The Challenge of Measuring Hunger through Survey

Forthcoming in Economic Development and Cultural Change

Joachim De Weerdt (IOB, University of Antwerp and LICOS, KU Leuven)

Kathleen Beegle (World Bank)

Jed Friedman (World Bank)

John Gibson (University of Waikato)

July 2015

ABSTRACT

There is widespread interest in estimating the number of hungry people in the world as well as trends in hunger. Current global counts rely on combining each country's total food balance with information on distribution patterns from household consumption expenditure surveys. Recent research has advocated for calculating hunger numbers directly from these same surveys, which are increasingly available in low-income countries. For either approach, embedded in this effort are a number of important details about how household surveys are designed and how these data are then used. Using a survey experiment in Tanzania, this study finds great fragility in hunger counts stemming from alternative survey designs. As such, caution should be taken in drawing inferences on hunger over time and space based on household surveys.

JEL Codes: C88, O12, Q18

Keywords: hunger prevalence, measurement error, consumption, survey design

This work was supported by the World Bank Research Committee. We would like to thank Carlo Cafiero, two anonymous referees and seminar participants at the University of Wollongong, the sixth International Conference on Agricultural Statistics (ICAS-VI), the 11th Midwest International Economic Development Conference, the Oxford annual conference of the Centre for the Study of African Economies, the University of Antwerp, and the KU Leuven for their comments and suggestions. All views are those of the authors and do not reflect the views of their employers including the World Bank or its member countries.

1 INTRODUCTION

At the World Food Summit in 1996 leaders from 183 countries committed to halving the number of people living in hunger; a commitment they renewed in 2009. Halving the proportion of hungry people between 1990 and 2015 is also part of the Millennium Development Goals (MDGs). Yet, according to the flagship publication of the Food and Agriculture Organization (FAO) of the United Nations, *The State of Food Insecurity in the World (SOFI),* there were still 805 million hungry people worldwide in 2012-14, down from slightly over 1 billion in 1990-92, The same report says developing regions (who account for 98% of the global hunger count) saw the proportion of hungry people drop from 23.4% to 13.5% over the same period (FAO 2014). This slow progress in reducing hunger contrasts with faster progress in reducing extreme poverty, which is another target of the first MDG.

The FAO estimates total yearly energy availability per country from food balance sheets (FBS) and uses survey data to determine how that energy is spread across the population. This information is used to parameterise a distribution function from which the share of the population falling below calorie requirements is determined. The FAO method has been the subject of a heated debate with criticisms from, for example, Smith (1998), Svedberg (1999, 2002) and de Haen et al. (2011) and rebuttals by the FAO in Naiken (2007, 2014) and Cafiero (2014).

More recently the criticism has focussed on the perceived fragility of the FAO's hunger numbers. Moore Lappé et al. (2013) overlay the trends in hunger reported in the SOFI 2010 report with those reported in SOFI 2012. The 2010 graph shows an estimated spike in hunger prevalence around the 2007-08 food-price hikes, which is no longer present in the 2012 graph after some methodological revisions. Furthermore, the 2012 graph increased the 1990-92 hunger rates, which form the basis to measure progress towards the MDG hunger target. The change obviously impacted positively the degree of portrayed optimism about reaching the MDG hunger target. The FAO defended the technical necessity of the change and argued that the initial spike was a projection relying on tentative assumptions about the evolution of food prices in developing countries (which turned out not to hold).

In the ensuing debate on what, if anything, could be an alternative to the FAO method, several suggestions have been made, each with its share of proponents and critics: 24-hour diet recall and weighted food records, subjective questions on self-assessed hunger, anthropometric measurements, a revealed preferences approach, and calculating hunger numbers directly from the food quantity data in household consumption and expenditure surveys. The focus of this paper is on comparing the latter, which we will refer to in shorthand as the HCES method, with the FAO method. While we will make no attempt at any formal deliberation regarding the merits of the other three methods, we will, in the next section, for the sake of completeness, outline each of these alternatives and refer to work that does make such broader comparisons.

The reasons for our focus on the HCES method as an alternative to the FAO method are two-fold. First, there has been an enormous expansion of HCES in low income countries. From 1990 onwards, there are at least 760 nationally representative household consumption expenditure surveys (HCES) available for 129 developing countries.[1] These HCES are already being used to monitor global poverty trends (Chen and Ravallion 2010) and hold the promise of allowing global hunger counts to be derived from them too (Smith 1998, Smith and Subandoro 2007, Fiedler et al. 2012b). This paper attempts to formally assess how well HCES are able to measure hunger and suggests concrete, low-cost improvements to make them more suitable for estimating hunger statistics (Smith et al, 2014). The second reason is purely pragmatic: the survey experiment that forms the basis of this paper contains a wealth of HCES data, but no weighted food consumption data, no anthropometric measurements and no responses to questions on hunger experiences, which excludes careful, experimental comparisons of these other alternatives to the FAO method.

The FAO and HCES methods both rely on household surveys.[2] In the case of the FAO method, the second and higher moments of the calorie distribution come from the surveys, while the HCES method relies on the surveys for all moments. Yet, the design of HCES varies over several key dimensions around the world. These dimensions include the method of data capture (diary versus recall questionnaires), the level of respondent (individual versus household), the reference period for which consumption is reported (anywhere from 24 hours to one year), and the degree of commodity detail (from less than 20 items to over 400 items). This variation in survey design has the potential to affect the comparability and reliability of hunger statistics across countries and over time. In this study we explore the implications of survey design on estimates of the number of hungry people.

We explore a unique survey experiment which randomly assigned seven different HCES designs to 3,520 households in Tanzania. This experiment covered urban and rural settings, and reflects the range of Sub-Saharan environments where, according to the FAO, the proportion of hungry people is highest and

[1] We use the generic term HCES to refer to a range of household survey efforts to capture total household consumption expenditures. This can include surveys described as household budget surveys, living standards surveys, or others.

[2] Neither method actually measure what individuals ate (as in a food intake survey) or ask about their perceptions of hunger (Thompson and Byek 1994, Radimer et al. 1990).

increasing. Using data from this experiment, we calculate hunger figures ranging from 19 to 68 percent in the same villages at the same time, depending on the survey method. The features of the survey experiment, described below, ensure that any differences in derived hunger numbers are solely attributable to survey design. The experiment did not collect anthropometric or hunger perceptions data, so cannot inform on how they compare or, in turn, depend on the data collection method.

Our results suggest that any comparative assessment of hunger prevalence using HCES should clearly take into account differences in survey design. However, hunger numbers cannot be adjusted through a simple scale correction to account for the various design effects: the differential likelihood that a household is counted as hungry through one survey design and as not hungry through another survey design is correlated with the household's size, wealth, location (urban or rural), and the education of its head. Comparing hunger numbers across survey designs is therefore not trivial. Furthermore, the ranking of socio-economic or geographical groups by hunger prevalence within them – an exercise that may be carried out with survey data in order to inform the targeting of nutrition interventions – will depend in part on the survey design.

The sensitivity of hunger estimates to survey design variations is greater than for other statistics derived from HCES, such as poverty counts and inequality measures (Beegle et al. 2012). One reason for this is that surveys differ the most in the ways that they go about measuring food consumption, whereas modules devoted to non-food consumption tend to be more standardized. Therefore, we advocate for more effort to be spent on harmonizing food consumption survey design, in order to obtain comparable hunger numbers within and between countries and over time. The existing idiosyncratic variation in survey designs is not inherent to measuring food consumption by household survey. Some countries manage to maintain consistent designs while others oscillate for unclear reasons. Across countries, there has been some success at regional harmonization in household surveys in Latin America through regional statistical initiatives, although surveys in that region typically focus on incomes. In other areas of socio-economic statistics, efforts to harmonize survey methods have progressed further – for example, cross-country data on fertility and maternal and child health is much more comparable because of the standardized approach to measurement taken by the Demographic and Health Surveys (DHS).

The remainder of the paper is organized as follows. Section 2 will discuss the different methods commonly used to measure hunger, while Section 3 walks through a number of errors that can be expected when measuring hunger directly from HCES and how some of these errors likely differ by survey design. That same section also lists some low-cost improvements to HCES that could make them more pertinent for hunger measurement. Section 4 introduces the data and experimental set-up. Section 5 quantifies differences in hunger numbers across the various arms of the survey experiment and verifies whether the magnitude of these differences is orthogonal to household characteristics. Section 6 presents a concluding discussion on how our results can be used to inform debates about measuring hunger through household surveys.

2 METHODS OF MEASURING HUNGER

The most widely publicized method of calculating global hunger is the one used by the Food and Agricultural Organisation (FAO) of the United Nations in a series of reports tracking world hunger, with FAO (2014) being the latest. The FAO indicator was established to track progress towards the first Millennium Development Goal of halving poverty and hunger by 2015. It relies on the assumption that the level of food energy consumption for the average individual in the population follows a log-normal distribution or skew-normal distribution.[3] These distributions are parameterised by the mean, the coefficient of variation (CV) and, in the case of the skew-normal, the skewness. The FAO calculates the mean from Food Balance Sheets (FBS), adding national food production and imports and subtracting exports, food losses, food used for seeds, animal feed, and stock changes to calculate the total availability of food in a country. Combining this with population data and accounting for losses at the retail level allows the FAO to estimate the total kilo calories available for human consumption per person, per country in a particular year. The CV and skewness are calculated from a limited number of HCES. For most countries the CV was kept constant across years and only the mean was revised.[4] Finally, the FAO estimates a range of age-sex specific daily energy requirements under the assumption of 'light work' and minimally adequate body mass for height. The required energy of the average individual in a population is taken as the minimum of the normal requirement range. These numbers are aggregated to yield the requirement of an average person. The area underneath the log-normal energy distribution

[3] The skew-normal distribution generalises the normal distribution to allow for skewing (Azzalini, 1985).

