

Hunger Safety Net Program (HSNP) and Index Based Livestock Insurance (IBLI) baseline comparison*

By Nathan Jensen^a, Megan Sheahan^a, Christopher B. Barrett^a, and Andrew Mude^b

^a Charles H. Dyson School of Applied Economics and Management, Cornell University

^b International Livestock Research Institute (ILRI)

November 2014

The Hunger Safety Net Program (HSNP), providing regular monthly cash transfers to beneficiaries, and Index Based Livestock Insurance (IBLI) Project, promoting and offering micro-insurance to protect pastoralists against livestock losses common in the face of drought, jointly operate in Marsabit district of northern Kenya. Because of the potential for confounding and interaction effects of both projects and given obvious room for research collaboration, the monitoring and evaluation (M&E) sub-component of the IBLI project, run by the International Livestock Research Institute (ILRI) in collaboration with Cornell University and other partners, was strategically designed to overlap and complement the HSNP M&E undertaken by Oxford Policy Management (OPM) and its partners.

As the data sets from both projects are now made publicly available and research findings about the impacts of these projects (separately, jointly, and comparatively) begin to be released, the Cornell-IBLI team thought it would be useful to provide a comparison of the two surveys, particularly since disparities may result in uneven interpretation of program impacts. Differences may arise from – among other things – variations in survey design, sampling strategy, data collection procedures, and data cleaning methodologies. In order to uncover these potential differences, we make a systematic study of the two surveys and compare baseline statistics across a number of indicators. The information underlying the discussion that follows is drawn from the publicly available codebooks and data sets for each project, unless otherwise noted.¹

Survey design and sampling procedures

Our first aim was to understand how any differences in survey design or household selection for the evaluation survey might lead to any observed differences in sample estimates. Table 1 provides a brief comparison of the sampling methodology, both for sampled sub-locations and for households within sub-locations. One obvious issue is that the full HSNP OPM sample covers a far broader geographic area than the IBLI sample covers: four districts for HSNP OPM (Mandera, Marsabit, Turkana, Wajir) relative to one for IBLI (Marsabit). Once we confine the HSNP OPM geographic sample to Marsabit district, however, differences in sub-location and household selection may still render Marsabit-level comparisons unreliable.

At the sub-location level, the IBLI team purposively selected 16 sub-locations based on agro-ecological and market access conditions, while the HSNP-OPM survey includes 12 Marsabit sub-locations in their sample that were chosen randomly, excluding “insecure” areas within the district (which were included

* The authors thank Casey Goldvale for research assistance, Fred Merttens for input on behalf of the HSNP OPM team, and Munenobu Ikegami for providing valuable feedback.

¹ The codebook, survey questionnaire, and data for the baseline IBLI Marsabit survey can be accessed here: <http://data.ilri.org/portal/dataset/ibli-marsabit-r1>. We use the publicly available codebook and baseline data for the HSNP OPM survey found here: <http://microdata.worldbank.org/index.php/catalog/1915> (downloaded on May 21, 2014).

in the IBLI sample). The HSNP baseline M&E report² states that “sub-locations were implicitly stratified by population density (households per square km), to ensure the sample was spread across both populous and sparsely populated sub-locations, and explicitly stratified by ‘old’ (greater) district” (p.10). For a list of the sub-locations included in each survey, see Table 2. This implies that while the HSNP OPM data can be considered representative of the population of those parts of Marsabit district that the OPM team deemed safe (the number of sub-locations excluded from the HSNP OPM sampling strategy due to safety considerations is unknown), the IBLI data is broadly characteristic of the range of conditions within the district. Importantly, neither sample can be considered fully representative of the Marsabit population since one excludes unknown sublocations and the other was purposively selected.

At the household level within sub-locations, the main difference in sampling seems to be that the IBLI survey randomly sampled households, stratified by wealth (using livestock holdings as a proxy), while the HSNP OPM survey samples randomly from HSNP eligible households in treatment areas and randomly from HSNP eligible but not yet participating households, stratified by settlement status, in control areas.³

Implementation of data collection and cleaning

Our second aim was to detect any differences in data collection or cleaning procedures that may lead to differences in the data across the two surveys. Table 3 includes our comparisons based on information specifically provided in the publicly available codebooks.

