WPS6736

POLICY RESEARCH WORKING PAPER   6736

The Challenge of Measuring Hunger

Joachim De Weerdt
Kathleen Beegle
Jed Friedman
John Gibson

Abstract

There is widespread interest in the number of hungry people in the world and 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. 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 a consequence, comparable and valid hunger numbers will be lacking until more effort is made to either harmonize survey designs or better understand the consequences of survey design variation.

The Challenge of Measuring Hunger

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

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. Yet there were still 870 million hungry people worldwide in 2010-2012, down only slightly from 1 billion in 1990-92, according to the flagship publication of the Food and Agriculture Organization (FAO) of the United Nations -- *The State of Food Insecurity in the World*. Over the same period, the slow fall in the proportion of hungry people, from 19% to 13% (FAO 2012), suggests that the goal of halving the hunger rate will not be met on time.

The FAO estimates of the hunger rate combine aggregate food balance sheets for every country with survey estimates of the variance in calorie availability. The population falling below calorie requirements is computed from the combination of the total food balance and its variance. This method, which we will refer to in shorthand notation as the FBS-CV method, has been heavily criticized by the international research community (Svedberg 1999, de Haen et al. 2011). As an alternative, some researchers have been advocating for calculating hunger numbers directly from the food quantity data in household surveys (Smith and Subandoro 2007, Fiedler et al. 2012b). Given the expansion of such surveys in low-income countries, this is increasingly feasible. From 1990 onward, 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. We will call this approach to calculating hunger the HCES-direct method. Of note is that neither the FBS-CV nor the HCES-direct method actually measures what individuals ate (as in a food intake survey) or ask about their perceptions of hunger (Thompson and Byek 1994, Radimer et al. 1990).

The FBS-CV and HCES-direct methods both rely on household surveys. In the case of the FBS-CV method, the second moment of the calorie distribution comes from the surveys, while the HCES-direct method relies on the surveys for all moments. Yet, the design of HCES varies over several key dimensions, such as 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

---

[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.

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 methods to 3,525 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 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.

Our results suggest that any comparative assessment of hunger prevalence using HCES should clearly take into account differences in survey design. However, hunger numbers are not simply scaled up or down by differences in survey design: 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 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). The reason is that the 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 spent on harmonizing 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 the survey approach to measurement. In other areas of socio-economic statistics, efforts have succeeded at harmonization in methods - cross-country data on fertility and health are much more comparable because of the standardized approach to measurement taken by the Demographic and Health Surveys.

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. 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 Agriculture Organization (FAO) of the United Nations in a series of reports tracking world hunger, with FAO (2012) being the latest. More recently, the FAO indicator is used to track progress toward the first Millennium Development Goal of halving poverty and hunger by 2015. It relies on the assumption that the supply of food energy follows a log-normal distribution, which can be parameterized by a mean and a coefficient of variation (CV) – to capture the food access distribution. It 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.[2] Combining this with population data allows the FAO to estimate the total kilo calories available for human consumption per person, per country in a particular year. The CV is calculated from a limited number of HCES.[3] For most countries the CV was kept constant across years and only the mean was revised.[4] Finally, the FAO estimates the required energy of a population by determining age-sex specific minimum daily energy requirements under the assumption of 'light work'. These numbers are aggregated to yield the requirement of an average person. The area underneath the log-normal energy

---

[2] Available at http://faostat3.fao.org/home/index.html (accessed 11 June 2013)

[3] 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.

[4] In 2012 the FAO revised its distribution to be the skew-normal distribution (Azzalini, 1985), which generalises the normal distribution to allow for skewing. The FAO now uses survey data to calculate the CV and coefficient of skewness. This revision also took into account the average physical stature of each age-sex group derived from DHS data. 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.

distribution which lies to the left of the energy requirement estimate 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). The criticisms reflect concerns about the reliability of the three components that go into the FBS-CV calculations, whether the CV is the best measure of dispersion to use, and whether it should be kept constant over time (de Haen et al. 2011). Further, the FBS-CV method relies on two different sources for two moments of the distribution (the mean and the spread). It is a method open to the same criticisms made about measures of global poverty that rely on GDP for the consumption mean and on household survey data for the variance (see the critique in Chen and Ravallion 2010).

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

The FAO calculates the CV of food availability directly for only for a limited number of countries and rarely updates this number (see footnotes 3 and 4). As such, differences in hunger estimates across countries and over time are mainly driven by FBS data. While the focus of this paper is not on the reliability of these national estimates of food availability, of note are three concerns about these estimates of food availability. First, food availability is a residual, 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 are 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).

An alternative approach to the FBS-CV method is to follow the lead of nutritionists, who typically rely on either an observed-weighed food method or a 24-hour recall (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 care giver.

