2019

IARIW-World Bank

Special IARIW-World Bank Conference "New Approaches to Defining and Measuring Poverty in a Growing World" Washington, DC, November 7-8, 2019

Improving Accurate Reporting of Food Consumption for Displaced Populations

Utz Johann Pape

Paper Prepared for the IARIW-World Bank Conference

Washington, DC, November 7-8, 2019

Improving accurate reporting of food consumption for displaced populations

Utz Johann Pape[1]

Abstract:

Misreporting is a well-known challenge for researchers in social sciences and even more so for policymakers who rely on accurate data to design effective relief strategies. This issue is especially prevalent if incentives for misreporting exist, for example, if there are perceived social or material implications to the answer reported. Due to the vulnerabilities that internally displaced people face and their dependency on aid support, this challenge is particularly acute when trying to measure economic welfare. In contexts of emergency, the need for accurate information is acute, but many approaches to improve data are intrusive or impractical. To address this challenge, this paper proposes a light touch method to improve the accuracy of consumption data: the inclusion of checks and primes to emphasize the importance of accurate reporting. The study assesses the effectiveness of this method, by randomly distributing the bundle of primes across a survey of internally displaced persons in South Sudan. In line with the main hypothesis, positive and significant effects arise for low consumption quantiles, especially consumption quantities that are more susceptible to manipulation. The findings suggest that light touch approaches to improving survey design can act as a cost-effective tool to induce more accurate reporting. Further research is needed to validate the measure against a gold standard approach and to understand what mechanisms are at play.

Keywords: Consumption Measurement, Poverty, Questionnaire Design, Behavioral Intervention

JEL: C83, D63, D90, I32

---

[1] Authors in alphabetically order. Corresponding author: Utz Pape (upape@worldbank.org). The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. This paper benefited largely from the valuable input provided by Lennart Kaplan, Julie Perng and Luca Parisotto as well as discussions with Quy-Toan Do.

1. Introduction

Accurate data on the key economic variables affecting people who have been forcibly displaced, such as consumption and assets, are essential to understanding their situation and to developing evidence based policies to support them. Poor information can lead to flawed diagnostics or incorrect assessments of impact. Data inaccuracies may lead policy makers to allocate funds to the wrong people or to the wrong programs. The standard way in which the World Bank and other policy organizations develop statistics is through individuals' responses to questions in economic surveys. Self-reported information is vulnerable to myriad reporting inaccuracies when social scientists ask personal or intrusive questions or when respondents anticipate social or material implications to the answers they provide. [2] This is of particular concern when respondents believe that misreporting may provide relief, both because of the of the sensitivity and the gravity of the policy challenge. In situations where it has been possible to compare survey responses to revealed economic behavior, striking disparities are sometimes found. In one investigation for example, Poterba and Summers (1986) report that misstatements regarding employment status in the Current Population Survey led to an underestimation of the duration of unemployment by up to 80 percent and even greater overestimates of the frequency of labor market entries and exits. In another study, Rosenfeld, Imai, and Shapiro (2016) look at voting behaviors in a sensitive anti-abortion referendum held in Mississippi in 2011. They compare actual county level vote shares against survey results from a sample frame of individuals who voted during the election (based on public records). Surveys that used direct questioning led to an underestimation of casting a "no" vote by more than 20 percentage points in the majority of counties.

There are a number of mechanisms through which the validity of self-reported information in surveys can be compromised. Some inaccuracies result from cognitive biases – for example, acquiescence or "yea-saying" (Bachman and O'Malley 1984; Hurd 1999), extreme responding (Cronbach 1946; Hamilton 1968), and question order bias (Sigelman 1981). One solution to problems such as question order bias is to randomize the order of questions (Warner 1965). Other inaccuracies emerge from conscious but not calculated behavior. Respondents may deliberately misreport information on sensitive subjects not to distort statistics but to maintain their reputation or to abide by political norms (Gilens, Sniderman, and Kuklinski 1998). A common solution to this is to enable participants to cloak their behaviors or beliefs. List experiments, endorsement experiments, and randomized experiments are commonly used techniques for this purpose (Rosenfeld, Imai, and Shapiro 2016).

[2] This is of particular concern, for example when asking about race (Kuklinski et al. 1997) or corruption (Gingerich 2010).

The explanations above assume that people intend to report accurately but are prevented from doing so due to aspects of the situation. In some contexts, individuals may misreport due to expectations about the implications of the results of the study. For example, individuals may misreport to increase earnings in a study context (Mazar, Amir, and Ariely 2008) or to shape the results of the study if they believe that it will inform policy. In situations where individuals wish to influence a particular research outcome, a guise of anonymity will not shift their behavior. It is important to note that our concern is not with the ethics of individual misreporting – this is a reasonable response to contexts of extreme vulnerability – but rather to ensure that policymakers have access to data that enables them to adequately serve the vulnerable population as a whole.

Behavioral science is increasingly being used as a policy tool to help policymakers create better policy and solve collective action problems more effectively (World Bank 2015). This is based on research illustrating that people make decisions on the basis of both external and internal reward mechanisms (Mazar and Ariely 2006). Even in cases where people have an extrinsic incentive to misreport, this may be overridden by a preference for remaining consistent with their values. One example of this is when individuals' beliefs regarding the consequences of misreporting affects their behavior. In an two-person experiment where one participant can increase her payoff by misreporting but at the expense to her counterpart, Gneezy (2005) finds that individuals' propensity to misreporting is sensitive to the costs it imposes on the other person. Contextual cues affect the salience of internal incentives (or intrinsic motivations) and thus the accuracy of responses. This psychological mechanism has been put to practical use in policy. In multiple contexts, normative messaging has been used to increase tax payments (Hallsworth et al. 2017; Hernandez et al. 2017) or reduce littering and environment theft (Cialdini 2003).

In this paper, we apply the tools of behavioral science to investigate the veracity of consumption reports by internally displaced persons (IDPs). In numerous rounds of data collection in Somalia and South Sudan, IDPs report significantly lower levels of consumption than non-IDP households. In previous survey rounds 45 percent of Somali IDP households report food consumption below subsistence levels and approximately 80 percent below recommended levels (refer Figure 5). While the data may be accurate, there are two reasons to suspect that it is not. First, such high levels of non consumption would be associated with high rates of mortality due to starvation. Although being high, the mortality rates among IDPs suggest that this is not happening systematically across the country at such a scale (FEWS NET, 2018).[3] Second, non-IDP households that are statistically similar on observable characteristics report higher levels of consumption than IDP households. While IDPs and non-IDPs may

[3] Although data from the USAID led Famine Early Warning Systems Network (FEWS NET) suggest high level of malnutrition, evidence on mortality across the counties is mixed (FEWS NET, 2018).

have different opportunities to generate income, it is unlikely that IDPs choose not to smooth their resources to balance between food and non-food consumption in a way that endangers their life.[4]

If it is the case that survey respondents misreport, the inaccuracies it generates in the data are highly problematic. At best, it makes the data spurious and unusable. At worst, it could lead to misallocations of aid, from more vulnerable areas to less vulnerable areas, or from solutions emphasizing sustainability to immediate relief where immediate relief is unnecessary. Due to the dangerous environment in South Sudan and Somalia, it is not currently possible to do use alternative data collection methods, for example ethnographic research, to investigate this puzzle in the data. The validity of alternative investigative methods such as food diaries is vulnerable to the same incentive to game as surveys.

One way to investigate whether people misreport is to test whether consumption rates change in response to nudges. If these primes are effective, they would be expected to particularly affect potentially underreporting, hence, poor households. Moreover, as vulnerable populations would have higher incentives to underreport, priming should be stronger for IDPs than for comparable non-IDP populations. We find the primes induce higher reporting in lower quintiles of reported consumption. This treatment pattern is driven by aid reliant IDPs and vanishes when considering the comparison group of non-IDPs. The results are especially strong for consumption quantities (items and kilograms), which are most easily subject to intentional misreporting. This suggests that IDPs are indeed misreporting. The paper has two main limitations. First, it can only compare the treated group against an estimate of the "true" consumption rates. Second, the intervention is bundled. For this reason, it is impossible to isolate the causal mechanism affecting the observed changes in reporting. Further work is needed to identify an estimate of the true level of consumption against which to compare the primed individuals and to isolate the causal mechanisms by which people are changing their behavior.

The paper proceeds as follows. Section 2 provides an overview about the underlying context and the compiled data. Section 3 provides an overview about the underlying methods, while Section 4 introduces the empirical approach, which builds the foundation for the results in Section 5. This is complemented by an assessment of robustness and potential channels in Section 6. Finally, findings are discussed and summarized in Section 7.