In terms of timing of data collection, the field work for the HSNP OPM survey was planned to span a 12 month period in order to be representative of the calendar year, however data collection actually took 14 months per the codebook (or 16 per the data, which appears to be an error). Per the “coversheet.dta” file in the HSNP OPM data, it appears that Marsabit households specifically were interviewed between August 2009 and October 2010, meaning also representative of the calendar year when limiting the sample to our district of interest (although based off of a very small sample of households within each month/year combination). The IBLI survey teams were fielded only in October and November of 2009, the short rainy season in Marsabit. The long span of the HSNP OPM data collection versus the relatively short span of the IBLI could result in big differences for some variables, especially those that commonly vary across time (income, expenditures, savings, herd sizes, etc.).

Data collection for the IBLI survey was handled by the IBLI research team within the International Livestock Research Institute (ILRI) alongside a small set of hired supervisors and large team of enumerators. All interviews for the baseline data were recorded via a paper questionnaire and hand-checked by supervisors before considering final. The HSNP OPM data collection was sub-contracted to Research Solutions, a local firm which administered the surveys via a computer-assisted device. No other information about enumeration or supervision of the HSNP OPM survey is available in the codebook.

² The “Kenya HSNP Monitoring and Evaluation Component Baseline Report” can be downloaded here: <http://microdata.worldbank.org/index.php/catalog/1915/download/29578>

³ HSNP eligibility was initially (before the baseline) determined for all households in all sublocations (control and treated) found in the HSNP survey. The treatment sublocations are those that started receiving transfers in the first wave of HSNP (2009), while the control sublocations do not begin receiving transfers until two years later, in 2011.

Entry for the IBLI data was done after the completion of field work by a supervised team at the ILRI office in Nairobi. Very basic data cleaning was handled by ILRI research staff and Cornell University graduate students, including renaming variables to be consistent over the full five year panel. No information has been made publicly available by the HSNP OPM team about the transfer between the computer-assisted software and usable Stata files and any important data cleaning considerations.

Since most information about the data collection, entry, and cleaning methods employed by OPM and its sub-contractor, Research Solutions, are not publicly available, we are unable to make claims about how any differences between the HSNP OPM and IBLI data sets may result from these operational differences, apart from the major difference in timing of data collection built into survey design.

Baseline summary statistics

Our third aim was to explore differences across descriptive baseline statistics over the full sample of households as described in Table 1 and especially among sub-samples of households that may be more directly comparable. Table 4 provides a tabulation of the chosen descriptive statistics. We expect that the samples used in columns 2 and 3, for HSNP OPM and IBLI, respectively, both representing the population of the surveyed sub-locations in Marsabit district, are likely the most comparable. T-tests were run to search for statistically significant differences in the means across the sub-groups (columns).⁴

Considerations with HSNP baseline statistics

All descriptive statistics using the baseline HSNP OPM data were computed using the included “hh_wt” variable, which should produce estimates representative of all households living in sub-locations covered by the evaluation. This is the same type of weight available through the IBLI survey. Most of the consumption and expenditure statistics were calculated using the aggregates found in a file contained in the baseline data set (see note to Table 4). Variables with high end outliers in the HSNP data were “winsorized,” whereby observations above the 99th percentile were replaced with the value at the 99th percentile. No other cleaning of the publicly available data was done.

Considerations with IBLI baseline statistics

Like the HSNP data, descriptive statistics derived from the IBLI data are weighted using the variable “weight,” which produces estimates representative of the sub-location and takes into account the distribution of herd size terciles therein. Note that the IBLI data uses self-reported HSNP participation at the household level and information from HSNP field staff in Marsabit to match each community with their appropriate targeting mechanism.⁵ According to those data, there is significant mis-targeting within HSNP communities (i.e., households meeting HSNP eligibility criteria but not receiving transfers) as well as a number of households that are not in HSNP targeted communities that report receiving HSNP transfers.⁶ The data were left unadjusted in both cases, and only minor cleaning of the publicly available data was done to produce all included statistics.

⁴ Stata do file documentation can be provided by the authors upon request.

