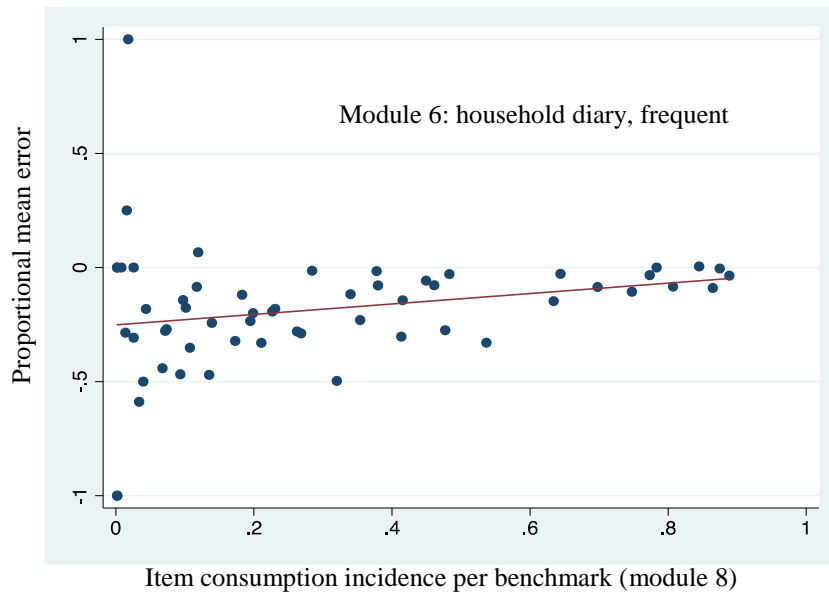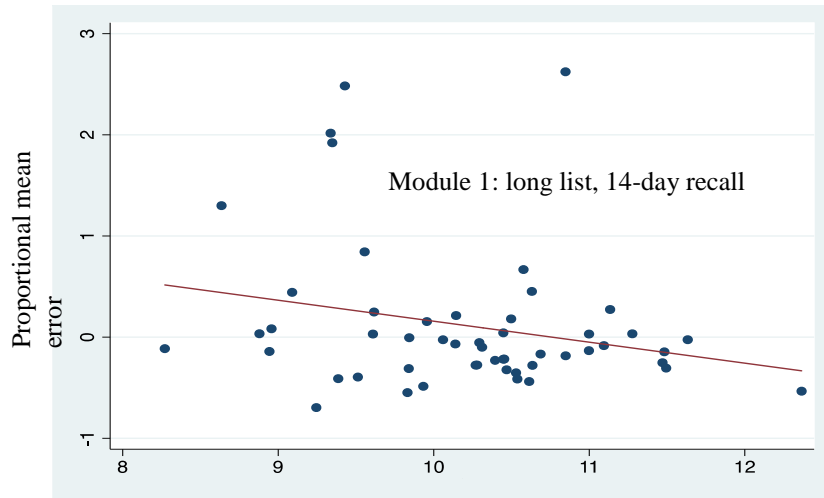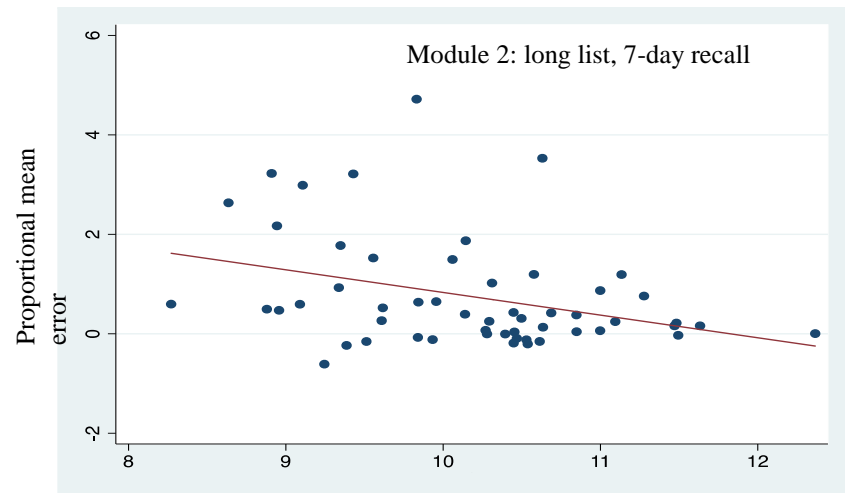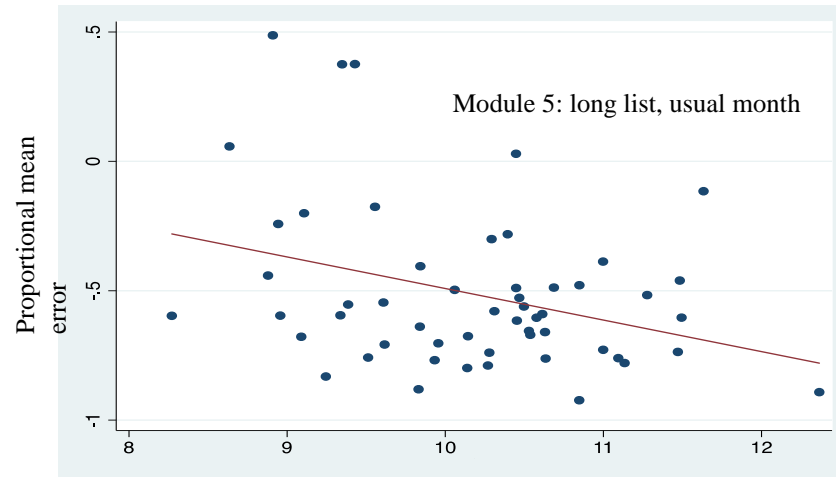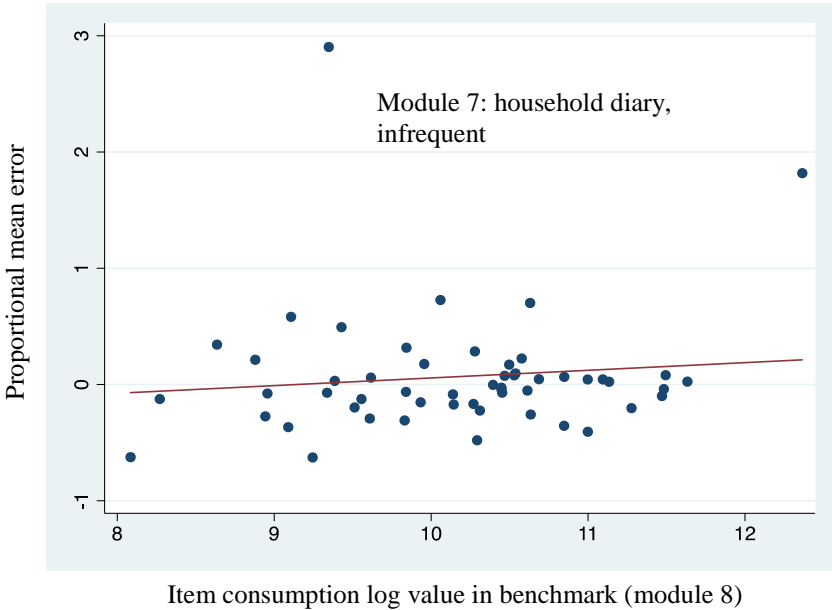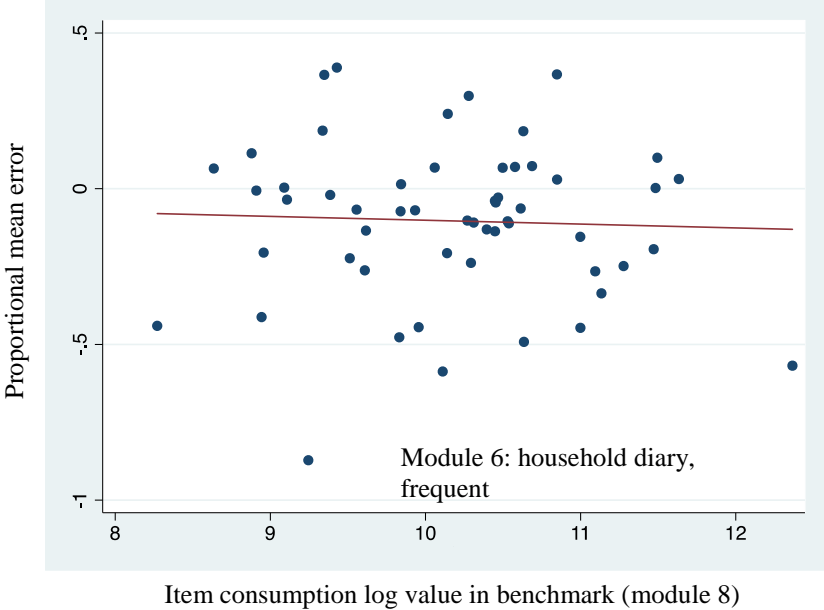**Appendix Table 1.** List of individual food items and subgroups used by the various survey designs

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, barlet 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, plaintains	✔
	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, avocadoes, 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	
	Fresh fish and seafood (including dagaa)	✔
	Dried/salted/canned fish and seafood (incl. dagaa)	
Milk and milk products	Fresh milk	✔
	Milk products (like cream, cheese, yoghurt etc., but excl. fats)	
	Canned milk/milk powder	
Oil, fats, spices	Cooking oil	✔
	Butter, margerine, ghee, and other fat products	
	Salt	
	Other spices	
Beverages	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	
	Wine and spirits	
Meals outside the household	Full meals (breakfast, lunch, dinner)	
	Barbecued meat, chips, roast bananas, and other snacks	
	Kibuku and other local brews	See note for details on this category
	Wine, commerical 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
Number of adults, 15 years or older (sample mean = 2.78 people)												
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
Number of children, 14 years or younger (sample mean = 2.50 people)												
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
Urban (proportion in sample = 0.34)												
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
Education of household head (sample mean = 4.73 years)												
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***
Household wealth index (normalized population mean = 0.001)												
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 indicatory 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
Number of adults, 15 years or older (sample mean = 2.78 people)												
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**
Number of children, 14 years or younger (sample mean = 2.50 people)												
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
Urban (proportion in sample = 0.34)												
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
Education of household head (sample mean = 4.73 years)												
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***
Household wealth index (normalized population mean = 0.001)												
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.