

# Improving Consumption Measurement and other Survey Data through CAPI: Evidence from a Randomized Experiment<sup>1</sup>

Accepted for Publication in Journal of Development Economics  
DOI: 10.1016/j.jdeveco.2011.12.001

December 2011

Bet Caeyers (University of Oxford)  
Neil Chalmers (Independent Consultant)  
Joachim De Weerd (EDI)

## ABSTRACT

*This paper reports on a randomized survey experiment among 1840 households, designed to compare pen-and-paper interviewing (PAPI) to computer-assisted personal interviewing (CAPI). We find that PAPI data contain a large number of errors, which can be avoided in CAPI. We show that error counts are not randomly distributed across the sample, but are correlated with household characteristics, potentially introducing sample bias in analysis if dubious observations need to be dropped. We demonstrate a tendency for the mean and spread of total measured consumption to be higher on paper compared to CAPI, translating into significantly lower measured poverty, higher measured inequality and higher income elasticity estimates. Investigating further the nature of PAPI's measurement error for consumption, we fail to reject the hypothesis that it is classical: it attenuates the coefficient on consumption when used as explanatory variable and we find no evidence of bias when consumption is used as dependent variable. Finally, CAPI and PAPI are compared in terms of interview length, costs and respondents' perceptions.*

---

<sup>1</sup> We gratefully acknowledge financial support from the World Bank's multi-year research agenda in survey methodology (LSMS Phase IV). We appreciate permission from the Millennium Challenge Corporation (MCC) to build on their existing survey in Pemba. We thank Kathleen Beegle, David McKenzie, Kinnon Scott and participants at the Conference on Survey Design and Measurement in Washington DC for feedback on the experiment's design and an earlier draft of this paper. The paper was substantially improved after incorporating suggestions made by the editors and two anonymous referees. A previous version circulated under the title "A Comparison of CAPI and PAPI through a Randomized Field Experiment". Leonard Kyaruzi, Deogratias Mitti and Mujobu Moyo lead the field teams, while Alessandro Romeo and Thaddaeus Rweyemamu took care of data entry of the paper questionnaires.

## 1. Introduction

Whilst the analysis of survey data has benefitted from the information technology revolution, most data collection in developing countries still uses traditional pen-and-paper interviewing (PAPI). In computer-assisted personal interviewing (CAPI) the interviewer reads questions from the screen of a handheld device, preloaded with the questionnaire, to the respondent. The respondent's answers are immediately entered into the device, which eliminates the need for manual re-keying of the data. The computer also automates the routing through the questionnaire and enables the interviewer to run a set of consistency checks *during* the interview, so that anomalies can be resolved with the respondent. These and numerous other features are believed to improve data quality, but it is unclear to what extent they actually do so and what effect this has on analysis. Furthermore, there is currently no empirical evidence from the developing world on how a switch from PAPI to CAPI would influence the length of the interview, respondents' perceptions, the cost of the survey, requirements on level of education of interviewers and so forth. This paper reports on a formal experiment, designed specifically to compare CAPI and PAPI along these and other lines.

The study was built on an existing LSMS-style CAPI survey of 1,200 households on the Island of Pemba in Tanzania. The experiment consisted of randomly sampling, within the same enumeration areas, 320 additional households to be interviewed using restricted CAPI (with disabled consistency checks) and 320 using PAPI. This design allows for a detailed comparison of errors, outliers, interview times, respondent's perceptions, interviewer effects and costs across the three methodologies. Special focus was given to improving the collection of consumption data, which utilises many of the powerful features of the computer, including complex validity checks and the ability to show pictures on the screen. The experiment lends itself to comparing simple poverty and inequality measures across the experiments.

While the first computer-assisted telephone interviews (CATI) were conducted by a US marketing firm in 1971, the first nation-wide CAPI survey occurred only in 1987 in the Netherlands (Nichols and de Leeuw, 1996). As CAPI became more popular for large-scale face-to face surveys in western countries, researchers became more aware of its impact on the survey process and outcomes. It was found that interviewers and respondents reacted favourably to the technology (Couper and Burt, 1994; de Leeuw and Nichols, 1996). Taylor (1998) shows that this remains true for respondents with, presumably, less exposure to modern technology, such as the elderly over 70 years of age. Banks and Lauri (2000) report that the attrition rate in the British Household Panel Survey was not affected when it switched from PAPI to CAPI in 1998. The literature also indicates the potential of CAPI to reduce routing and other errors (de Leeuw, 2008). There has been a number of CAPI surveys in the developing world, an enumeration of which is beyond the scope of this paper. Apart from the paper by Fafchamps et al (2010), however, we are not aware of any systematic attempt to study the effect on data quality and analysis.

The lack of evidence on how to reduce errors in surveys in developing countries stands in stark contrast to how much is known about the effects of measurement error in analysis

(Bound et al., 2001; Chesher and Schluter, 2002). Classical measurement error is defined by Bound et al. (2001) as an error in the measurement of a particular variable which is uncorrelated with the true value of that variable, the true values of other variables in the model, and any errors in measuring those variables. As we do not have independent, validation data in this experiment, we cannot directly measure the error to analyse its nature. We are, however, able to set up two testable hypotheses that should hold if measurement error is classical: in regression analysis, classical measurement error causes no bias when just the dependent variable has error, but attenuates the estimated coefficient on a single error-ridden explanatory variable. We fail to reject the hypothesis that the introduced measurement error is classical, at least for consumption measurements and based on these two tests. There is some consolation in this finding, as non-random, mean-reverting errors negatively correlated with true values bias regression coefficients even when just the dependent variable has error. When an explanatory variable has such error, its coefficient may be biased either toward or away from zero (Gibson and Kim, 2007). Moreover, the main correction for measurement error bias – instrumental variables (IV) – is inconsistent when errors are correlated with true values (Black, Berger and Scott, 2000).

The next section describes the design of the experiment and the differences we hypothesise to exist between CAPI and PAPI. Section 3 discusses results pertaining to errors and sample size reduction. It shows that CAPI significantly reduces the number of inconsistencies per survey. Some of these errors may require observations to be omitted from analysis, which could bias the sample because missing variables are not randomly distributed. Section 4 analyses the nature of measurement error in consumption aggregates. It first compares nutrition, consumption, poverty and inequality data across the three experiments. It then hypothesises that error is introduced through PAPI and sets up two testable predictions to verify whether this measurement error is classical. The first is that regression coefficients on consumption as an independent variable should be attenuated. The second is that there is no bias in a model where the error-ridden variable is used as a dependent variable. We find that, despite the fact that error counts are higher in certain types of households, we cannot reject that (after cleaning) the introduced measurement error is classical. Section 5 looks at other dimensions of comparison, such as cost, length of the interview and respondents' perceptions. Section 6 discusses some concluding observations.

## **2. Experimental set-up and hypothesised effects**

### *2.1. Set-up*

The experiment was run alongside an existing household survey on Pemba Island (which is part of Zanzibar, Tanzania). The main survey was conducted in July and August 2009 on behalf of MCA-T (Millennium Challenge Account Tanzania) as a baseline to evaluate their rural roads upgrade programme. In total 1,200 households were interviewed - 15 in each of the 80 Enumeration Areas (EAs). All households were administered a full CAPI questionnaire using an Ultra Mobile Personal Computer (UMPC), which is a handheld device with a 7" touch screen (a screen smaller than that of a laptop, but larger than that of a PDA). In a first experiment, we randomly selected 4 additional households per EA

(320 in total) who were interviewed with the same CAPI questionnaire, but with one important CAPI feature disabled: the system of consistency checks. The purpose of this experiment was to isolate the effect of consistency checks, which are believed to have important impact on data quality, especially in the consumption data. In the remainder of this paper we will refer to this application as ‘restricted CAPI’, in order to distinguish it from the unrestricted ‘full CAPI’ application which included the system of consistency checks. To investigate all other CAPI effects, as a bundle, a second experiment randomly selected another 4 households per EA to be interviewed using PAPI. The PAPI data were transferred to computer using two pass verification to minimize keystroke errors. Each of the four interviewers in a team conducted one restricted CAPI, one PAPI and three or four full CAPI interviews per cluster. For the restricted CAPI and PAPI, interviewers were allocated a specific household to interview at a specific time within the team’s two-day visit to the EA. This was done to ensure that questionnaires were not clustered per interviewer or in time.

All experimental questionnaires were conducted by the same 20 interviewers working on the main MCA survey. This increased the likelihood of contamination within the experiment, though it is hard to know the direction of the bias *a priori*. On the one hand, interviewers could learn about the kind of checks CAPI implements (something they may not have done had the questionnaire been purely on paper), but on the other hand interviewers could unlearn the practice of carefully verifying a questionnaire at the end of the day as they get used to the computer doing it for them. We tackle this contamination bias in two different ways. First, during training and fieldwork, interviewers were repeatedly instructed to check questionnaires at the end of the interview and again before submitting them to the supervisor. The supervisor, in turn, would check the questionnaires for errors. Questionnaires with errors that could not be resolved at base camp were returned to the interviewer, who was then required to revisit the household. Second, we have data to control for the number of months of experience that interviewers had using paper questionnaires and using electronic questionnaires.

The experimental questionnaire took, on average, 84 minutes to administer and included the following sections: Control data, GPS-coordinates, household head details, household member roster, demographics, education, health, amenities, assets, livestock, agriculture and consumption.<sup>2</sup>

A few days after the electronic questionnaire was conducted, a separate team of locally recruited interviewers returned to 4 households per experiment to ask 13 simple questions on the experience of the respondent in participating in the survey.

---

<sup>2</sup> The main questionnaire, implemented on behalf of MCA-Tanzania, included some additional sections on prices, transfers, shocks, credit, self-help groups and the like. To avoid these sections interfering with the experiment they were placed at the end of the main questionnaire. The full questionnaire is available from the authors upon request.

## 2.2. Experiment 1: the effect of validation checks

The full electronic questionnaire included a comprehensive system of internal validation checks.<sup>3</sup> The first experiment was set up to isolate the effects of these checks by comparing full CAPI to restricted CAPI. The checks are believed to lead to more accurate data capture, because they were run *during* the interview, at a time when they could still be resolved with the respondent. The check procedure does not run automatically, but is activated by the interviewer by manually clicking check-buttons. They are run at various stages during the interview, typically after completing all the questions on one screen. A final, global check can be run at the end of the interview. The checking procedure was repeated by the supervisor at the end of each survey day, and once more by the data processing team at headquarters after data transfer (usually the day after data collection). The full CAPI application contained 366 consistency checks. These fall into three broad categories, depending on whether they were designed to detect routing errors (248 checks), unlikely entries (61 checks) or impossible entries (57 checks). We will discuss each of these checks in turn.

Over two thirds of the checks aimed at detecting violations of the questionnaire's routing scheme. Routing errors occur by answering a question that is supposed to be skipped, or by skipping a question that is supposed to be answered. The questionnaire had a total of 152 variables, out of which 100 were dependent on previous answers and 52 were unconditional. Each unconditional question had a single check detecting missing entry, while each conditional question had two checks: one detecting missing entry and one detecting an entry made in a disabled field. Four routing checks turned out to have malfunctioned, leading to a total of 248 routing checks. Answers such as 'don't know' or 'refused' were not recorded as missing, but had their own codes.

Another 16% of the checks constituted checks detecting impossible entries or impossible combinations of entries. Some were simple range checks on a single variable, for example verifying that the number of days a person reported to be ill for in the past 4 weeks did not exceed 28 or ensuring the value for a consumed quantity was not negative. Others checked consistency across variables, highlighting, for example, situations where the age someone started school at exceeded his current age, or a member's relation to the head of the household was 'spouse', but the head's marital status was 'never married', or a male person had pregnancy related problems. Some of these checks could have been avoided by restricting the range of permissible responses in the first place (more on this below).

The remaining 17% of the checks constituted checks detecting possible, but unlikely entries, such as an uncommon number of cows, or an uncommon expenditure value. Verifications for unlikely combinations of entries could trigger warning messages such as "nobody in the household is older than 15 years", "the main activity of person is full-time student but person is not currently in school", or "a house with a thatched roof is unlikely to have electricity, please verify". If an unlikely entry was detected, the interviewer was

---

<sup>3</sup> Examples of screen shots of the electronic questionnaire, including check buttons, are available in on-line Appendix 1. The complete list of consistency checks is given in on-line Appendix 2.

obliged to verify with the respondent, and, if the unlikely entry turned out to be correct, to comment on the situation to reassure the analyst that the data point was indeed correct.

Besides the system of 366 consistency checks, the full electronic questionnaire also included a report summarizing the total calorific intake and its sources, as implied by the entries in the consumption section, allowing the interviewer to verify the plausibility of the consumption data.<sup>4</sup> This consumption report was also part of full CAPI and omitted in restricted CAPI, but it will be more completely discussed in Section 4 below.

Finally, as respondents partake in resolving errors and inconsistencies, one could hypothesise that attitudinal factors, such as belief in the accuracy and usefulness of the survey, are affected by consistency checks.

### *2.3. Experiment 2: bundle of other CAPI features*

Experiment 2 consisted of adding a further 320 PAPI questionnaires to the sample. Because of the random nature of the questionnaire allocation, any difference between restricted CAPI and PAPI can only be due to the bundled effect of all CAPI features, excluding checks.

In line with most CAPI applications, we incorporated automated routing. The literature stresses automated routing as one of the most important error reducing features of CAPI. For example, Banks and Laurie (2000) note that reducing errors related to complex routing in a 45 minute questionnaire was the main justification for migrating the British Panel Household Survey to CAPI in 1999. Automated routing avoids asking a question that should have been skipped, which may decrease the length of the interview, avoids asking irrelevant questions (which confuses respondents and may lower the regard they hold for the survey and its results) and decreases time spent correcting data after the fieldwork. Automated routing also avoids the converse: skipping questions that should have been asked and may therefore prevent dropping observations during analysis. In this CAPI application, automated routing did not eliminate the need for routing checks. Unlike other existing CAPI surveys, our experiment displayed multiple questions and sections per screen and allowed the interviewer to continue the survey even if a required field/section was left blank. We made a conscious decision to set the programme up like that in order to allow the interviewer to return to a question later if, for example, the most knowledgeable person was not around. If an interviewer backtracked to change a response that determines subsequent routing, then an entry in a disabled field occurred. Again, we could have set it up so that the computer deletes entries in disabled fields automatically, but we were worried that that could lead to unintended data loss, especially if gateway questions are accidentally changed after completing a section. The experiment allows us to disentangle the effects of checks from those of automated routing.

The data were stored in a relational database, using a record structure which eliminates redundancy. Key identifiers were used to link the various data tables in a manner that ensures the referential integrity of the complete dataset (this means, for example that a

---

<sup>4</sup> On-line Appendix 3 gives an example of a consumption summary report.

household asset cannot exist without a related household, the identifier key being common to both data tables). Answers to most questions were selected from pre-coded drop-down menus or made use of radio-buttons. In some cases, drop-down menus were altered dynamically, depending on previous responses, so that the interviewer was never presented with an impossible response code. For example, when linking a woman to the ID of the husband the drop-down menu was restricted to married men within the household based on the previously filled in marital status and sex variables.<sup>5</sup> GPS coordinates and start and end times of the interview were captured automatically by the computer, eliminating any scope for interviewer error. In PAPI, the interviewer needed to copy the GPS coordinates from a GPS receiver and record start and end time of the interview in the appropriate fields.<sup>6</sup> Finally, PAPI had a data entry stage where paper forms were re-keyed into the computer. There were numerous other smaller features that could all add up to a cleaner dataset. The experiment was not set up to isolate the effect of each of these features separately, so we can only identify them as a bundle of effects driving the difference between restricted CAPI and PAPI.

Just like the system of consistency checks, also the bundle of other CAPI features may contribute to the respondent's attitude towards the survey. For instance, noticing that the interviewer is using a computer device instead of pen and paper may increase the respondent's perception of survey reliability.

#### *2.4. Implications for sample bias and analysis*

One likely consequence of the survey errors as described is that they generate missing variables and so reduce the effective sample size available for analysis. A questionnaire with missing or obviously erroneous data may lead the analyst to drop the observation entirely. If observations are randomly dropped, then one could simply increase the sample size of a PAPI survey to compensate. If, however, such mistakes are correlated with household characteristics otherwise of interest to the data user, then the analysis could be affected. We set up a formal test for this in Section 3. Alternatively, an analyst may decide to make assumptions about the problematic observations in order to avoid dropping them from the sample. These assumptions may then introduce measurement error. Section 4 analyses the nature of that measurement error and its effects on analysis. The remainder of this section gives more detail on the types of checks full CAPI included.