2. Context and data

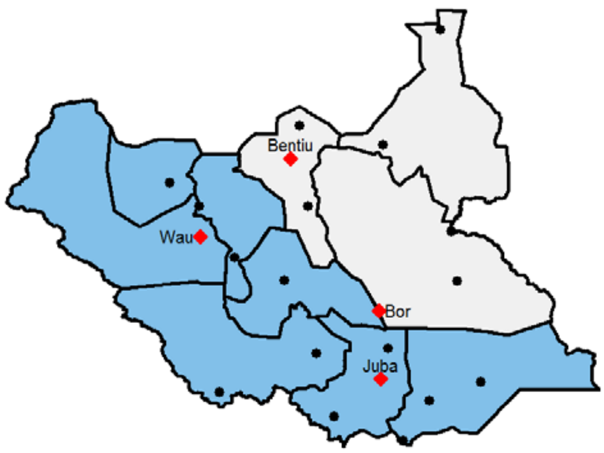
On July 9 2011 South Sudan became the 55th African independent state after seceding peacefully from the Republic of Sudan. Facing a history of a 50 year lasting conflict South Sudan slid back to instability after its peaceful independence process. This led to an internal displacement of circa two million, more

[4] The underlying survey data of this study discussed at a later stage actually indicates that IDPs have a more calorie intensive food consumption profile (refer Figure 7).

than 15 percent of South Sudan's population (UN OCHA, 2017). Moreover, the conflict contributed to a deterioration of South Sudanese economic outcomes, with poverty rates reaching 82 percent in 2016, widespread severe food shortages and famine being declared in some counties in 2017 (The World Bank, 2016; Devi, 2017). This makes well-targeted crisis response and aid allocation highly important.

The experiment sample includes 4145 IDP and 781 non-IDP households interviewed in 2017 in South Sudan across the High Frequency South Sudan Survey (HFSSS), the Crisis Recovery Survey (CRS), and the IDP Census and Sampling Study (IDPCSS). The CRS interviewed a representative sample of IDPs in IDP camps across South Sudan. In the same period the HFS conducted interviews across urban centers in seven of the ten former states (Figure 1). The IDPCSS conducted a census of all households in Juba POC1. The consumption modules in questionnaires administered to respondents in the three surveys were built in exactly the same manner so as to ensure comparability, and the fieldwork was implemented by the same organization. The only difference across the three surveys is the population that was sampled.

Figure 1: HFS and CRS coverage.



Note: The HFS interviewed a representative sample of households in urban centers in the states colored in blue in the map above. The CRS interviewed households in 4 of the largest IDP camps in South Sudan, denoted by red diamonds in the map. Major urban areas are indicated via black dots. The IDPCSS was conducted in the Juba POC1.

The conditions in camps do not allow for standard household surveys, hence, an alternative survey approach based on the Rapid Consumption Methodology was applied (Pape & Mistiaen, 2015). Here, only 30 / 25 food and non-food items are administered to all households. Additional 20 food and non-food items vary between households. More specifically, households are pre-assigned to one out of four sub-modules for food and non-food consumption (each containing 20 items). Neither the enumerators nor the respondents see the structure of the sub-modules, but the assigned items are asked in a categorically meaningful way (like cereals, fruits, etc).[5]

The data is used to construct four outcome measures. The surveys collect information on quantities in terms of (i) number of consumption items and (ii) kilograms. The quantities can be used to construct measures of (iii) monetary and (iv) caloric food consumption scaling the quantities with data on average prices and energy levels.[6] Though we are mainly interested in evaluating the impact of the nudges on the total consumption value - both in terms of money and food intake - these variables are difficult for respondents to falsify because these are *second-order* values that are calculated as a function of other variables, including consumption quantities and calories or prices that are in turn deflated. All of this adds noise to the answer provided by the respondent, and they depend in part on variables over which the respondent has no control. The consumption quantity in kilograms is a more direct measure of the quantity consumed as expressed by the respondent, and may lead to more accurate estimation of the impact of the nudges. Finally, counting the number of items may lead to an even more accurate measure, since the variable does not undergo any cleaning at all and is taken at face value. Furthermore, omitting an item is likely to be the easiest and quickest way for respondents to reduce the true value of the household's consumption.[7]

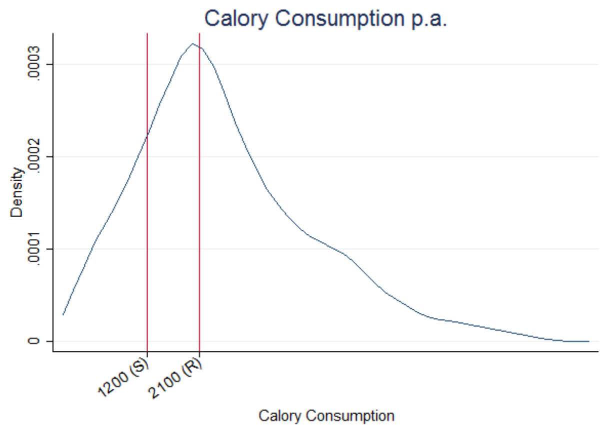
Poverty amongst IDP households is high, and 9 in 10 IDP households across South Sudan live under $1.90 USD PPP (2011) per capita per day in 2017. IDP households in the sample interviewed for the experiment consume on average 333 SSP (2017) per capita per day. IDP households reported on average 6.63 core consumption items. These figures represent about 20 percent of core items asked to the households. Figure 2 visualizes that 39 percent of households report consumption below the recommended daily intake of 2,100 kcal (R) and 16 percent below the subsistence level of 1,200 kcal (S) (Ravallion & Bidani, 1994).

[5] Due to the survey method applied CRS surveys contain the core consumption module and one additional consumption module. The share in imputed consumption is on average 99.9 percent. IDP surveys contain due to the previously outlined time constraints only core consumption items. However, by design these items capture the lion's share of consumption (on average approx. 94 percent of total consumption in more comprehensive CRS surveys).

[6] For a description of the caloric intake measure, please consult the appendix.

[7] Note that the number of consumption items is not reported per capita as it does not increase proportionally with household size.

Figure 2: Density plot of value of core food consumption.



Note: Estimates presented in the figure above are not weighted and are representative only of IDP and non-IDP households surveyed in the study sample [8]

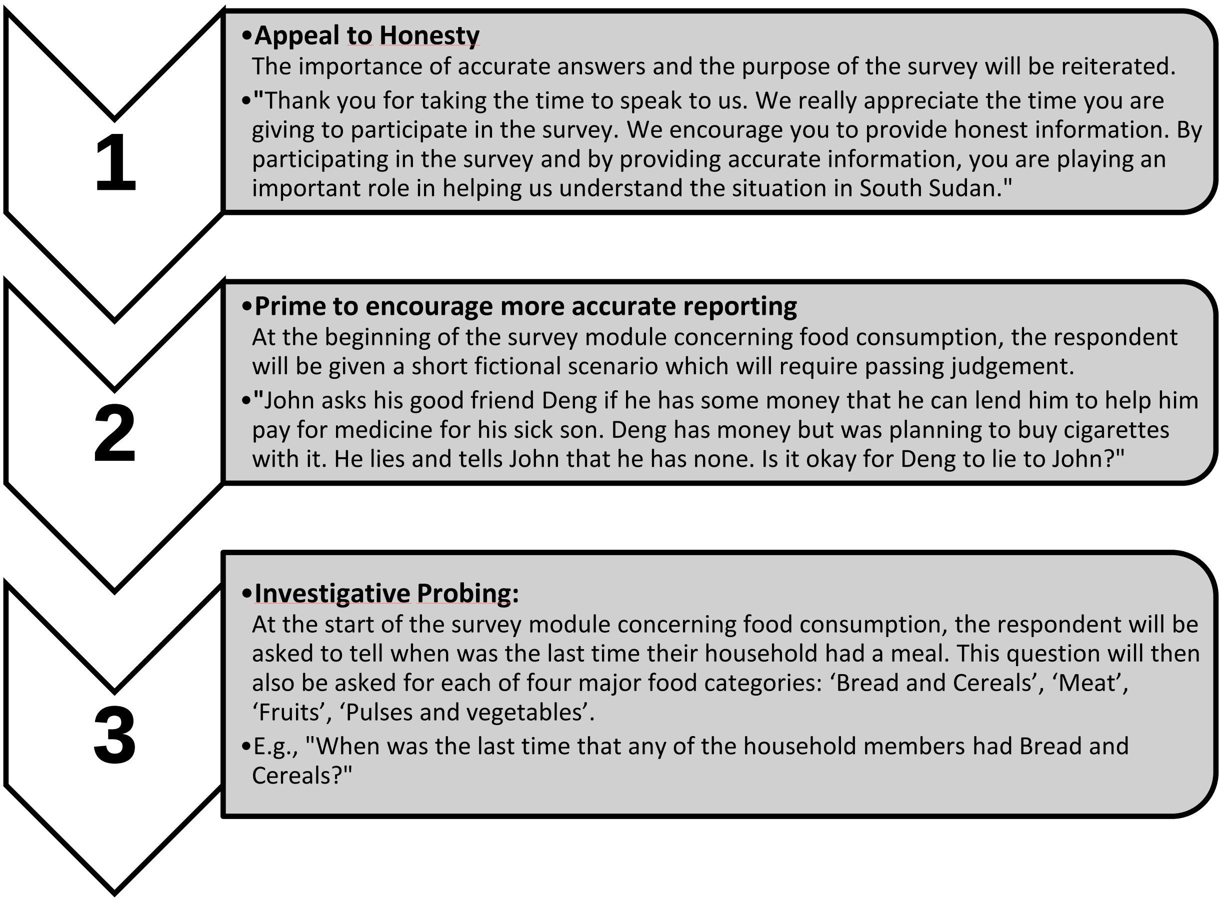
3. Approach and randomization

Our light-touch method introduce exogenous variation into the consumption module to try and tease out whether consumption might be underreported in IDP households. A prime is an environmental cue that unconsciously induces a subsequent cognition or behavior. For example, in studies with prisoners and bankers, participants who engage in activities that prime their identity report less accurately in behavioral experiments than participants who have not participated in priming activities (Cohn, Fehr, and Maréchal 2014; Cohn, Maréchal, and Noll 2010).[9] Nudges have been found to elicit more accurate responses during questionnaires (Rasinski et al. 2005; Vinski and Watter 2012).