[4] Smith (1998) reports that, at the time, for 18 out of the 99 countries, the CVs are estimated based on analysis of nationally representative HCES. The rest of the countries' CVs are predicted either from measures of income distribution or as the mean CV estimated for other countries in the same region. The CVs were then also assumed not to change over the twenty-year period for which undernourishment estimates are undertaken. In 2012 the FAO updated the CV estimate for 37 countries, and, for the other countries (where they did not obtain HCES), they used the same CV as in the past. In 2014 a new model for indirect estimates of the CV was introduced that linked CVs to macro-economic indicators.

which lies to the left of the energy requirement is the FAO's estimate of the proportion of the population with inadequate access to food.

The FAO method has been widely critiqued by the research community (Svedberg, 1999, de Haen et al. 2011, Smith and Subandoro 2007, Fiedler et al. 2012a) and defended by Naiken (2007, 2014) and Cafiero (2014). The debate mainly reflects concerns about the reliability of three components that go into the FAO's calculations: the mean, the spread and the requirement threshold.

With regard to the first two components, the FAO method relies on two different sources for these two moments of the distribution (the mean and the spread). In some ways, this is comparable to the lively discussion by Chen and Ravallion (2010) in the context of measurements of global poverty that rely on national accounts for the consumption mean and on household survey data for the variance. The FAO calculates the spread (CV) and more recently also the skewness of food availability directly for only a limited number of countries and updates are infrequent (See footnote 4). As such, differences in hunger estimates across countries and over time are mainly driven by FBS data. The concern is that this puts too much emphasis on food availability as opposed to access to food (Sen, 1981).

There are a range of concerns about the reliability of the FBS data. First, average food availability is a residual in the FBS-method, so any errors in reported production, trade, and stocks will affect the estimates of national food availability. Second, for grain crops the production and trade data are potentially reliable, since it is feasible to measure production with sample plots, with satellite and aerial mapping and so forth, but the same is not true for root crops (potatoes, sweet potatoes, and cassava are especially important food sources for the poor in some countries) whose yield cannot be observed remotely. Moreover, there are complex relationships between production and what is fed to animals and what is retained for seed, which affect the amount left over for human consumption. Studies suggest that there can be substantial errors in root crop food balance sheet data (Horton 1988). Finally, among the grain crops, storage data is especially problematic for rice, which is mostly privately stored (including by producers on farm), making it very difficult to ascertain the amount of rice available for human consumption in a particular period (Timmer 2009).

Regarding the third component, the hunger threshold, Svedberg (2002) has argued that there exists a positive correlation between per capita calorie intake and per capita calorie requirement across households, implying that in order to determine the share of the population that is hungry one needs to consider the joint distribution of availability and requirement, not a single cut-off point. Naiken (2007) defends the FAO approach and this issue continues to be a point of debate.

Methodological problems aside, the FAO method explicitly aims to compare only across countries, not within countries, and so does not help national governments determine which areas or population groups are at risk of hunger.

An alternative approach to the FAO method is to follow the lead of nutritionists, who typically rely on either an observed-weighed food method or a 24-hour recall survey (Gibson 2005). The latter uses multiple passes over the consumed items, which allows for better recollection and an increasing amount of data to be collected at each stage. There can also be a verification pass, in which the respondent is asked to confirm the answers recorded. For example, the respondent may start off by giving a broad overview of the food eaten in the past 24 hours, after which there is a more detailed description of each food (including preparation method and ingredients for prepared meals). Attention is given to quantifying the volume or weight of the consumed food with techniques such as pictures, food replicas, weighing, or volumetric estimation using local measures. Consumption by children is sought either from or in the presence of the main adult caretaker.

While 24-hour diet recall and weighed food records are trusted by nutritionists, such surveys are few in number and drawn from insufficiently representative samples to provide valid evidence on the prevalence or depth of hunger for entire populations over extended periods (Fiedler 2013). This is at least in part due to the fact that they are time consuming and therefore expensive to collect. There is only scant evidence on how they compare with HCES and other methods (Zo Rambeloson et al. 2012). At least three other less resource-intensive alternatives to the FAO approach have been suggested to derive hunger numbers: anthropometric data, self-assessments, revealed preferences and direct use of HCES.[5]

Anthropometric outcomes are highly correlated with energy intake and are an attractive alternative for the measurement of hunger. With just a few variables (age, sex, height and weight) a score can be calculated at the individual level. As a summary measure, anthropometrics will only yield the net effect of all macro- and micro-nutrients, as well as disease exposure, on growth and cannot inform on shortfalls in the consumption of individual nutrients. Further, just as an individual's basal metabolic rate

[5] For discussions on other measures, including composite indices, we refer to Barrett (2010), Masset (2011) and Jones et al. (2013)

varies, so does the nourished “benchmark” by which an individual should be assessed. Another challenge is to ensure that anthropometric standards sufficiently reflect genetic or gender variation, and to define indicators for and across all age ranges. In a review of the feasibility of international growth standards for children, Butte et al (2007, p.155) note that “it cannot be ruled out that some of the observed differences in linear growth across ethnic groups reflect true differences in genetic potential rather than the sole influence of environmental factors.” Another useful summary of the problems associated with the international reference standard is provided by Klasen (2008). It particularly seems to be Asiatic populations whose growth patterns reflect a different genetic potential, while too little is known about genetic potential of other some other populations, such as in the Pacific Islands (Ulijaszek, 1994). Consequently, anthropometric analysis in such settings may use internal growth reference standards (e.g, Mueller et al, 2001) which is fine for within-country targeting but highlights the problems of cross-country comparisons.

It is a priori unclear whether, all else equal, anthropometric data are cheaper or more expensive to collect than food quantity data. There are important details regarding survey implementation that are essential for reliable data, for example whether the child’s height is measured recumbent or standing, whether shoes are taken off, whether note is taken of clothing worn, the frequency with which anthropometric equipment is recalibrated and how recalibration services can be reached in rural areas. The fact that anthropometrics are individual-specific is a great advantage, but from a practical perspective also requires that all members of the household are present at the time of interview, which can substantially complicate survey logistics. Any interviews during work or school hours may lead to a systematically selected population with anthropometric data recorded, because only those not in school or not at work are available to be measured.

Despite these challenges anthropometric data are a useful and powerful tool for analysing hunger, and comparable data, mostly for children under 5, from the MICS and DHS surveys are already widely cited. At the same time there has been an explosion in the number of HCES offering a potential new source of data on hunger. HCES also expand the scope of what can be analysed and modelled beyond what is possible with MICS and DHS surveys. The latter are rich in data on child and maternal health and fertility, but limited in terms of data on incomes, prices, and other economic determinants of hunger. HCES are primarily designed to collect this information. While the two approaches – HCES and anthropometry – are clearly complementary, anthropometrics were not collected as part of the survey experiment analyzed here so we cannot make any formal comparisons in this paper.

The second alternative is to use data on self-reported concerns and experiences of inability to access food in adequate quantities or of adequate quality. The Gallup World Poll has previously asked 'Have there been times in the past 12 months when you did not have enough money to buy the food that you or your family needed?' (Headey 2013). The FAO's Voices of the Hungry project is developing a food insecurity experience scale (FIES), which entails eight questions about conditions experienced in the last 12 months (all yes or no responses). These questions range from whether the respondent has felt any anxiety about having enough food at any time during the previous 12 months to whether they felt hungry but didn't eat because there was not enough money or other resources for food. Beginning in 2014, the FIES was included in the Gallup World Poll and the FAO now also promotes national surveys to include the short module. Such brief questions are quicker and cheaper to collect than full HCES efforts. However, how well they correlate with other measures (like food consumption) is unclear. Migotto et al. (2007) analyze data from 4 countries and find that subjective perceptions of food consumption adequacy are, at best, weakly correlated with calorie consumption, dietary diversity, and anthropometric measures. Gunderson and Ribar (2011) investigate the correlation between food expenditures and two widely-used self-reported food hardship indicators in the U.S. They conclude that there are serious concerns about the external validity of the self-reported measures. Our study did not collect subjective perception data and so cannot contrast such data with HCES derived hunger estimates.