The share of questionnaires that have at least one impossible or missing entry potentially leading to missing values in our dataset amounts to 2%, 40% and 83% in respectively full

---

<sup>5</sup> As pointed out by one referee, some of the checks could have been alternatively implemented by restricting answer options. The spouse drop-down and the item-specific unit list in the consumption section (described in Section 4.1) are two example of where we opted for this approach. In many places, however, we preferred checks as it could confuse an interviewer if he or she fails to locate an expected response option from the drop-down without any indication of which previous answer triggered its elimination from the list.

<sup>6</sup> Time data are notoriously difficult to collect in Tanzania, because Swahili time is counted differently. 7 am is considered the first hour of the day and called "1 o'clock". Time during the day is counted upwards from there till 6 pm, which is called "12 o'clock" (the 12<sup>th</sup> hour of the day). After that the first hour of the night is 7 pm and so forth. English and Swahili times are often mixed up in the same questionnaires.

CAPI, restricted CAPI and PAPI. Whether or not the analyst will drop an observation, however, will probably depend on the willingness to make assumptions and the type of analysis conducted. Table A1.1 in Appendix 1 lists the 15 most commonly occurring missing values in any section in PAPI, excluding the consumption section (discussed separately below and in Section 4).<sup>7</sup> The most frequent errors are nonsensical survey durations, which occur in 24% of PAPI questionnaire, but in virtually no CAPI questionnaires. One could think that interviewers were more negligent recording time stamps, because they did not consider them an important focus of the study. The questionnaire was implemented in the context of a rural upgrade project and thus any questions on transport were especially important in the study. Despite this, many PAPI questionnaires have problematic transport data. Appendix 1 shows that 9% of questionnaires miss the amount paid to transport at least one sold agricultural item, 7% miss data on the amount spent on transport to school for at least one member, 6% on the one-way fare to school and 7% on the location at which crops sold fetched the highest price. In practice an analyst may assume that by leaving the value blank the interviewer may have wanted to indicate that they were supposed to be zero. Another analyst may decide the interviewer made a mistake and place the value at the cluster or sample median. Neither will have much basis for that decision. Robustness analysis for these and hundreds of similar data cleaning decisions that need to be made in a typical dataset is unlikely to be feasible. Assuming that a purist would want to drop any household that has any of the four transport related question missing, then that would imply dropping 20% of observations. The other potentially missing variables listed in Table A2.1 occur in core variables, which are key to calculating statistics like fertility rates, literacy rates, the number of people living with a disability and the number of landless households.

Table A2.2 in Appendix 1 lists the ten most common consumption related (potentially) missing values. In terms of the share of questionnaires in which the error occurs at least once, the most common consumption related error concerns food items for which the three sources ('purchases', 'home production' and 'gifts') do not sum to the indicated total. This error occurred at least once in more than 17% of all PAPI surveys. In comparison, this error occurred only in 3 % of restricted CAPI households and in close to none of the full CAPI.<sup>8</sup> In terms of the average frequency per questionnaire, the top error concerns the question "In the past 7 days did household consume any [Food Item]?", which was missing for 4 food items on average (out of a total of 53 items per survey) in about 6 % of all PAPI surveys, 1 time on average in about 9 % of all restricted CAPI surveys, and zero times in full CAPI surveys. In Section 4, we will determine whether these inconsistencies lead to different analytical conclusions.

---

<sup>7</sup> Note that we look at the 15 most common missing values in PAPI, as opposed to the 15 most common missing values over all three applications. The main purpose of this table is to inform us on the type of errors made in PAPI, and not necessarily to compare the frequency of missing values across the three applications.

<sup>8</sup> The most likely reason why this error did not occur as frequently in restricted CAPI as in PAPI is that CAPI displayed the total amount consumed coming from the three different sources on the screen, allowing the interviewer to check. If, despite this, the sum was still not correct, then a consistency check in full CAPI warned the interviewer of the mistake.

### 2.5. Interviewer effects

The quality of survey data depends to a large extent on both the technical capacity and the integrity of the interviewers. We expect education level and previous survey experience to improve the quality of survey data. In CAPI, the use of new survey technology might pose additional challenges to the interviewers on the one hand. On the other hand, we expect some CAPI features, such as automatic routing and the elaborate system of validation checks, to assist the interviewers, possibly compensating for lower education and experience. In PAPI, it is likely that interviewers make less routing and consistency errors as the field work progresses, because they receive feedback from their supervisors at the end of each survey day.

## 3. Errors and Sample Size Reduction

### 3.1. Methods of Analysis

To investigate the effects of CAPI on errors and potential sample size reduction more formally, we start by estimating  $Y_{ijc}$  (simply written as  $Y_i$  in what follows), which is a count of the number of problematic variables (some of which may potentially have to be dropped in analysis) in the questionnaire of household  $i$ , interviewed by interviewer  $j$  in community  $c$ :

$$(1) \quad Y_{ijc} = \alpha + \beta C_i + \gamma V_i + \varepsilon_i$$

where  $C_i$  indicates a CAPI questionnaire (a dummy equal to one for both full and restricted CAPI and zero for PAPI) and  $V_i$  is a dummy set to one if the interviewer had access to validation checks during the interview, which was only the case in full CAPI, but not in restricted CAPI and PAPI. In equation (1)  $\gamma$  measures the effect of the validation checks on the dependent variable, while  $\beta$  is an estimate of the bundled effect of all other CAPI features that could influence the number of errors in a questionnaire.

If error counts depend on household characteristics otherwise of interest to the data user, then the dropping observations with erroneous variables could introduce sample bias.<sup>9</sup> To investigate this, we check whether the number of problematic values in a questionnaire depends on household characteristics  $X_i$  and whether CAPI can correct for this.

Therefore, we are particularly interested in the level effect of  $X_i$  as well as its interactions with  $C_i$  and  $V_i$ :

$$(2) \quad Y_{ijc} = \alpha + \beta C_i + \gamma V_i + \delta X_i + \phi X_i \cdot C_i + \rho X_i \cdot V_i + \varepsilon_i$$

Where  $Y_{ijc}$  is a count of the number of variables that potentially have to be dropped or cleaned in household  $i$ .

---

<sup>9</sup> Observations may not need to be dropped if cleaning assumptions are made. This may introduce measurement error, the nature of which is the subject of Section 4.

In a final specification we will estimate interaction effects of  $C_i$  and  $V_i$  with interviewer's characteristics such as months of experience with CAPI, months of experience with PAPI, and years of education. This will allow us to verify whether the measured effects differ with experience in either type of questionnaires, as well as with education level.

Although the set-up ensured that questionnaires were equally and randomly spread over interviewers, clusters and time, we also verified that all results were robust to the controls for additional factors that may influence the number of errors in an interview: characteristics of the respondent (age, sex, literacy, whether a head of household), characteristics of the interview (conducted on day one or two of the team's visit), the interviewer and the location. The latter two effects are included as cluster ( $\mu_c$ ) and interviewer ( $\lambda_j$ ) fixed effects. We find that all estimations are robust to these further controls, so will not report this further.

### *3.2. Routing Errors*

Our measure for the number of routing errors is a simple count of the number of times an unconditional variable was missing or a conditional variable mistakenly entered or missing (dependent on previous responses). It should be noted that a single error early on can sometimes have a cascading effect, creating a large number of routing errors throughout the questionnaire. Table 1 shows that PAPI contained an average of 10 routing errors per survey, restricted CAPI 0.6 and full CAPI 0.0.

Column 1 and Column 2 in the first panel of Table 2 show that restricted CAPI significantly reduces the total number of routing errors by almost 10 per questionnaire compared to PAPI. Column 1 shows that there are on average 4 missing entries in required fields (the constant in the regression without controls), out of which 3.5 are eliminated through CAPI. The remaining 0.5 errors are wiped out by adding checks to CAPI. All of the 6.3 entries made in fields that ought to have been skipped, on average in PAPI, are eliminated by CAPI, with no additional effect of the checks. The latter type of error is perhaps less problematic than the former one, but such ambiguity in the data is nevertheless best avoided and will, in any case, add time to the interview (see below). Taken together, this shows that 94% of routing errors are avoided through the automated routing system and that the checks eliminate almost all those that remain. Appendix 2 shows that this does not lead to respondents reporting a smoother survey experience. It is unlikely that this result stems from interviewers leaving 'don't know' responses blank. First, there were specific codes for such a response and the interviewers were trained extensively on this matter. Secondly, a comparison of the occurrence of 'don't know' answers across the three different experiments does not show any significant differences. CAPI lends itself to the use of unfolding brackets to reduce 'don't know' answers, but this specific experiment did not make use of them.

### *3.3. Unlikely and Impossible Entries*

Column 3 in the first panel of Table 2 shows that restricted CAPI reduces the number of impossible entries by 0.34 per questionnaire compared to PAPI and adding checks further reduces this number by 0.15, to almost zero. This means that in a dataset of 1200 households, moving from PAPI to full CAPI would reduce the number of impossible

entries by 588 in total. The bundled effect of ‘all other CAPI features’ on the occurrence of impossible entries, as discussed in Section 2, seems to be larger than that of the checks.

The last column in panel 1 of Table 2 shows that CAPI significantly reduces the number of unlikely entries by 0.26 per household survey. This effect is even greater when checks are available, with number of unlikely entries falling from 1.35 in PAPI to 0.63 in full CAPI. This result suggests that, although some unlikely entries remain (once confirmed to be correct by the interviewer), full CAPI successfully assists the interviewer in detecting unusual entries that turn out to be incorrect after confirmation. Furthermore, because the programme flags these entries and reminds the interviewer to comment, the analyst is reassured that the data point is indeed correct.

Appendix 2 further shows that the techniques introduced by CAPI to avoid these errors do not increase the credibility or usefulness of the results in the eyes of the respondent.

An unintended natural experiment occurred within the experiment. It was realised, during analysis for this paper, that 13 validation rules had been erroneously omitted from the programme. Tabulating the number of times each of these malfunctioning checks was violated in the resulting dataset, we find no significant differences across the three types of questionnaires. This suggests that CAPI is only as good as the features that get built into it. Without checks or other error reducing features, CAPI has no impact on impossible entries.

Panel 2 in Table 2 shows that there are 24% of households that had problematic interview duration calculations in PAPI, but CAPI reduces this to virtually 0. The same panel further shows that PAPI has 6.6% problematic GPS locations, which are largely eliminated through CAPI’s automatic GPS capture. Enumeration Areas were very small in Pemba and we can be confident that any location farther than 1 km away from the cluster centre is problematic. One may argue that any analysis requiring the use of time stamps or GPS locations should simply increase its sample size to account for this. But, as will be shown next, these missing observations occur more frequently in certain types of households.

#### *3.4. Implications for Sample Size Reduction and Sample Bias*

A missing or an impossible entry may cause an observation to be dropped, which may lead to biased estimates if the missing values are non-randomly distributed across the sample. To investigate this, Table 3 shows estimates of Equation (2) for four different left hand side variables (the uninteracted results for these four regressions are shown in the first two columns of panel 2 in Table 2). The dependent variables in the first two columns are simple sums of the number of missing entries in required fields and the number of impossible entries. This sum is first made for entries in any part of the questionnaire, excluding the consumption section (column 1) and then separately for entries in the consumption section (column 2). Both of these can lead to either dropping the observation in question or making an *ad-hoc* data cleaning decision about what is going on. The third dependent variable indicates whether or not there was a problem with the

time stamps and the fourth whether or not there was a problem with the GPS co-ordinates. We do not use the information on entries in disabled fields or unlikely observations as these two types of errors would likely not lead to dropping the observation. Unlikely observations may introduce error and affect analysis, but that will be the subject of Section 4 below.

Table 3 shows that the sum of the number of missing entries in required fields and the number of impossible entries (as picked up by the validation rules) are dependent on household characteristics. The first column shows how large, female headed and non-farm households are more likely to have non-consumption related entries that could cause the observation to be dropped or the entry to be altered by the analyst. The second column shows a different pattern when focussing on the consumption section. We see here that rich households make more errors, possibly due to their more complicated consumption patterns. As expected, farming households have more problematic consumption data, as a larger share needs to be estimated from home production, often using subjective units of measurement. The coefficient of household size is now significantly negative. The effect is not large – increasing household size with 5 members would reduce the number of consumption errors in a questionnaire by an average of 0.4 – but still significant and could be explained by the fact that larger households (more than 9 members) have only an average of 1.8 more consumed items compared to small households (1-3 members), while smaller households are 40% more likely to use decimals in their quantity estimation, which generally are more prone to erroneous entries. Furthermore, while there is no difference in the types of consumption items consumed by smaller households, the sources from which they obtain them are different: small households have more consumption from gifts and may therefore be less familiar with objective units of measurement found in the market place. The coefficient of female headed household is no longer significant. Importantly, once the interaction with CAPI and any of the characteristics discussed above is made, the effects disappear: the sum of the level and interaction effects is never statistically significant (verified by the authors). Interactions with age and education of head were not found to be significant.

Surprisingly, we find that even problems with time stamps and GPS locations are not independent of household characteristics. In particular, they occur more frequently in large households. This could be because large households have a much longer interview time, as the questionnaire contains many roster questions that are repeated for each member. Median interviewing times on paper are 53, 82 and 113 minutes for a 1, 3 and 8 person household, respectively. It took 141 minutes to interview the one 13 member household in the survey. This increase in interviewer workload may reduce concentration when copying time or GPS co-ordinates.<sup>10</sup> We confirmed by a formal statistical test that CAPI undoes the negative effect of household size on problematic GPS and interview duration measurements.

---

<sup>10</sup> Lengthy questionnaires have more non-consumption errors in general. This is confirmed by the results of a regression of the number of non-consumption related missing/impossible entries on the duration of interview (not reported), which shows a significantly positive correlation between the two. Interestingly, the interview length does not influence the number of missing/impossible entries in the consumption section, which squares with the finding of the negative coefficient on household size here.

### *3.5. Interaction effects with interviewer characteristics and survey period*

Table 4 shows the interaction effects of our main variables of interest (CAPI and checks) with total number of years of formal schooling and number of months of CAPI and PAPI survey experience of the interviewer. We find that both PAPI survey experience and education significantly reduce the total number of alerts (routing + impossible + unlikely entries) in PAPI surveys. Interestingly, the number of years of CAPI survey experience seems to significantly *increase* the number of errors on paper, for a given number of years of PAPI experience. This suggests some unlearning of best-practice PAPI skills once interviewers switch from paper to the computer. Both experience and education effects disappear once CAPI is used (confirmed by formal statistical tests). Banks and Laurie (2000) noted how PAPI interviewers can be easily re-trained to conduct CAPI. This result suggests that CAPI can, to some extent, compensate for lower education and experience level of interviewers, mainly because of automated routing. The interaction effects of checks with education and experience are not significant.

Table 5 provides data on whether error rates drop as the survey progresses over time and, if so, whether the pattern is different for CAPI and PAPI. To do this, we split the 37 survey days up into quartiles and include dummies for each period, both as levels and as interactions with the CAPI and checks dummies. The results suggest that error rates do indeed drop for PAPI, but not for CAPI. Compared to the first quartile, the total number of alerts is significantly lower in subsequent survey quartiles for PAPI, with almost 10 alerts less per PAPI household survey in the last 9 days of the survey. Once the interaction effects with CAPI are added to the level coefficients, the effect of the quartile disappears (confirmed by formal statistical tests). In other words, there is no similar learning effect for CAPI. One reason for this could be that the average number of alerts in the first quartile of CAPI survey work is very low (0.8) relative to PAPI (18) and therefore there is much less scope for improvements under CAPI than under PAPI (Table 1). Interactions with checks are insignificant. Taken together, the results from Table 4 and 5 suggest that CAPI is less dependent on interviewer experience, education and interviewers learning over time as the survey progresses.

## **4. Measuring Consumption**

### *4.1. Food Consumption*

Estimates of poverty and welfare in developing countries are frequently calculated using a consumption recall module in a household questionnaire. While the largest share of consumption is related to food, it is exactly food consumption that is most problematic to measure accurately. The typical food recall module will have the interviewer go over a list of food items in two iterations. In the first iteration, each item consumed by the household over the recall period is flagged. A second iteration then goes through the list of flagged items and, for each, asks total household consumption and its decomposition into sources (purchases, home production and other sources).

