

**EGERTON UNIVERSITY  
TEGEMEO INSTITUTE OF AGRICULTURAL POLICY AND  
DEVELOPMENT**

**&**

**MICHIGAN STATE UNIVERSITY**

**TEGEMEO AGRICULTURAL POLICY RESEARCH ANALYSIS (TAPRA)  
PROJECT**

**HOUSEHOLD SURVEY 2010 DATA DOCUMENTATION**

**Support for this study was provided under the Tegemeo Agricultural Policy and Analysis (TAPRA) project, supported by the United States Agency for International Development / Kenya. Supplementary support for this study is provided by the Office of Sustainable Development, Africa Bureau, and USAID/Washington.**

**2010**

## **SAMPLING METHOD**

This TAPRA sample is only composed of TAPRA households that were interviewed in 2007.

The sampling method used was similar across all the sites and is described below:

1. Within the designated area of study (considering AEZs and other criteria), all the villages/sub-areas were listed with the help of the administration or chief.

AEZ, population, and whether the district belonged to the "original" KAMPAP districts (districts where Tegemeo had conducted much research before and had some supplementary data and information on) were some of the key factors in this exercise.

The first step was to identify the spatial distribution of AEZ in the district. The idea was to capture as much of the diverse conditions as possible in our sampling. From this step we were able to classify certain areas within AEZ with the help of the Ministry of Agriculture officers. Each district was in turn divided into divisions, locations and sub-locations and then villages/wards. From the district level we were able to pick representative divisions with the help of the district officers. I believe that we also took into account the populations and AEZ conditions within these areas to help us select these divisions. Because not all divisions could possibly be visited we picked a random sample of these divisions for further follow-up. These were selected with the idea of incorporating the diversities that were inherent in each district that we visited (a representative sample).

At the division level, a similar exercise was carried out with the help of the Ministry officials. Then the locations were selected randomly. This was followed by sub-locations and then finally the villages/clusters below.

2. From this list (and considering the sample size required from the area) a number of villages were randomly selected by picking from the list above.
3. For the selected villages, and with the help of the administration and key informants, we listed all household units within the village by head of household.
4. In most cases the list above exceeded the sample size requirements for the area. Accordingly we used the 'universal' KAMPAP sampling technique to select households for interview.

Universal KAMPAP sampling technique description: Most village elders/chiefs have a pretty comprehensive list of householders' names. Suppose we had a total list of 76 households for a village or cluster from the chief (numbered from 1 to 76). Assume too that all we needed was to interview 12 households from this village. The objective was to randomly select every sixth household to get the 12 we needed (approx  $76/12=6$ ). The question is, on a numerical list of 1 to 76 where do you start the selection (is it 1,2,3,4,5 or 6)? We wrote the numbers 1 to 6 on different pieces of paper of similar size, folded and mixed them up. Then we asked a villager or the chief to pick one of these papers and reveal the number. Suppose the number picked is 3; then we proceeded to pick the households starting from the third on the list, i.e. 3,9,15,21,27 etc.

5. It happened that in some areas some of the selected households within a village had household heads that were related by marriage or some other kinship relationship (though the samples had been selected randomly in the first place). In such instances one could find cousins, brothers, uncles, etc who had bought farms in the same area and over the years subdivided their farms to their children, etc but all these were clearly separate households with different management styles and approached their household decisions separately. Relationships among households do not necessarily imply joint decision-making.

6. In conclusion the samples were as random as was possible and the data should be able to express this random nature despite some pockets here and there of 'relationships', if one may.

**SUMMARY OF HOUSEHOLDS SURVEYED**

Out of the 2010 Tampa survey sample of 1372 households, there were 1309 households that were interviewed. There were 30 households that were not interviewed in 2007 for various reasons (but were not dissolved or moved away). Those households were not included in the sample for the 2010.

Turkana and Garissa were not interviewed. The argument was that the original sample was not typical of the area. Garissa for example, had households who were engaged in irrigation which gave an indication that the area was highly productive. Turkana district did not give the typical scenario of a nomadic pastoralist household. Moreover, in Turkana, it was difficult to generate panel data due to the nomadic nature of the household.

It is important to note that there was no replacement of households in the TAPRA sample for this survey.

**intview Why HH is not able to participate in interview**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Completed	1309	95.4	97.5	97.5
	1 Head & spouse dead	5	.4	.4	97.9
	2 Head & spouse separated	1	.1	.1	98.0
	3 Refused	3	.2	.2	98.2
	4 HH mems cannot be found	3	.2	.2	98.4
	5 Family commitments (burial, wedding...)	1	.1	.1	98.5
	6 HH moved from area	12	.9	.9	99.4
	7 HH mems working outside area	1	.1	.1	99.5
	8 Displaced by post election violence	5	.4	.4	99.9
	10 HH dissolved	2	.1	.1	100.0
	Total	1342	97.8	100.0	
Missing	-7 Not interviewed	30	2.2		
Total		1372	100.0		

The data for page one of the survey instrument are contained in two files: allhhid10.sav and hhidfinal10.sav. The first file (allhhid10.sav) contains all the original selected households to be interviewed. The second file (hhidfinal10.sav) contains only those households that completed the interview for this 2010 survey (1309 hhids) of the TAPRA sample. This file should be used to merge the identifying characteristics to the other files as needed.