To investigate whether consumption might be underreported in IDP households, we introduce exogenous variation into the consumption module. Households are randomly exposed to a bundle of light-touch measures. These include an emphasis on the importance of accurate answers at the beginning of the survey, a short fictional scenario which will require passing judgment on the behavior of one of the characters, and additional questions to tell when was the last time their household had a meal, forcing the respondents to explicitly report that they have not eaten in the last week.

[8] We do not use weights throughout the study as the research hypothesis relates not to the average treatment effect, but particularly the pirmes' effectiveness at the tails of the distribution.

[9] Questionnaires confirmed that participants associate their identity with dishonesty.



Households are randomly exposed to behavioral treatments, in the form of a prime for more accurate reporting and investigative probing, to try and elicit more truthful answers from respondents. This way they do not constrain the choice frame, but rather alter the anchoring towards more truthful reporting(The World Bank, 2015).

The bundle of primes addresses different behavioral processes. (1) Appeals to honesty are a standard tool in surveys to increase data accuracy by relying on social approval (Talwar, Arruda, & Yachison, 2015). (2) Primes to encourage more accurate reporting induce unconscious cognitions, which are intended to affect subsequent behavior. When facing incentives to misreport, respondents would answer more accurately to sustain self-consistency. (3) Investigative probing puts a higher salience on the question. By asking for broader categories first, subsequent sub categories are put under more scrutiny. Self-consistency is reinforced by relating to a longer recall period of seven days.[10] While the appeal to honesty and the prime target intentional misreporting, investigative probing is addressing classical measurement error.

The sample was randomly selected into each treatment arms in two groups of approximately 50 percent, with 2,467 households in the control group and 2,459 in the treatment group. The

[10] The methodological appendix provides an overview of the relevant questions in the food consumption module.

randomization process was built into the CAPI (Computer Assisted Personal Interview) questionnaires administered in the surveys. As our research hypothesis suggests stronger effects of nudges for more vulnerable populations, we focus on IDPs for the main analysis. The availability of the HFS sample provides a comparison group of non-IDP households for the experiment, which will be used for robustness checks. The treatment and control groups are relatively balanced. There is a higher share of male headed households in the treatment group, which have also more members, though in practical terms these differences are relatively small. As gender of the household head and household size are potentially correlated with poverty, these variables are included in the regression models and interacted with the treatment to control for potential impacts (Lanjouw & Ravallion, 1995).

Table 1: Balance across treatment and control arms (IDP sample).

	Control	Treatment	Difference, p-value
Household size	4.835	5.098	0.003***
	(0.060)	(0.064)	
Gender of household head	0.492	0.448	0.005***
	(0.011)	(0.011)	
Literacy of household head	0.507	0.529	0.155
	(0.011)	(0.011)	
Household head completed some primary school	0.540	0.563	0.133
	(0.011)	(0.011)	
Is the household head employed	0.328	0.319	0.555
	(0.010)	(0.010)	
Share of children in household	0.364	0.373	0.309
	(0.006)	(0.006)	
Share of elderly in household	0.011	0.010	0.582
	(0.002)	(0.001)	
First Component of Asset Principal Component Analysis	-0.126	-0.194	0.162
	(0.037)	(0.032)	
N	2079	2066	
Proportion	0.502	0.498	

*Standard errors in parentheses; **p<0.05, ***p<0.01*

4. Empirical Strategy

To assess the effect of our prime on reporting behavior, we can formulate following simple regression equation.

$$Y_i = \beta_0 + \beta_1 T_i + \boldsymbol{T_i} * \boldsymbol{X_i}\beta_2 + X_i\beta_3 + \gamma_s + \alpha_t + \varepsilon_i, \tag{3}$$

where Y_i is the log of the outcome variable. Across different models we estimate the effect for (i) the number of consumption items consumed [referred to in the regression equation as Cons. Num.], (ii) consumption quantity per adult equivalent (in kilograms) [Cons. Quant.], (iii) monetary consumption value per adult equivalent [Cons. Val.] and (iv) daily caloric intake per adult equivalent [Cons. Cal.].

Our main treatment variable T_i is a dummy variable which takes the value of 1 if the household i was assigned to the treatment group. γ_s indicates a set of camp fixed effects, α_t are month fixed effects, and ε_i is the idiosyncratic error term. X_i denotes a vector of control variables generally associated with consumption, including household size, the gender of the household head, and the proportion of children (under 18) in the household. Moreover, we add an asset index based on the first component of a principal component analysis (Filmer & Pritchett, 2001; McKenzie, 2005).[11] The model will be estimated with and without controls to check the impact they may have. As the treatment might interact with the unbalanced covariates, it makes sense to add to the regression $T_i * X_i$, the interaction of the unbalanced controls with the treatment variable (Lin & Green, 2016; Baranov, Bhalotra, Biroli, & Maselko, 2017).

It is expected that the respondents who will be affected by the treatment are respondents that would otherwise misreport and, hence, a more likely to be at the extremes of the distribution.[12] Therefore, we complement our analysis with a quantile regression approach. The idea of the quantile regression framework, which was introduced by Koenker & Bassett Jr (1978), is to take the entire distribution of the dependent variable into account by estimating several regressions, which put more weight to the quantile of interest. The underlying minimization problem can be stated as follows:

$$Q(\theta) = arg \min_{\tau} \sum_{i:y_i>\tau} \theta|y_i - \tau| + \sum_{i:y_i\leq\tau} (1-\theta)|y_i - \tau|, \tag{4}$$

where θi is the quantile of interest and the weighted sums of deviations $|y_i - \tau|$ of the outcome per quantile. Minimizing the latter, differential effects conditional on the quantile of the dependent variable are obtained. Further, it has the advantage of being less prone to outliers and non-normality of the error term. For our purpose, quantile regressions offer the advantage that they are more flexible than simple interactions with poverty lines, which would be endogenous to consumption levels.

[11] As assets (bikes, fans, rickshaws etc.) can be more easily surveyed by enumerators, those are likely to capture parts of the household wealth.

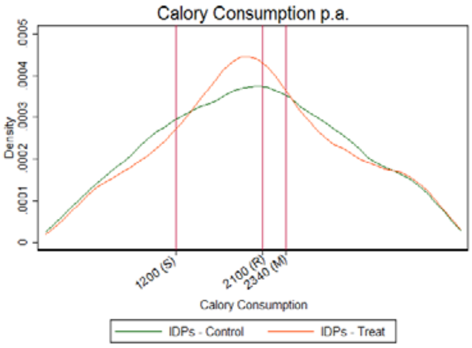
[12] Although these hypotheses were not pre-registered, they are based on theoretical considerations about the mechanisms of the underlying behavioral primes.

5 Results

There is a slight indication that the treatment may have worked, based on consumption distributions across treatment and control group. The consumption distribution shown in The median of calorie consumption is well above the recommended daily intake. However, still a substantial part of the distribution of 16 percent reports below the subsistence level and 40 percent below the recommended daily intake. Hence, the prime would also be relevant in the adult equivalent setting to achieve more precise reporting, which is analyzed in our regression framework subsequently. Taking into account the finding that consumption levels are lower than to be expected, the most relevant treatment effects can be expected at the left tail of the distribution.

Figure 3 *shows a slight difference in caloric consumption between IDP households in the treatment group and the control group, though this is apparent only at lower levels of consumption, i.e. below the subsistence level of 1,200 kcal.* The median of calorie consumption is well above the recommended daily intake. However, still a substantial part of the distribution of 16 percent reports below the subsistence level and 40 percent below the recommended daily intake.[13] Hence, the prime would also be relevant in the adult equivalent setting to achieve more precise reporting, which is analyzed in our regression framework subsequently. Taking into account the finding that consumption levels are lower than to be expected, the most relevant treatment effects can be expected at the left tail of the distribution.

Figure 3: Caloric consumption p.c. (adult equivalents).



Note: The underlying data is based on per adult equivalents. Caloric consumption levels are labeled in the following graph as **S** subsistence equivalent (1200 kcal p.c.), **R** recommended daily intake (2100 kcal p.c.) and **M** the median (2340 kcal p.c.).

[13] Compared to the monetary consumption levels, the calory consumption p.a. seems rather high. This is partly attributable to the fact that IDP's consumption focuses on energy intensive products, where cooking oil and sorghum constitute 45 percent of food expenditure. If we contrast the consumption shares with non-IDPs, we find that although the diet of non-IDPs is less energy intensive, it comprises a higher variety (see Figure 7).