The revealed preferences approach conjectures that hungry people will have very different consumption patterns compared to those who are not hungry. The extreme physical and psychological discomfort of feeling hungry raises the marginal utility of additional calories dramatically, such that those food items with the lowest cost per calorie will dominate the consumption basket. Only once hunger is satisfied will concerns such as palatability come into play. Jensen and Miller (2010) provide a strong theoretical foundation for this method and suggest that the percentage of calories consumed from the staple food source, the cheapest source of calories, is a good indicator. Heady and Ecker (2013) review a number of food security indicators and argues that dietary diversity measures hold the strongest potential for comparisons over time and space. They argue that such measures could be embedded at fairly low cost into existing national household surveys, such as HCES and DHS. More research is clearly needed to validate these methods and look at, for example, the possibility that preferences in certain areas are such that the traditional diet remains the preferred diet even at relatively high levels of wealth. If this is the case then an increase in quantity consumed would not come with a change in diet composition. Questions can also be raised about its appropriateness in contexts where hunger suppressants, such as betel nut, qat or coca leaves are in wide use. Furthermore, it is unclear how much potential there is to use these methods to look at other micro-nutrient deficiencies (Ruel et al. 2012). While there is clearly a promising correlation between dietary diversity and food security, it is unclear whether the threshold needed to distinguish those who are hungry from those who are not can always be readily identified.

The fourth approach, and the one that we focus on in this paper, is to use information on food quantities from HCES to calculate calorie consumption and derive hunger statistics (Svedberg 1999, de Haen et al. 2011, Smith 1998, Smith et al. 2006). The arguments are compelling. HCES are positioned between the single subjective hunger question and the intensive 24-hour recall. They are the work-horse of monetary poverty measurement (in regards to global poverty estimates, see Chen and Ravallion 2010) and are now also ubiquitous – with an explosion in availability across developing countries in the last two decades.

Because of these reasons, the HCES approach has been suggested as an alternative to the FAO's method. The remainder of this paper will therefore be dedicated to juxtaposing the FAO and HCES methods, and considering their robustness to variations in survey design. We shed light on these issues through our survey experiment, but before moving on to a description of the experiment, the next section briefly discusses which sources of error we should expect in HCES and why we should expect a subset of these errors to depend on survey design.

3 POTENTIAL SOURCES OF ERROR IN HCES

While the potential usefulness of HCES for measuring hunger is apparent, there are several recording, processing, and analytic steps to take before the entries from a HCES can be converted into meaningful measures of the adequacy of caloric intake for a survey sample or before the CV needed for the FAO method can be calculated.[6] This section outlines these steps and, for each one, discusses the errors that can be introduced in the resulting measurement of calories.

Understanding the sources of such errors is important since some errors plausibly depend on survey design in non-trivial ways. These design effects matter because the extent of variation in survey design across countries – and even within countries as statistical agencies modify questionnaires over time – is extremely large. Fiedler et al. (2012b) present a useful list of various HCES in low and middle income countries highlighting substantial differences in their design across a select number of dimensions. Smith

[6] Smith and Subandoro (2007) provide a practical guide.

et al. (2014) have designed a very detailed metadata survey that tries to categorise HCES in all their relevant dimensions (requiring a 22-page form to cover all variations). Drawing on surveys from 100 different countries, it is apparent that there is large variation in terms of survey mode (diaries vs. recall), in terms of the length of the food item list and across the recall periods used. For example, while the modal recall period for food used by the surveys covered in their study is 7 days, this is used in only 31% of the cases.

Our experiment informs primarily on reporting errors (the first group of errors discussed below), but it is reasonable to assume that variation in the other types of errors will have similar effects.

3.1 Reporting Error

Reporting error occurs when the information relayed by the respondent to the interviewer is not accurate. Perhaps the most common error in this category is recall error, such as a householder under-reports true consumption over the period of recall due to faulty memory. Presumably the longer the period of recall, the greater the cognitive demand on the respondent and the greater the divergence between reported and actual consumption. Several studies have documented that, all else equal, the longer the period of recall, the lower the reported consumption per standardized unit of time. Closely related to recall error is telescoping, where a household compresses consumption that occurred over a longer period of time into the reference period asked and thus reports consumption greater than the actual value. A third important source of reporting error is the inability to accurately capture individual consumption by household members that occurs outside the purview of the survey respondent. Clearly this inability may be more significant for certain types of food, such as snacks or meals taken outside the home. The degree of inaccuracy is likely to increase with the number of adult household members and with the diversity of their activities outside the home as typically there is only one survey respondent per household.

We can expect diaries to suffer less from recall or telescoping errors, since consumption is able to be recorded soon after it occurs, although the extent to which diaries are supervised will remain important to ensure they are filled in frequently. Unsupervised diaries may end up being effectively like self-administered recall modules with endogenous recall periods if some types of respondents do not fill them in every day. Diaries administered at the individual level should also be better at capturing the individual consumption outside the household, whereas such inaccuracies may persist in household-level diaries.

Other sources of reporting error with no obvious direction of bias include rounding error and cognitive errors that result from consideration of hypothetical consumption constructs such as questions about consumption in a “usual” month. This type of question may present additional cognitive demands compared to a definitive recall period in the immediate past. There can be intentional misreporting in the light of respondent fatigue. Whether the respondent is presented with a long or a short list of consumption items may therefore influence the quality of the responses.[7] Finally, misreporting may arise from social desirability bias. The respondent may wish to exaggerate or understate her consumption in order to appear poorer or richer, perhaps due to a belief that the responses given may determine eligibility for some future program.

Consequently, HCES with different methods of data capture (diary versus recall questionnaires), levels of respondent (individual versus household), recall period or degree of commodity detail may not be comparable. The survey experiment we use in this paper was designed in part to assess the extent to which variation across these dimensions alter measures of calorie availability.

3.2 Survey Design Error

The appropriateness of HCES data for nutritional assessments is frequently hampered by survey design. There are small design changes that are unlikely to impact the survey costs, but could provide substantial improvements to ensure that the resulting data are as useful as possible for nutritional assessments. Two examples are given below and we refer interested readers to a more comprehensive treatment of this issue in Smith et al. (2014).

First, many HCES omit details on meals consumed outside the home by household members (at most collecting expenditure, but not the quantity information that is needed to estimate calorie content). Conversely, meals within the household that were shared with non-household members are not typically enumerated, and may wrongly get treated as being eaten by the householders. The frequency with which family members eat outside the home or share household meals with outsiders is unlikely to be orthogonal to household characteristics, such as wealth. The possible pitfalls of failing to account for inter-household transfers of food have been well recognized in the literature. For example Bouis (1994) shows how this can bias estimates of the income elasticity of calories, which summarizes the variation in calorie availability or intake across income groups.

[7] Beegle et al. (2012) find a reduction from 49 to 41 minutes when reducing the list of food consumed within the household over the past 7 days from 58 to 17 items. Interview times increased to 76 minutes when the list had 58 items and the more cognitively demanding ‘typical month’ phrasing was used in the question.

Second, HCES sometimes ask about food acquisition rather than consumption. As food stocks are consumed and replenished, what was acquired over the recall period may not be an accurate reflection of food consumption.[8] Deaton and Zaidi (2002), the most common reference for designing HCES and calculating consumption aggregates, explicitly recommend probing for food consumed and not food acquired and to record food from all sources, including meals taken outside the house.

3.3 Interviewer or Data Entry Error

Intentional error could also stem from interviewers subtly guiding respondents to give answers that minimize interview length, or who rush to complete the questionnaire. We can assume that such errors become more likely as questionnaires get longer and if supervision is limited. Extensive enumerator training and field supervision should minimize these errors. The questionnaires or diary booklets need to be key-punched into a computer by data entry clerks. This is a particularly tedious process and a potential source of mistakes. While the advent of computer-assisted personal interviewing (CAPI) holds the potential to reduce such errors (Caeyers et al. 2012) in the short-run, for a number of reasons paper-based surveys will continue to be a common method used in low-income countries.

3.4 Unit Conversion Error

Throughout most of the developing world, households do not typically purchase, harvest, or consume their food in standardized units (kilograms or litres). Some surveys force reporting in standardized units (an example would be the Tanzania National Panel survey) but there are doubts about the accuracy of these reports when made by people who never transact in metric units. A typical HCES consumption module will allow the respondent to report in local units, such as bunches, heaps, tins, buckets, or bundles. In order to quantify food consumption, local units must be converted into standardized units. If cheap calorie-rich staple foods are more likely to be reported in non-standardized units (pieces of cassava, bunches of bananas) and conversion factors are inadequate, then unit conversion error could significantly distort the resulting calorie estimates.

One relatively low-cost addition to existing HCES efforts would be to ensure that such conversion factors, which are often geographically specific, are systematically collected by survey teams.

[8] Gibson and Kim (2012) use a HCES with direct measures of consumption from food stocks and find an error of up to 300 KCal per person per day from ignoring destocking of one major calorie-source (rice) that is subject to bulk buying and storage.

3.5 Food Heterogeneity Error

Having obtained estimates of weight or volume of each food item consumed, pre-existing food tables are then used to determine energy content. This happens in two steps. First the edible proportion of the food item is estimated, subtracting from the total the expected amount of stems, peels, bones or other inedible parts of the food. Then, the energy content of the edible part of the food can be calculated. For this study we relied on Lukmanj et al. (2008) and USDA (2002) to calculate the edible portion within each quantity and the amount of energy it contained, expressed in KCal. An important feature here is the level of specificity of food item in the list. To the extent that the list of food items contains grouped foods, calories may be measured with error.[9]

3.6 Errors in Calorie Requirement

Once the food reported in an HCES is converted into calories, the household's calorie intake is compared to its need. This estimation presents another potential source of error. James and Schofield (1990) and FAO/WHO/UNU (1985) note how energy requirements will depend on a wide range of factors such as metabolism, age, gender, weight, height, activity level and for women on whether or not they are breastfeeding. While HCES will capture the demographic composition of the household (age and gender of household members), other information is often not collected (whether women in the household are pregnant or breastfeeding, body weight, height and levels of physical activity).