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 (Fiedler 2013). This is at least in part due to the fact that they are time consuming and therefore expensive to collect. At least three other less resource-intensive alternatives to the FAO approach have been suggested to derive hunger numbers: anthropometric data, self-assessments, and direct use of HCES.

Anthropometric data are an indirect measure, which is highly correlated with energy intake. The obvious criticism is that anthropometrics are also influenced by other factors, such as disease, and it is not clear that anthropometric standards sufficiently reflect either genetic or gender variation. The second alternative is subjective questions where the respondent self-assesses food adequacy. The Gallup World Poll, for example, asks '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). These one-shot 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 four countries and find that subjective perceptions of food consumption adequacy are, at best, weakly correlated with calorie consumption, dietary diversity, and anthropometric measures.

The third approach, and the one that we concentrate on here, is to use HCES to 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 respected 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-direct approach has been propagated by the international research community as the most viable alternative to the FAO's FBS-CV method. The remainder of this paper will therefore be dedicated to juxtaposing the FBS-CV and HCES-direct methods, and considering their robustness to variations in survey design. We can shed some 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 calorific intake for a survey sample or before the CV needed for the FAO method can be calculated.[5] 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 of 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. Dupriez et al. (2013) have designed a very detailed metadata survey that tries to categorize HCES in all their relevant dimensions (requiring a 22-page form to cover all variations). Drawing on surveys from 100 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 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.

---

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

REPORTING ERROR IN CONSUMPTION

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.[6] Finally, misreporting may

---

[6] 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 itemsand the more cognitively demanding 'typical month' phrasing was used in the question.

arise from social desirability bias. The respondent may think the responses given will come to bear on some future intervention and may wish to exaggerate or understate his consumption with that in mind.

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.

QUESTIONNAIRE DESIGN ERRORS

Aside from the various reporting errors, other criticisms of using HCES for nutritional assessments relate to easily avoidable design mistakes. 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.

Another example is that 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.[7] These two examples are perhaps the lowest hanging fruit when it comes to the harmonization of HCES for measuring food consumption. In fact, 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.

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

---

[7] 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.

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 will continue to be a common survey method used in low-income countries.

UNIT CONVERSION ERROR

Throughout most of the developing world, households do not typically purchase, harvest, or consume their food in standardized units (kilograms or liters). 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. There is no systematic approach across countries to convert to standard units.

FOOD HETEROGENEITY ERROR

Having obtained estimates of weight or volume of each food item, pre-existing food tables can be used to determine energy content. This happens in two steps. Frist 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 the list of food items contains grouped foods, calories may be measured with error.[8]

ERRORS IN CALORIE REQUIREMENTS

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

---

[8] 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

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. DATA

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 eight 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). An eighth module is excluded from the current analysis as it did not capture food quantities. The eight designs were strategically selected to reflect the most 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 food quantities 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. 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 24 households were randomly selected to participate and three households were randomly assigned to each of the eight 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. This, in addition to dropping the eighth experimental arm as explained above, yields a final sample size of 3,525 households.[9]

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 eight 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).

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 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.

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 liters. Where district-level conversion rates were not available, the sample median was used and where this was not available, measurements at the survey's

---

[9] We have almost no item non-response (in that the respondent does not answer whether the household consumed the specific item in the specified period) in the recall modules. We do not observe any distinct patterns in non-response across our survey designs, or within a recall design by the location of the item on the list.

headquarters were taken after the fieldwork was done.[10] 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.
In regards to household calorie needs, we control for age, gender, and whether a woman is breastfeeding, but do not have data on body weight, height, or level of physical activity. The average daily energy requirement is 2068 kcal per capita, averaged across all households. Since households were administered one of the eight modules through random selection, we do not see any differences across modules in this aspect. A household is categorized as food energy deficient or hungry if total dietary energy available is lower than the energy requirement for that household.

5. RESULTS

We utilize this experiment to explore the survey design implications for measuring hunger. We consider both the HCES-direct method that is advocated by numerous researchers and the FBS-CV method that is used by the FAO when making global hunger counts. 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.

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 1793 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 is again measured with module 1 (68.4% 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

---

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

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).

One of the arguments for favoring the HCES-direct method over the FBS-CV method is that it allows for within-country comparisons across geographical zones or socio-economic or demographic groups, which is a key concern for policy makers at the national level. A natural question to ask then is whether such intra-country comparisons might also be affected by survey design variations. 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 categorized 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. We estimate Equation (1) separately for each of four selected household characteristics, with results reported in Table 3.

Many of the interaction terms are statistically significant, which suggests that estimated within-country patterns of hunger are sensitive to the type of HCES that is used. For example, for each standard deviation increase in the asset index, a household given the usual month survey method will have a 13 percentage point lower chance of being measured hungry compared to a household at the same level of wealth but assessed through personal diary (this result is of course conditional on the mean effect of the questionnaire design, which is captured by the M_k term). 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.