Regression results

In order to test for the influence of control variables, the regressions are estimated with and without control variables. When not conditioning on control variables, the results indicate only a significant treatment effects for the number of consumption items in Column (1). This outcome measure would be easiest to falsify as it does not undergo further cleaning, e.g., in terms of deflation or calorie scaling. When adding further controls, coefficients turn larger and imply treatment effects of 6-14 percent. The interactions of the treatment and the asset index as well as household size have negative and significant coefficients in line previous work. For example, larger households are on average more prone to consumption poverty and might react differentially (Lanjouw & Ravallion, 1995). The simple treatment indicators also turn significant for the kilogram consumption quantities in Column (4) and the monetary consumption value in Column (6).[14] Yet, our main indicator of interest, the caloric food consumption remains unaffected. This is in line with our hypothesis that the average treatment effect should be limited and rather uninformative as the primes are expected to particularly affect misreporting at the tails. For this purpose, a quantile regression analysis is taken out to provide more nuanced estimates, subsequently.

Table 2: Results from baseline estimation, model (1).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ln(Cons. Num.)	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Value)	ln(Cons: Cal.)	ln(Cons. Cal.)
Treatment	0.035** (0.016)	0.061* (0.033)	0.028 (0.018)	0.137*** (0.042)	-0.018 (0.018)	0.081* (0.039)	0.019 (0.028)	0.001 (0.067)
Observations	3,955	3,955	3,955	3,955	3,955	3,955	3,955	3,955
R-squared	0.001	0.273	0.001	0.070	0.000	0.078	0.000	0.123
State FE	NO	YES	NO	YES	NO	YES	NO	YES
Month FE	NO	YES	NO	YES	NO	YES	NO	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
Controls Interacted	NO	YES	NO	YES	NO	YES	NO	YES

Note: Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Columns (1-2) are measured on the household level. Columns (3-8) refer to per capita OECD adult equivalents. A full set of coefficients for control variables can be found in Table 14 the appendix.

[14] Unintuitively, with regard to the monetary consumption values in column (5), negative coefficients are estimated, contradicting a higher consumption quantity. In line with other studies, this could be explained by larger households buying larger quantities and, hence, consuming more while paying lower bulk purchasing prices (Deaton & Paxson, 1998).[14]

To capture this heterogeneity across consumption levels, quantile regressions are applied. Results are shown in Figure 4 and Table 3.

Table 3: Results from quantile regressions of different outcome variables.

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.165**	0.342***	0.079	0.235*
	(0.064)	(0.079)	(0.068)	(0.127)
Q0.25	0.058**	0.201***	0.198***	0.140*
	(0.028)	(0.067)	(0.053)	(0.080)
Q0.5	0.018	0.136**	0.119**	0.042
	(0.032)	(0.056)	(0.050)	(0.062)
Q0.75	0.047	0.114**	0.071	0.032
	(0.034)	(0.050)	(0.051)	(0.067)
Q0.9	-0.016	0.049	-0.015	0.013
	(0.028)	(0.050)	(0.054)	(0.064)
Observations	3,955	3,955	3,955	3,955
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

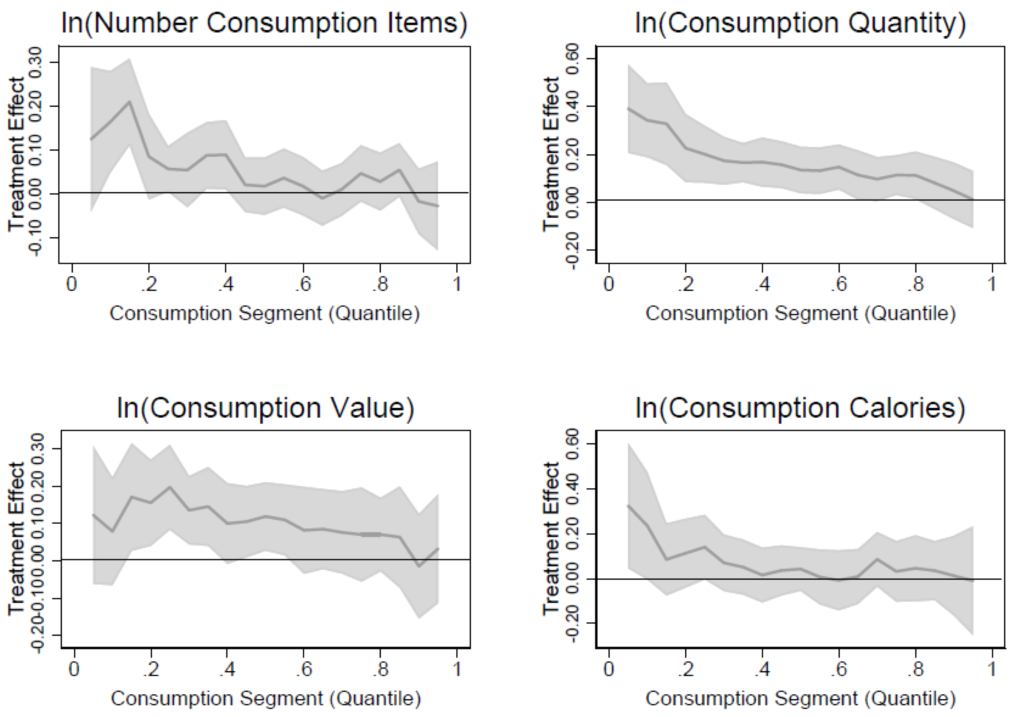
Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

The priming significantly increases reported consumption in lower quintiles. Significant treatment effects occur mainly for the number of consumption items and the quantities in kilogram.[15] Monetary and caloric consumption measures are less strongly affected (Figure 4). The latter might also be less susceptible to deliberate misreporting as they depend in part on variables over which the respondent has no control as the pure consumption quantities are scaled by calorie levels or deflated.[16]

[15] For an overview of the point estimates please consult Table 3 in the analytical appendix.

[16] Conditional quantile regressions are sometimes considered as uninformative as they describe the effect on the distribution rather than on the individual. Hence, we also consider unconditional quantile regressions in the appendix. Results are robust and support an upward shift in lower quantiles of the outcome variables (Table 12).

Figure 4: Treatment effects across quantiles.



Note: Treatment effects and confidence intervals plotted for different quantiles.

Ultimately, we are interested in the question if the prime is sufficiently strong to shift a significant share of the distribution to more credible consumption levels both in terms of monetary and caloric food consumption. For this purpose, we construct four dichotomous indicators. Those are equal to one if (i) respondent households surpass the caloric subsistence level of 1200 kcal or (ii) the recommended level of caloric intake of 2100 kcal. Two further dummies are created at (iii) 66.66 percent and (iv) 100 percent of a normalized poverty line, which is scaled by the fact that only core consumption items were assessed consistently across all surveys. Table 4 depicts results for the three threshold using model (3). Although the coefficients are mostly positive, only two coefficients turn significant in Column (2) and (3). Therefore, the results stress the nuanced effect of the prime, which only affects certain strata of the population.

Table 4: Results using poverty thresholds, model (2) and (3).

	(1)	(2)	(3)	(4)
VARIABLES	**>1200kcal**	**>2100kcal**	$>\left(\frac{2}{3}\right)$ ***Poverty Line***	**>Poverty Line**
Treatment	0.010 (0.027)	0.069* (0.037)	0.063* (0.037)	0.029 (0.036)
Observations	3,955	3,955	3,955	3,955
R-squared	0.067	0.098	0.118	0.135

State FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Controls Interacted	YES	YES	YES	YES

5. Treatment heterogeneity and robustness

Heterogenous effects:

If the primes would reduce misreporting, stronger effects are to be expected among subpopulations that have higher incentives to misreport, e.g., aid-reliant IDPs. In order to assess this channel more thoroughly, (i) heterogenous effects are estimated contingent on aid reliance and (ii) the sample is compared to a non-IDP comparison group.

Parts of the respondents from the CRS and HFS were also interviewed with regard to their previous support through UN agencies. This dummy indicator can be used for an assessment of heterogenous effects.[17] The model is analogous to equation (3), where we add UN assistance as a further control variable as well as an interaction term of UN assistance with the behavioral treatment. The results are depicted in Table 5 and indicate no clear pattern. Only for the number of consumption items a positive significant coefficient is found. The significant positive interaction of the treatment and previous aid exposure could be treated as some weak evidence that the prime is more effective for aid exposed IDPs, but should not be overstated due to the non-significance for the other three outcomes of interest.

Table 5: Channel – UN assistance.

	(1) ln(Cons. Num.)	(2) ln(Cons. Quant.)	(3) ln(Cons. Val)	(4) ln(Cons. Cal.)
Treatment	0.100	0.195**	0.171*	0.105
	(0.066)	(0.080)	(0.081)	(0.087)
UN Assistance	-0.028	-0.065	-0.152***	-0.143***
	(0.038)	(0.045)	(0.043)	(0.046)
Treatment*UN Assistance	**0.104****	**-0.059**	**0.016**	**0.011**
	(0.051)	**(0.060)**	**(0.061)**	**(0.064)**
Observations	2,204	2,204	2,204	2,204
R-squared	0.38	0.086	0.098	0.108
State FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES

[17] The results can only be interpreted as an explorative analysis as UN assistance was not balanced across treatment and control groups, where treatment households have a higher probability of being previously exposed to aid.

Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

The non-IDP subsample offers an interesting opportunity to assess the robustness of the results. Constraining the sample only on non-IDPs, we can estimate our results analogous to Table 3. The pattern of positive and significant treatment coefficients in the lower quantiles vanishes, except for Column (1). This could be interpreted as evidence that the light-touch method applied are more efficient for the vulnerable IDP population, which has higher incentives to indicate need than the non-IDPs. This would be in line with previous studies (e.g., Cilliers, Dube, & Siddiqi, 2015) suggesting a high degree of social desirability bias in the setting of foreign assistance. Specifically, the populations exposed to development aid, in our setting the IDPs, would be more likely to provide socially desirable answers to signal their “worthiness” for assistance. This corresponds to Table 5, providing some weak evidence that the primes are more effective for respondents relying on UN aid. It would be of particular interest to examine those heterogenous effects based on more fine-grained data on neediness and degree of aid reliance of recipients. For this purpose, however, a “true” benchmark would be needed. As administrative data is non-existent or of poor quality, an alternative for future research might be to build on measures from qualitative work as suggested by Blattman, Jamison, Koroknay-Palicz, Rodrigues, & Sheridan (2016). Moreover, one should be careful to draw too strong conclusions from these results as the number of observations is limited in this comparatively small sub-sample.

Table 6: Quantile Regressions – reduced sample (only non-IDPs).

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	-0.027	-0.069	-0.026	0.032
	(0.079)	(0.102)	(0.110)	(0.113)
Q0.25	0.148**	-0.052	0.012	-0.057
	(0.073)	(0.095)	(0.107)	(0.122)
Q0.5	0.067	-0.041	-0.032	0.044
	(0.067)	(0.081)	(0.100)	(0.100)
Q0.75	-0.071	-0.072	-0.015	-0.052
	(0.054)	(0.080)	(0.092)	(0.080)
Q0.9	-0.041	0.157	0.074	0.119
	(0.047)	(0.105)	(0.144)	(0.127)
Observations	780	780	780	770

Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

Robustness:

In line with hardly credible low consumption levels, misreporting could be considered to be more prevalent at the tails of the distribution, hence, among the extreme values. On the one hand, it makes, thus, sense to consider those outliers. On the other hand, it is problematic to base the inference mainly on those extreme values. Ideally, one would know how to distinguish the intentionally misreported outliers and the ones that are caused by errors in reporting or data entry. The log normalization in the main analysis is chosen as a compromise of keeping most data possible, but making estimates less susceptible to outliers. This suggests two natural robustness checks: (i) in a more liberal setting, the outcomes in levels are used and (ii) in a more conservative setting, the outliers at the 5th and 95th percentile are discarded. Regression results using the levels are depicted in Table 7.[18]

Table 7: Quantile Regressions – outcomes in levels.

	(1)	(2)	(3)	(4)
Outcome Variables	Cons. Num.	Cons. Quant.	Cons. Val.	Cons. Cal.
Q0.1	0.544**	0.741***	9.440	229.126*
	(0.254)	(0.173)	(13.305)	(136.013)
Q0.25	0.298*	0.675**	60.585***	179.447
	(0.156)	(0.224)	(16.261)	(157.584)
Q0.5	0.151	0.638*	49.404**	197.589
	(0.194)	(0.280)	(20.477)	(192.043)
Q0.75	0.341	0.700**	19.499	281.317
	(0.246)	(0.339)	(30.762)	(286.250)
Q0.9	-0.077	0.609	-30.117	-279.159
	(0.333)	(0.540)	(41.581)	(284.569)
Observations	3,955	3,955	3,955	3,955
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES

[18] As scaling of the outcome variables is different – e.g., the outliers with regard to consumption quantity in kilograms might not correspond to the consumption quantity in calories – the outliers for one measure do not always correspond to outliers in the other measure. In order to guarantee that we still base the inference on the same observations, outliers from all corresponding variables are dropped, which explains that the resulting sample is smaller than 90 percent of the full sample.

Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

Table 8 depicts the results without outliers and indicates a slightly less nuanced pattern. In line with our hypothesis of stronger misreporting tendencies on the extremes, Column (1) indicates significant treatment effects at the 10th and 25th percentile. Although significant treatment effects among higher quintiles can be found in Column (2) and (3), the coefficients for the 25th percentile are quantitatively larger. Finally, with regard to caloric consumption in Column (4) statistical significance vanishes, but the largest coefficient is to be found in the 10th percentile. Hence, although the pattern gets weakened when excluding outliers, the prime still significantly affects the reported consumption quantities with stronger effects in the lower quantiles.

Table 8: Quantile Regressions – without outliers.

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.124**	0.106	0.085	0.058
	(0.049)	(0.067)	(0.064)	(0.091)
Q0.25	0.045*	0.139**	0.162***	0.042
	(0.027)	(0.055)	(0.044)	(0.077)
Q0.5	0.000	0.065	0.119**	0.037
	(0.032)	(0.050)	(0.046)	(0.059)
Q0.75	0.028	0.077*	0.086*	0.049
	(0.032)	(0.043)	(0.048)	(0.063)
Q0.9	-0.027	0.064	0.027	0.039
	(0.023)	(0.039)	(0.049)	(0.051)
Observations	3,711	3,605	3,576	3,500
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

Regression techniques, which are based on assumptions for large samples drawn from finite populations, are often not suitable in the context of randomized experiments (Heß, 2017). The uncertainty is in this case not coming from the sampled units observed, but from the fact that we can only observe one of the potential outcomes, which is due to the treatment applied to the different units (Athey & Imbens, 2017). One approach would be to take the randomization explicitly into account and follow R.A. Fisher's idea of statistical inference via permutation tests of treatment allocation (Young, 2016). The idea is to assume uncertainty about the treatment allocation and compare the actual treatment allocation to re-randomizations. The results of this exercise are depicted in Table 11, underscoring the robustness of the main results.

Table 9: Results from baseline estimation, model (2), with random-inference based p-values.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ln(Cons. Num.)	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Val.)	ln(Cons. Cal.)	ln(Cons. cal.)
Treatment	0.0348** (0.0200)	0.0614** (0.0560)	0.0281 (0.1300)	0.1374*** (0.0020)	-0.0178 (0.2820)	0.0812** (0.0340)	0.0189 (0.4980)	0.0007 (0.9940)
Observations	3,995	3,995	3,995	3,995	3,995	3,995	3,995	3,995
R-squared	0.0012	0.2744	0.0003	0.0805	0.0006	0.0725	0.0001	0.1232
State FE	NO	YES	NO	YES	NO	YES	NO	YES
Month FE	NO	YES	NO	YES	NO	YES	NO	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
Controls Interacted	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1.

6. Discussion and Conclusion

The conflict in South Sudan displaced circa two million persons, constituting more than 15 percent of the country's population (UN OCHA, 2017). Moreover, the majority of population is living in extreme poverty (The World Bank, 2016). Humanitarian crises like the one in South Sudan ask for well targeted policy responses, which address the population strata with the highest need first. This, however, is no arbitrary task as aid allocation mechanisms might set adverse incentives to underreport. Even given the extreme context, surveyed consumption levels indicate an unusually high share below subsistence levels.

For this purpose, this study assesses the effectiveness of a bundle of light-touch measures. In line with our hypothesis we find significant treatment effects, which cluster in lower (potentially underreported) consumption quintiles. Moreover, effects are stronger for the number of consumption items than for monetary consumption quantities, where former are more susceptible to deliberate misreporting. Furthermore, the significant treatment effects are driven mainly by the vulnerable IDP subpopulation, which are more likely to be in need for foreign aid. Primes can, hence, help to improve data accuracy and inform policy to develop durable solutions. However, results should be taken with a grain of salt as it is not possible to compare the reported consumption outcomes to more objective consumption data. Although the mortality rates among IDPs suggest that starvation is not happening systematically across the country, the precarious situation calls for further scrutiny.[19] Before adjusting poverty estimates a thorough comparison with more "objective" data from administrative, anthropometric or observational sources is needed. While this type of data was not available in IDP camps due to the fragile context, future research could validate this finding in other settings.

Moreover, unbundling the primes in different treatment arms could help to shed light on the underlying causal mechanisms. The underlying design of one treatment and control arm does not allow for further disentangling the results. However, if classical measurement error would be affected only, treatment effects of the primes should be uniform. In contrast, heterogenous effects across quantiles suggest that the targeting of intentional misreporting via the appeal to honesty and prime to report more accurately would be the driver of our results. In order to design more effective primes, disentangling the pathways and trying different combinations could be a beneficial way forward. Our research can be considered as an early step to employ priming for better targeted policy responses in challenging contexts, which might not only be applicable in South Sudan, but also in other contexts facing humanitarian crises.

[19] FEWS NET (2018).