4 THE SURVEY EXPERIMENT

While the potential sources of mismeasurement in HCES are numerous, this study systematically explores the net effect of questionnaire design on arguably the major category of error, reporting error, using a survey experiment conducted in Tanzania. There were a total of seven alternate designs, which differ by method of data capture, level of respondent, length of reference period, number of items in the recall list, and nature of the cognitive task required of the respondent. These alternative designs were randomly assigned to a national sample of over 4,000 households. Modules 1-4 are recall designs and modules 5-7 are diaries (Table 1).[10] The seven designs were strategically selected to reflect the most

[9] Behrman and Deolalikar (1987) find that estimates of nutrient availability based on expenditure patterns can be upwardly biased when there is a high level of aggregation in food groups in expenditure surveys. Households replace cheap calories with more expensive ones as income goes up -- even though this extra expenditure yields no increase in calories (instead increasing attributes like taste and convenience). This increased expenditure can wrongly be equated with more calories if this replacement happens within a food group since quantities (and calories) rise at a lower rate than expenditures.

[10] The survey experiment included an 8^{th} module design which we excluded from the current analysis as it did not capture food quantities.

common methods utilized in low income countries and are informative of the kind of variation one is likely to find in the type of consumption and expenditure surveys used in the countries where concerns about hunger are most pressing.

In the food recall modules, households report the quantity consumed from three sources (purchases, home production, and gifts/payments). Modules 1 and 2 contain a list of 58 food groups; module 3 has a subset list that consists of the 17 most important food groups that constitute, on average, 77 percent of food consumption expenditure in Tanzania based on the Household Budget Survey 2000/01. To make module 3 comparable, we scale up calorie availability for that module (by 1/0.77). Among the recall modules, module 4 deviates from a reporting of actual expenditure over a specified time period. Instead it asks for "usual" consumption, following a recommendation in Deaton and Grosh (2000), where households report the number of months in which the food item is typically consumed by the household, the quantity usually consumed in those months, and the average value of what is consumed in those months. These questions aim to measure permanent rather than transitory living standards, without interviewing the same households repeatedly throughout the year. Hence, module 4 introduces two key differences from the other recall modules: a longer time frame and a different (and more complicated) cognitive task required of respondents.

The three diary modules are of the standard "acquisition type." Specifically, they add everything that came into the household through harvests, purchases, gifts, and stock reductions and subtract everything that went out of the household through sales, gifts, and stock increases. Modules 5 and 6 are household diaries in which a single diary is used to record all household consumption activities. The two household diaries differed by the frequency of supervision that each received from trained survey staff. The infrequent diary received supervisory visits weekly while the frequent diary was supervised every other day.

Module 7 is a personal diary, where each adult member keeps their own diary while children were placed on the diaries of the adults who knew most about their daily activities. Diary entries are specific to an individual and should leave no scope for double-counting purchases or self-produced goods. It is possible that a "gift" could be given to the household and accidentally recorded by two individuals. However, interviewers were trained to cross-check individual diaries for similar items purchased, produced, or gifted that occur on the same day and to query these during the checks. In many cases, one person will acquire food for the household (such as buying 5 kilograms of rice), which is entered in the diary of the person acquiring the food. So the personal diary is a not an individual's record of food consumption. Rather, it records the food brought into the household by each member even if for several members to consume (as well as food consumed outside the household). Each individual respondent with a diary was supervised every other day. This intensive supervision of the personal diary sample would be impractical for most surveys but these investments were made in order to establish a benchmark for analytic comparisons. We view module 7 as akin to a 24-hour food-intake approach, not only because of the intensity of supervision but also because of the detailed cross-checks on meals to check for food in-flows and out-flows that were otherwise missed. Module 7 arguably provides the most accurate estimate of actual food consumption and calorie availability.

The field work was conducted from September 2007 to August 2008 in villages and urban areas from seven districts across Tanzania: one district from each of the regions of Dodoma, Pwani, Dar es Salaam, Manyara, and Shinyanga and two districts in the Kagera Region. The districts were purposively selected to capture variations between urban and rural areas as well as across other socio-economic dimensions to inform survey design related to labor statistics and consumption expenditure for low-income settings. In each district, 24 communities were randomly selected from the 2002 Census, with probability-proportional-to-size (PPS). Within communities, a random sub-village (enumeration area, EA) was chosen and all households therein were listed. A total of 21 households per sub-village were randomly selected to participate and three households were randomly assigned to each of the seven modules. Among the original households selected for the survey and assigned to a module, there were 13 replacements due to refusals. Three households that started a diary were dropped because they did not complete their final interview. Another five households were dropped due to missing data on some of their key household characteristics, yielding a final sample size of 3,520 households.[11]

The basic characteristics of the sampled households generally match those from the nationally representative 2006-07 Household Budget Survey (the comparison results are not presented here but are available from the authors upon request). The randomized assignment of households to the seven different questionnaire variants appears successful in terms of balance across characteristics relevant for consumption and consumption measurement when examining a set of core household characteristics (Beegle et al. 2012).

[11] It is worth noting that we have almost no item non-response in the consumption section of the recall modules, i.e. all respondents answer all questions for all consumption items.

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

The survey was administered on paper. To minimize data entry errors, all questionnaires were entered twice and discrepancies were adjudicated. As non-standard units are common in Tanzania, the experiment collected conversion factors during a community price survey conducted by the field supervisors in each sample community. Supervisors used a food weighing scale to obtain a metric value of food-specific non-standard unit combinations. Median district-level metric conversion rates were used to convert non-metric units into kilograms or litres. Where district-level conversion rates were not available, the sample median was used and where this was not available, measurements at the survey's headquarters were taken after the fieldwork was done.[12] More details on the experiment are described by Beegle et al. (2012) who use the same experiment to compare consumption, poverty and inequality numbers across the different methods.

The food quantity estimates were transformed into food energy availability using food composition tables (Lukmanj et al. 2008). The total food energy available was converted to per capita daily averages, adjusting for meals eaten out of the home and meals shared with non-household members.

[12] See Capéau and Dercon (2005) for an econometric approach that can be used when direct measurements are not available.

5 RESULTS

We utilize our survey experiment to explore the survey design implications for measuring hunger. The next two sections consider, respectively, the HCES method that is advocated by numerous researchers and the FAO method. Since the HCES modules we use are typical of those found in low-income countries where concerns about hunger are most apparent, and our survey setting is also typical of these conditions, the results should be broadly informative about the degree of sensitivity of hunger statistics to variation in survey design. The FAO method, however, does not derive solely from survey data, but our survey experiment can still inform on sensitivity of this measure in two dimensions of variation: variations in the CV (which are derived from surveys) and variations in the mean (which are derived from FBS).

5.1 The HCES method

In order to derive hunger numbers from HCES we follow the comprehensive guidelines from Smith et al. (2006) and Smith and Subandoro (2007). The previous section already described how food quantity availability at the household level was calculated. That availability is contrasted against the energy required for basal metabolic function and light activity of all household members, adjusted for age, gender, and the likelihood a woman is breastfeeding. We do not have data on body weight, height, or level of physical activity. Using the recommended daily intake values from Smith et al. (2006, p 25), the average daily energy requirement is 2068 kcal per capita, averaged across all households in our survey experiment. Since all seven modules shared a common household roster design, we do not see any differences across modules in this aspect. A household is categorised as food energy deficient or hungry if total dietary energy available is lower than the energy requirement for that household.

Table 2 presents hunger estimates derived from HCES alone. The calorie measure displays a great amount of variability across the different survey methods. The estimated amount of daily per capita kilocalories available varies from 1794 Kcals (module 1, the long list of 58 food items with 14 day recall) to 2677 Kcals for the resource intensive personal diary (module 7). As a consequence, the estimates of hunger prevalence (the proportion of the individuals living in hungry households) vary by a factor of 3.6 and range from 18.8 to 68.4 percent, depending on the survey design. Following the pattern for calorie levels, the highest hunger rate is again measured with module 1 (68.3% of the population) and the lowest with module 7, with a prevalence of 18.8%. In general, recall modules record lower per capita calories and hence higher hunger prevalence. The usual month approach (module 4) records the second highest hunger prevalence (almost 60%), while the household diaries report hunger just slightly higher than with the individual diary (23%-27% depending on the level of supervision, compared with 19% for the individual diary).

These hunger estimates show more variation across survey design than the poverty estimates from the same survey experiment discussed in Beegle et al (2012), for a number of reasons. First, the poverty estimates include non-food expenditures that are collected the same way across designs and so this dampens the cross-design variation in monetary-based measures such as poverty. Second, low-price, calorie-dense food consumed from household stocks may be susceptible to under-reporting in recall surveys, making the variation in calorie-based measures greater than the variation in expenditure-based measures. Finally, these same staple foods are frequently reported in non-standardized units (pieces of cassava, bunches of bananas). The resulting conversion error will have a larger impact on estimates of calories (because they are calorie rich), but less so on expenditures (because they are cheap).