⁵ Phase I of HSNP used three targeting mechanisms (age, dependency ratio, and community based targeting), which were randomly assigned at the community level.

⁶ Jensen, Mude, & Barrett (2014) "Index Insurance and Cash Transfers: A Comparative Analysis from Northern Kenya" contains a detailed analysis of HSNP adherence in the IBLI data, with respect to HSNP participation. This paper is not yet publicly available, but a presentation with an early version of these results can be found here: http://dyson.cornell.edu/faculty_sites/cbb2/presentations/AAEAIndexInsuranceCashTransfers.pptx.

Main findings from comparison of descriptive statistics

Based on the t-test of differences in average statistics across the two surveys and sub-groups, we note the following:

- Religion and language variables indicate significant differences in the sample households for each survey.
- Household members in the HSNP survey claim have a higher level of education and literacy than the average IBLI-surveyed household member, both in the full Marsabit and HSNP participant samples. That would be consistent with the hypothesis that educational attainment is lower in areas deemed insecure that were surveyed by the IBLI team but not by the OPM/Research Solutions team.
- IBLI-surveyed households are far more likely to own livestock and have larger herd sizes than HSNP-surveyed households. The HSNP OPM data, however, only includes information on owned livestock while the IBLI data includes herd size regardless of ownership status. Moreover, a report on the extent of pastoralism in the region by Krätli and Swift (2014) states that there was “strong under-reporting of livestock holdings” in the HSNP OPM data set which is “considered unavoidable as respondents knew that the goals of the survey was to identify poor households eligible for financial support” (p. 9).⁷ Given both the incongruence in what is included in the aggregate as well as this statement, it is difficult to know to what extent the differences we observe are guided primarily by true differences in sampled households, survey design, or the incentive structure underlying the responses. Again, these differences are consistent with the hypothesis that more secure areas depend less on livestock-based livelihoods and thus the differences in sampling protocols accounts for the difference in herd sizes.
- IBLI-surveyed households are less likely to farm or have agricultural land than HSNP-surveyed households. This is likely to be a reflection of the sampling methods used: the HSNP OPM surveyed sub-locations represent a random stratified sample of “secure” sub-locations within Marsabit while the IBLI survey sublocations were identified specifically to represent a range of market access and ecological conditions.
- There exists a difference in mean total savings values between IBLI-surveyed and HSNP-surveyed, where HSNP-surveyed households have more reported savings than their IBLI-surveyed counterparts. While there may be many reasons that households may misreport savings values, there is no indication of why this difference should necessarily exist between the two samples unless one believes that households in insecure areas are less likely to hold cash savings.
- IBLI-surveyed households also appear poorer than HSNP-surveyed households with respect to levels of cash income and overall consumption expenditures. IBLI-surveyed households, more specifically, have lower levels of food consumption expenditures and education expenditures. Differences in food consumption expenditure levels may be explained by the timing of data collection and recall period, however education expenditures should be annually representative. HSNP-surveyed households, on the other hand, have lower levels of non-food expenditures, which may be the result of smaller herd sizes and, therefore, livestock-related expenditures. Without knowledge of how the HSNP OPM consumption and expenditure aggregates were assembled, there is no way to know if the included categories drive the differences or if the differences come from actual difference in household wealth level or in consumption patterns due to the differences in spatial coverage (i.e., inclusion or exclusion of insecure areas).

⁷ Krätli S. and Swift J. (2014) “‘Counting Pastoralists’ in Kenya,” DLCI/REGLAP, Nairobi. Available for download here: http://www.disasterriskreduction.net/fileadmin/user_upload/drought/docs/1_Counting%20Pastoralists%20FINAL%2030%20April%202014.pdf

Discussion

This comparison reveals that the households sampled as part of the HSNP OPM and IBLI impact evaluation survey efforts are not identical, which comes through clearly in the small set of descriptive statistics provided. These differences could come from discrepancies in sampling strategy, survey weight construction, timing of data collection, and/or incentives to answer questions in particular ways, as discussed, or actual differences in survey questionnaire/instrument design or the presentation of the questions to respondents by enumerators, not discussed.

