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

Jed Friedman[a,*], Kathleen Beegle[a], Joachim De Weerdt[b], John Gibson[c]

[a] The World Bank, USA
[b] University of Antwerp and KU Leuven, Belgium
[c] University of Waikato, New Zealand

## ARTICLE INFO

## ABSTRACT

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

## 1. Introduction

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

This study focuses on food consumption measurement, leveraging a recent survey experiment to study the performance of commonly used consumption survey modules in order to shed light on the nature of reporting error in consumption data. The experiment randomly allocated to a nationally representative sample of Tanzanian households one of eight consumption survey modules. A daily supervised individual diary with dependent household members assigned to a unique adult respondent is taken as the benchmark, or "gold standard", survey approach given the resources and care survey teams devoted to it (described in detail below). The accuracy of the other modules is assessed with respect to this benchmark. Previous work from the same experiment (Beegle et al., 2012; Gibson et al., 2015; De Weerdt et al., 2016) explores relative performance of the eight modules in terms of mean consumption, inequality, poverty,

http://dx.doi.org/10.1016/j.foodpol.2017.08.016

and the prevalence of hunger. These studies concentrate on household-level total aggregates for food, non-food, and total consumption and do not consider variation in performance amongst individual items, as is done here. Moreover, variation in mean consumption by module – which is up to 27 percent of total value in these studies – conveys the net effect of all possible types of reporting error including opposing impacts of recall and telescoping error, as well as the difficulty of fully capturing individual consumption opportunities outside the home.

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

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

## 2. Errors in consumption measurement: A brief taxonomy

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

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

Recall error. A main concern is that respondents forget the occurrence of a consumption event resulting in what is termed recall error. Lower salience and longer recall periods make it more likely for respondents to forget (Sudman and Bradburn, 1973). Several studies show that, all else equal, the longer the period of recall, the lower the reported consumption per standardized unit of time (Scott and Amenuvegbe, 1991; Grosh et al., 1995).

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

Rule of thumb error. Respondents may not always recall and count events (Menon, 1993). Especially for longer recall periods that typically have more transactions, respondents may cease trying to enumerate each one and instead use rules of thumb to estimate them (Blair and Burton, 1987; Gibson and Kim, 2007; de Nicola and Giné, 2014; Arthi et al., 2016). In this case, what is termed rule-of-thumb error likely depends on the frequency and regularity of the transaction with less frequent items reported with more error. Whereas recall error downwardly biases the consumption estimate, and telescoping creates upward bias, there is no obvious direction of bias for responses that resort to rule of thumb instead

of enumeration. We may expect this error to be particularly pertinent for hypothetical consumption constructs such as questions about consumption in a "usual" month. "Usual" month consumption is an explicit attempt to abstract away from seasonal considerations in consumption, however this type of question may present additional cognitive demands compared to a definitive recall period in the immediate past.

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

Other error types. While the analysis in this study will focus on the four types of reporting error listed above, misreporting can also arise from other sources such as rounding error, social desirability bias, and strategic responses. An example of the latter is a respondent who exaggerates (or understates) her consumption in order to appear poorer (or richer) due to a belief that the responses given may determine eligibility for some future social program. There may also be intentional misreporting arising from respondent fatigue, so whether the respondent is presented with a long or a short list of consumption items can influence the overall quality of the responses.[2]

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

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

## 3. The Tanzania survey experiment

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

---