Turning to the implications of survey design for the FAO's FBS-CV method, we use the 7 modules from our experiment to calculate a CV of calorie availability, following FAO (1996, Appendix 3).[11] Specifically, we collapse the data to 10 deciles of daily per capita kilocalories available and then calculate the CV across the medians of these 10 groups, subtracting 0.05 from that CV to account for other errors. The FAO motivates these manipulations, which serve to lower the CV, by the desire to purge the CV of random variation, seasonal variation, and measurement error. Furthermore the FAO forces any CV to lie between 0.20 and 0.35, setting any outliers to their nearest acceptable value.

In addition to the CV, we need an estimate of the mean of calorie availability per capita in order to fully parameterize the log-normal distribution. We use two approaches for the mean estimate. First, we take the mean from the personal diary (μ^*= 2677 kcal per person per day) as the estimate which is likely to be closest to the truth. Second, we set μ^* to the module-specific mean of calorie availability. The motivation for this second approach, which is purely illustrative, is that the FAO derives this mean from the FBS, which also differ in methodology and implementation across countries, so we would like to reflect this sensitivity in estimates of the average food balance.

The third and final component needed to replicate the FAO calculations is to use the sample average of daily energy requirement (2068 kcal per person per day). Using these estimates, the FAO approach produces a mean (μ), a standard deviation (σ) and a z-score (z):

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

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

$$z = \frac{\ln(2068) - \mu}{\sigma}$$

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 2068 kcal/day. This number is the percentage of people who fall to the left of 2068 kcal/day assuming a log-normal distributional form parameterized by μ^* and CV. When hunger rates estimated in this same manner are aggregated across countries they yield the widely reported estimates of 870 million hungry people that is noted in the introduction of this paper.

---

[11] This is the last FAO report on which we have a detailed description of the exact mechanisms used. The methodology used has been modified slightly since then, as explained Annex 2 of FAO (2012) and footnote 4.

Table 4 presents two sets of results. A common mean for the calorie distribution, which comes from module 7 – the personal diary, is used in Panel A, while module-specific means are used in Panel B. The last column shows the estimates of the proportion of the population that are hungry. Holding the mean calorie availability fixed at the value estimated when using the personal diary method, the rate of malnourishment ranges from 17% to 28%. This is a difference in the estimate of hungry people in Tanzania of nearly 5 million people. And it is due solely to differences in the calculated CV across survey methods.

At present, the FAO hunger estimates rely on CVs which are not always country specific, do not vary over time within countries, and have several adjustments made to suppress variability. As shown in Panel A, abstracting from the challenge of measuring mean food availability from FBS, the variation in CV estimates driven only by survey design can result in large differences in estimates of hunger even when adjustments are made to suppress variability. When we also allow for variation in the measure of average calorie availability, the variation in apparent hunger is even larger. Panel B shows that if setting the mean to the specific value of each module, the proportion of undernourished people varies greatly: ranging from 19% when the personal diary is used to 72% when the estimates come from an HCES that relies on a 14 day recall with a long list of food items. That would constitute a difference of 24m people in a country like Tanzania, which has a total population of roughly 45m. In general the patterns of variation in panel B of Table 4 are broadly similar to those in Table 2, which reflects the importance of the first moment of the food distribution in the resulting hunger numbers.

6. CONCLUDING DISCUSSION

There is a push by the international research community to calculate hunger numbers directly from household surveys (HCES-direct method), as opposed to calculating them from a combination of food balance sheets and household surveys (FBS-CV method) as currently done by the FAO. The FBS-CV method has the advantage of more frequent availability, but there are clear concerns with this method as well. Whereas we increasingly know and understand survey errors, we have very little handle on errors in FBS data. Furthermore, the FBS-CV method cannot allow for the analysis of food insecurity patterns within countries.

Still, despite these critiques, the evidence in this paper cautions against a naive switch to the HCES-direct method. In our survey experiment, we calculate hunger to range between 19 and 68 percent –this is a difference of more than 23 million people in Tanzania (a country with a population of 45 million

according to the 2012 census). These differences are solely driven by differences in survey design, presenting a strong challenge to both the HCES-direct and the FBS-CV methods. The FBS-CV is additionally dependent on variations in the FBS data collection, which are plausibly greater and the consequences of design variation less understood than those for surveys.

More research effort is needed to understand the impact of survey methodology on resulting hunger numbers. Note, though, that this undertaking needs to go beyond the determination of simple mean correction factors for each survey type as relative module performance in part depends on household characteristics.

While we support the calculation of hunger numbers directly from household surveys, we believe that this may only be feasible once much more effort is put into harmonizing survey design. Our findings strongly support the current drive toward such harmonization as expressed, for example, in Carletto et al. (2012).