Apart from shedding light on these differences, it is difficult to say which data set should be favored for analysis, generally speaking, in this region. And although there are, at times, large differences between the data, in many cases the data are very similar. In the end, the decision likely should be motivated by the nature of the question to be addressed and the coverage of that information in each survey. But, when it comes to making comparisons of the two programs – HSNP and IBLI – one should remain cognizant of the many dissimilarities between the two complementary survey efforts and how those differences may drive results.

Table 1: Comparison of survey design and household selection

	HSNP OPM data (2009-10)	IBLI data (2009)
Districts	Mandera, Marsabit, Turkana, Wajir	Marsabit
Number of sub-locations	48	16
Sub-location selection method	12 sub-locations within each of the four districts were randomly selected using probability proportion to size (PPS) and implicit stratification by population density, including both treatment and control sub-locations. Only “secure” sub-locations considered. Further stratification within treatment sub-locations based on 3 HSNP targeting mechanisms (community-based, dependency ratio, age/social pension).	Purposively chosen sub-locations to represent a broad variation of livestock production system, agro-ecology, market accessibility, and ethnic composition. Half of the sub-locations (8) were targeted by HSNP during their first phase. All three targeting mechanisms were captured in the survey: age/social pension (3), dependency ratio (2), community based targeting (3). In addition, a 9 th community (Bubisa) started receiving HSNP transfers in 2011. We include HSNP households in Bubisa in our sample of HSNP households in Table 4.
Number of surveyed households	5,108	924
Surveyed household selection method	Using HSNP administrative records for Groups A and B (below). 66 sampled households per sub-location: - <i>Group A: Households in the treatment sub-locations selected for inclusion in HSNP;</i> - <i>Group B: Households in control sub-locations selected for inclusion in HSNP but with delayed payments.</i> Using household listings undertaken in a sample of three settlements within each sub-location for Groups C and D (below), with settlement stratification based on the residency status of settlements: - <i>Group C: Households in treatment sub-locations that were not selected for inclusion in HSNP;</i> - <i>Group D: Households in control sub-locations that were not selected for inclusion in HSNP.</i>	Number of households in each sub-location chosen to be proportional to 1999 household population census (with 30 minimum and 100 maximum restrictions) Random sample of households within a sub-location was stratified by wealth (i.e., livestock holding) tercile as determined by key informants’ knowledge of households in the area
Survey weight methodology	Two different survey weights are made available with the data, those that produce estimates representative of (1) all households living in sub-locations covered by the evaluation and (2) the entire population of secure sub-locations within each district.	Produce estimates representative of the population and livestock holding distribution at the sub-location level.

Note: All details of the HSNP sampling strategy are taken directly from the publicly available baseline codebook.

Table 2: Sublocations sampled in each survey

HSNP OPM data (2009-10) <i>Marsabit only</i>	IBLI data (2009)
Badasa	Dakabaricha
Mata Arba	Dirib Gombo
North Horr	Sagante
Maikona	Bubisa
Laisamis	El Gade
Kamboye	Kalacha
Hulahula	Turbi
Majengo	Karare
Lonyoripichau	Kargi
Korr	Kurkum
Marsabit Township	Logologo
Wabera	Illaut
	Lontolio
	Loyangalani
	Ngurunit
	South Horr

Source: The HSNP sublocations can be found in section A.3 of the “HSNP Impact Evaluation Final Report: 2009-2013” available at

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/284251/Kenya-Hunger-Safety-Net-Programme-2009-2012.pdf.