## **DATA FILE DESCRIPTIONS FOR RURAL HOUSEHOLD SURVEY**

Directory Structure: - In the subdirectory where you keep all your files you should create a directory called “Kenya”. The next level is called “Kenyahh2010”. There are several subdirectories off this directory:

C:\...

  \Kenya

    \Kenyahh2010

      \anal- analysis files and syntax.

      \augdata- final data files to be used for analysis

      \docs- documentation of all files including the survey instruments and enumerator manual

      \lookup- lookup data files and syntaxes.

      \NewVars- files and syntaxes that have been computed and ready for analysis

        \demog – adults equivalents and household size

        \income – income variables

      \tmp- used to store temporary files that the analyst does not plan to retain.

Variables to identify location:

  aez - agricultural ecological zones

  aezsmall - aez subdivided into more specific zones

  zone – habitat zones

  prov (province)

  dist (district),

  div (division),

  loc (location),

  subloc (sub-location),

  vil (village)

In addition to the identifying variables listed above GPS coordinates were collected and recorded for all the households that were interviewed. The GPS coordinates were collected in decimal degrees for this survey, whereas in the 2007 survey they were collected in degrees, minutes and seconds.

## DATA FILES

Directory: C:\...\Kenya\Kenyahh2010\augdata

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
Household identification	hhidfinal10.sav	hhid	1,309		All households that <b>completed</b> the interview – <b>use this file to merge in location variables</b>
Household level questions	+hh10.sav	hhid	1,309		General household level questions.
<p><b>Notes on hh10 file:</b>            GPS coordinates were collected in decimal degrees for this survey. In 2007 the data were collected degrees, minutes and seconds. There are several cases where the hh does not know how far the nearest NCPB depot is. The enumerator did not then ask if they sold to the NCPB and if not, why not.</p>					
Household	allhhid10.sav	hhid	1,342		All households that were – <b>use only if want to know</b> <b>households were not in</b> Crop inventory- field crop vegetables (tc = tissue cu Field level data - acreage preparation types and co
Inventory of crops	incrop10.sav	hhid, crop	15,406		
Field level information	field10.sav	hhid, harvest, field	8,735		
Cropping patterns	croplev10.sav	hhid, harvest, field, crop	20,791	kgseed = kgs of seed planted; kgharv = kgs harvested; kgsold = kgs sold; kgsspol = kgs spoiled	Crop level data - crops grown, seed information, harvest, sales & buyers, amount spoiled for fruits and vegetables
Fertilizer used	fert10.sav	hhid, harvest, field, ferttype	8,433	Ferttotal – amount used was standardized to kgs Fertcost – cost of fertilizer	Types and amounts of fertilizer used in each field Price of fertilizer is calculated using PriceFert.sps, File is at fertilizer type, fertilizer unit level (fertqty*pfert). Manure and compost are not valued.

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
Type of maize seed used	maizeseed10.sav	hhid, harvest, field, crop, sdvar, sdobtain, units	2,739	kgseed = kgs of seed obtained; totval = total value of seed obtained	Seed type – sdvar = 22 (DH2) is a different seed from sdvar = 56 (DH 02)
Non-agricultural credit	nagcred10.sav	Hhid, crduse, ctype, crdsor	503		
crop inputs purchased with own cash or credit	input10.sav	hhid, inputype, mcrop, numpur, punit, inputpr, inpsorce	4,852		Fertilizer and other inputs purchased/hired. Transport costs for manure were not collected
Fertilizer subsidies received over the last 3 years	fertsubsidy10.sav	hhid, sfert, subsidyr, sbunit	259	sbkg – kgs of fertilizer received as a subsidy	A respondent would say the fertilizer was given by government simply because it was handed out by the chief or assistant chief (government). It's not always possible for the farmer to know the actual source. There could be an issue of confusing the year the subsidy was given.
Availability of fertilizer in last 3 years	Fertaval10.sav	hhid, fyear	134		
Labour inputs	labour10.sav	hhid, activity	9,387		Labour inputs for largest monocropped maize field. Some monocropped fields will have vegetables and fruits listed in the field.
<b>Notes on labour:</b> Where harvest is missing the household generally harvested green maize as they weeded. An assumption was made during cleaning with respect to hired labor – the household could either hire labor or they could hire as a contract, but not both. This issue should be clarified in future panel surveys. Some low costs or hours were justified by notes indicating the person was supervising the activity. New categories were created for a combination of tasks where the respondent could not break down the hours to individual tasks.					
Who makes the decisions on production, marketing, and income use	decision10.sav	hhid, enterp	7,854	partentr – added during data cleaning as a yes/no question to permit 6 cases per	If the HH did not have the enterprise in the reference period, they could have practiced the same earlier hence all HH were to respond to the six enterprises. However if a HH has never engaged in that enterprise it