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	504
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	504
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	503
6	14 day household diaries with infrequent visits[b]	14 days	503
7	14 day individual diaries with frequent visits[a]	14 days	503
			3,525

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	1793	0.684
	(1722-1864)	(0.640-0.728)
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	2413	0.268
	(2341-2486)	(0.223-0.313)
6. 14 day household diaries with infrequent visits	2520	0.230
	(2446-2594)	(0.186-0.274)
7. 14 day individual diaries with frequent visits	2677	0.188
	(2600-2756)	(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.

Table 3: Interaction effects for module type and selected household characteristics

Personal diary omitted	Household size	Asset index	Urban	Education of hh head
PANEL A: OLS, ln(kcal pc)				
1. Long list, 14 day recall	-0.021**	0.043*	0.000	-0.000
2. Long list, 7 day recall	-0.009	0.071***	0.003	0.088
3. Short list, 7 day recall	0.003	0.026	0.001	0.003
4. Long list, usual month	-0.044***	0.142***	0.025***	0.148***
5. HH Diary, frequent	0.004	-0.003	0.001	-0.069
6. HH Diary, infrequent	0.005	-0.026	0.003	-0.113**
PANEL B: LPM, hungry				
1. Long list, 14 day recall	0.037***	-0.052*	0.003	-0.068
2. Long list, 7 day recall	0.028***	-0.109***	-0.006	-0.146**
3. Short list, 7 day recall	0.023**	-0.066**	-0.009	-0.073
4. Long list, usual month	0.043***	-0.130***	-0.026***	-0.171***
5. HH Diary, frequent	0.004	0.005	-0.001	0.005
6. HH Diary, infrequent	0.001	0.032	0.003	0.094

Note: estimates of Equation (1). Each column represents the results of a (separate) regression (OLS or LPM) of a selected HCES-derived measure (mentioned in the titles of the panels) on 7 module assignment dummies, a single selected household characteristic (mentioned in the column headings) and 7 interaction terms of that household characteristic with the module assignment dummies. Only the interaction terms are reported. The personal diary is the omitted category. Level effects and 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.

Table 4: The sensitivity of the FAO FBS-CV method to module type

PANEL A: MEAN FIXED AT PERSONAL DIARY LEVEL

Module	mean (μ*)	CV	z-score Eq (1)	proportion hungry*
1. Long list of 58 food items; 14 day recall	2,677	0.35	-0.59	0.28
2. Long list of 58 food items; 7 day recall	2,677	0.32	-0.67	0.25
3. Short list of 17 food items; 7 day recall^	2,677	0.26	-0.87	0.19
4. Long list of 58 food items; usual 12 month recall	2,677	0.35	-0.59	0.28
5. 14 day household diaries with frequent visits	2,677	0.27	-0.85	0.20
6. 14 day household diaries with infrequent visits	2,677	0.25	-0.94	0.17
7. 14 day individual diaries with frequent visits	2,677	0.26	-0.89	0.19

* The proportion hungry is derived as the *p*-value from the *z*-score, i.e. the area to the left of 2068 kcal/day in the cumulative standard normal distribution.

PANEL B: MODULE-SPECIFIC MEAN

Module	mean (μ*)	CV	z-score Eq (1)	proportion hungry*
1. Long list of 58 food items; 14 day recall	1,793	0.35	0.59	0.72
2. Long list of 58 food items; 7 day recall	2,129	0.32	0.06	0.53
3. Short list of 17 food items; 7 day recall^	2,066	0.26	0.13	0.55
4. Long list of 58 food items; usual 12 month recall	1,909	0.35	0.41	0.66
5. 14 day household diaries with frequent visits	2,413	0.27	-0.45	0.33
6. 14 day household diaries with infrequent visits	2,520	0.25	-0.69	0.24
7. 14 day individual diaries with frequent visits	2,677	0.26	-0.89	0.19

* The proportion hungry is derived as the *p*-value from the *z*-score, i.e. the area to the left of 2068 kcal/day in the cumulative standard normal distribution.

REFERENCES

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

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.

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. and 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.

Dupriez, O., Smith, L. and Troubat, N. 2013. Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys. FAO, IHSN and World Bank, accessed on 12 September 2013 at http://www.ihsn.org/home/node/34

FAO, 1996. The Sixth World Food Survey. FAO, Rome.

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/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. 2012. Towards Overcoming the Food Consumption Information Gap: Strengthening Housheold Consumption and Information Surveys for Food and Nutrition Policymaking. *Global Food Security* 2: 56-63.

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.

Headey, D. 2013. The impact of The Global Food crisis on Self-Assessed Food Security. Policy Research

   Working Paper Number 6329, The World Bank.

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.

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).

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.

Radimer, K L, Olson C M, Campbell C C. 1990. Development of Indicators to Assess Hunger. *Journal of

   Nutrition* 120:1544-1548.

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.

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

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.

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.