Now that we have established that average daily per capita kilocalories and hunger numbers differ substantially by module assignment, an obvious next question is whether the correlates of hunger also differ by module. This should be of concern for anyone analysing hunger and its determinants. Furthermore, one of the arguments for favouring the HCES method over the FAO method is that it allows for within-country comparisons across geographical zones or socio-economic or demographic groups, which constitutes important information for policy makers at the national level. If the within-module relative ranking of households changes depending on the household's characteristics then intra-country comparisons of different population groups may not be consistent across the different modules. To investigate the likely extent of this we adopt the following framework:

(1) $$Y_{ik} = \beta_k M_k + \beta_x X_{ik} + \gamma_k M_k X_{ik} + e_{ik}$$

where Y_{ik} is either the (log) of kilocalories per capita (estimated with an Ordinary Least Squares regression – OLS) or a variable indicating that the household is categorised as hungry (estimated with a Linear Probability Model – LPM). Households are indexed *i* and questionnaire assignments *k*, with M_k a vector of six dummy variables for module type, omitting the resource-intensive personal diary which is the base category, and X_{ik} is a single household characteristic. Randomization of module assignment ensures that the error term, e_{ik}, is orthogonal to both M_k and X_{ik} and to their interaction. In order to explore the interactions between correlates of hunger and module effects, our main focus is on estimating Equation (1) separately for each of four selected household characteristics: total household size, urban location, the number of years of formal education of the household head, and an asset index as a measure of household resources (asset-wealth) derived from housing conditions and household durable goods (Filmer and Pritchett, 2001). We also present results in Appendix I that jointly estimate these interactions for all four household characteristics at once, in order to consider the conditional effects.

Tables 3 and 4 show the coefficients of the six level effects of the module assignment β_k (module 7 is the omitted category), the level effect of the single household characteristic included in the regression β_x (the variable name is indicated in the column heading) and six interaction terms γ_k of the module assignment with the household characteristic in question, from regressions on calorie availability and a hunger dummy, respectively.

The level effects of the module assignment indicate large differences in the averages for calorie availability and hunger, confirming the results of Table 2. The level effects of the household characteristics show that calorie availability, as measured by module 7 (the base), is higher in smaller, urban, asset-rich households with educated heads. Much of the shifts in calorie availability happen within households above the nutrition threshold, so that actual hunger numbers are somewhat less sensitive to these characteristics in module 7. That may partly be a reflection of the fact that our base, module 7, displays the lowest hunger prevalence.

Our main interest lies in the interaction effects γ_k, which show that the effect of each of these household characteristics depends critically on the module used. For example, for each standard deviation increase in the asset index (the index is normalized to mean zero and standard deviation of unity), a household given the usual month survey method will have a 13 percentage point lower chance of being measured hungry compared to an identical household assessed through personal diary. Other recall modules have a similar sign but the interaction effect is smaller in magnitude. In other words, recall modules would increasingly underestimate hunger prevalence as the household grows richer.

On the other hand, recall modules tend to overstate hunger vis-à-vis diaries with respect to household size. For every one-person increase in the number of household members, the relative likelihood of a hunger diagnosis with recall surveys increases by 2-4 percentage points.

Module 4 uses the usual month approach, which intends to provide a measure of consumption across a whole year and therefore free from any concerns about seasonality. This is such a desirable property for survey analysts that this module is frequently put forward as a best-practice in survey research (Deaton and Grosh, 2000). One could also conjecture it to be the most abstract module and for that reason also the most cognitively burdensome one. The results in Tables 3 and 4 would certainly be consistent with such a hypothesis. We see that head's education only influences the household's relative position compared to the personal diary when module 4, the typical month approach, is used. For each additional year of education of the household head this module inflates the measurement of per capita calorie availability in the household by 2.5 percentage points. It would take 13 years of education, or the completion of secondary school, for the interaction effect of head's education to cancel out the negative level effects of the module (-0.457) and head's education (-0.011), although in practice only 2% of the heads in our sample have attained that level of education.

Hunger prevalence depends critically on location in module 2 and 4. An additional wedge of 15 and 17 percentage points, respectively, is driven between rural and urban households in these modules, compared to the personal diaries. This means that key hunger statistics, such as whether hunger is a rural or urban phenomenon, and those policies that result from them, will potentially depend on survey design.

When these interaction effects are jointly considered, in Appendix Table 1, the urban location and education interactions show up at statistically significant mostly in the calorie regressions and not in the hunger models, while the wealth effects and the household size effects show up both in the calorie estimates and in the hunger estimates. Conditional on all of the other covariates, a 1-person increase in household size raises the likelihood that the household would be considered hungry by 3-6 percentage points when the recall method is used compared with when the diary method is used. Thus whether larger households (who tend to be younger) or smaller ones (who tend to be older or headed by a single adult) are more likely to be measured as hungry depends on the type of HCES used. A reverse, and larger, effect occurs with respect to wealth; wealthier households are less likely to show up as hungry when using the recall method compared with the diary method. Thus survey data used as evidence on potentially contentious issues, such as the effect of economic growth and higher incomes on hunger (Ravallion, 1990), may not be easily compared across studies that differ in terms of using the recall method rather than the diary method.

5.2 The FAO method

We now turn to the FAO method, which depends on survey design for the calculation of the CV. It also critically depends on the design of the collection of FBS data, as well as on a number of assumptions (e.g. on trade, food wastage and the like). The FAO methodology is described in the technical annexes of

SOFI reports and in Wanner et al. (2014). The data that underlie these hunger measurements is available in the public domain.[13] These data put average gross calorie availability in Tanzania at 2161 kcal per person per day at the year of the survey. From that is deducted an estimated 3.46% wastage at the retail level, giving an average net availability of μ=2086 kcal per person per day, which is close the values we obtain from the recall modules. The FAO's average kcal requirement per person per day is 2103, which is also close to our estimate of 2068. However, it does not use the average requirements as the cut-off value, relying instead on the *minimum* requirement λ=1691 kcal per person per day in Tanzania at the time of our survey. This lower minimum level reflects requirements for people at the 5th percentile of body mass index distribution conducting light activity. Finally the FAO reports a CV of 0.36.

We can use these values to calculate a mean (μ*), a standard deviation (σ*) and a *z*-score (*z**) for the log-normal distribution as follows[14]:

$$\sigma^* = \sqrt{\ln(CV^2) + 1}$$

$$\mu^* = \ln(\mu) - \frac{\sigma^{*2}}{2}$$

$$z^* = \frac{\ln(\lambda) - \mu^*}{\sigma^*}$$

From the *z*-score, *p*-values are derived in the cumulative standard normal distribution to indicate the area to the left of the average requirement of 1691 kcal/day. This number is the percentage of people who fall to the left of 1691 kcal/day assuming a log-normal distributional form parameterized by μ* and σ*. Plugging in the above numbers puts hunger prevalence in Tanzania at 33%, or 14 million people out of a total population of 41.1 million (at the time of the survey).

A first experiment is to keep the FAO's mean (μ=2086) and cut-off (λ=1691) for Tanzania fixed, and let only the CV vary by survey module and verify how that influences the resulting hunger prevalence. For the calculation of the CV we remain as close as possible to the latest method used by the FAO (FAO 2014, p47). We proceed as follows. First we regress per capita calorie availability at the household level on a constant, the log of income (proxied by total household expenditures) and seasonal dummies. We then predict, for each observation, that part of the per capita calorie availability that is due to income

[13] http://bit.ly/14FRxGV accessed on 13 February 2015

[14] This seems most appropriate as in countries with low average energy availability the FAO uses the log-normal distribution.

alone and calculate the CV across those predictions. The resulting CV is denoted $CV|y$. To this is added a component to account for variation in calories orthogonal to income – $CV|r$ – to obtain the final *CV* in the following manner:

$$CV = \sqrt{(CV|y)^2 + (CV|r)^2}$$

For most countries the FAO estimates of $CV|r$ are close to 0.2, which is what we will set it to here. The third column in Table 5 shows that, calculated in this way, *CV* varies from 0.22 to 0.28, with lower values from the diaries compared to the recall modules and Module 4 standing out with the highest *CV* of 0.28.[15] All of the values are, however, lower than the FAO value of 0.36.

The results, presented in Table 5, show hunger prevalence varying in lockstep with *CV* from around 1 in 5 hungry in the diaries to 1 in 4 hungry when measured through recall. This disparity is solely due to differences in the calculated CV across survey methods. The largest difference in measured hunger prevalence solely due to differences in calculated CV is 6 percentage points, between the household diary with frequent visits (20%) and the typical month module (26%), which would represent 2.7 million people in Tanzania. If we drop the adjustments made to suppress variability, and calculate the CV immediately from the kcal data, then the CV ranges from 0.32-0.45 and the hunger numbers from 30% to 39% (not shown in table), i.e. an order of magnitude higher but still within a more narrow range than the HCES estimates. Note also that in practice the current estimates of *CV* are not always country specific and do not, typically, vary over time within countries.