References

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31(2)*, 3-32.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-Saying, Nay-Saying and Going to Extremes: Black-White Differences in Response Styles. *Public Opinion Quarterly 48(2)*, pp. 491-509.

Baranov, V., Bhalotra, S., Biroli, P., & Maselko, J. (2017). *Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Large Randomized Control Trial.* Bonn: Institute for the Study of Labor (IZA).

Blattman, C., Jamison, J., Koroknay-Palicz, T., Rodrigues, K., & Sheridan, M. (2016). Measuring the measurement error: A method to qualitatively validate. *Journal of Development Economics, 120*, 99-112.

Cilliers, J., Dube, O., & Siddiqi, B. (2015). The white-man effect: How foreigner presence affects behavior in experiments. *Journal of Economic Behavior and Organization (118)*, pp. 397-414.

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). "Business culture and dishonesty in the banking industry.". *Nature 516.7529*, 86.

Cohn, A., Maréchal, M. A., & Noll, T. (2010). Bad Boys: How Criminal Identity Salience Affects Rule Violation. *Review of Economic Studies 82(4)*, 1289–1308.

Cronbach, L. J. (1946). Response Sets and Validity. *Educational and psychological measurement (6)*, pp. 672-83.

Deaton, A., & Paxson, C. (1998). Economies of Scale, Household Size, and the Demand for Food. *Journal of Political Economy, 106(5)*, 897-930.

Deaton, A., & Zaidi, S. (2002). *Living Standards Measurement Study Working Paper No. 135.* Washington DC: The World Bank.

Devi, S. (2017, May 20). Famine in South Sudan. *The Lancet, 389(10083)*, 1967–1970.

Euler, M., Krishna, V., Schwarze, S., Siregar, H., & Qaim, M. (2017). Oil Palm Adoption, Household Welfare, and Nutrition Among Smallholder Farmers in Indonesia. *World Development*, 219-235.

FEWS NET. (2018). *Famine Early Warning Systems Network.* Retrieved March 22, 2018, from South Sudan Food Security Outlook - October 2017 to May 2018:

http://www.fews.net/sites/default/files/documents/reports/SOUTH%20SUDAN%20Food%20Security%20Outlook_102017_0.pdf

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography*, 115-132.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica (77:3)*, 953-973.

Gilens, M., Sniderman, P. M., & Kuklinski, J. H. (1998). Affirmative Action and the Politics of Realignment. *British Journal of Political Science 28(1)*, pp. 159-83.

Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review 95(1)*, pp. 384-94.

Hamilton, D. L. (1968). Personality Attributes Associated with Extreme Response Style. *Psychological Bulleting 69(3)*, pp. 192-203.

Haughton, J., & Khandker, S. R. (2009). *Handbook on Poverty & Inequality.* Washington DC: World Bank Publications.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *Stata Journal*, 630-651.

Hurd, M. D. (1999). Anchoring and Acquiescence Bias in Measuring Assets in Household Surveys. *Journal of Risk and Uncertainty 19(1/3)*, pp. 111-36.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica, 46(1)*, 33-50.

Kuklinski, J. H. (1997). Racial Prejudice and Attitudes Toward Affirmative Action. *American Journal of Political Science 41(2)*, pp. 402-19.

Lanjouw, P., & Ravallion, M. (1995). Poverty and household size. *The Economic Journal*, 1415-1434.

Lin, W., & Green, D. P. (2016). Standard Operating Procedures: A Safety Net for Pre-Analysis Plans. *PS: Political Science & Politics*, 495-500.

Mazar, N., & Ariely, D. (2006). Dishonesty in Everyday Life and Its Policy Implications. *Journal of Public Policy & Marketing 25(1)*, pp. 117-26.

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research 45(6)*, pp. 633-44.

McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Population Economics, 18(2)*, 229–260.

Pape, U., & Mistiaen, J. (2015). *Measuring Household Consumption and Poverty in 60 Minutes: The Mogadishu.* Washington DC: World Bank / Proceedings of ABCA Conference 2015.

Rasinski, K. A., Visser, P. S., Zagatsky, M., & Rickett, E. M. (2005). Using implicit goal priming to improve the quality of self-report data. *Journal of Experimental Social Psychology, 41(3)*, 321-327.

Ravallion, M., & Bidani, B. (1994). How robust is a poverty profile? *The world bank economic review, 8(1)*, 75-102.

Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions. *American Journal of Political Science 60(3)*, pp. 783-802.

Siegelman, L. (1981). Question-Order Effects on Presidential Popularity. *Public Opinion Qarterly*, pp. 199-207.

Stecklov, G., Weinreb, A., & Carletto, C. (2017). Can incentives improve survey data quality in developing countries? Results from a field experiment in India. *Series A - Statistics in Society*, 1-24.

Talwar, V., Arruda, C., & Yachison, S. (2015). The effects of punishment and appeals for honesty on children's truth-telling behavior. *Journal of Experimental Child Psychology*, 209-217.

The World Bank. (2015). *World Development Report: Mind, Society, and Behavior.* Washington DC: The World Bank.

The World Bank. (2016, October 20). *The World Bank in South Sudan*. Retrieved February 24, 2018, from South Sudan Overview: http://www.worldbank.org/en/country/southsudan/overview

UN OCHA. (2017, December). *United Nations Office for the Coordination of Humanitarian Affairs*. Retrieved February 24, 2018, from South Sudan: http://www.unocha.org/south-sudan

United Nations. (2017). *UN News*. Retrieved March 15, 2018, from Famine declared in region of South Sudan – UN: https://news.un.org/en/story/2017/02/551812-famine-declared-region-south-sudan-un

Van Den Broeck, G., & Maertens, M. (2017). Moving Up or Moving Out? Insights into Rural Development and Poverty Reduction in Senegal. *World Development*, 95-109.

Vinski, M., & Watter, S. (2012). Priming Honesty Reduces Subjective Bias in Self-Report Measures of Mind Wandering. *Consciousness & Cognition 21(1)*, 451-55.

Voors, M. J., Nillesen, E. E., Verwimp, P., Bulte, E. H., Lensink, R., & Soest, D. P. (2012). Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review, 102(2)*, 941-964.

Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association 60(309)*, pp. 63-69.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 817-38.

Young, A. (2016). *Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results.* London: London School of Economics, Working Paper.

Appendix:

Technical appendix

<u>Construction of the caloric food intake measure:</u>

While monetary poverty lines are a key metric, when identifying the poor, caloric food poverty headcounts are of equal relevance in our context. We create a food intake approximation by multiplying the quantities of food items from the core consumption survey with average caloric values of these products. The caloric intake c_i of household i is estimated as follows: $\text{caloric intake}_i = \frac{1}{\text{hhsize}_i} \sum \text{item}_j * \text{calories}_j * \text{quantity}_{ij}$.

Forty-three percent of household members are children, who naturally have lower consumption levels than adults. We can account for this by using adult equivalents (AE) and rely on OECD scales, which scales consumption of additional adults per household by factor 0.7 and of children by factor 0.5 (Haughton & Khandker, 2009).

$$AE = 1 + 0.7(N_{adult} - 1) + 0.5N_{children} \tag{1}$$

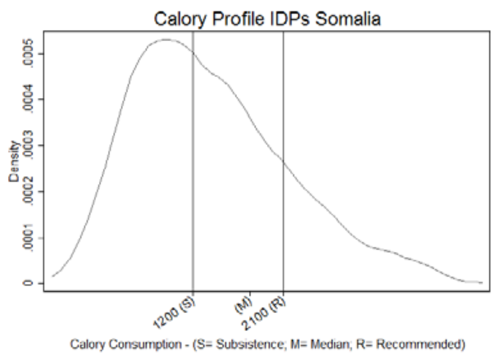
$$AE = 1 + 0.7(N_{adult} - 1) + 0.5N_{children}$$

Analytical Appendix

<u>Caloric food poverty in Somalia:</u>

Using the same approach, we derive caloric food intake measures, which motivated the notion that misreporting might be prevalent.

Figure 5: Calorie consumption - IDPs Somalia.



Source: Authors' calculations using HFS *Somalia Wave 1.*

Balance across survey strata:

Table 10: Treatment distribution by survey strata.

	State/Camp	Treatment with light-touch measures Control No.	Treatment No.	Total No.
CRS	JubaPOC	223	263	486
	Wau	294	284	578
	Bor	292	257	549
	Bentiu	294	297	591
IDPCSS	Juba POC1 – IDPCSS	976	965	1,941
HFS - Wave 4	Warrap	60	60	120
	Northern Bahr el Ghazal	50	61	111
	Western Bahr el Ghazal	62	58	120
	Lakes	50	54	104
	Western Equatoria	54	50	104
	Central Equatoria	38	40	78
	Eastern Equatoria	74	70	144
	Total	2467	2459	4926

Reaction to the light-touch method:

An overwhelming majority of respondents answered in a positive manner to the fictional scenario. Less than 10 percent of respondents answered that it is ok for the character in the fictional scenario to lie to his friend. [20]

[Prime to encourage more accurate reporting]: *I will give you a little scenario and would like to know what you think: John asks his good friend Deng if he has some money that he can lend him to help him pay for medicine for his sick son. Deng has money but was planning to buy cigarettes with it. He lies and tells John that he has none. Is it okay for Deng to lie to John?*

	Percent	N
Yes, it is okay for Deng to lie to John.	8.8	217
No, it is not okay for Deng to lie to John.	91.2	2,240
Total	100	2,457

This might be interpreted in two ways. First, it might point to a low fraction of respondents, who would be willing to lie, which would reduce the potential of finding significant treatment effects. Second, it could indicate that the prime would increase the propensity to report truthfully. However, as studies suggest a high social desirability bias in the aid allocation setting (Stecklov, Weinreb, & Carletto, 2017;

[20] The respondents, who find a lie inappropriate, have a higher share of male and unemployed household heads. Moreover, IDPs have a significantly lower probability to find a lie acceptable. *For a more detailed description of characteristics between respondents, who affirmed and rejected the lie, please see Table 11.*

Cilliers, Dube, & Siddiqi, 2015), implications should not be drawn too early and will be discussed in subsequent sections.