Three important problems arise. First, quantities are expressed in imprecise units; households report consumption as “pieces” of cassava, “bundles” of spinach or

“bunches” of bananas (Capéau and Dercon, 2006). This leads to ambiguous item-unit combinations. While the size of such units is subject to interpretation (large versus small), the analyst needs a clear mapping onto metric units. Second, the list of units is uniform for each consumption item, even though some units in the list do not apply to all items (e.g. “litre” for “potatoes”). This causes conflicting item-unit combinations, usually detected only much later during data analysis. Third, a completed consumption module represents a rather unwieldy matrix making it hard for an interviewer to maintain an overview of the consumption pattern of the household. Therefore, obvious errors and irregularities in the reported consumption are only highlighted several months later when researchers start analysing the data. At that point, the only solution is to either make an *ad-hoc* assumption about what is meant, or omit the observation from the sample.

In CAPI the screen of the handheld device can be used to display pictures of vague units, such as “bundle” or “bunch”, so that they can be more precisely mapped onto metric units.<sup>11</sup> The application can also tailor the list of units to be specific to the item, making it impossible to, for instance, express potato consumption in litres or cooking oil in bags. Finally, by mapping each item-unit combination to its calorific value, the computer can summarize, in a report, the calorific intake pattern of the household<sup>12</sup>. This allows the interviewer to carry out a report-based check during the interview, to verify whether the total Kcal per AEU lies within reasonable boundaries and that the sources of calories are sensible given the context in which the interviewer is conducting the work. We refer to this report as the ‘consumption report’ in what follows. Additionally the automated routing and the consistency checks, discussed in detail in section 2, are expected to improve data quality. Some of these features could, in principle, be implemented in paper questionnaires, although the logistics are more complicated here. This is especially true for the automated routing, the consistency checks and the consumption report, which rely on complex matrix manipulations and look-up tables. In this experiment full CAPI had all three features, restricted CAPI omitted the checks (which we mean to include both the validation checks and the consumption report on total food energy intake and its sources), while PAPI also omitted pictures and item-specific units (e.g. we would just have ‘bunch’, or ‘litres’, as a possible unit code for reporting banana consumption).

Table 6 shows that in 1% of PAPI cases the item unit combination made no sense; in these cases the calorie value was replaced by the median EA-level value in the subsequent analysis. A further 42% of the item-unit combinations in PAPI were ambiguous (pieces, bunches, bundles, heaps, etc.) and, in order to obtain a precise conversion to Kcal values, an assumption about the exact size of the ambiguous unit needed to be made. We used lower and upper bound estimates of the unit conversion rates, as well as a mid-range value (a typical user would have likely used this mid-range estimate). While upper- and lower-bound conversion rates were quite reasonably set<sup>13</sup>, Table 6 shows that changing the assumptions on unit conversions from lower- to upper-bound estimates raises calorific intake per AEU per day from 2,478 to 4,362. There is

---

<sup>11</sup> Examples of pictures used are displayed in the on-line appendix 4.

<sup>12</sup> On-line Appendix 3 gives an example of such a report.

<sup>13</sup> In many cases the units could be matched to the CAPI pictures and the lower-bound was taken as the size of the smallest depicted unit and the upper-bound the size of the largest unit.

also a substantial increase in the standard deviation as one goes from full CAPI (655) to restricted CAPI (1,177) and PAPI (1,644 – 3,379). The number of outlier observations, with values over 4,000 Kcal per adult per day, is 1% for full CAPI, 8% for restricted CAPI and 7%, 20% and 35% for PAPI, depending on the conversion assumption.

These results suggest that the effect of the ‘other CAPI features’, probably pictures and item-specific units in this case, depends on how far off the *ad-hoc* assumptions on unit size in PAPI are from reality, while the effect of the checks is independent of this. In fact, we do see that in CAPI the pictures of the smaller units were 14 times more likely to be chosen than those of large units and nearly 2.5 times more likely than mid-range units. Equipped with this knowledge we can adapt the unit conversions, but it is fair to say that most similar surveys would base their unit conversions on much thinner data. Because we know the small unit assumptions are closest to the truth, we will use these in the remainder of the text. In this way we expect any differences between PAPI and CAPI to be lower bound estimates. Finally, there are a number of other small data cleaning decisions that needed to be made with regard to all the violated consistency checks.<sup>14</sup>

Would our assessment of the food situation have changed, depending on whether we did a survey on paper or electronically? The answer depends on the calorific intake threshold we consider when defining malnutrition. Had we done the survey on paper, we would have concluded that 21% of households live on less than 1,500 KCals per AEU per day. Had the same survey been conducted in full CAPI then the conclusion would have been that 8% of households live below this threshold. This difference is statistically significant at well under 0.01%. Table 6 further shows that restricted CAPI puts the same figure at 14%, implying that, on average, 6 percentage points of the difference between full CAPI and PAPI is due to checks, while 7 percentage points are due to other CAPI features. Raising the threshold to 1,800 Kcal/AEU/day still shows a significant difference between full CAPI and PAPI ( $p < 0.01\%$ ), but the effect is completely due to the checks and not to the bundle of other CAPI features. When we consider 2,200 Kcal/AEU/day as the malnutrition threshold, we do not find any significant differences between CAPI and PAPI.

#### 4.2. Total Consumption, Poverty and Inequality

To arrive at a consumption measure, we place a monetary (Tanzanian Shilling (TZS)) value on food consumption. For purchased items this comes directly from the respondent’s assessment of the value, while for gifts and home produced items unit prices are used to convert the quantity estimations into monetary values. To this food consumption is added a non-food consumption component, which was asked directly in monetary value both on paper and in CAPI. As Pemba is a very small island and the survey was concentrated in the Northern half only, we do not correct for prices. The regression analysis, however, will always verify robustness to inclusion of cluster fixed effects.

Table 7 shows that average consumption increases with 9% as one moves from full CAPI to restricted CAPI and another 15% when moving to PAPI, creating a jump of 25% from

---

<sup>14</sup> For example, we assumed that a missing source entry indicated zero consumption for that particular source, as discussed in Section 3.4.

full CAPI to PAPI. These mean differences also translate into very different conclusions regarding the number of people that live below the basic need poverty line, with the poverty headcount going down from 83% in full CAPI to 68% in PAPI.<sup>15</sup> Note that the 2005 Zanzibar Budget Household Survey (conducted on paper) reported a poverty headcount of 72.54 on average for the region we consider (Wete and Micheweni districts). Because both the differences between full and restricted CAPI, as well as between restricted CAPI and PAPI are significant, we conclude that both the checks, as well as the bundle of other CAPI features are important. Interestingly, the CDF drawn in Figure 1 show that the effect of the checks depends on where one draws the poverty line. For poor households there is no effect of the checks, while the effects start appearing from around TZS 400,000 onwards. This evidence is consistent with the fact that rich households have more complicated consumption patterns, where the power of the computer is important to summarise the information in an intelligible way. With poorer households, it is possible that consumption patterns are relatively uncomplicated so that even without the consumption reports an interviewer has an intuitive sense of whether the entries, as a whole, are reasonable.

Next, we investigate what this means for inequality. Consistent with expectations, we see that full CAPI lowers the Gini coefficient to 0.24 from 0.30 in PAPI ( $p < 0.01$ ), a difference almost entirely attributable to the checks and not the bundle of other CAPI features. At least some of the variation picked up in restricted CAPI and PAPI is actually measurement error rather than real inequality. Finally, we estimate Engel curves (log food consumption regressed on a constant and the log of total consumption) within each of the three different data sets and show, in Table 7, what the implied differences are with respect to the calculation of the income elasticity of food consumption. We see, as with the Gini coefficient, pronounced differences when moving from full CAPI to restricted CAPI, but not from restricted CAPI to PAPI. We conclude that it is mainly the checks that explain the difference between paper and electronic questionnaires with respect to Gini coefficients and income elasticities.

#### 4.3. Testing for Classical Measurement Error in Consumption

The previous results demonstrate a tendency for the mean and spread of total consumption to go up on paper, compared to full CAPI, translating in lower measured poverty and higher measured inequality in PAPI. We have also seen that this can lead to significantly different coefficients on the total consumption variable when estimating Engel curves. We also know from Section 3 that simple error counts in consumption data depend on wealth, household size and the occupation of the head. In this section, we further analyse the nature of the measurement error (after cleaning, see footnote 17). We closely follow Bound et al. (2001) by considering a true linear model  $y^* = X^* \beta + \varepsilon$ . Rather than observing the true scalar  $y^*$  and the true vector  $X^*$ , we observe  $X = X^* + \rho$  and  $y = y^* + \varphi$ . The measurement error is called classical if  $\rho$  and  $\varphi$  are uncorrelated to

---

<sup>15</sup> Our poverty line is set at a value of TZS 580,832 of annual consumption per aeu. To construct this poverty line, we started from the basic need poverty lines of Wete and Micheweni (where this experiment was conducted) as defined by the Zanzibar Household Budget Survey 2005. We then adjusted this poverty line for inflation and differences in survey methods by multiplying it by a factor reflecting the difference in median consumption between the 2005 HBS dataset and our own 2009 dataset.

$X^*$ ,  $y^*$  or  $\varepsilon$ , i.e. when the error is uncorrelated with the true value of the measured variable, the true values of other variables in the model, and the stochastic disturbance term. Lacking validation data, i.e. information on the true values  $X^*$  and  $y^*$ , it is difficult to test these assumptions underlying the classical measurement error model directly. However, we can test whether our data is consistent with standard classical measurement error results: classical measurement error causes no bias when just the dependent variable has error and attenuates the estimated coefficient on a single error-ridden explanatory variable. The remainder of this section uses that insight to set up two tests for classical measurement error.

In a first test, we exploit the insight that an explanatory variable, measured with classical error, will lead to attenuation bias. In order to test the hypothesis of zero attenuation bias, we estimate the following estimation equation for three different outcomes  $O_i$  that can be explained by consumption:

$$(3) \quad O_i = \alpha + \beta C_i + \gamma V_i + \delta \text{LnCons}_i + \phi \text{LnCons}_i \cdot C_i + \rho \text{LnCons}_i \cdot V_i + \varepsilon_i$$

Where  $C_i$  and  $V_i$  are defined as before and  $\text{LnCons}_i$  is the log of total consumption per AEU of household  $i$ . Finding  $|\delta| < |\delta + \phi|$  and/or  $|\delta| < |\delta + \phi + \rho|$ , would be consistent with attenuation bias with PAPI. Table 8 shows three regressions. First one where  $i$  is a child between 7 and 14 years old and  $O_i$  is the number of years of formal schooling completed by the child, controlling for age fixed effects to compare only children of the same age. In a second specification  $i$  is a school-going child in the sample and  $O_i$  is the amount spent by the household on his or her education (in which case we drop the education expenditure components from the consumption aggregate to avoid spurious correlation). The third is a regression where  $O_i$  is a dummy for whether the child (0 to 14 years old) slept under a treated mosquito net the night prior to the survey, again controlling for age fixed effects. In all cases we would expect family wealth and its correlates, measured through consumption, to influence the outcome in question, while children's outcomes to have no influence on consumption.

The results, displayed in Table 8, show that, on paper, the number of years of schooling a child has attained is independent of household consumption. Once checks are included, however, a positive association emerges, consistent with attenuation bias in the PAPI and restricted CAPI results. Total consumption does explain schooling expenditures for children at school and whether or not a child slept under a bednet, but the size of the effect is estimated at around half of what full CAPI estimates, again consistent with attenuation bias. The coefficient on CAPI is not significant, indicating that it is the consumption reports and validation checks that are responsible for these differences and not the bundle of other CAPI features.

One critique on these regressions could be that the left hand side variables are themselves measured more accurately in CAPI compared to PAPI. The computer verified during the interview, for example, that the grade attained was sensible given the age of the child, flagged zero education expenditures for school-going children and expenditures over 3 million Tanzanian Shillings as unlikely and cross-checked the transport costs for going to

school with other entries in the transport section. Finally, there was cross-validation check between bednet use and ownership of a bednet in the assets section. However, estimating  $O_i = \alpha + \beta C_i + \gamma V_i + \varepsilon_i$  gave insignificant coefficients on  $C_i$  and  $V_i$ , giving some confidence that the results are not affected by this possibility.

We also estimated all coefficients restricting ourselves to within-sample estimates and artificially (and randomly) reducing the full CAPI sample to have the same observations as restricted CAPI and PAPI. All results remained intact after this exercise. Results are also robust to inclusion of cluster fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy), household characteristics (household size), interviewer fixed effects and interview characteristics (conducted on first day of team's visit). In 13 PAPI child records at least one of these variables was missing, in which case we included a dummy indicating this.

The second test starts from the standard result that there is no bias if just the dependent variable suffers from classical measurement error. To investigate whether the evidence of this experiment is consistent with the latter property of classical measurement error, we run a regression explaining log total consumption per AEU with various factors typically included in analysis of determinants of household level poverty and interaction effects between each of them with  $C_i$  and  $V_i$ . Under conditions further specified below, significant interaction effects would prove, by contraposition, that measurement error is not classical. In specifying our set of explanatory variables, we follow the guidelines provided by Haughton and Khandker (2009). The explanatory variables (household characteristics) included in the first regression in Table 9 are: household size, dependency ratio (counting adults as people aged 15 to 65), a dummy for whether the household head is female, average number of years of education of adult HH members (aged 15 and above), number of days head was ill in the last 4 weeks, dummy for whether the household owns its house, dummy for whether the house has a modern roof (made of iron, concrete/cement, tiles or asbestos), proportion of employed adults in the household and the acres of land owned. The regressions also control for cluster fixed effects.

We verified that all these variables are themselves not suffering from attenuation bias by regressing each one on  $C_i$  and  $V_i$  and ensuring both coefficients are insignificant. We considered, but omitted other variables, such as a dummy equal to 1 if head is *not* employed in the agricultural sector, the number of productive assets and the number of livestock owned by the household, because they were dependent on  $C_i$  and  $V_i$  and so could suffer from attenuation bias. In this case, a significant interaction effect may simply result from attenuation bias, and lead us to falsely reject the classical measurement hypothesis.

Column 1 in Table 9 displays the results of this regression and finds no significant interaction effects of CAPI and checks with determinants of poverty. This finding is consistent with the dependent variable having classical measurement error. Dropping the insignificant regressors from the analysis does not alter the results. For the sake of completeness, column 2 adds the problematic regressors and finds their interactions are indeed significant. This does not lead us to reject the classical measurement hypothesis,

as these results could simply be driven by attenuation bias on the PAPI estimates (insignificant level effects), which gets alleviated through CAPI (significant interaction effects).

Finally, we looked at the joint significance of all interaction terms with  $C_i$ , with  $V_i$  and both groups of interaction terms together and reject all three times the null that they are not significantly different from zero in column 1 of Table 9. We also ran more parsimonious regressions – one for each of the regressors – and find that, with the exception of the regression that includes the proportion of household members who are employed, all interaction effects are insignificant in all regressions. We do not consider our one exception sufficient evidence to reject the null of classical measurement error: considering the number of coefficients we are scrutinizing, we would expect some false positives to be present.

### **5. Further Dimensions of Comparison: Cost, Length of Interview and Respondents' perceptions.**

CAPI has a larger fixed cost component: up-front outlays for the development of the software and the purchase of the hardware. But many of the variable costs (per interview) that are incurred for PAPI, such as printing and data entry, are eliminated in CAPI. This makes CAPI, budget-wise, a more viable option for larger surveys. As a smooth-functioning rental market for hand-held devices does not exist, surveys with fewer interviewers, spread over a longer time will be cheaper, as fewer machines need to be bought. Organisations that regularly conduct surveys could share machines across projects to overcome this problem. The software used in this experiment (and the baseline survey on which it was based) is based on Microsoft Access and had been under development for 2 years. However, it still required about 50 consultancy days each of a senior and junior programme developer to adapt it for this specific survey. The survey had roughly 80% similarity with other surveys already conducted with the same programme. This comes to a fixed cost of USD 40,000 in consultancy fees for making the programme. A data entry application could be developed for around USD 4,000 using consultants at similar rates. The UMPCs used in this experiment, including peripherals such as extra batteries, a replacement battery after 2 years of use, GPS units, bags, charging equipment, transport, were about USD 1,800 a piece and have an estimated life time of 600 interviewing days or roughly 3 dollars per day. Interviewers conducted 3 interviews per day in this survey, so the variable UMPC cost per questionnaire was roughly 1 dollar per questionnaire. The variable costs per paper questionnaire was about USD 4 for data entry clerks and desktop computers, 4 dollars per questionnaire for data entry management and supervision (including adjudication of errors) and USD 2 for printing a single questionnaire. Thus in the context of this survey solving  $40,000 + X = 4000 + 10X$  for  $X$  gives a break-even point of 4,000 questionnaires. Below this, paper is cheaper and above this CAPI is cheaper. For example, a survey of 2,500 households would be USD 13,500 more expensive on CAPI, while a survey of 10,000 households would be USD 54,000 cheaper on CAPI. Adding wasted observations to this changes this number. For example, if a paper survey needs to collect 10% more observations, then the break-even point drops to 3,600 questionnaires. The break-even point can drop even

further if one considers reduced interview length (see below) and reduced data cleaning efforts.