A second experiment is to keep the CV and threshold fixed, but let the mean vary by survey module. Of course the FAO does not actually take the mean from the survey data, so this experiment is only informative if one is willing to believe that the amount of variation in the FBS data that is due to design is (at least) of the same order of magnitude as that present in surveys.[16] Table 6 shows that setting the mean to the specific value of each module in turn leads to wide variation in the proportion of undernourished people. Panel A shows that if we take *CV* from the personal diary, the prevalence of

---

[15] The FAO's method for determining *CV* has changed over time. Previously, ten deciles of daily per capita kilocalories (FAO 1996, appendix 3) or income (Naiken, 2002) were defined. The CV was then calculated across the medians of these ten groups, with 0.05 subtracted to account for other errors. The motivation for these manipulations, which lower the CV, was the desire to purge the CV of random variation, seasonal variation, and measurement error. The FAO forced the resulting *CV* to lie between 0.20 and 0.35, setting any outliers to their nearest acceptable value. When applying these two methods we obtain *CV* values of between 0.24-0.35 and 0.25-0.35, respectively.

[16] Given the wide array of techniques used to collect FBS data and the problems mentioned in Section 2, our presumption is that variability is higher. We know of no research that investigates this, however.

hunger ranges from 3% when the personal diary is used to 44% when the estimates come from an HCES that relies on a 14-day recall with a long list of food items. Panel B shows the range is from 13% to 50% if we assume the FAO *CV*, which is higher at 0.36. That would constitute a difference of 18 to 20 million people in a country like Tanzania, which currently has a total population of roughly 49 million. The results highlight the importance of the first moment of the food distribution in the resulting hunger numbers.

6 CONCLUDING DISCUSSION

There is a push by some in the international research community to calculate hunger numbers directly from household surveys, as opposed to calculating them from a combination of food balance sheets and household surveys as currently done by the FAO. The FAO method has the advantage of more frequent availability, but there are clear concerns with this method as well. While we are beginning to understand the nature of measurement errors in HCES data (see, for example, Gibson et al., 2015), we have little handle on errors in FBS data. Furthermore, the FAO method cannot allow for the analysis of food insecurity patterns within countries. Can household consumption surveys be a better source?

The evidence in this paper cautions against a naive switch to the HCES method. In our survey experiment, we calculate the prevalence of hunger from the HCES method to range between 19 and 68 percent. This is a difference of more than 24 million people in Tanzania, a country with a population of 49 million according to the 2012 census, and it is solely driven by differences in survey design. Our additional analyses of the FAO method suggest that it is varying estimates of the first moment of the distribution that contribute the most to the variation in hunger numbers and both the HCES method and the FAO method are vulnerable to this effect.

One primary concern for survey implementers and funders are the costs associated with each of the HCES modules. The means derived from the household diaries lie relatively close to our gold standard module 7, but it is worth noting that even the cheapest version of diary implementation, with only infrequent visits, still costs three times more than a recall module. Recall modules are relatively inexpensive and there are smaller cost differences among them. It is worth noting that moving from a short list of 17 items to a longer list of 58 items increased the time required for this module from 41 to 49 minutes, so that there is relatively little cost savings to be made by reducing the number of items on the list. Module 4, with the more abstract typical month recall period, takes 76 minutes to administer and does not perform well across the board when it comes to measuring or analysing hunger. More details on the time and cost implications of each of the modules can be found in Beegle et al. (2012). These authors also analyse the implications of survey design on poverty and inequality measures, which is something that most multi-purpose surveys will need to balance with effects on hunger measurement.[17]

Across the various survey designs, hunger statistics for individual households or sub-groups do not move up and down in lockstep. Instead, relative positions can change from one module to the other depending on the household or group's characteristics. Therefore simple mean correction factors for each module would not be sufficient when comparisons need to be made across population groups with different characteristics. Potentially, the regression coefficients from our Equation (1) could be used to make corrections, although it would be difficult to measure and control for all relevant characteristics.[18] Perhaps the bigger challenge is to what extent any correction factor estimated from this study can apply to different contexts and countries (e.g. settings with different income levels or staple foods).

HCES data are ubiquitous and hold the potential of being a useful complement to existing measures of hunger. Their appropriateness for hunger measurement could be improved through a series of low-cost design changes and more thoughtful harmonization of survey design (Smith et al., 2014; Carletto et al., 2012). But until more is done to undertake these design changes, and to understand the various sources of error and how they differ between survey methods, and at a broader level also between the HCES method and the FAO method, we would advocate for caution in drawing inferences from comparing survey-based hunger estimates over time and space.

[17] For example, the usual month method (Module 4) not only takes the longest time to field, it also shows the most inequality and one of the highest poverty rates of modules in the comparisons by Beegle et al. (2012). These comparisons use a constant poverty line for all modules, but if cost of basic needs (CBN) poverty lines are calculated separately by module, the poverty line from the usual month data would be higher than for the other modules, causing the poverty rate to seem even higher. This sensitivity of the CBN poverty line occurs because the food share is lower with the usual month method, and a lower food share matters because a food Engel curve is used to scale up from cost of the food poverty to the cost of the total poverty. In general, the estimated coefficients for food Engel curves appear to be sensitive to variation across the survey modules (Gibson et al, 2015).

[18] In a very different context, Dollar and Kraay (2002) and Gruen and Klasen (2007) apply correction factors to countries' Gini coefficients in order to correct for cross-country differences in inequality estimates (e.g. either based on income or consumption).

Table 1: Survey experiment consumption modules

Module	Consumption measurement	Recall/survey period	N households
1	Long list of 58 food items; 14 day recall	14 day	503
2	Long list of 58 food items; 7 day recall	7 day	504
3	Short list of 17 food items; 7 day recall	7 day	503
4	Long list of 58 food items; usual 12 month recall	Usual 12 months	504
5	14 day household diaries with frequent visits[a]	14 days	502
6	14 day household diaries with infrequent visits[b]	14 days	501
7	14 day individual diaries with frequent visits[a]	14 days	503
			3,520

Notes: An 8th module was included in the experiment but not used in the analysis as it collected expenditure but no quantities. (a) Frequent visits entailed daily visits by the local assistant and visits every other day by the survey enumerator for the duration of the 2-week diary.
(b) Infrequent visits entail 3 visits: to deliver the diary (day 1), to pick up week 1 diary and drop off week 2 diary (day 8), and to pick up week 2 diary (day 15). Households assigned to the infrequent diary but who had no literate members (about 18 percent of the 503 households) were visited every other day by the local assistant and the enumerator.

Table 2: Calorie availability and hunger prevalence

Module	Mean Kcal per capita (95% CI)	Hunger Prevalence (95% CI)
1. Long list of 58 food items; 14 day recall	1794	0.683
	(1723-1865)	(0.639-0.728)
2. Long list of 58 food items; 7 day recall	2129	0.481
	(2055-2203)	(0.432-0.531)
3. Short list of 17 food items; 7 day recall^	2066	0.484
	(2001-2131)	(0.435-0.534)
4. Long list of 58 food items; usual 12 month recall	1909	0.594
	(1823-1995)	(0.546-0.642)
5. 14 day household diaries with frequent visits	2412	0.268
	(2340-2485)	(0.223-0.313)
6. 14 day household diaries with infrequent visits	2517	0.230
	(2443-2591)	(0.186-0.275)
7. 14 day individual diaries with frequent visits	2677	0.188
	(2599-2755)	(0.148-0.228)

(^)The 17 foods account for 77 percent of the food budget, so calorie availability is scaled up by 1/0.77. *N*=3,520.

Table 3: Interaction effects between module type and selected household characteristics (OLS), average caloric availability.

OLS, LHS = ln(kcal pc)	Household size	Urban	Education of hh head (years)	Asset index
Level Effects (personal diary omitted)				
1. Long list, 14 day recall	-0.301***	-0.407***	-0.407***	-0.410***
2. Long list, 7 day recall	-0.196***	-0.271***	-0.256***	-0.244***
3. Short list, 7 day recall	-0.286***	-0.270***	-0.272***	-0.270***
4. Long list, usual month	-0.104**	-0.390***	-0.457***	-0.344***
5. HH Diary, frequent	-0.135***	-0.092***	-0.120***	-0.116***
6. HH Diary, infrequent	-0.083	-0.021	-0.074*	-0.061**
HH char. mentioned in column heading	-0.034***	0.112***	0.011**	0.056***
Interaction Effects				
(HH char. mentioned in column heading) *				
(1. Long list, 14 day recall)	-0.021**	0.001	0.000	0.043*
(2. Long list, 7 day recall)	-0.009	0.088	0.003	0.071***
(3. Short list, 7 day recall)	0.003	0.004	0.001	0.026
(4. Long list, usual month)	-0.044***	0.148***	0.025***	0.142***
(5. HH Diary, frequent)	0.004	-0.069	0.001	-0.003
(6. HH Diary, infrequent)	0.004	-0.114**	0.003	-0.026
Sample mean HH char. mentioned in column heading	5.249	.344	4.726	.001

Note: estimates of Equation (1). Each column represents the results of a (separate) OLS regression of ln(kcal pc) on 6 module assignment dummies, a single selected household characteristic (mentioned in the column headings), and 6 interaction terms of that household characteristic with the module assignment dummies. The personal diary is the omitted category. Standard errors are omitted to improve readability, but available upon request from the authors. *** indicates significance at 1 percent; ** at 5 percent; and * at 10 percent. *N*=3,520.