Appropriateness of lying:

It is puzzling that IDPs have on average a lower probability to report that they would find a lie appropriate when compared to non-IDPs (see Table 10). This is in line with more pro-social preferences of conflict affected populations found by Voors, Nillesen, Verwimp, Bulte, Lensink, & Soest (2012). However, this might be misleading, as the analysis of channels indicates that the significant treatment effects are attributable to the IDP subsample, which seem to be more likely to misreport.

Table 11: Distribution of respondents, who would find a lie (in-)appropriate.

	Yes, it is okay for Deng to lie to John.	No, it is not okay for Deng to lie to John.	Overall	(1) vs. (2), p-value
Household size	5.041	5.123	5.119	0.696
	(0.228)	(0.061)	(0.059)	
Gender of household head	0.327	0.456	0.445	0.000***
	(0.032)	(0.011)	(0.010)	
Literacy of household head	0.544	0.532	0.533	0.734
	(0.034)	(0.011)	(0.010)	
Household head completed some primary school	0.565	0.568	0.568	0.919
	(0.034)	(0.010)	(0.010)	
Is the household head employed	0.184	0.279	0.270	0.003***
	(0.026)	(0.009)	(0.009)	
Share of children in household	0.315	0.356	0.353	0.042*
	(0.019)	(0.006)	(0.006)	
Share of elderly in household	0.014	0.015	0.015	0.890
	(0.007)	(0.002)	(0.002)	
Level of Education of Household Head	2.060	1.967	1.975	0.205
	(0.075)	(0.022)	(0.021)	
non-IDP Population	0.212	0.155	0.160	0.029*
	(0.028)	(0.008)	(0.007)	
N	217	2238	2455	
Proportion	0.088	0.912	1.000	

*** p<0.01, ** p<0.05, * p<0.1.

Robustness of results using an unconditional quantile regression:

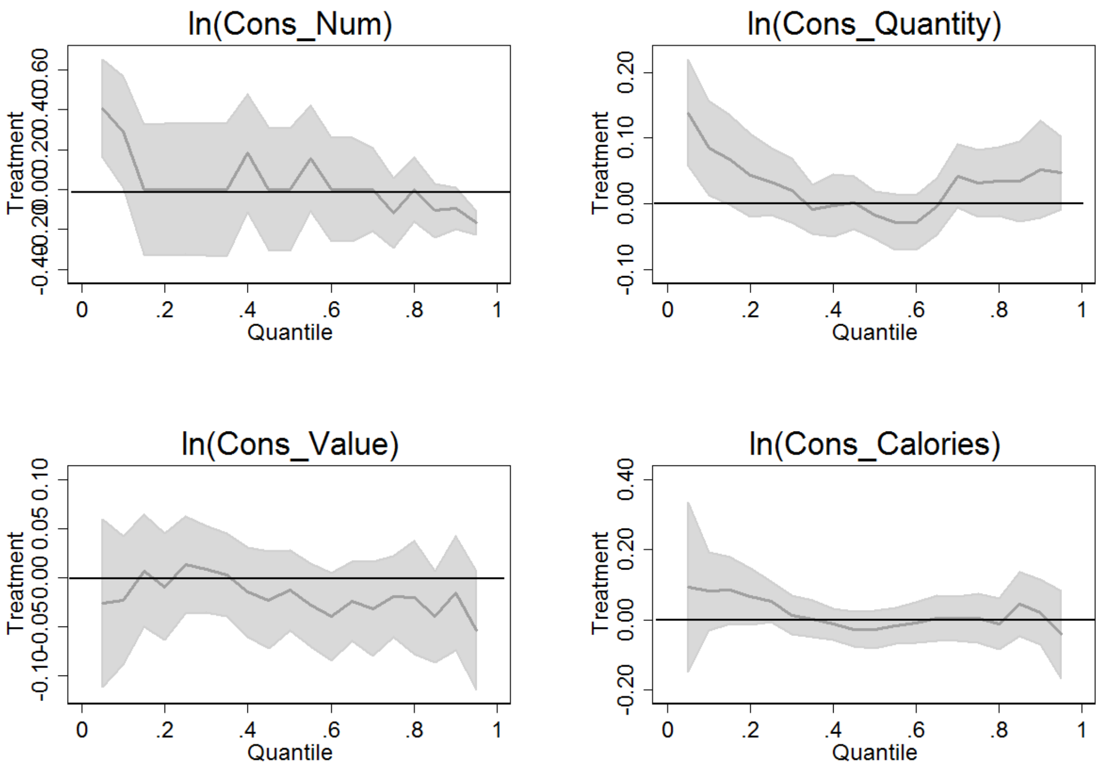
Conditional quantile regressions are sometimes critiqued on the ground that they would consider the treatment effect conditional on the distribution and not on the individual ranking. Therefore, we also replicate the main regressions within an unconditional quantile regression framework (Firpo, Fortin, & Lemieux, 2009). Table 12 depicts the results of unconditional quantile regressions. The results indicate a comparable pattern to Table 3. Especially, the quantities of consumption items and kilograms experience positive treatment effects in lower quantiles. Although higher quantiles are affected as well in Column (2), the largest effects can be found in the 10% quantile, which would be consistent with the hypothesis of more accurate answers among potentially under reporting households.

Table 12: Results from unconditional quantile regressions of different outcome variables.

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.105**	0.259***	0.076	0.134
	(0.046)	(0.090)	(0.079)	(0.145)
Q0.25	0.078**	0.210***	0.169***	0.075
	(0.032)	(0.067)	(0.062)	(0.077)
Q0.5	0.004	0.104**	0.118**	0.071
	(0.035)	(0.053)	(0.056)	(0.063)
Q0.75	-0.012	0.132**	0.067	0.025
	(0.040)	(0.066)	(0.059)	(0.089)
Q0.9	0.024	0.075	-0.003	0.062
	(0.044)	(0.087)	(0.077)	(0.119)
Observations	3,955	3,955	3,955	3,955
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1. Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

Figure 6: Treatment effects across quantiles (unconditional quantile regressions).



Note: Treatment effects and confidence intervals plotted for different quantiles.

<u>Robustness of results in extended IDP and non-IDP subsample</u>

Table 11 reports the corresponding results of a quantile regression for an extended sample of IDPs and Non-IDPs. Results correspond to the previously found larger coefficients in the lower quintiles. Coefficients are of similar size and the pattern remains qualitatively similar. However, statistical significance is reduced in column (3) and (4) with regard to the indicators that are measured with more noise (e.g., monetary consumption values and caloric consumption).

Table 13: Quantile Regressions – extended sample IDPs and Non-IDPs.

	(1)	(2)	(3)	(4)
Outcome Variables	ln(Cons. Num.)	ln(Cons. Quant.)	ln(Cons. Val.)	ln(Cons. Cal.)
Q0.1	0.136***	0.254***	0.072	0.153
	(0.049)	(0.058)	(0.067)	(0.094)
Q0.25	0.085***	0.123**	0.085	0.044
	(0.031)	(0.049)	(0.052)	(0.064)
Q0.5	0.024	0.088*	0.092**	0.037
	(0.029)	(0.049)	(0.043)	(0.053)
Q0.75	0.018	0.094**	0.052	0.028

	(0.031)	(0.042)	(0.044)	(0.052)
Q0.9	-0.019	0.058	-0.026	0.035
	(0.024)	(0.050)	(0.048)	(0.050)
Observations	4,735	4,735	4,735	4,735
Month FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Interacted Controls	YES	YES	YES	YES

Robust standard errors in parentheses (White, 1980): *** p<0.01, ** p<0.05, * p<0.1.
Column (1) is measured on the household level. Columns (2-4) refer to per capita OECD adult equivalents.