Table 3: Comparison of survey implementation and data cleaning methods

	HSNP OPM data (2009-10)	IBLI data (2009)
Timing of survey	Designed to take 12 months, but the codebook states it was actually done in 14 months. However, a tabulation in the data shows that data was actually collected between August 2009 and December 2010 (with no data collection in December 2009), which actually spans 16 months. The sequence in which the sampled evaluation sub-locations are targeted and surveyed was determined randomly.	October and November 2009
Medium of data collection	Computer Assisted Personal Interview (CAPI)	Paper questionnaires
Management of data collection (field work)	Sub-contracted to Research Solutions	Implemented by a mixed team of IBLI-staff and locally-hired supervisors with enumerators mostly from Marsabit.
Experiences from data collection (field work)	Delays and challenges (insecurity, migrations, and impassable roads) made the data collection two months longer than anticipated.	Despite what could have been 7 sub-teams, only 2 drivers were available to the full field team, which limited the extent to which supervisors could engage in supervising enumerators and checking completed surveys in a timely manner.
Management of data entry	Because household interviews were conducted with CAPI, no data entry was required. Each sub-location's data was uploaded by a team leader at the end of each month.	For the first two rounds data was collected data using paper questionnaires which entered using Microsoft Access by four data entry clerks under the supervision of an ILRI staff member in the Nairobi office after the completion of field work. In the following rounds, data was collected using SurveyBe software and CAPI.
Management of data cleaning	The downloaded CAPI data was subject to an iterative data cleaning process between the headquarters in Nairobi and the Oxford-based team.	The data received basic cleaning by a team of ILRI staff members and Cornell University graduate students after each round. After round 5, a substantial cleaning effort was undertaken between April and September 2014 so that the first five rounds could more easily be merged for panel data analysis. This involved creating consistent variable names and labels, value, value label, and missing value coding over survey rounds. No data were otherwise altered during this process.

Table 4: Comparison of descriptive statistics using HSNP and IBLI data over a number of sub-samples

Sample or sub-sample	HSNP OPM data (2009-10)		IBLI data (2009)			
	(1) Overall sample	(2) Marsabit househol ds only	(3) Overall sample	(4) HSNP households only	(5) Non-HSNP area, HH meets ≥1 HSNP criteria ¹	(6) Non-HSNP area, predicted participation criteria ²
Number of households	5,109	1,300	924	384	311	215
Household characteristics						
Main household religion=Muslim (%)	0.54	0.38	0.19 ***	0.21 ***	0.25 ***	0.23 ***
Main household religion=Catholic (%)	0.21	0.36	0.31 *	0.24 ***	0.22 ***	0.18 ***
Main household language=Somali (%)	0.36	0.01	0.00 **	0.00 ***	0.00 ***	0.00 ***
Main household language=Rendille (%)	0.08	0.33	0.22 ***	0.38	0.32	0.37
Percent of household that is male (%)	0.50	0.50	0.50	0.50	0.51	0.51
Household size (AE) ^{3,4}	4.26	4.10	4.34 **	4.35	4.57 ***	4.60 ***
Number of adults (yrs>=18) in household	2.42	2.45	2.41	2.45	2.48	2.62
Number of children (yrs<17) in household	2.98	2.69	2.97 ***	2.91 *	3.17 ***	3.00 **
Number of literate household members	1.48	1.78	0.95 ***	0.93 ***	1.08 ***	1.14 ***
Number of members with any education	2.06	2.35	1.52 ***	1.47 ***	1.71 ***	1.66 ***
Household member education in years	1.71	2.82	1.52 ***	1.30 ***	1.42 ***	1.42 ***
Household member education in years >0	5.29	6.19	4.96 ***	4.75 ***	4.69 ***	4.99 ***
Household head education in years	1.84	3.18	1.08 ***	0.45 ***	0.69 ***	0.35 ***
Household head education in years >0	9.22	10.04	8.17 ***	7.12 ***	8.58 **	7.50 **
Number of members currently in school	1.22	1.23	1.14	1.16	1.36	1.34
Average household age	22.9	23.4	22.0 **	22.7	23.0	25.2 **
Average head age	43.2	42.8	45.3 ***	46.0 ***	48.1 ***	50.1 ***
Household received HSNP transfer	0.51	0.49	0.37 ***	1.00 ***	0.63 ***	0.73 ***
Household livestock						
Households that keep any livestock (ratio)	0.70	0.67	0.88 ***	0.88 ***	0.91 ***	0.89 ***