Table 4: Interaction effects between module type and selected household characteristics (LPM), hunger estimates.

LPM, LHS = hungry (dummy)	Household size	Urban	Education of hh head (years)	Asset index
Level Effects (personal diary omitted)				
1. Long list, 14 day recall	0.241***	0.458***	0.419***	0.435***
2. Long list, 7 day recall	0.114*	0.306***	0.285***	0.258***
3. Short list, 7 day recall	0.147**	0.291***	0.312***	0.265***
4. Long list, usual month	0.112*	0.400***	0.464***	0.344***
5. HH Diary, frequent	0.048	0.070**	0.078*	0.072**
6. HH Diary, infrequent	0.032	0.006	0.025	0.039
HH char. mentioned in column heading	0.012*	-0.002	-0.005	-0.010
Interaction Effects				
(HH char. mentioned in column heading) *				
(1. Long list, 14 day recall)	0.037***	-0.070	0.003	-0.052*
(2. Long list, 7 day recall)	0.028***	-0.146**	-0.006	-0.109***
(3. Short list, 7 day recall)	0.023**	-0.077	-0.010	-0.066**
(4. Long list, usual month)	0.043***	-0.171***	-0.026***	-0.130***
(5. HH Diary, frequent)	0.004	0.005	-0.001	0.005
(6. HH Diary, infrequent)	0.001	0.095	0.003	0.032
Sample mean HH char. mentioned in column heading	5.249	.344	4.726	.001

Note: estimates of Equation (1). Each column represents the results of a (separate) LPM regression of a dummy indicating the household is categorised as hungry on 6 module assignment dummies, a single selected household characteristic (mentioned in the column headings), and 6 interaction terms of that household characteristic with the module assignment dummies. The personal diary is the omitted category. Standard errors are omitted to improve readability, but available upon request from the authors. *** indicates significance at 1 percent; ** at 5 percent; and * at 10 percent. *N*=3,520.

Table 5: The sensitivity of the FAO method to variation in *CV*

Module	mean (μ)	CV	z-score Eq (1)	proportion hungry*
1. Long list of 58 food items; 14 day recall	2086	0.26	-0.69	0.25
2. Long list of 58 food items; 7 day recall	2086	0.25	-0.73	0.23
3. Short list of 17 food items; 7 day recall	2086	0.24	-0.75	0.23
4. Long list of 58 food items; usual 12 month recall	2086	0.28	-0.64	0.26
5. 14 day household diaries with frequent visits	2086	0.22	-0.85	0.20
6. 14 day household diaries with infrequent visits	2086	0.23	-0.80	0.21
7. 14 day individual diaries with frequent visits	2086	0.23	-0.80	0.21

(*) The proportion hungry is derived as the *p*-value from the *z*-score, i.e. the area to the left of 1691 kcal/day in the cumulative standard normal distribution. *N*=3,520.

Table 6: The sensitivity of the FAO method to variation in μ

PANEL A: using personal diary *CV* = 0.23

Module	mean (μ)	CV	z-score Eq (1)	proportion hungry*
1. Long list of 58 food items; 14 day recall	1794	0.23	-0.14	0.44
2. Long list of 58 food items; 7 day recall	2129	0.23	-0.89	0.19
3. Short list of 17 food items; 7 day recall	2066	0.23	-0.76	0.22
4. Long list of 58 food items; usual 12 month recall	1909	0.23	-0.42	0.34
5. 14 day household diaries with frequent visits	2412	0.23	-1.44	0.08
6. 14 day household diaries with infrequent visits	2517	0.23	-1.62	0.05
7. 14 day individual diaries with frequent visits	2677	0.23	-1.89	0.03

PANEL B: using FAO *CV* = 0.36

Module	mean (μ)	CV	z-score Eq (1)	proportion hungry*
8. Long list of 58 food items; 14 day recall	1794	0.36	0.00	0.50
9. Long list of 58 food items; 7 day recall	2129	0.36	-0.49	0.31
10. Short list of 17 food items; 7 day recall	2066	0.36	-0.40	0.34
11. Long list of 58 food items; usual 12 month recall	1909	0.36	-0.17	0.43
12. 14 day household diaries with frequent visits	2412	0.36	-0.84	0.20
13. 14 day household diaries with infrequent visits	2517	0.36	-0.96	0.17
14. 14 day individual diaries with frequent visits	2677	0.36	-1.14	0.13

(*) The proportion hungry is derived as the *p*-value from the *z*-score, i.e. the area to the left of 1691 kcal/day in the cumulative standard normal distribution. *N*=3,520.

REFERENCES

Azzalini A. 1985. A Class of Distributions Which Includes the Normal Ones*. Scandinavian Journal of Statistics* 12: 171-178.

Barrett C. 2010. Measuring food insecurity. *Science* 327: 825-828.

Beegle K, De Weerdt J, Friedman J, Gibson J. 2012. Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania. *Journal of Development Economics* 98: 19–33.

Behrman J, Deolalikar A. 1987. Will Developing Country Nutrition Improve with Income? A Case Study for Rural South India. *Journal of Political Economy* 95(3): 492-507.

Bouis H. 1994. The effect of income on demand for food in poor countries: Are our food consumption databases giving us reliable estimates? *Journal of Development Economics* 44:199-226.

Butte N, Garza C, de Onis M. 2007. Evaluation of the Feasibility of International Growth Standards for School-Aged Children and Adolescents. *Journal of Nutrition* 137: 153-157.

Cafiero C. 2014. Advances in hunger measurement: traditional FAO methods and recent innovations. FAO Statistics Division Working Paper Series No. 14-04.

Capéau B, Dercon S. 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.

Caeyers B, Chalmers N, De Weerdt J. 2012. Improving Consumption Measurement and other Survey Data through CAPI: Evidence from a Randomized Experiment *Journal of Development Economics* 98:19-33.

Carletto C, Zezza A, Banarjee R. 2012. Towards Better Measurements of Household Food Security: Harmonizing Indicators and the Role of Household Surveys. *Global Food Security* 2:30-40.

Chen S, Ravallion M. 2010. The Developing World is Poorer than We Thought, But No Less Successful in the Fight Against Poverty. *Quarterly Journal of Economics* 125(4): 1577-1625.

de Haen H, Klasen S, Qaim M. 2011. What do We Really Know? Metrics for Food Insecurity and Malnutrition. *Food Policy* 36(6): 760–769.

Deaton A, Grosh M. 2000. Consumption. In Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study, Margaret Grosh and Paul Glewwe, eds. New York: Oxford University Press (for World Bank).

Deaton A, Zaidi S. 2002. Guidelines for Constructing Consumption Aggregates for Welfare Analysis. World Bank LSMS Working Paper 135.

Dollar D, Kraay A. 2002. Growth is good for the poor. *Journal of Economic Growth* 7(3):195-225.

FAO. 1996. The Sixth World Food Survey. Rome: Food and Agriculture Organization of the United Nations.

FAO. 2012. The State of Food Insecurity in the World: Economic Growth is Necessary but Not Sufficient to Accelerate Reduction of Hunger and Malnutrition. Rome: Food and Agriculture Organization of the United Nations.

FAO. 2014. The State of Food Insecurity in the World: Strengthening the enabling environment for food security and nutrition. Rome: Food and Agriculture Organization of the United Nations.

FAO/WHO/UNU (Food and Agriculture Organization of the United Nations/World Health Organization/United Nations University). 1985. Protein and Energy Requirements. Rome: Food and Agriculture Organization of the United Nations.

Fiedler J, Lividini K, Bermudez O, Smitz M. 2012a. Household Consumption and Expenditures surveys (HCES): a Primer for Food and Nutrition Analysts. *Food and Nutrition Bulletin* 33(3): S235-S241.

Fiedler J, Carletto C, Dupriez O. 2012b. Still Waiting for Godot? Improving Household Consumption and Expenditures Surveys (HCES) to Enable More Evidence-based Nutrition Policies. *Food and Nutrition Bulletin* 33(3): S242-S251.

Fiedler J. 2013. Towards Overcoming the Food Consumption Information Gap: Strengthening Household Consumption and Information Surveys for Food and Nutrition Policymaking. *Global Food Security* 2: 56-63.

Filmer D, Pritchett L. 2001. Estimating wealth effect without expenditure data or tears: an application to educational enrollments in states of India. *Demography* 38 (1), 115–132.

Gibson R. 2005. Principles of Nutritional Assessment. 2nd edition. New York: Oxford University Press.

Gibson J, Kim B. 2012. Testing the Infrequent Purchases Model Using Direct Measurement of Hidden Consumption from Food Stocks. *American Journal of Agricultural Economics* 94(1): 257-270.

Gibson J, Beegle K, De Weerdt J, Friedman J. 2015. What Does Variation in Survey Design Reveal About the Nature of Measurement Errors in Household Consumption? *Oxford Bulletin of Economics and Statistics* 77(3): 466-474.