It is likely that by the time this paper is published these figures have already changed substantially. Driven by the popularity of Apple's iPad, many hardware manufacturers are now developing their own UMPCs and it seems likely that the price of a machine will go down to around USD 300. Also in terms of software, this project used a hard-coded questionnaire that can only be adapted by experienced software engineers. This system was used because existing products (e.g. the programme used by Fafchamps et al (2010)) did not allow us to build in the complexity we needed. Once the market develops off-the-shelf products that do not require software engineers to be involved, there should be no reason to believe that making an electronic questionnaire takes longer than making a data entry programme. Fafchamps et al (2010), cite 75 hours of researcher time to build an enterprise survey and 20 hours to programme the follow-up questionnaires; without hiring the services of a software specialist.

Table 1 shows that interview time is reduced in CAPI by 10% (and by 14% in the restricted CAPI without validation checks). Examples of CAPI features that may be responsible for this reduction are the automatic routing system and the use of drop down menus to select responses from (instead of codes listed in a box somewhere else on the page). Finally, we see virtually no differences in the respondent's perception (e.g. degree of intimidation, perception of confidentiality, etc.) between CAPI and PAPI (see Appendix 2).

## **6. Concluding Discussion**

Many researchers and survey implementers are keen to switch from paper to electronic surveys, but there is currently little quantitative, empirical information available to inform that choice. This paper uses data from a survey experiment to identify differences between PAPI and CAPI and finds that errors leading to missing variables in PAPI are virtually eliminated in CAPI. A simple, compensatory increase in sample size on paper cannot adequately deal with this problem, because observations are not randomly dropped. The effect of CAPI is particularly evident when measuring consumption. We find that paper questionnaires can lead to estimates of higher mean consumption, lower poverty and higher inequality. We performed a number of regression analyses using the consumption aggregates as dependent and explanatory variables. Paper and restricted CAPI suffer from attenuation bias when consumption aggregates are used as regressors. We do not, however, find evidence of bias when consumption aggregates are used as the left hand side variable of a regression model. Hence, our evidence is consistent with classical measurement error. While there is scope for mimicking some of the CAPI features on paper, this seems unfeasible for the checks and reports made available to the interviewer during the interview. Results show that these two features play a key role, especially for reducing errors in consumption data. We further explain why this specific CAPI product leads to higher fixed up-front costs, but lower variable per-questionnaire costs. Finally, we show that interview times are significantly lower in CAPI, while there is no change in respondents' experience.

Some analytical caveats remain, however. First, the experiment does not provide an independent validation of the data as in, for example, Bound and Krueger (1991) and Bound et al. (1994). One could argue that in the analysis in Section 3, where we use error counts at the left hand side of the regression models, any dependency between the checks and the error count variables could bias the results.

A second critique is that an (apparently) error-free questionnaire is not the same as one that reflects reality. One may worry that interviewers can now simply enter any data that the computer is willing to accept. An unscrupulous interviewer could simply change a value to anything that suppresses the error message rather than make the effort to obtain the correct value. This survey was subject to intense quality control: supervisors did daily direct observations and brief re-interviews of respondents to check the validity of the data. An independent quality controller went to a random subset of 18 EAs to conduct re-interviews. The interviewers were aware that these random quality control visits would take place and no false data were detected in either PAPI or CAPI. As members of staff of a survey company (EDI), most interviewers intend to be on board for several surveys and several years.

Third, CAPI is not a panacea. We have found that the success of CAPI depends substantially on the effort spent programming, piloting and testing the application, as well as on careful consideration to the underlying data management and transfer systems. This specific experiment required considerable resources refining the application before taking it to the field. Without such preparation PAPI may well outperform CAPI. More generally, one needs to keep in mind that this experiment had no variation with respect to CAPI applications.

The literature from the developed world, 10-15 years ago, talks about the inevitability of the switch from CAPI to PAPI; despite the challenges the benefits are so attractive that a switch seems irresistible. It is our contention that a similar desire is growing amongst survey practitioners working in lower income countries, and with good reason, as this paper illustrates.

## REFERENCES

- Banks, Randy and Laurie, Heather (2000). "From PAPI to CAPI: The Case of the British Household Panel Survey." *Social Science Computer Review*, 18(4): 397-406.
- Black, Dan A., Berger, Mark C., and Scott, Frank A. (2000). "Bounding Parameter Estimates with Nonclassical Measurement Error." *Journal of the American Statistical Association*, 95(451): 739-748.
- Bound, John, Brown, Charles, Duncan, Greg J., and Rodgers, Willard L. (1994). "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, 12(3): 345-368.
- Bound, John, Brown, Charles, and Mathiowetz, Nancy (2001). "Measurement Error in Survey Data." In *Handbook of Econometrics* Vol. 5. J. Heckman and E. Leamer (eds), Elsevier, pp. 3705-3843.
- Bound, John and Krueger, Alan B. (1991). "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9(1): 1-24.
- Capéau, Bart, and Dercon, Stefan (2006). "Prices, Unit Values and Local Measurement Units in Rural Surveys: an Econometric Approach with an Application to Poverty Measurement in Ethiopia." *Journal of African Economies*, 15(2): 181-211.
- Chesher, Andrew, and Schluter, Christian (2002). "Welfare Measurement and Measurement Error." *Review of Economic Studies*, 69(2): 357-378.
- Couper, Mick P. and Burt, Geraldine (1994). "Interviewer Attitudes Toward Computer-Assisted Personal Interviewing (CAPI)." *Social Science Computer Review*, 12(1): 38-54.
- de Leeuw, Edith D. (2008). "The Effect of Computer-Assisted Interviewing on Data Quality: A Review of the Evidence.", mimeo, Department of Methodology and Statistics, Utrecht University. Accessed at <http://www.xs4all.nl/~edithl/surveyhandbook/deleeuw-cai-overview-updated.pdf>.
- de Leeuw, E. and Nicholls II, William L. (1996). "Technological Innovations in Data Collection: Acceptance, Data Quality and Costs." *Sociological Research Online*, 1(4). Accessed at <http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>.
- Fafchamps, Marcel, McKenzie, David, Quinn, Simon, Woodruff, Christopher (2010). "Using PDA Consistency Checks to Increase the Precision of Profits and Sales Measurement in Panels." CSAE Working Paper 2010-19.

- Gibson, John, and Kim, Bonggeun (2007). "Measurement Error in Recall Surveys and the Relationship between Household Size and Food Demand." *American Journal of Agricultural Economics*, 89(2): 473-489.
- Haughton, Jonathan and Khandker, Shahidur R. (2009). "Understanding the Determinants of Poverty.", Chapter 8 in *Handbook on Poverty and Inequality*, Vol. 1, World Bank, pp. 145-160.
- Office of Chief Government Statistician Zanzibar, Zanzibar Household Budget Survey 2004/2005, Zanzibar, Tanzania.
- Nicholls II, W. and de Leeuw, E. (1996). "Factors in Acceptance of Computer-Assisted Interviewing Methods: A Conceptual and Historical Review.", Proceedings of the Survey Research Methods Section, American Statistical Association. Available at [http://www.amstat.org/sections/SRMS/Proceedings/papers/1996\\_130.pdf](http://www.amstat.org/sections/SRMS/Proceedings/papers/1996_130.pdf).
- Taylor, Sue (1998). "Setting up Computer-Assisted Personal Interviewing in the Australian Longitudinal Study of Ageing." *Statistical Science*, 13(1): 14-18.

**Table 1: Summary statistics on errors, interviewer and survey characteristics and sample size**

		<b>Full CAPI</b>	<b>Restricted CAPI</b>	<b>PAPI</b>	
<b>Routing errors</b>	Average number of routing errors per HH (total)	0.0 (0.1)	0.6 (1.1)	10.4 (11.9)	
	Average nr of entries in to be skipped fields per HH	0.0 (0.0)	0.1 (0.6)	6.3 (9.3)	
	Average nr of missing entries in required fields per HH	0.0 (0.1)	0.5 (0.9)	4.0 (5.6)	
<b>Impossible/ Unlikely entries</b>	Average nr of impossible entries per HH	0.0 (0.0)	0.2 (0.4)	0.5 (1.1)	
	Average nr of unlikely entries per HH	0.6 (1.0)	1.1 (1.2)	1.4 (1.4)	
<b>Errors/unlikely entries per survey period quartile</b>	Average nr of routing errors + impossible entries + unlikely entries per survey period quartile (37 survey days in total)	1st	0.8 (1.1)	2.3 (2.2)	17.9 (16.6)
		2nd	0.8 (1.0)	2.1 (1.8)	12.2 (11.6)
		3rd	0.5 (0.9)	1.7 (1.6)	9.4 (10.2)
		4th	0.4 (0.7)	1.2 (1.2)	8.4 (7.4)
<b>GPS data</b>	% HHs > 1 km from cluster centre (likely outliers given the small size of the EAs)	0.6	1.3	6.6	
<b>Time stamp data</b>	% surveys with problematic time stamps	0.9	0.3	23.8	
	% surveys conducted on day 1 of the cluster visit	45.6	50.0	49.1	
	Average survey duration <sup>1</sup>	81 (24)	78 (23)	89 (25)	
<b>Interviewer characteristics</b>	Average PAPI survey experience (months)	5.7 (8.9)	5.7 (8.9)	5.7 (8.9)	
	Average CAPI survey experience (months)	7.4 (4.6)	7.3 (4.6)	7.4 (4.6)	
	Average education (years)	13.3 (1.3)	13.6 (1.3)	13.4 (1.3)	
	Total nr of interviewers	20			
<b>Sample size</b>	Total nr of clusters	80			
	Total nr of households per cluster	1200	320	320	

Notes: Values in parentheses are standard deviations

1. For full CAPI only the modules overlapping with the experiment were counted towards the survey duration (see Section 2 for details)

**Table 2: Effect of CAPI and Checks on data quality**

LHS	<b>Routing errors: Missing entries in required fields</b>	<b>Routing errors: Entries in fields that should have be been skipped</b>	<b>Impossible entries</b>	<b>Unlikely entries</b>
<b>Panel 1</b>				
	<b>OLS</b>	<b>OLS</b>	<b>OLS</b>	<b>OLS</b>
CAPI	-3.544*** (0.187)	-6.225*** (0.306)	-0.337*** (0.040)	-0.259*** (0.087)
Checks	-0.490*** (0.149)	-0.117 (0.244)	-0.148*** (0.032)	-0.461*** (0.069)
Const.	4.038*** (0.132)	6.344*** (0.216)	0.487*** (0.028)	1.347*** (0.061)
<b>Panel 2</b>				
<b>LHS</b>	<b>Potentially missing values in non- consumption sections (OLS)</b>	<b>Potentially missing values in consumption section (OLS)</b>	<b>Time Stamp Problems (LPM)</b>	<b>GPS Problems (LPM)</b>
CAPI	-2.800*** (0.142)	-1.081*** (0.121)	-0.234*** (0.015)	-0.053*** (0.010)
Checks	-0.287** (0.113)	-0.351*** (0.096)	0.006 (0.012)	-0.007 (0.008)
Const.	3.091*** (0.101)	1.434*** (0.085)	0.238*** (0.011)	0.066*** (0.007)

Notes: N=1840. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

1. Standard errors are shown in parentheses

2. All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy), household characteristics (No. of household members) and interview characteristics (conducted on first day of team's visit). In 16 PAPI households at least one of these variables was missing, in which case we included a dummy indicating this. Potentially missing values are measured as the sum of missing and impossible entries.

**Table 3: Interaction effects with household characteristics**

	Missing/ Impossible entries in non- consumption sections (OLS) coef/se	Missing/ Impossible entries in consumption section (OLS) coef/se	Time Stamp Problems (LPM) coef/se	GPS Problems (LPM) coef/se
CAPI	0.197 (0.434)	-1.369*** (0.379)	-0.163*** (0.049)	0.013 (0.032)
Checks	-0.203 (0.333)	-0.446 (0.291)	0.005 (0.037)	-0.022 (0.025)
Dummy = 1 if household head is female	1.367*** (0.271)	-0.044 (0.237)	-0.010 (0.030)	0.003 (0.020)
Household size	0.421*** (0.040)	-0.084** (0.035)	0.014*** (0.005)	0.008*** (0.003)
Dummy = 1 if head <i>not</i> employed in agriculture	1.149*** (0.209)	-0.408** (0.183)	0.013 (0.023)	0.015 (0.015)
Dummy = 1 if HH belongs to richest 25th percentile	-0.049 (0.212)	0.701*** (0.186)	-0.034 (0.024)	0.014 (0.016)
<b><i>CAPI interacted with:</i></b>				
Dummy = 1 if household head is female	-1.322*** (0.370)	-0.084 (0.323)	0.006 (0.041)	-0.020 (0.027)
Household size	-0.410*** (0.055)	0.079 (0.048)	-0.014** (0.006)	-0.010** (0.004)
Dummy = 1 if head <i>not</i> employed in agriculture	-1.069*** (0.297)	0.456* (0.259)	-0.019 (0.033)	-0.007 (0.022)
Dummy = 1 if HH belongs to richest 25th percentile	0.027 (0.310)	-0.882*** (0.271)	0.046 (0.035)	-0.005 (0.023)
<b><i>Checks interacted with:</i></b>				
Dummy = 1 if household head is female	-0.047 (0.284)	0.128 (0.249)	-0.002 (0.032)	0.027 (0.021)
Household size	-0.009 (0.043)	0.005 (0.037)	0.001 (0.005)	0.002 (0.003)
Dummy = 1 if head <i>not</i> employed in agriculture	-0.080 (0.235)	-0.044 (0.206)	0.004 (0.026)	-0.006 (0.017)
Dummy = 1 if HH belongs to richest 25th percentile	0.030 (0.260)	0.184 (0.227)	-0.008 (0.029)	-0.006 (0.019)
<i>_cons</i>	0.002 (0.320)	1.813*** (0.280)	0.166*** (0.036)	0.010 (0.024)
N	1,840	1,840	1,840	1,840

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

1. Standard errors are shown in parentheses

2. All estimates are robust to cluster fixed effects, controlling for interview characteristics (interviewer ID, day of team's visit).

**Table 4: Interaction effects with interviewer characteristics**

	TOTAL nr of alerts	TOTAL nr of alerts
CAPI	-10.366*** (0.424)	-30.967*** (4.865)
Checks	-1.217*** (0.337)	-5.048 (3.872)
PAPI experience (months)		-0.315*** (0.037)
CAPI experience (months)		0.199*** (0.065)
Formal education (years)		-1.833*** (0.252)
<i>CAPI interacted with:</i>		
PAPI experience (months)		0.284*** (0.052)
CAPI experience (months)		-0.188** (0.092)
Education (years)		1.526*** (0.356)
<i>Checks interacted with:</i>		
PAPI experience (months)		0.027 (0.042)
CAPI experience (months)		0.003 (0.073)
Education (years)		0.274 (0.283)
_cons	12.216*** (0.300)	37.015*** (3.444)
<i>N</i>	1,840	1,840

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

1. Standard errors are shown in parentheses

2. All estimates are robust to cluster fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy), household characteristics (No. of household members) and interview characteristics (conducted on first day of team's visit).

3. Total Nr. of alerts is the sum of routing errors and impossible and unlikely entries.

**Table 5: Interaction effects with survey period**

	TOTAL nr of alerts	TOTAL nr of alerts
CAPI	-10.366*** (0.424)	-15.657*** (0.782)
Checks	-1.217*** (0.337)	-1.528** (0.616)
Survey days 10-18		-5.723*** (0.784)
Survey days 19-27		-8.509*** (0.817)
Survey days 28-37		-9.522*** (0.817)
<i>CAPI interacted with:</i>		
Survey days 10-18		5.568*** (1.107)
Survey days 19-27		7.878*** (1.149)
Survey days 28-37		8.433*** (1.156)
<i>Checks interacted with:</i>		
Survey days 10-18		0.156 (0.879)
Survey days 19-27		0.418 (0.908)
Survey days 28-37		0.756 (0.922)
_cons	12.216*** (0.300)	17.941*** (0.558)
<i>N</i>	1,840	1,840

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

1. Standard errors are shown in parentheses

2. All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy) and household characteristics (No. of household members) and interview characteristics (conducted on first day of team's visit).