Robustness to per capita instead of per adult equivalents:

There is some uncertainty about the per adult equivalent scaling in the data. Ideally the distribution might be estimated from more fine-grained data on the intra-household consumption distribution. This is often not available, and, as Deaton & Zaidi (2002) summarize, “no satisfactory” scaling method is identified so far. Therefore, the OECD scaling methodology is still frequently used (e.g., Euler, Krishna, Schwarze, Siregar, & Qaim, 2017; Van Den Broeck & Maertens, 2017). Yet, one might be concerned that the main results are not robust to different scaling. Therefore, we construct our outcome measure alternatively using agnostic per capita scales. In line with the low consumption levels, the median of per capita calorie intake (1,589 kcal. per day) is well below the recommended daily intake of 2,100 kcal (Ravallion & Bidani, 1994). Almost one third of respondents (30.1 percent) report a calorie intake below the daily subsistence level of 1,200 kcal per day. In contrast, several respondents report overly high consumption levels, which surpass conventional consumption levels by far (> 4,000 kcal. per day). This supports previous evidence that misreporting is prevalent. As with the number of consumption items, the graph indicates that there is a slight shift in reported consumption among the treated regarding very low consumption levels.

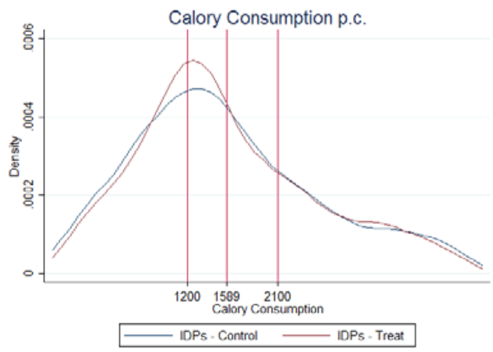


Figure 7: Calory Consumption p.c.

Analogous to the estimation in Table 3, Table 14 depicts the results of a quantile regression using agnostic per capita scales. The estimates indicate that the treatment effects remain stable and respondents would report statistically significantly higher quantities in Column (1) and Column (2) if treated. Hence, scaling does not explain our results, but is a factor to take into account, when interpreting the outcomes.

Table 14: Results from quantile regressions of different outcome variables (pc scales).

	(1)	(2)	(3)
Outcome Variables	ln(Cons. Quant. p.c..)	ln(Cons. Val. p.c.)	ln(Cons. Cal. p.c.)
Q0.1	0.358***	0.040	0.207
	(0.087)	(0.068)	(0.135)
Q0.25	0.161***	0.160**	0.076
	(0.059)	(0.053)	(0.081)
Q0.5	0.124***	0.079	0.073
	(0.057)	(0.054)	(0.066)
Q0.75	0.050	0.055	0.021
	(0.049)	(0.054)	(0.071)
Q0.9	0.057	-0.003	0.027
	(0.063)	(0.051)	(0.081)
Observations	3,955	3,955	3,955
Month FE	YES	YES	YES
State FE	YES	YES	YES
Controls	YES	YES	YES
Interacted Controls	YES	YES	YES

Table 15: Results – full set of (interacted) controls.

VARIABLES	(1) ln(Cons. Num.)	(2) ln(Cons. Quant.)	(3) ln(Cons. Val.)	(4) ln(Cons. Cal.)
Treatment	0.061*	0.137***	0.081**	0.001
	(0.033)	(0.042)	(0.039)	(0.067)
Household size	0.033***	-0.024***	-0.044***	-0.105***
	(0.005)	(0.005)	(0.005)	(0.008)
Female Gender of household head	0.009	0.043*	0.022	-0.047
	(0.018)	(0.026)	(0.025)	(0.039)
Share of children in household	0.106**	0.243***	0.190***	0.027
	(0.046)	(0.053)	(0.054)	(0.085)
1st component of asset PCA	0.026***	0.013	0.022***	0.039***
	(0.005)	(0.008)	(0.008)	(0.012)
Treatment * Household Size	-0.008	-0.015**	-0.009	0.006
	(0.006)	(0.007)	(0.007)	(0.011)
Female Gender of household head #0b.treat	-0.020	-0.007	-0.008	-0.059
	(0.028)	(0.036)	(0.036)	(0.054)
1.treat# Share of children in household	-0.001	-0.066	-0.107	-0.016
	(0.059)	(0.074)	(0.073)	(0.116)
1.treat#1st component of asset PCA	-0.017**	-0.003	-0.007	-0.003
	(0.007)	(0.011)	(0.010)	(0.016)
Observations	3,955	3,955	3,955	3,955
R-squared	0.274	0.073	0.080	0.123
State FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

<u>Household Size and purchasing prices per kilo price:</u>

In order to find out if the data bores out the pattern that larger households pay lower prices, e.g., due to bulk purchasing, we regress the log of the reported price on household size, state, month and consumption good specific fixed effects.

$$\ln(\text{price}_i) = \propto + \beta_1 \text{hhsize}_i + \gamma_s + \delta_t + \theta_g + \varepsilon_i \quad (5)$$

$$\ln(\text{price}_i) = \propto + \beta_1 \text{hhsize}_i + \gamma_s + \delta_t + \theta_g + \varepsilon_i$$

The results are depicted in Table 12 and indicate a negative average correlation. This supports the choice of interacting unbalanced controls with the treatment indicator.

Table 16: Correlation of household size and purchasing prices per kilo.

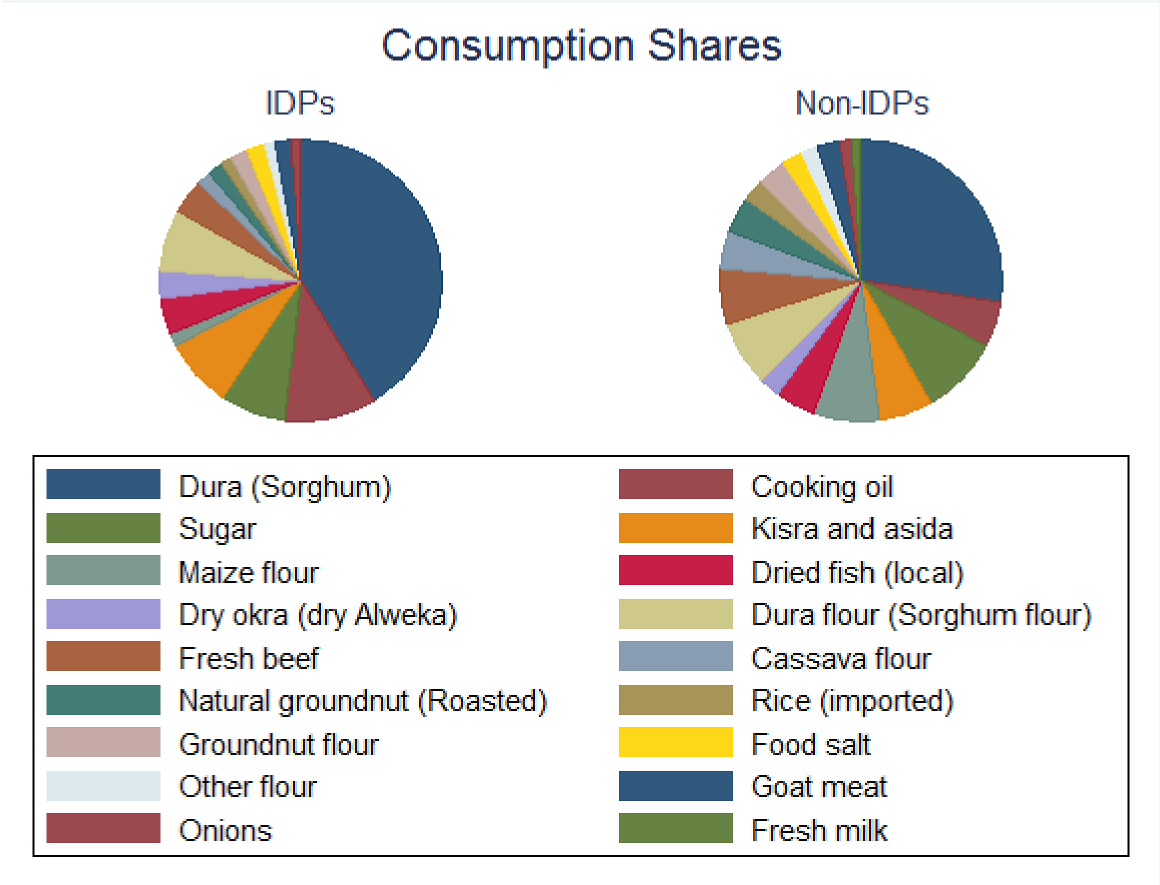
VARIABLES	(1) lnprice
Household size	-0.003**
	(0.001)
Observations	24,409
R-squared	0.548
State FE	YES
Month FE	YES
Item FE	Yes

Robust standard errors in parentheses (White, 1980):
*** p<0.01, ** p<0.05, * p<0.1

Consumption shares of IDP and non-IDP populations:

Figure 9 describes the consumption shares of IDPs and non-IDPs. While the figure shows that the diet of IDPs is slightly less diverse than the diet of non-IDPs, it is also revealed that large shares of IDP budget are spent on goods, which offer a high caloric intake per SSP spent, e.g., sorghum and cooking oil. The high energy content of IDP's food consumption also corresponds to the counter intuitive pattern found in the data, where IDPs consume less than non-IDPs in terms of monetary value, but more in terms of caloric food intake.

Figure 8: Consumption Shares (SSP values).



Note: The figure lists the consumption shares of items, which constitute at least 1% of household consumption.