Gruen C., Klasen S. 2008. Growth, inequality, and welfare: comparisons across space and time. *Oxford Economic Papers* 60:212-236.

Gunderson C, Ribar D. 2011. Food Insecurity and Insufficiency at Low Levels of Food Expenditures. *Review of Income and Wealth* 57(4): 704-726.

Headey D. 2013. The impact of The Global Food crisis on Self-Assessed Food Security. Policy Research Working Paper Number 6329, The World Bank.

Headey D, Ecker O. 2013. Rethinking the Measurement of Food Security: From First Principles to Best Practice. *Food Security* 5:327–343.

Horton D. 1988. Underground Crops and Long-term Trends in Production of Roots and Tubers. Winrock International Institute for Agricultural Development.

James W, Schofield E. 1990. Human Energy Requirements: A Manual for Planners and Nutritionists. Oxford, U.K.: Oxford University Press.

Jensen R, Miller N. 2010. A Revealed Preference Approach to Measuring Hunger and Undernutrition. mimeo.

Jones A, Ngure F, Pelto G, Young S. 2013. What are we assessing when we measure food security? A compendium and review of current metrics. *Advances in Nutrition* 4:481-505.

Klasen S. 2008. Poverty, undernutrition, and child mortality: Some inter-regional puzzles and their implications for research and policy. *Journal of Economic Inequality* 6(1):89–115.

Lukmanji Z, Hertzmark E, Mlingi N, Assey V, Ndossi G, Fawzi W. 2008. Tanzania Food Composition Tables. Dar es Salaam, Tanzania, Muhimbili University of Health and Allied Sciences (MUHAS), Tanzania Food and Nutrition Centre (TFNC), Harvard School of Public Health (HSPH).

Masset E. 2011. A review of hunger indices and methods to monitor country commitment to fighting hunger. *Food Policy* 36: S102-S108.

Migotto M, Davis B, Carletto C Beegle K. 2007. Measuring Food Security Using Respondents' Perception of Food Consumption Adequacy. in Food Security: Indicators, Measurement, and the Impact of Trade Openness. B. Guha-Khasnobis, S.S. Archaya and B. Davis, Eds. UNU-WIDER and ICSSR.

Moore Lappé F, Clapp J, Anderson M, Broad R, Messer E, Pogge T, Wise T. How We Count Hunger Matters, *Ethics and International Affairs* 27(3): 251-259.

Mueller I, Vounatsou P, Allen B, Smith T. 2001. Spatial Patterns of Child Growth in Papua New Guinea and their Relation to Environment, Diet, Socio-economic Status and Subsistence Activities. *Annals of Human Biology* 28(3): 263-280.

Naiken L. 2002. FAO methodology for estimating the prevalence of undernourishment. Keynote paper for the International Scientific Symposium on Measurement and Assessment of Food Deprivation and Undernutrition. Rome, 26 - 28 June 2002. FAO, Rome.

Naiken L. 2007. The Probability Distribution Framework for Estimating the Prevalence of Undernourishment: Exploding the Myth of the Bivariate Distribution. Statistics Division Working Paper Series No. ESS/ESSG/009e. FAO Rome.

Naiken L. 2014. Methodological issues in the estimation of the prevalence of undernourishment based on dietary energy consumption data: a review and clarification. FAO Statistics Division Working Paper Series ESS / 14-03.

Radimer K, Olson C, Campbell C. 1990. Development of Indicators to Assess Hunger. *Journal of Nutrition* 120:1544-1548.

Ravallion, M. 1990. Income effects on undernutrition. *Economic Development and Cultural Change* 38(3): 489-515.

Ruel M, Harris J, Cunningham K. 2012. Measuring dietary quality in developing countries: A review of the usefulness of individual dietary diversity indicators. Chapter 18 in V. R. Preedy (Ed.), Diet quality: An evidence-based approach. New York: Springer.

Sen A. 1981. Poverty and famines: an essay on entitlement and deprivation. New York: Oxford University Press.

Smith L, Subandoro A. 2007. Measuring Food Security Using Household Expenditure Surveys. Food Security in Practice Technical Guide Series. Washington, D.C.: International Food Policy Research Institute.

Smith L. 1998. Can FAO's Measure of Chronic Undernourishment be Strengthened? *Food Policy* 23(5): 425-445.

Smith L, Alderman H, Aduayom D. 2006. Food Insecurity in Sub-Saharan Africa. New Estimates from Household Expenditure Surveys.

Smith L, Subandoro A. 2007. Measuring Food Security Using Household Expenditure Surveys. Food Security in Practice Technical Guide Series. Washington, D.C.: International Food Policy Research Institute.

Smith L, Dupriez O and Troubat N (2014) Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys. IHSN Working Paper No. 008

Svedberg P. 1999. 841 Million Undernourished? *World Development* 27(12): 2081-2098.

Svedberg P. 2002. Undernutrition Overestimated. *Economic Development and Cultural Change* 51(1):5-36.

Thompson F, Byek T. 1994. Dietary Assessment Resource Manual. *Journal of Nutrition*. 124: 2245S-2317S.

Timmer P. 2009. Rice Price Formation in the Short Run and the Long Run: The Role of Market Structure in Explaining Volatility. Center for Global Development Working Paper 172.

Ulijaszek S. 1994. Between-Population Variation in Pre-Adolescent Growth. *European Journal of Clinical Nutrition* 48(1): 5-13.

U.S. Department of Agriculture (USDA). 2002. "Composition of Foods Raw, Processed, Prepared", USDA Nutrient Database for Standard Reference, Release 15, USDA, Maryland, United States.

Wanner N, Cafiero C, Troubat N, Conforti P. 2014. Refinements to the FAO methodology for estimating the prevalence of undernourishment indicator. FAO Statistics Division Working Paper Series ESSS 14-05.

Zo Rambeloson J, Dary O, Fiedler J, Franklin N. 2012. Comparison of estimates of the nutrient density of the diet of women and children in Uganda by Household Consumption and Expenditures Surveys (HCES) and 24-hour recall. *Food and Nutrition Bulletin* 33(3): S199-S207

APPENDIX

Table A1: Regressions from Table 2 and Table 3, with all covariates considered jointly.

	OLS ln (kcal pc)		LPM Hungry dummy	
	Coef.	**SE**	**Coef.**	**SE**
MODULE LEVEL EFFECTS				
1. Long list, 14 day recall	0.069	0.048	-0.083	0.055
2. Long list, 7 day recall	0.094*	0.050	-0.027	0.057
3. Short list, 7 day recall	-0.028	0.049	0.059	0.056
4. Long list, usual month	0.108**	0.049	-0.013	0.056
5. HH Diary, frequent	0.021	0.048	-0.023	0.055
6. HH Diary, infrequent	-0.005	0.048	0.090	0.055
HOUSEHOLD CHARACTERISTICS AND THEIR INTERACTIONS				
Household size	-0.020***	0.005	0.004	0.006
* (1. Long list, 14 day recall)	-0.040***	0.007	0.053***	0.008
* (2. Long list, 7 day recall)	-0.022***	0.007	0.031***	0.008
* (3. Short list, 7 day recall)	-0.017**	0.007	0.033***	0.008
* (4. Long list, usual month)	-0.053***	0.007	0.055***	0.008
* (5. HH Diary, frequent)	-0.005	0.007	0.010	0.007
* (6. HH Diary, infrequent)	-0.003	0.007	0.002	0.007
Urban location dummy	0.120***	0.044	-0.025	0.050
* (1. Long list, 14 day recall)	-0.264***	0.061	0.155**	0.070
* (2. Long list, 7 day recall)	-0.101	0.062	0.053	0.070
* (3. Short list, 7 day recall)	-0.170***	0.063	0.103	0.072
* (4. Long list, usual month)	-0.183***	0.063	0.115	0.072
* (5. HH Diary, frequent)	-0.166***	0.062	0.033	0.070
* (6. HH Diary, infrequent)	-0.170***	0.062	0.076	0.070
Education of HH head (years)	0.017***	0.005	-0.011**	0.005
* (1. Long list, 14 day recall)	-0.020***	0.006	0.021***	0.007
* (2. Long list, 7 day recall)	-0.018***	0.007	0.014*	0.007
* (3. Short list, 7 day recall)	-0.019***	0.007	0.005	0.008
* (4. Long list, usual month)	0.001	0.007	-0.003	0.007
* (5. HH Diary, frequent)	-0.005	0.006	0.001	0.007
* (6. HH Diary, infrequent)	0.004	0.007	-0.003	0.007

Asset index	-0.024	0.024	0.021	0.027
* (1. Long list, 14 day recall)	0.148***	0.035	-0.120***	0.040
* (2. Long list, 7 day recall)	0.142***	0.036	-0.143***	0.041
* (3. Short list, 7 day recall)	0.102***	0.036	-0.084**	0.041
* (4. Long list, usual month)	0.188***	0.035	-0.134***	0.040
* (5. HH Diary, frequent)	0.063*	0.035	-0.011	0.040
* (6. HH Diary, infrequent)	0.016	0.035	0.031	0.040
CONSTANT	7.858***	0.018	0.224***	0.021

The personal diary is the omitted category. *** indicates significance at 1 percent; ** at 5 percent; and * at 10 percent. *N*=3,520.