3. Total Nr. of alerts is the sum of routing errors and impossible and unlikely entries.

**Table 6: Summary statistics on calorific intake per AEU per day (per HH)**

		Full CAPI	Restricted CAPI	PAPI		
Item – unit combinations in food consumption section (%)	No problem	98.4	98.0	56.8		
	Ambiguous	1.6	2.0	42.0		
	Non valid	0.0	0.0	1.2		
<b>Unit size assumption</b>		<b>N/A</b>	<b>N/A</b>	<b>Small</b>	<b>Medium</b>	<b>Large</b>
Calorific intake per AEU per day (Kcal) <sup>4</sup>	Mean	2,297 (655)	2,471 (1,177)	2,478 (1,644)	3,145 (2,038)	4,362 (3,379)
	Min	413	235	232	232	232
	Max	5,117	10,528	18,969	20,181	23,057
Outliers in the distribution of calorific intake (Kcal) per AEU per day (%)	< 1000	1.8	3.4	3.8	1.9	1.3
	> 4000	0.8	7.8	7.2	19.7	35.4
Malnutrition: % of households under threshold of calorific intake (Kcal) per AEU per day. <sup>5</sup>	<1,500	8.3***	14.4**	20.6		
	<1,800	23.1**	29.4	29.4		
	<2,200	47.5	50.3	51.9		
N		1200	320	318	318	318

- Notes:
- Standard errors are shown in parentheses
  - Calorific intake values include meals taken outside of the household by household members. AEU positively adjusted for nr of meals taken by guests in the household.
  - N/A = Not applicable
  - Two PAPI observations with calorific intake values based on small unit assumption over 23,570 Kcal per AEU per day were excluded from the analysis to avoid these outliers driving the results.
  - Asterisks indicate *p*-values for a one-sided *t*-test whether value is significantly different compared to PAPI.  
\*\*\* *p*<0.01, \*\* *p*<0.05, \* *p*<0.1

**Table 7: Consumption, Poverty and Inequality**

				significance of <i>t</i> -test		
				(1) = (2)	(2) = (3)	(1) = (3)
	<b>(1) Full CAPI</b>	<b>(2) Restricted CAPI</b>	<b>(3) PAPI</b>			
Total Consumption per AEU mean and (standard deviation)	435,251 (222,881)	475,889 (279,349)	546,223 (337,538)	***	***	***
Poverty Headcount	83.0	74.6	68.3	***	*	***
Gini (95% CI)	.24 (.22-.25)	.29 (.26-.32)	.30 (.27-.32)	**	=	***
Income elasticity of food consumption	.89 (.01)	.95 (.02)	.94 (.02)	*	=	***
<i>N</i>	1200	319	319			

- Notes:
1. Standard errors are shown in parentheses; Standard errors of the Gini coefficients were calculated using the jackknife method;
  2. Calorific intake values include meals taken outside of the household by household members. AEU positively adjusted for nr of meals taken by guests in the household;
  3. N/A = Not applicable;
  4. Two PAPI outlier observations with total consumption over 3,000,000 are dropped;
  5. Asterisks indicate *p*-values for a one-sided *t*-test whether value is significantly different compared to PAPI. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$  and =  $p \geq 0.1$ ;
  6. PAPI food consumption is calculated using the small unit assumption;
  7. Income elasticities are those of the interaction term between log total consumption and a dummy indicating which sample the observations is from in a regression that includes only the 2 samples being tested;

**Table 8: Estimates of Equation (3): Attenuation Bias**

	No. of years of schooling (children aged 7-14)	Schooling expenditures on school-going children <sup>4</sup>	Child slept under a treated bednet night before survey (children aged 0-14)
	Age FE	Age FE	LPM Age FE
CAPI ( $\beta$ )	-0.451 (2.316)	39,915 (28,616)	-0.035 (0.629)
Checks ( $\gamma$ )	-3.808** (1.842)	-82,896*** (22,376)	-1.963*** (0.509)
Log total consumption per aeu ( $\delta$ )	0.204 (0.135)	8,000*** (1,676)	0.134*** (0.036)
Interaction of log total consumption per aeu with:			
CAPI ( $\phi$ )	0.054 (0.180)	-2,958 (2,219)	0.004 (0.049)
Checks ( $\rho$ )	0.292** (0.144)	6,477*** (1,745)	0.154*** (0.040)
<i>N</i>	2,683	2,137	5,148

*p*-value of *F*-test  
( $\delta + \phi + \rho = 0$ )

0.00

0.00

0.00

Notes: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

1. Standard errors are shown in parentheses;

2. All regressions control for age FE;

3. All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics (age, head of household, sex, literacy) and household characteristics (No. of household members) and interview characteristics (conducted on first day of team's visit).

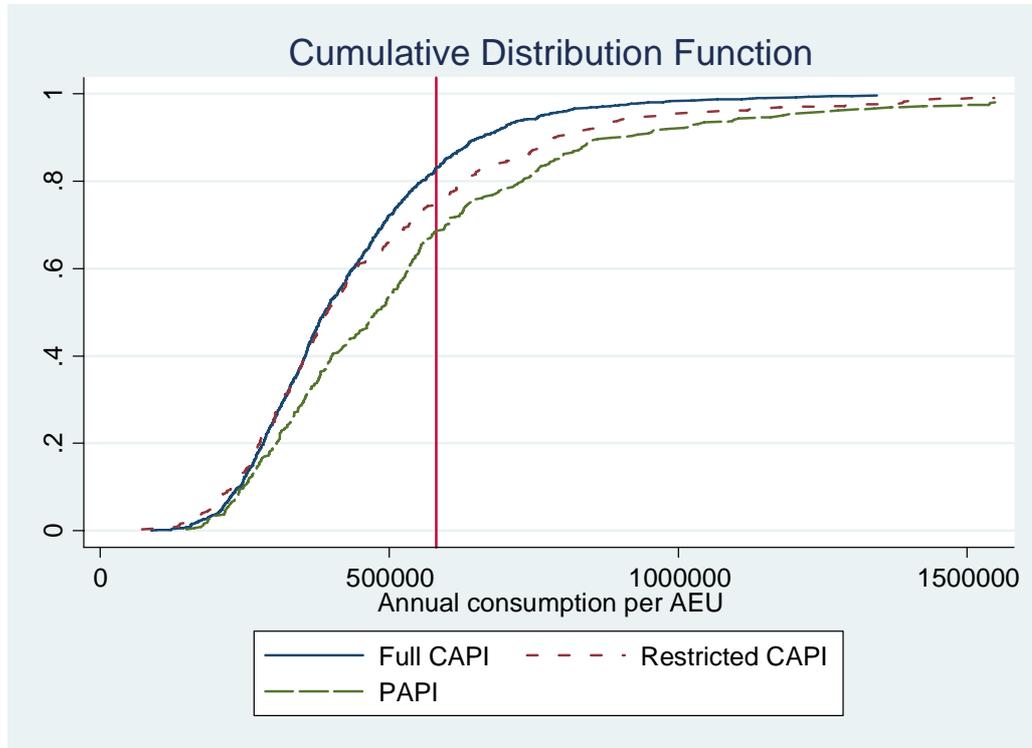
4. To avoid spurious correlation in the results of column 2, we exclude the education component from the consumption aggregate used in that regression.

**Table 9: Implications of measurement error in the dependent variable**

	Log total consumption/aeu		Log total consumption/aeu	
	Final set of regressors		Initial set of regressors	
	(1)	(1)	(2)	(2)
CAPI	-0.203	(0.196)	-0.279	(0.209)
Checks	0.055	(0.154)	0.124	(0.163)
Household size	-0.077***	(0.011)	-0.075***	(0.011)
Dependency ratio (adults 15-65 years old)	0.020	(0.023)	0.021	(0.023)
Dummy = 1 if household head is female	-0.233***	(0.066)	-0.212***	(0.067)
Average education (years) adults (15+)	0.018**	(0.008)	0.015*	(0.008)
Number of days head ill in last 4 weeks	-0.007**	(0.003)	-0.007**	(0.003)
Dummy = 1 if HH owns its house	0.105	(0.079)	0.092	(0.079)
Dummy = 1 if house has robust roof	0.085*	(0.049)	0.063	(0.050)
Proportion of employed adults in the HH	0.211*	(0.128)	0.227*	(0.130)
Acres of land owned	0.028*	(0.015)	0.029*	(0.015)
Dummy = 1 if head <i>not</i> employed in agriculture			0.062	(0.052)
Number of productive assets owned by the HH			0.056	(0.037)
Number of livestock owned by the HH			-0.000	(0.002)
<b>CAPI interacted with:</b>				
Household size	0.010	(0.015)	0.002	(0.015)
Dependency ratio (adults 15-65 years old)	0.015	(0.034)	0.013	(0.033)
Dummy = 1 if household head is female	0.068	(0.088)	0.098	(0.089)
Average education (years) adults (15+)	0.012	(0.011)	0.012	(0.012)
Number of days head ill in last 4 weeks	0.007	(0.004)	0.007	(0.004)
Dummy = 1 if HH owns its house	-0.055	(0.112)	-0.048	(0.111)
Dummy = 1 if house has robust roof	-0.028	(0.069)	-0.046	(0.069)
Proportion of employed adults in the HH	-0.109	(0.184)	-0.095	(0.187)
Acres of land owned	0.020	(0.023)	-0.003	(0.023)
Dummy = 1 if head <i>not</i> employed in agriculture			0.020	(0.072)
Number of productive assets owned by the HH			0.047	(0.050)
Number of livestock owned by the HH			0.017***	(0.005)
<b>Checks interacted with:</b>				
Household size	0.008	(0.011)	0.016	(0.011)
Dependency ratio (adults 15-65 years old)	-0.023	(0.027)	-0.024	(0.027)
Dummy = 1 if household head is female	-0.048	(0.066)	-0.061	(0.067)
Average education (years) adults (15+)	-0.007	(0.009)	-0.009	(0.010)
Number of days head ill in last 4 weeks	-0.002	(0.003)	-0.002	(0.003)
Dummy = 1 if HH owns its house	-0.053	(0.089)	-0.052	(0.087)
Dummy = 1 if house has robust roof	0.036	(0.054)	0.064	(0.054)
Proportion of employed adults in the HH	-0.073	(0.147)	-0.067	(0.150)
Acres of land owned	-0.023	(0.019)	-0.003	(0.019)
Dummy = 1 if head <i>not</i> employed in agriculture			0.000	(0.057)
Number of productive assets owned by the HH			-0.078**	(0.039)
Number of livestock owned by the HH			-0.012***	(0.004)
<i>cons</i>	13.242***	(0.143)	13.164***	(0.151)
<i>N</i>	1,837		1,837	

Notes: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Standard errors are shown in parentheses; Both equations control for cluster fixed effects; Robust roof materials are considered iron, concrete/cement, tiles or asbestos; Robust to different measures of household health (e.g. average health household members) and household educational level (e.g. average level of education of all HH members); The choice of measures used in table 9 was based on minimizing the difference in means of the measures between CAPI and PAPI to avoid attenuation bias affecting the results; Three observations with outliers in land area owned are dropped from the sample.

**Figure 1: Cumulative Distribution Function**



Notes: 15 values over 1,600,000 are not displayed on the graph for a better presentation. The vertical line presents the poverty line.

**Appendix 1: Most common missing values (missing + erroneous entries).**

**Table A1.1: 15 most common missing values outside the consumption section in PAPI**

Section/Level	Validation Check Message	% of surveys where error occurred at least once			Average freq. of error in surveys where error occurred		
		Full CAPI	Restricted CAPI	PAPI	Full CAPI	Restricted CAPI	PAPI
Start & finish/HH	Error in interview duration calculation	0.9%	0.3%	23.8 %			
Agriculture/ Crop Item	Missing: Over the past year, how much did you spend on transport in order to sell the products from [Crop Item]?	0.0%	0.0%	9.4%	0.0	0.0	1.4
Health/ HH Member	Missing: In the past 5 years, has [Household Member] given birth to children (including children dead)?	0.1%	3.1%	8.1%	1.0	1.0	1.1
Education/ HH Member	Missing: How much was spent by the members of your household on [Household Member]'s transport to/from school?	0.0%	0.0%	7.2%	0.0	0.0	1.3
Agriculture/ Crop Item	Missing: In which of those 2 places (i.e. selling places) do you fetch the highest price per unit of [Crop Item].	0.0%	0.3%	6.9%	0.0	1.0	1.4
GPS/HH	Error with GPS location	0.6%	1.3%	6.6%			
Education/ HH Member	Missing: Has [Household Member] ever attended school?	0.0%	0.0%	6.3%	0.0	0.0	1.3
Health/ HH Member	Missing: Is [Household Member] permanently physically or mentally disabled in any way which limits or prevents normal daily activities or work?	0.0%	0.0%	5.6%	0.0	0.0	2.4
Education/ Transport Item	Missing: What is the one way fare to go to school using [Transportation Mode]?	0.0%	0.3%	5.6%	0.0	1.0	1.3
Agriculture/HH	Missing: Do you own any agricultural land/farm (including grazing and fallow land)?"	0.0%	0.0%	5.3%			
Agriculture/ Crop Item	Missing: Rank the most important crops for generating cash income	0.0%	0.0%	4.7%	0.0	0.0	1.6
Demographics/ HH Member	Missing: What is [Household Member]'s marital status?	0.0%	0.6%	4.4%	0.0	1.0	1.6
Education/ HH Member	Missing: Can [Household Member] read and write?	0.0%	0.0%	4.4%	0.0	0.0	1.0
Demographics/ HH Member	Missing: Is [Household Member] an actual member of this household (satisfying some membership criteria, such as being present at the household for at least 9 out of 12 past months)?	0.0%	0.0%	4.1%	0.0	0.0	1.2
Health/ HH Member	If the number of days person was too ill to perform his/her normal daily activities in the past 4 weeks is greater than 0, the answer to "In the past 12 months was [Household Member] too ill to perform his/her normal daily activities?" cannot be "no"..	0.0%	0.0%	3.8%	0.0	0.0	1.5

**Table A1.2: 10 most common missing values in consumption section in PAPI**

Section/Level	Validation Check Message	% of surveys where error occurred at least once			Average freq. of error in surveys where error occurred		
		Full CAPI	Restricted CAPI	PAPI	Full CAPI	Restricted CAPI	PAPI
Consumption/ Food Item	The total amount of consumption of this food item differs from the sum ('How much came from purchases' + 'How much came from own production' + 'How much came from gifts and other sources')	0.1%	2.8%	17.2 %	1.0	1.0	1.5
Consumption/ Food Item	Missing: How much [Food Item] came from 'gifts and other sources' in the past [Recall Period]?	0.0%	5.0%	14.4 %	0.0	1.0	1.3
Consumption/ Food Item	Missing: How much [Food ITEM] came from 'own production' out of what was spent in the last [Recall Period]?	0.0%	3.4%	10.3 %	0.0	1.0	1.2
Consumption/ Food Item	Missing: UNIT of 'How much [Food Item] did your HH consume in the past [Recall Period]'?	0.0%	0.0%	10.0 %	0.0	0.0	1.2
Consumption/ Food Item	Missing: How much [Food Item] came from 'purchases' out of what was spent in the last [Recall Period]?	0.0%	0.6%	7.2%	0.0	1.0	1.9
Consumption/ HH Member	Missing: How much expenditure information for [Household Member] is not captured in what you have mentioned to me?	0.1%	0.9%	6.3%	1.0	1.3	2.0
Consumption/ Food Item	Missing: In the past 7 days did household consume any [Food Item]?	0.0%	9.1%	5.6%	0.0	1.3	3.8
Consumption/ Food Item	Missing: How much [Food Item] did your household consume in the past [Recall Period]?	0.0%	0.3%	5.0%	0.0	1.0	1.0
Consumption/ Non Food Item	Missing: In the past [Recall Period] did household consume/purchase any [Non Food Item]?	0.0%	5.6%	4.1%	0.0	1.2	1.1
Consumption/ Food Item	The household has consumed [Food Item], but no source for obtaining this item (Purchased/Production/gifts) has been selected	0.0%	0.0%	1.9%	0.0	0.0	1.0