Sample or sub-sample	HSNP OPM data (2009-10)		IBLI data (2009)			
	(1) Overall sample	(2) Marsabit househol ds only	(3) Overall sample	(4) HSNP households only	(5) Non-HSNP area, HH meets ≥1 HSNP criteria ¹	(6) Non-HSNP area, predicted participation criteria ²
Number of households	5,109	1,300	924	384	311	215
Total current herd size (TLUs)	7.38	7.45	14.74 ***	20.47 ***	20.33 ***	23.71 ***
Total current herd size (TLUs) >0	10.72	11.24	16.79 ***	23.13 ***	22.26 ***	26.76 ***
Ratio of cattle to total herd (not TLU-based)	0.07	0.14	0.14	0.07 ***	0.10 **	0.10 *
Ratio of camels to total herd (not TLU-based)	0.08	0.06	0.07	0.08 **	0.08 **	0.08 **
Household agriculture						
Households that own or farm agricultural land (ratio)	0.09	0.15	0.12 *	0.08 ***	0.09 ***	0.08 ***
Total irrigated land owned by household (hectares)	0.16	0.00	0.00	0.00	0.00	0.00
Total non-irrigated land owned by household (hectares)	0.23	0.75	1.21	0.72	1.05	0.76
Total land owned by household (hectares)	0.39	0.75	1.21	0.72	1.05	0.76
Household expenditures, income and savings (KSh)⁵						
Total savings	3,459	11,679	5,430 ***	1,879 ***	8,196	2,244 ***
Total cash income over the past 12 months	43,606	47,338	42,635	30,519 ***	46,683	41,757
Total amount owed at time of survey	2,422	3,068	2,824	1,794	4,468	5,202
Total monthly consumption expenditure ³	8,205	7,582	5,458 ***	5,285 ***	6,439 **	5,844 ***
Total monthly consumption expenditure per AE ³	2,043	2,010	1,424 ***	1,435 ***	1,572 ***	1,489 ***
Monthly food consumption expenditure per AE ³	1,544	1,505	974 ***	1,022 ***	1,044 ***	1,046 ***
Food share of consumption expenditure (%) ³	0.78	0.78	0.72 ***	0.69 ***	0.69 ***	0.68 ***
Total monthly non-food	1,386	1,338	1,870 **	1,672 **	2,408 **	1,967 ***

	HSNP OPM data (2009-10)		IBLI data (2009)			
	(1) Overall sample	(2) Marsabit househol ds only	(3) Overall sample	(4) HSNP households only	(5) Non-HSNP area, HH meets ≥ 1 HSNP criteria ¹	(6) Non-HSNP area, predicted participation criteria ²
Number of households	5,109	1,300	924	384	311	215
consumption *						
Mean monthly education expenditure per child (6- 17 years) ³	95	146	90 ***	100	116	142
Total education expenditure ³	212	279	236	229	282	317

Notes: The asterisks represent the statistical significance of t-tests performed between the means from column 2 (HSNP OPM data with sample from Marsabit district only) and a range of comparable sub-groups in the IBLI data (columns 3 through 6). *** 0.01, ** 0.05, * 0.1.

¹ HSNP randomly assigned one of three targeting mechanisms to each HSNP community. In column (5), we restrict the sample to those households that are not in HSNP communities but do meet at least one of the three HSNP targeting mechanisms.

² In column (6), we simulate HSNP eligibility using a probit model to estimate the relationship between HSNP participation (as the dependent variable) and 15 covariates, including the social pension and dependency ratio eligibility criteria. Those estimates are used to predict likelihood of participation. Within each community, the households that fall within the top 50% highest predicted likelihood of participation are designated 'predicted participants'. Households that meet that criteria and are from non-HSNP communities are the sample used for these statistics.

³ Variables are taken from the file "disaggregation data" for the HSNP OPM statistics and are not calculated by the authors.

⁴ Adult equivalents (AE) are calculated according to age in years as the following: AE=1 if 15<age<61, AE=0.7 if 4<age<16 or age>60, AE=0.5 if age<5.

⁵ Because the raw data used to calculate the "disaggregation data" for the HSNP OPM data is not available, we are restricted to using the variables that exist within that dataset. We have attempted to construct similar variables using the IBLI data, but are not always able to do so. For the data presented under the heading "Household expenditures, income and savings (KSh)", the variable descriptions (in the left hand column) are drawn directly from the HSNP data file when HSNP data exists. To construct equivalent variables with the IBLI data, we estimate monthly averages as 1/12th the annual consumption. Variables including the words cash or expenditures use only data related to cash exchanges.