## APPENDIX 2: Respondent's perception, per survey method<sup>1</sup>

Question	Response options	Full CAPI <sup>1</sup> (%)	Restricted CAPI (%)	PAPI (%)
What did you think of the duration of the interview? <sup>1,2</sup>	Short (1)	26.3	35.1	37.5
	Long (2)	73.7	64.9	62.5
	Standard deviation	0.4	0.5	0.5
Did you enjoy participating in the interview?	Yes (1)	6.8	9.1	10.2
	No (2)	93.2	90.9	89.8
	Standard deviation	0.3	0.3	0.3
How smooth did you find the interview terms of flow of questions?	Bad (1)	1.2	2.0	1.0
	Normal (2)	29.3	30.2	27.3
	Rather good (3)	61.7	60.1	63.5
	Very good (4)	7.7	7.8	8.3
	Standard deviation	0.6	0.6	0.6
Do you believe your answers will get used for policy making?	Yes (1)	88.8	87.8	87.4
	No (2)	11.2	12.2	12.6
	Standard deviation	0.3	0.3	0.3
We are not talking about this specific interview, but do you think that results of these types of surveys are generally reliable?	Not reliable at all (1)	2.5	2.9	2.2
	Rather reliable (2)	90.7	91.2	90.5
	Very reliable (3)	6.8	5.8	7.0
	Standard deviation	0.3	0.3	0.3
Do you believe that the information you provided in the interview is 100 % confidential?	Yes (1)	95.1	98.1	96.1
	No (2)	4.9	2.0	3.8
	Standard deviation	0.2	0.1	0.2
If we went through the survey again, do you think any answers would change?	Yes (1)	25.0	20.2	20.1
	No (2)	75.0	79.8	79.9
	Standard deviation	0.4	0.4	0.4
Did you feel comfortable talking to the interviewer?	Not at all (1)	4.3	2.0	3.5
	A little (2)	34.0	27.6	32.5
	Very much (3)	61.7	70.5	64.0
	Standard deviation	0.6	0.5	0.6
How nervous did you feel during the interview?	Not at all (1)	86.7	90.9	86.7
	A little (2)	11.1	8.1	9.8
	A lot (3)	2.2	1.0	3.5
	Standard deviation	0.4	0.3	0.5
How difficult did you find the questions? <sup>3</sup>	Easy (1)	38.3	29.6	35.7
	Difficult (2)	61.7	70.5	64.3
	Standard deviation	0.5	0.5	0.5
If we were not recording the answers (just talking to you), do you think you would have answered anything differently?	Yes (1)	29.6	31.8	27.7
	No (2)	70.4	68.2	72.3
	Standard deviation	0.5	0.5	0.5
Would you participate to this survey again next year?	Yes (1)	95.4	97.0	98.1
	No (2)	3.1	1.7	1.3
	Maybe (3)	1.5	1.3	0.6
	Standard deviation	0.3	0.3	0.2

1. Full CAPI covered more modules than restricted CAPI and PAPI and hence had a longer interview time, which may affect responses (See Section 2 for details).

2. Answers grouped as Short = 'short' + 'just fine'; Long = 'Long' + 'Too long' (with no change to results)

3. Answers grouped as Easy = 'Very easy' + 'Rather easy'; Difficult = 'Rather difficult' + 'Very difficult' (with no change to results)

## ON-LINE APPENDICES

The following appendices are not intended to be printed on paper, but could be available on-line.

### Online Appendix 1: Screen shots of CAPI

Screen shot of section 2: HH head info

**T** Household **540107/1001** Roster   

1. Start 2. HH head info 3. Roster 4. Various (1) 5. Consumption data 6. Various (2) 8. Finish 9. Comments

1. Name of HH head (as per listing)?

2. What is the name of the head of this HH?

3. Where was the head raised? Read all responses

4. In which region was the head raised?

5. In which district was the head raised?

Screen shot of section 3: Roster

**T** Household **540107/1001** Roster   

1. Start 2. HH head info 3. Roster 4. Various (1) 5. Consumption data 6. Various (2) 8. Finish 9. Comments

[Instructions] 5. Roster number of the main respondent

1. ID	2. Name	3. Sex	4. Age	Demo-graphics	Education	Health
1	<input type="text" value="Bernard Bunanka"/>	Male [1] <input type="text" value=""/>	<input type="text" value="23"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>
2	<input type="text" value="Mary Bunanka"/>	Female [2] <input type="text" value=""/>	<input type="text" value="21"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>
3	<input type="text" value="John Bunanka"/>	<input type="text" value=""/> Male [1] Female [2]	<input type="text" value="2"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>	<input type="button" value="a"/> <input type="button" value="b"/> <input type="button" value="c"/>

**On-line Appendix 2: List of all validation checks used in full CAPI experiment  
(grouped by section and by warning type)**

<b>Section</b>	<b>Type</b>	<b>Warning message</b>
Agriculture	Error	Entry in disabled question 2a: How much land do you farm? AREA
Agriculture	Error	Duplicate crops have been selected
Agriculture	Error	Entry in disabled question 5: Do you have a title deed/offer letter for any land that you own?
Agriculture	Error	Entry in disabled question 4a: What is the total area you own? AREA
Agriculture	Error	Entry in disabled question 12: Did you or any other HH member use any manure on your farm in the past 12 months?
Agriculture	Error	Entry in disabled question 10: Did you or any other HH member irrigate any of your fields in the past 12 months?
Agriculture	Error	Entry in disabled question 13: Did you or any other HH member use any hybrid seeds on your farm in the past 12 months?
Agriculture	Error	Entry in disabled question 14: Have you or any other HH member spoken to a government agricultural extension officer in the past 12 months?
Agriculture	Error	Entry in disabled question 4b: What is the total area you own? AREA UNIT
Agriculture	Error	Entry in disabled question 9: Rank the most important crops for generating CASH income - sold for cash (1= most important; 3 = least important)
Agriculture	Error	Entry in disabled question 8: For what purpose is this crop? READ ALL RESPONSES
Agriculture	Error	Entry in disabled question 2b: How much land do you farm? AREA UNIT
Agriculture	Error	Entry in disabled question 6b: which are the 3 most important crops or other agricultural activities this household depends on most? CROP DESCRIPTION
Agriculture	Error	Entry in disabled question 11: Did you or any other HH member use any chemical fertilizer on any of your fields in the past 12 months?
Agriculture	Missing	Missing question 8: For what purpose is this crop? READ ALL RESPONSES
Agriculture	Missing	Missing question 9: Rank the most important crops for generating CASH income - sold for cash (1= most important; 3 = least important)
Agriculture	Missing	Missing question 11: Did you or any other HH member use any chemical fertilizer on any of your fields in the past 12 months?
Agriculture	Missing	Missing question 14: Have you or any other HH member spoken to a government agricultural extension officer in the past 12 months?
Agriculture	Missing	Missing question 10: Did you or any other HH member irrigate any of your fields in the past 12 months?
Agriculture	Missing	Missing: Does anyone in this household conduct farming activities?
Agriculture	Missing	Missing question 5: Do you have a title deed/offer letter for any land that you own?
Agriculture	Missing	Missing question 12: Did you or any other HH member use any manure on your farm in the past 12 months?
Agriculture	Missing	Missing question 6b: which are the 3 most important crops or other agricultural activities this household depends on most? CROP DESCRIPTION
Agriculture	Missing	Missing question 13: Did you or any other HH member use any hybrid seeds on your farm in the past 12 months?
Agriculture	Missing	Missing question 4a: What is the total area you own? AREA
Agriculture	Missing	Missing question 2a: How much land do you farm? AREA
Agriculture	Missing	Missing: Do you own any agricultural land/farm (including grazing and fallow

Section	Type	Warning message
		land)?
Agriculture	Missing	Missing question 2b: How much land do you farm? AREA UNIT
Agriculture	Missing	Missing question 4b: What is the total area you own? AREA UNIT
Amenities	Error	The entry for the number of rooms is not allowed to be negative
Amenities	Error	Entry in disabled question 6c: How many habitable rooms in THE OTHER DWELLINGS does this household occupy? DO NOT COUNT BATHROOMS, TOILETS, STOREROOMS, OR GARAGE
Amenities	Error	Entry in disabled question 5: Do you have any documentation of ownership of the dwelling?
Amenities	Error	Entry in disabled question 11: How long does it take to get water from drinking water source to this dwelling in the dry season? GO AND RETURN TRIP INCLUDE WAITING TIME - MINUTES
Amenities	Error	Entry in disabled question 8: How long does it take to get water from drinking water source to this dwelling in the rainy season? GO AND RETURN TRIP INCLUDE WAITING TIME - MINUTES
Amenities	Error	Entry in disabled question 12: Out of these [READ] minutes, how long do you spend waiting?
Amenities	Error	Entry in disabled question 9: Out of these [READ] minutes, how long do you spend waiting? MINUTES
Amenities	Missing	Missing question 8: How long does it take to get water from drinking water source to this dwelling in the rainy season? GO AND RETURN TRIP INCLUDE WAITING TIME - MINUTES
Amenities	Missing	Missing: How many habitable rooms in THE MAIN DWELLING does this household occupy? DO NOT COUNT BATHROOMS, TOILETS, STOREROOMS, OR GARAGE
Amenities	Missing	Missing: What material is the floor of this house made of?
Amenities	Missing	Missing question 6c: How many habitable rooms in THE OTHER DWELLINGS does this household occupy? DO NOT COUNT BATHROOMS, TOILETS, STOREROOMS, OR GARAGE
Amenities	Missing	Missing: What material is the roof of this house made of?
Amenities	Missing	Missing question 11: How long does it take to get water from drinking water source to this dwelling in the dry season? GO AND RETURN TRIP INCLUDE WAITING TIME - MINUTES
Amenities	Missing	Missing: What is the major fuel used for cooking?
Amenities	Missing	Missing question 12: Out of these [READ] minutes, how long do you spend waiting?
Amenities	Missing	Missing: What is the tenure status of the main residence? READ ALL RESPONSES
Amenities	Missing	Missing: Is there any other dwelling that the HH uses?
Amenities	Missing	Missing: What is the main type of toilet used by this HH?
Amenities	Missing	Missing: What material are the walls of this house made of?
Amenities	Missing	Missing: What is the household's main source of drinking water in the dry season?
Amenities	Missing	Missing: What is the main source of energy used for lighting?
Amenities	Missing	Missing: What is the household's main source of drinking water in the rainy season?
Amenities	Missing	Missing question 5: Do you have any documentation of ownership of the dwelling?
Amenities	Missing	Missing question 9: Out of these [READ] minutes, how long do you spend

Section	Type	Warning message
		waiting? MINUTES
Amenities	Warning	It is uncommon to have electricity in a thatch roofed house. Please verify whether this information is correct.
Amenities	Warning	It is uncommon to have electricity in a mud roofed house. Please verify whether this information is correct.
Amenities	Warning	It is uncommon to have electricity in a wood roofed house. Please verify whether this information is correct.
Amenities	Warning	It is uncommon to have such a high number of rooms in a dwelling, are you sure the given information is correct? Please verify whether this information is correct.
Amenities	Warning	It is uncommon to have such a high number of rooms in a dwelling, are you sure the given information is correct? Please verify whether this information is correct.
Amenities	Warning	It is uncommon to have tiles in a mud walled house
Amenities	Warning	It is uncommon to have tiles in a wooden walled house. Please verify whether this information is correct.
Assets	Error	Entry in disabled question 2: How many (total quantity) FUNCTIONING [ITEM] does your household own?
Assets	Missing	Missing question 2: How many (total quantity) FUNCTIONING [ITEM] does your household own?
Assets	Missing	Missing: Asset code
Assets	Missing	Missing: Do you, or anyone else in your household, own a functioning [ASSET]?
Assets	Warning	Are you sure the household owns a TV antenna or Video/DVD but does not own a TV?
Assets	Warning	HH has at least one HH member whose main activity is fishing. It is unlikely that someone with main activity self-employed fishing does not own any fishing equipment at all (neither lantern, nor boat/canoe, nor other fishing equipment). Please verify.
Assets	Warning	None of the HH members slept under a mosquito net last night (T3CQ14) although the HH does claim to have at least one mosquito net (T4BQ1). This is possible, but please verify.
Assets	Warning	At least one HH member slept under a mosquito net last night (T3CQ14) although the HH claims not to have any mosquito nets (T4BQ1). This is possible, but please verify.
Assets	Warning	The HH's main source of lighting is said to be candles (T4AQ14). Hence it is uncommon that the HH owns this asset. It is possible, however. Please double check.
Assets	Warning	HH has at least one HH member whose main activity is farming. It is unlikely that someone with main activity self-employed farming does not own any farming equipment at all (whealbarrow/cart, nor harrow/plough, nor other farming equipment). Please verify.
Consumption Data	Error	Respondent to the consumption questions should not be less than 12 years of age
Consumption Data	Error	The answer to "What is the FIRST main source of cash income?" is not allowed to be
Consumption Data	Error	The FIRST and SECOND main source of cash income is not allowed to be the same
Consumption Data	Missing	Missing: Respondent on consumption questions
Consumption Data	Missing	Missing: What are your HH's 2 main sources of cash income, starting with the most important one? FIRST
Consumption Data	Missing	Missing: What are your HH's 2 main sources of cash income, starting with the

Section	Type	Warning message
		most important one? SECOND
Details of Missing Information Based on HH Roster	Missing	Missing: How much expenditure information for [NAME?] is not captured in what you have mentioned to me?
Details of Missing Information Based on HH Roster	Warning	It is uncommon that NONE of the consumption expenditures is captured in information given to you. Note that this question is in NEGATION form. Are you sure the given information is correct?
Finish	Error	Entry in disabled question 6: Why is the interview only partially completed?
Finish	Error	Entry in disabled question 2: How proficient was the respondent in Swahili?
Finish	Missing	Missing: Date and time of interview finish
Finish	Missing	Missing: Number of visits required to complete the interview
Finish	Missing	Missing question 2: How proficient was the respondent in Swahili?
Finish	Missing	Missing question 6: Why is the interview only partially completed?
Finish	Missing	Missing: Interview result
Food Consumption Details	Error	Entry in disabled question 1a: How much [ITEM] did your HH CONSUME in the past [PERIOD]?
Food Consumption Details	Error	Entry in disabled question 1b: How much [ITEM] did your HH CONSUME in the past [PERIOD]?
Food Consumption Details	Error	Entry in disabled question 3: How much did you spend?
Food Consumption Details	Error	The value HOME PRODUCED quantity is not allowed to be negative
Food Consumption Details	Error	Entry in disabled question 2a: How much [ITEM] came from PURCHASES out of what was spent in the last [PERIOD]?
Food Consumption Details	Error	Entry in disabled question 4a: How much [ITEM] came from OWN PRODUCTION out of what was spent in the last [PERIOD]?
Food Consumption Details	Error	Entry in disabled question 5: How much would it have cost you if you had purchased this HOME PRODUCED quantity in the main market or store in this village?
Food Consumption Details	Error	The entry for the consumed quantity that came from purchases is not allowed to be negative (except for -99)
Food Consumption Details	Error	Entry in disabled question 6a: How much came from GIFTS AND OTHER SOURCES in the past [PERIOD]?
Food Consumption Details	Error	The household has consumed this item, but no source for obtaining this consumption item (Purchased/Produced/Other sources) has been selected
Food Consumption Details	Error	The entry for the consumed quantity that came from own production is not allowed to be negative (except for -99)
Food Consumption Details	Error	The entry for the consumed quantity is not allowed to be zero
Food Consumption Details	Error	The entry for the consumed quantity that came from gifts and other sources is not allowed to be negative (except for -99)
Food Consumption Details	Error	The total amount consumed has to be greater than zero for this consumption item given the information entered by you in question field T6eaQ2.
Food Consumption Details	Error	The value GIFT quantity is not allowed to be negative
Food Consumption Details	Error	The entry for the consumed quantity is not allowed to be negative

Section	Type	Warning message
Food Consumption Details	Error	The total amount of consumption must be equal to (How much came from purchases + How much came from own production + How much came from gifts and other sources)
Food Consumption Details	Error	Entry in disabled question 7: How much would it have cost you if you had purchased this GIFT quantity in the main market or store in this village?
Food Consumption Details	Missing	Missing question 2a: How much [ITEM] came from PURCHASES out of what was spent in the last [PERIOD]?
Food Consumption Details	Missing	Missing question 1b: How much [ITEM] did your HH CONSUME in the past [PERIOD]?
Food Consumption Details	Missing	Missing question 1a: How much [ITEM] did your HH CONSUME in the past [PERIOD]?
Food Consumption Details	Missing	Missing question 7: How much would it have cost you if you had purchased this GIFT quantity in the main market or store in this village?
Food Consumption Details	Missing	Missing question 6a: How much came from GIFTS AND OTHER SOURCES in the past [PERIOD]?
Food Consumption Details	Missing	Missing question 4a: How much [ITEM] came from OWN PRODUCTION out of what was spent in the last [PERIOD]?
Food Consumption Details	Missing	Missing question 3: How much did you spend?
Food Consumption Details	Missing	Missing question 5: How much would it have cost you if you had purchased this HOME PRODUCED quantity in the main market or store in this village?
Food Consumption Details	Warning	It is very unlikely that a household in Tanzania cooks with such a small amount of cooking oil. Are you sure the given information is correct?
Frequent Non-food Expenditures (1)	Missing	MISSING: Expenditure
Frequent Non-food Expenditures (1)	Warning	It is uncommon to have an expenditure value of less than 100 TSH. Are you sure the given information is correct?
Frequent Non-food Expenditures (2)	Missing	MISSING: Expenditure
Frequent Non-food Expenditures (2)	Warning	It is uncommon to have an expenditure value of less than 100 TSH. Are you sure the given information is correct?
HH Crop details	Error	Second most common location should be different from first most common location
HH Crop details	Error	The location where the HH fetches the highest price for this crop should be either the location mentioned in Q1 or the one in Q3.
HH Crop details	Error	Entry in disabled question 6: What is the total cash value of the sale of [CROP] by this household during the last year? IN TSHS
HH Crop details	Error	Entry in disabled question 5: In which of those two places do you fetch the highest price per unit of [CROP]?
HH Crop details	Error	Entry in disabled question 1: Over the past 12 months, where did the household sell most of its produce of [CROP]? MOST COMMON LOCATION
HH Crop details	Error	Entry in disabled question 3: Over the past 12 months, where did the household sell most of its produce of [CROP]? SECOND MOST COMMON LOCATION
HH Crop details	Error	Entry in disabled question 2: How do you transport [CROP] to the location where it is sold? MOST COMMON LOCATION
HH Crop details	Error	Entry in disabled question 4: How do you transport [CROP] to the location where it is sold? SECOND MOST COMMON LOCATION
HH Crop details	Error	Entry in disabled question 7: Over the past year, how much did you spend on

Section	Type	Warning message
		transport in order to sell the products from [CROP] (all destinations)?
HH Crop details	Missing	Missing question 1: Over the past 12 months, where did the household sell most of its produce of [CROP]? MOST COMMON LOCATION
HH Crop details	Missing	Missing question 2: How do you transport [CROP] to the location where it is sold? MOST COMMON LOCATION
HH Crop details	Missing	Missing question 7: Over the past year, how much did you spend on transport in order to sell the products from [CROP] (all destinations)?
HH Crop details	Missing	Missing question 4: How do you transport [CROP] to the location where it is sold? SECOND MOST COMMON LOCATION
HH Crop details	Missing	Missing question 6: What is the total cash value of the sale of [CROP] by this household during the last year? IN TSHS
HH Crop details	Missing	Missing question 5: In which of those two places do you fetch the highest price per unit of [CROP]?
HH Crop details	Missing	Missing question 3: Over the past 12 months, where did the household sell most of its produce of [CROP]? SECOND MOST COMMON LOCATION
HH Head Info	Error	Entry in disabled question 5: In which district was the head raised?
HH Head Info	Error	Entry in disabled question 4: In which region was the head raised?
HH Head Info	Missing	Missing: What is the name of the head of this HH?
HH Head Info	Missing	Missing question 5: In which district was the head raised?
HH Head Info	Missing	Missing question 4: In which region was the head raised?
HH Head Info	Missing	Missing: Where was the head raised? READ ALL RESPONSES
Household Member - Demographics	Error	Entry in disabled question 2: What is [NAME] marital status?
Household Member - Demographics	Error	Entry in disabled question 3: Is the spouse of [NAME] living in household?
Household Member - Demographics	Error	Entry in disabled question 6: Do you expect that [NAME] will be residing here in 6 months from now?
Household Member - Demographics	Error	Household member has ID 1 but it is not the household head
Household Member - Demographics	Error	Only household member with ID 1 can be household head
Household Member - Demographics	Error	The relationship of [NAME] is Wife/Husband, although the marital status is not married or no non-formal union
Household Member - Demographics	Error	Entry in disabled question 4: Who is [NAME] husband?
Household Member - Demographics	Error	Entry in disabled question 7: What is [NAME] main daily activity?
Household Member - Demographics	Error	The person selected as being the husband of [NAME] is the head of the household, whilst the relationship to head selected in Q1 differs from 'spouse'.
Household Member - Demographics	Missing	Missing question 2: What is [NAME] marital status?
Household Member - Demographics	Missing	Missing: For how long was [NAME] absent during the last 12 months?
Household Member - Demographics	Missing	Missing question 3: Is the spouse of [NAME] living in household?
Household Member - Demographics	Missing	Missing question 6: Do you expect that [NAME] will be residing here in 6 months from now?

Section	Type	Warning message
Household Member - Demographics	Missing	Missing question 7: What is [NAME] main daily activity?
Household Member - Demographics	Missing	Missing question 4: Who is [NAME] husband?
Household Member - Demographics	Missing	Missing: What is the relationship of [NAME] to the head of the household?
Household Member - Demographics	Warning	This person is very old to be in boarding school. Are you sure this information is correct? Please verify.
Household Member - Demographics	Warning	It is very rare that a person this young is retired. Please verify.
Household Member - Demographics	Warning	It is unusual that a person is over 7 years old and neither a full-time student nor performing any type of work. Please verify.
Household Member - Demographics	Warning	[NAME] is very young for this activity. Are you sure this is accurate?
Household Member - Demographics	Warning	Person is said to be in boarding school (T3AQ5) but its main occupation is said to be different than 'student'. Please verify.
Household Member - Demographics	Warning	A child younger than 12 years old is unlikely to have the relationship to head that is selected. Are you sure that the given information is correct?
Household Member - Demographics	Warning	It is very uncommon that a child of less than 6 years old is not present all year long. Are you sure the given information is correct?
Household Member - Demographics	Warning	The man is a polygamist, hence it is unlikely that he is present all year long, unless all his wives live in the same household. Please verify.
Household Member - Demographics	Warning	It is uncommon that the head of the household or his/her spouse/husband (T3AQ1) is a boarding school child (T3AQ5). Please double check this information.
Household Member - Education	Error	Entry in disabled question 21: What is the ONE-WAY fare to go to school using this mode of transportation (in Tshs)?
Household Member - Education	Error	Entry in disabled question 19: How does [NAME] usually go to school?
Household Member - Education	Error	The answer to 'How many times has [NAME] repeated grades?' is not allowed to be negative
Household Member - Education	Error	Entry in disabled question 20b: MINUTES
Household Member - Education	Error	Entry in disabled question 22c: SCHOOL BOOKS AND MATERIALS
Household Member - Education	Error	Entry in disabled question 12: Has [NAME] successfully passed this exam?
Household Member - Education	Error	Entry in disabled question 10: Did [NAME] ever sit for a national examination from which results are out?
Household Member - Education	Error	Entry in disabled question 1: Can [NAME] read and write?
Household Member - Education	Error	Entry in disabled question 17: Has [NAME] missed school in the last schooling week?
Household Member - Education	Error	The entry for 'In total how much was spent on [NAME]'s education in the last 12 months ...?' is not allowed to be negative.
Household Member - Education	Error	Educational expenditures on OTHER CONTRIBUTIONS are not allowed to be negative
Household Member	Error	Educational expenditures on MEALS are not allowed to be negative

Section	Type	Warning message
- Education		
Household Member - Education	Error	Educational expenditures on FEES are not allowed to be negative
Household Member - Education	Error	Educational expenditures on EXTRA TUITION are not allowed to be negative
Household Member - Education	Error	Educational expenditures on BOOKS are not allowed to be negative
Household Member - Education	Error	The entry to 'what is the one way fare to go to school' is not allowed to be negative
Household Member - Education	Error	The last grade of examination cannot be higher than 1 grade higher than the highest level of completed education
Household Member - Education	Error	Entry in disabled question 2: Has [NAME] ever attended school?
Household Member - Education	Error	Entry in disabled question 18: Why was [NAME] absent from school?
Household Member - Education	Error	Entry in disabled question 4: How old was [NAME] when he/she started primary school?
Household Member - Education	Error	Entry in disabled question 20a: How long does it take [NAME] to go to school using this mode of transportation? HOURS
Household Member - Education	Error	Entry in disabled question 15: Was [NAME] in school in the last 12 months?
Household Member - Education	Error	The answer to 'how long does it take to go to school - MINS' is not allowed to be negative
Household Member - Education	Error	The entry to 'how long does it take to go to school - HOURS' is not allowed to be negative
Household Member - Education	Error	Entry in disabled question 22h: TOTAL
Household Member - Education	Error	Educational expenditures on TRANSPORT are not allowed to be negative
Household Member - Education	Error	Entry in disabled question 22d: UNIFORMS
Household Member - Education	Error	The age when he/she started school cannot exceed his/her current age
Household Member - Education	Error	Entry in disabled question 9: What is the year of [NAME] last grade repetition?
Household Member - Education	Error	Entry in disabled question 13: What division did [NAME] score on the examination?
Household Member - Education	Error	Educational expenditures on UNIFORMS are not allowed to be negative
Household Member - Education	Error	Entry in disabled question 11: For which level was the last examination that [NAME] took?
Household Member - Education	Error	Entry in disabled question 14: Is [NAME] currently in school?
Household Member - Education	Error	The last grade of repetition is not allowed to be 'none', because the answer to 'Has [NAME] ever repeated a grade?' is said to be 'yes'.
Household Member - Education	Error	Entry in disabled question 5: What is the highest level of COMPLETED education of [NAME]?

Section	Type	Warning message
Household Member - Education	Error	Entry in disabled question 8: What is the last grade [NAME] has repeated?
Household Member - Education	Error	Entry in disabled question 6: Has [NAME] ever repeated a grade?
Household Member - Education	Error	The number of times that [NAME] has repeated grade is not allowed to be 0, because the answer to 'Has [NAME] ever repeated a grade?' is said to be 'yes'.
Household Member - Education	Error	Entry in disabled question 7: How many times has [NAME] repeated grades?
Household Member - Education	Error	Entry in disabled question 22e: EXTRA TUITION
Household Member - Education	Error	Entry in disabled question 16: Who runs/manages school [NAME] is attending?
Household Member - Education	Error	Entry in disabled question 3: What type of school has [NAME] attended? READ ALL RESPONSES
Household Member - Education	Error	Entry in disabled question 22g: SCHOOL MEALS
Household Member - Education	Error	Entry in disabled question 22f: OTHER CONTRIBUTIONS FOR EDUCATION
Household Member - Education	Error	Entry in disabled question 22a: How much was spent in the last 12 months by the members of your HH on [NAME]: TRANSPORT TO/FROM SCHOOL
Household Member - Education	Error	Entry in disabled question 22b: SCHOOL FEES
Household Member - Education	Missing	Missing question 22d: UNIFORMS
Household Member - Education	Missing	Missing question 22a: How much was spent in the last 12 months by the members of your HH on [NAME]: TRANSPORT TO/FROM SCHOOL
Household Member - Education	Missing	Missing question 22f: OTHER CONTRIBUTIONS FOR EDUCATION
Household Member - Education	Missing	Missing question 1: Can [NAME] read and write?
Household Member - Education	Missing	Missing question 22g: SCHOOL MEALS
Household Member - Education	Missing	Missing question 22c: SCHOOL BOOKS AND MATERIALS
Household Member - Education	Missing	Missing question 22e: EXTRA TUITION
Household Member - Education	Missing	Missing question 7: How many times has [NAME] repeated grades?
Household Member - Education	Missing	Missing question 6: Has [NAME] ever repeated a grade?
Household Member - Education	Missing	Missing question 3: What type of school has [NAME] attended? READ ALL RESPONSES
Household Member - Education	Missing	Missing question 12: Has [NAME] successfully passed this exam?
Household Member - Education	Missing	Missing question 10: Did [NAME] ever sit for a national examination from which results are out?
Household Member - Education	Missing	Missing question 2: Has [NAME] ever attended school?

Section	Type	Warning message
Household Member - Education	Missing	Missing question 22h: TOTAL
Household Member - Education	Missing	Missing question 22b: SCHOOL FEES
Household Member - Education	Missing	Missing question 19: How does [NAME] usually go to school?
Household Member - Education	Missing	Missing question 20a: How long does it take [NAME] to go to school using this mode of transportation? HOURS
Household Member - Education	Missing	Missing question 20b: MINUTES
Household Member - Education	Missing	Missing question 21: What is the ONE-WAY fare to go to school using this mode of transportation (in Tshs)?
Household Member - Education	Missing	Missing question 17: Has [NAME] missed school in the last schooling week?
Household Member - Education	Missing	Missing question 5: What is the highest level of COMPLETED education of [NAME]?
Household Member - Education	Missing	Missing question 8: What is the last grade [NAME] has repeated?
Household Member - Education	Missing	Missing question 4: How old was [NAME] when he/she started primary school?
Household Member - Education	Missing	Missing question 15: Was [NAME] in school in the last 12 months?
Household Member - Education	Missing	Missing question 11: For which level was the last examination that [NAME] took?
Household Member - Education	Missing	Missing question 13: What division did [NAME] score on the examination?
Household Member - Education	Missing	Missing question 9: What is the year of [NAME] last grade repetition?
Household Member - Education	Missing	Missing question 14: Is [NAME] currently in school?
Household Member - Education	Missing	Missing question 16: Who runs/manages school [NAME] is attending?
Household Member - Education	Missing	Missing question 18: Why was [NAME] absent from school?
Household Member - Education	Warning	It is very uncommon that the answer to question '22h - Amount spent in the last 12 months by the HH on [NAME]'s education: TOTAL - AUTOMATICALLY CALCULATED BASED ON RESPONSES TO 21a-21g. NOT EDITABLE BY INTERVIEWER. ' is greater than 3000000. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Education	Warning	Main activity was recorded as full-time student but person is not currently in school? Please verify.
Household Member - Education	Warning	Given the entered age at which this person started education, it is uncommon that he/she has yet reached the stated 'highest level of education'. Please double check.
Household Member - Education	Warning	It is uncommon to have a repetition rate higher than 5. Are you sure the given information is correct?
Household Member - Education	Warning	A person of less than 12 years old is unlikely to have reached the identified level of education. Are you sure the given is correct?

Section	Type	Warning message
Household Member - Education	Warning	A person of less than 5 years old is unlikely to have reached the identified level of education. Are you sure the given information is correct?
Household Member - Education	Warning	A person of less than 16 years old is unlikely to have reached the identified level of education. Are you sure the given information is correct?
Household Member - Education	Warning	A person of less than 18 years old is unlikely to have reached the identified level of education. Are you sure the given information is correct?
Household Member - Education	Warning	It is very uncommon that the answer to question '20a - How long does it take [NAME] to go to school using this mode of transportation? HOURS ' is greater than 3. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Education	Warning	It is very uncommon that the answer to question '20b - How long does it take [NAME] to go to school using this mode of transportation? MINUTES ' is greater than 59. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Education	Warning	It is very uncommon that the answer to question '21 - What is the ONE-WAY fare to go to school using this mode of transportation (in Tshs)? ENTER ZERO IF NONE ' is greater than 150000. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Education	Warning	It is very uncommon that the answer to question '4 - How old was [NAME] when he/she started school? ' is greater than 11. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Education	Warning	It is very uncommon that the answer to question '4 - How old was [NAME] when he/she started school? ' is smaller than 6. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Education	Warning	The age when he/she started school is not allowed to be negative
Household Member - Education	Warning	The answer to 'How many times has [NAME] repeated grades?' is DK. Make sure a comment about this DK is made in the comment box of this question.
Household Member - Education	Warning	Last year's total transport costs to go to school for [NAME] are somehow low given the one way fare cost of the most commonly used transport method used by [NAME], see Q21. Please double check.
Household Member - Education	Warning	Only in special cases the last grade of repetition will be higher than 1 grade higher than the highest level of completed education. Are you sure the given information is correct?
Household Member - Education	Warning	It is uncommon to spend such a high amount on education of a child. Please double check whether the given information is correct.
Household Member - Education	Warning	It is uncommon that there is nothing spent at all on the child's education if it was in school in the last 12 months. Are you sure the given information is correct?
Household Member - Education	Warning	The answer to 'In total how much was spent on [NAME]'s education in the last 12 months ...?' is DK. A comment about this DK MUST be made.
Household Member - Health	Error	Entry in disabled question 20: Did [NAME] regularly go to a health clinic when he/she was pregnant with his/her last child?
Household Member - Health	Error	Entry in disabled question 18: Of the children who died, how many died before their first birthday?
Household Member - Health	Error	Contradictory information: As incapacitated was selected as [NAME]'s main daily activity, [NAME] must have a physical or mental disability that limits or prevents normal daily activities or work. Please correct either main daily activity in T3A or disability status in T3C.
Household Member	Error	Entry in disabled question 16: In the past 5 years, how many children did [NAME]

Section	Type	Warning message
- Health		give birth to (including children who were born dead)?
Household Member - Health	Error	the answer to 'How much did it cost?' is not allowed to be negative
Household Member - Health	Error	Entry in disabled question 23: Was this birth registered?
Household Member - Health	Error	Entry in disabled question 21: Where did [NAME] deliver his/her last child?
Household Member - Health	Error	Entry in disabled question 22: Who delivered this child?
Household Member - Health	Error	Entry in disabled question 5: For the last 4 weeks was [NAME] hospitalized or had overnight stay(s) in medical facility?
Household Member - Health	Error	Entry in disabled question 4: What was the most important kind of health provider that [NAME] visited?
Household Member - Health	Error	It is not possible that a male person has pregnancy related problems.
Household Member - Health	Error	Entry in disabled question 6: How was treatment mainly financed?
Household Member - Health	Error	Entry in disabled question 19: Of the children who died, how many died between their first and their fifth birthday?
Household Member - Health	Error	Entry in disabled question 11: Is [NAME] PERMANENTLY physically or mentally disabled in any way which limits or prevents normal daily activities or work?
Household Member - Health	Error	Entry in disabled question 12: What type of disability does [NAME] have?
Household Member - Health	Error	Entry in disabled question 13: How is the impact of [NAME] disability on his/her daily activities compared to 12 months ago?
Household Member - Health	Error	Entry in disabled question 15: In the past 5 years, has [NAME] given birth to children (including children born dead)?
Household Member - Health	Error	Entry in disabled question 3: For how many days in the last 4 weeks has [NAME] suffered from this main health problem?
Household Member - Health	Error	Entry in disabled question 10: Estimate the total number of days [NAME] was not able to perform his/her daily activities due to illness for the past 12 months?
Household Member - Health	Error	Entry in disabled question 9: In the past 12 months have there been any episodes in which [NAME] was too ill to perform his/her normal daily activities?
Household Member - Health	Error	Entry in disabled question 8: In the past 4 weeks, for how many days was [NAME] unable to perform his/her normal daily activities due to the illness/injury? ENTER 0 IF NONE
Household Member - Health	Error	Entry in disabled question 2: What was the main health problem [NAME] was suffering from (in the last 4 weeks)?
Household Member - Health	Error	Entry in disabled question 7: How much did it cost?
Household Member - Health	Error	Entry in disabled question 17: How many of those children are still alive now?
Household Member - Health	Error	The entry for "Estimate the total number of days in the past 12 months?" is not allowed to be negative.
Household Member - Health	Error	The entry for "Estimate the total number of days in the past 12 months?" cannot exceed 365 (i.e. the number of days in 1 year).
Household Member - Health	Error	The days ill in 4 weeks cannot exceed 28

Section	Type	Warning message
Household Member - Health	Error	The nr of days ill in the past 12 months cannot be smaller than the nr of days ill in the past 4 weeks
Household Member - Health	Error	Entry in disabled question 1: Was [NAME] sick or injured in the last 4 weeks?
Household Member - Health	Error	If the number of days the person was too ill to perform his/her normal daily activities in the past 4 weeks is greater than 0, the answer to "In the past 12 months ... too ill to perform his/her normal daily activities?" cannot be "no".
Household Member - Health	Error	Entry in disabled question 14: Did [NAME] sleep under a bednet yesterday?
Household Member - Health	Error	The number of children born in the past 5 years who are still alive, plus the number of children who died before their first birthday, plus the number of children who died between their first and fifth birthday have to add up to the total number of children given birth to.
Household Member - Health	Error	The entry for "For how many days was [NAME] suffering from this main health problem?" is not allowed to be negative.
Household Member - Health	Error	The entry for "In the past 4 weeks, for how many days was [NAME] unable to perform his/her normal daily activities due to the illness/injury?" is not allowed to be negative.
Household Member - Health	Error	The entry for "For how many days was [NAME] suffering from this main health problem?" is not allowed to be zero, since the respondent claimed that [NAME] did suffer from the disease.
Household Member - Health	Error	The entry for "In the past 4 weeks, for how many days was [NAME] unable to perform his/her normal daily activities due to the illness/injury?" is not allowed to be higher than 28 days (since it has to be within 4 weeks interval).
Household Member - Health	Missing	Missing question 10: Estimate the total number of days [NAME] was not able to perform his/her daily activities due to illness for the past 12 months?
Household Member - Health	Missing	Missing question 13: How is the impact of [NAME] disability on his/her daily activities compared to 12 months ago?
Household Member - Health	Missing	Missing question 15: In the past 5 years, has [NAME] given birth to children (including children born dead)?
Household Member - Health	Missing	Missing question 4: What was the most important kind of health provider that [NAME] visited?
Household Member - Health	Missing	Missing question 17: How many of those children are still alive now?
Household Member - Health	Missing	Missing question 19: Of the children who died, how many died between their first and their fifth birthday?
Household Member - Health	Missing	Missing question 16: In the past 5 years, how many children did [NAME] give birth to (including children who were born dead)?
Household Member - Health	Missing	Missing question 18: Of the children who died, how many died before their first birthday?
Household Member - Health	Missing	Missing question 9: In the past 12 months have there been any episodes in which [NAME] was too ill to perform his/her normal daily activities?
Household Member - Health	Missing	Missing question 3: For how many days in the last 4 weeks has [NAME] suffered from this main health problem?
Household Member - Health	Missing	Missing question 8: In the past 4 weeks, for how many days was [NAME] unable to perform his/her normal daily activities due to the illness/injury? ENTER 0 IF NONE
Household Member	Missing	Missing question 5: For the last 4 weeks was [NAME] hospitalized or had

Section	Type	Warning message
- Health		overnight stay(s) in medical facility?
Household Member - Health	Missing	Missing question 2: What was the main health problem [NAME] was suffering from (in the last 4 weeks)?
Household Member - Health	Missing	Missing question 6: How was treatment mainly financed?
Household Member - Health	Missing	Missing question 22: Who delivered this child?
Household Member - Health	Missing	Missing question 14: Did [NAME] sleep under a bednet yesterday?
Household Member - Health	Missing	Missing question 21: Where did [NAME] deliver his/her last child?
Household Member - Health	Missing	Missing question 20: Did [NAME] regularly go to a health clinic when he/she was pregnant with his/her last child?
Household Member - Health	Missing	Missing question 11: Is [NAME] PERMANENTLY physically or mentally disabled in any way which limits or prevents normal daily activities or work?
Household Member - Health	Missing	Missing question 23: Was this birth registered?
Household Member - Health	Missing	Missing question 12: What type of disability does [NAME] have?
Household Member - Health	Missing	Missing question 7: How much did it cost?
Household Member - Health	Missing	Missing question 1: Was [NAME] sick or injured in the last 4 weeks?
Household Member - Health	Warning	It is very uncommon that a person of this age has pregnancy related problems. Please double check whether the information you entered is the actual information given by the respondent.
Household Member - Health	Warning	Since [NAME] was hospitalised due to illness it is unlikely that the number of days he/she was unable to perform his/her normal daily activities due to illness is zero. Are you sure the given information is correct?
Household Member - Health	Warning	The number of children of the head that are born in last 5 years and that are still alive is smaller than the number of biological children counted in the household roster. It is very possible that the child is no longer a household member, but please double check this.
Household Member - Health	Warning	In T3BQ18 it is stated that the child was absent from school last week because of illness. However, T3CQ1 states that child has not been sick in last 4 weeks. Please double check.
Household Member - Health	Warning	It is very uncommon that the answer to question '7 - How much did it cost? IN TSH ' is greater than 300000. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Health	Warning	It is very uncommon that the answer to question '7 - How much did it cost? IN TSH ' is smaller than 500. Please double check whether the information you entered is the actual information given by the respondent
Household Member - Health	Warning	The person is said to be 'sick' or 'incapacitated' in T3AQ7, hence it is strange that the answer to 'sick in last 4 weeks?' is said to be 'no'. Please double check.
Less Frequent Expenditure	Missing	MISSING: Expenditure
Less Frequent Expenditure	Warning	It is uncommon to have an expenditure value of less than 100 TSH. Are you sure the given information is correct?

Section	Type	Warning message
Livestock	Error	Entry in disabled question 2: How many [LIVESTOCK TYPE] does the HH own TODAY?
Livestock	Error	The number of livestock owned is not allowed to be negative
Livestock	Missing	Missing: Do you, or anyone else in your household, own [LIVESTOCK TYPE]?
Livestock	Missing	Missing question 2: How many [LIVESTOCK TYPE] does the HH own TODAY?
Livestock	Missing	Missing: Code for type of animal
Livestock	Warning	It is uncommon to own such a high number of rabbits. Are you sure the given information is correct?
Livestock	Warning	It is uncommon to own such a high number of ducks/turkeys. Are you sure the given information is correct?
Livestock	Warning	It is uncommon to own such a high number of chickens. Are you sure the given information is correct?
Livestock	Warning	It is uncommon to own such a high number of cows. Are you sure the given information is correct?
Livestock	Warning	It is uncommon to own such a high number of sheep. Are you sure the given information is correct?
Livestock	Warning	It is uncommon to own such a high number of goats. Are you sure the given information is correct?
Meals taken by guests + Meals taken outside HH	Error	Entry in disabled question 3: How many meals did [NAME] miss in the last 7 days? (assume 2 meals per day; hence a number between 1 and 14)
Meals taken by guests + Meals taken outside HH	Error	The answer to question: '3 - Number of meals that [NAME] missed (between 1 and 14) ' should not be greater than 14.
Meals taken by guests + Meals taken outside HH	Error	The answer to question:'3 - Number of meals that [NAME] missed (between 1 and 14) ' should be at least 1.
Meals taken by guests + Meals taken outside HH	Missing	Missing: How many person-meals were consumed by guests over the last 7 days? (e.g. if 2 extra people shared 3 HH meals, enter 6 person meals): people aged 5-12
Meals taken by guests + Meals taken outside HH	Missing	Missing: How many person-meals were consumed by guests over the last 7 days? (e.g. if 2 extra people shared 3 HH meals, enter 6 SHARES): people aged 0-4)
Meals taken by guests + Meals taken outside HH	Missing	Missing: How many person-meals were consumed by guests over the last 7 days? (e.g. if 2 extra people shared 3 HH meals, enter 6 person meals): people aged 13-18
Meals taken by guests + Meals taken outside HH	Missing	Missing: How many person-meals were consumed by guests over the last 7 days? (e.g. if 2 extra people shared 3 HH meals, enter 6 person meals): people aged 19-59
Meals taken by guests + Meals taken outside HH	Missing	Missing: How many person-meals were consumed by guests over the last 7 days? (e.g. if 2 extra people shared 3 HH meals, enter 6 person meals): people aged 60+
Meals taken by guests + Meals taken outside HH	Missing	Missing question 3: How many meals did [NAME] miss in the last 7 days? (assume 2 meals per day; hence a number between 1 and 14)
Meals taken by guests + Meals taken outside HH	Missing	Missing: Did [NAME] miss any HH meals in the last 7 days?

Section	Type	Warning message
Meals taken by guests + Meals taken outside HH	Warning	It is very uncommon that the answer to question '3 - Number of meals that [NAME] missed (between 1 and 14) ' is greater than 14. Please double check whether the information you entered is the actual information given by the respondent
Meals taken by guests + Meals taken outside HH	Warning	It is very uncommon that the answer to question '3 - Number of meals that [NAME] missed (between 1 and 14) ' is smaller than 1. Please double check whether the information you entered is the actual information given by the respondent
Outside Food and Drink	Missing	Missing: Expenditure
Outside Food and Drink	Warning	It is uncommon to have an expenditure value of less than 100 TSH. Are you sure the given information is correct?
Roster	Error	Main respondent should not be less than 12 years of age
Roster	Error	HH head claims to be unmarried but there is at least one wife.
Roster	Error	Inconsistency between marital status of husband and wife
Roster	Error	HH head claims to have monogamous marriage but there is more than one wife.
Roster	Error	The household roster is empty.
Roster	Missing	Missing: Is this person a HH member?
Roster	Missing	Missing: Roster number of the main respondent
Roster	Missing	Missing: What was the age of [NAME] at last birthday (in completed years)? IF LESS THAN 1 YEAR, ENTER 0
Roster	Missing	Missing: Is [NAME] male or female?
Roster	Missing	Missing: HH member name
Roster	Missing	Missing: Member ID (automatically generated by ticking the "add new HH member" command button on form)
Roster	Warning	It is uncommon not to have any member older than 15 years old in a household. Are you sure the given information is correct?
Roster	Warning	Main respondent is not the head or spouse. A reason for this MUST be entered in the comments.
Roster	Warning	It is uncommon to have a person of this age. Are you sure the given information is correct?
Roster	Warning	The age difference between parent and child is less than 14 years. Please verify.
Roster	Warning	The same name appears more than once in the HH member roster. Are you sure the given information is correct?
Roster Sampling	Error	Entry in disabled question 2: Why the member is not available for interview
Roster Sampling	Missing	Missing: Member is available for interview
Roster Sampling	Missing	Missing: Has [NAME] been sampled for the interview of TRANSPORT section?
Roster Sampling	Missing	Missing question 2: Why the member is not available for interview
Roster Sampling	Missing	Missing: Has [NAME] been sampled for the interview of LABOUR section?
Select Consumption Items	Missing	Missing: In the past [RECALL] did household consume/purchase any [ITEM]?
Select Consumption Items	Missing	Missing: In the past [RECALL] did household consume/purchase any [ITEM]?
Start	Missing	Missing: Language of interview?
Start	Missing	Missing: Was an interpreter used?
Start	Missing	Missing: Date and time of interview start?
Start	Warning	An interpreter is used, hence a comment MUST be made about how smooth the interview was conducted, whether using an interpreter caused any difficulties, etc.

**On-line Appendix 3: Example of the consumption data summary report used by the enumerators whilst verifying data collected**

*Food energy consumption data*

<b>540201 Bahareni</b>				
<b>540201/0101</b>				
Item	Amount consumed over reporting period		Kcals/day	
0201 Cassava fresh	75	Small piece	4,457	38%
0102 Rice (husked)	4	Kilograms	2,040	17%
0106 Wheat flour	2	Kilograms	1,040	9%
0301 Sugar	1.8	Kilograms	964	8%
0504 Coconuts (mature)	7	Piece/Number	896	8%
0814 Fresh fish (excl. dagaa)	4	Medium fish	644	5%
0105 Maize (flour) Sembe	1	Kilograms	520	4%
0807 Chicken, other poultry and wild birds (incl. pigeons, kwarara, sabile)	2	Piece/Number	420	4%
0701 Ripe bananas	3	Hand	253	2%
0401 Dried beans	0.5	Kilograms	229	2%
0503 Coconuts (immature)	2	Piece/Number	84	1%
0110 Boflo	2	Piece/Number	77	1%
0811 Eggs	6	Piece/Number	68	1%
0818 Fresh dagaa	1	1/4 KG Tin	16	0%
0601 Tomatoes	6	Small piece	11	0%
0602 Onions	1	Small piece	1	0%
Total Kcalories per day			<b>11,720</b>	<b>99%</b>
Total number of Adult Equivalent Units (AEU) in the household			<b>4.72</b>	
Total Kcalories consumed per AEU per day			<b>2,506</b>	

**On-line Appendix 4: Examples of pictures used in the consumption section:**