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
				household	would be Not Applicable for that HH.
Land transactions for last 10 years	landmkt10.sav	hhid, pid	182		
Livestock	livestock10.sav	hhid, livecode	4,461	Vpurch – value of purchases Vsold – value of sold Vsoldnet – value of net sales	Livestock inventory and sales. Standardized median values of type of animal sold and type of animal purchased were computed and used to value the animals sold and purchased. (PricePurchLS.sps and PriceLS.sps)
All cow milk production	cowmilk10.sav	hhid, milk	1,660	totmilk = kgs of milk (produced, sold); totmilkv = Value of milk (Ksh)	Standardized median value of milk was computed to value milk production and sales (priceMP.sps)
Livestock products	liveprod10.sav	hhid, liveprod	1,384	vprod – value of production (Ksh) vsales – value of sales (Ksh)	Livestock production and sales, standardized price calculated using priceLP.sps
Livestock costs	livescost10.sav	hhid, animsp	1,624	totcost – total livestock expenses	4 specific groups of animals
Inputs for livestock received on credit	livestinput10.sav	hhid, input	149		Credit (cash or in kind) received for livestock care. <i>The training instructions were that if feed or any other item to do with livestock was received on credit (cash or in kind), then it would appear in the <u>livestinput10</u> file, but not in the <u>livescost10</u> to avoid double counting</i>
Extension advice	extension10.sav	hhid, serv	2,526		Amount willing to pay for 3 hours of extension advice for new technology
household members from	demog10.sav	hhid, mem	8,919	Age – actual age	household members listed in 2007 except

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
previous survey				subtracting birthdate from 2010. notmem07 – variable to indicate whether the member returned.	those who had died.
<b>Notes on demog10.sav</b>	<p>Only those members who had died were left out of the listing of members from 2007 to be used for the 2010 survey. Members who were no longer in the household in 2007 but had been a member at some time were also not listed. There was no breakdown in the listing for 2010 between those who were present and those who were no longer members in 2007. Thus, the enumerator did not know if the member had left and was now returning and they also could not identify previous members before 2007 who were returning. There are 457 members present in 2010 who were not members in 2007. <u>The reason for returning to the household was not collected.</u></p> <p>There are 13 cases where the person listed was not a member in 2007, has died and spent no time in the household the last 12 months. There are 7 people who were not in the household in 2007, but spent time in the household for this survey and have died. No data were collected as to why these 20 people had returned to the household.</p>				
Additional members	demogA10.sav	hhid, mem	945	Age – actual age subtracting birthdate from 2010.	Adult household members not listed in 2007. New members start with the number 91 (905 cases). Some of the people listed here are not new members, but returning members who had left before 2007 (40 cases). The member numbers for these people are less than 91.
Mortality since 2007	mortality10.sav	hhid, pdmem	162		Previous deaths, cause, symptoms, sex, year and month died, relation to head, level of education
Business / informal labour income	business10.sav	hhid, mem, activity	1,246	Low, medium, high=# of low, medium and high income months	Business and informal labour activities
Salaries and pensions	salwg10.sav	hhid, mem, activity	1,233	totsal = total salary	Salaries / permanent employment-pensions



Type of data	File name	Key variables	Number of cases	Computed variables	Comments
				for the year	and remittances
Savings accounts	savings10.sav	Hhid, mem, saves, acopen	1,502	da03 (relationship to head) – merged from demog10 file	The member name was not recorded so the member number could not be verified, there could be data entry errors.
Home consumption purchases	purch10.sav	hhid, purch	4,756	kg1, kg2, kg3, kg12, totkgpch = kgs purchased	Purchases for home consumption by 4-month periods
Weather patterns	climate10.sav	hhid, weather	1,705		Affct - Has this affected your farming? if no, the rest of the questions will be N/A for that change for that HH
Post-election violence (PEV)	pev10.sav	hhid, deffct	120		
Mobile phone usage	cellphone10.sav	hhid, usephone	7,795		
Household assets	asset10.sav	hhid, asset	9,187	assetval = total value of assets (Ksh)	This file should only contain those assets that the household owns that are usable/repairable
Storage of grains	store10.sav	hhid, store, grain	1,113	kgsstore – kgs stored, kgsloss – kgs lost in storage	

C:\...\Kenya\Kenyahh2010\lookup

Type of data	File name	File to be used with	Key	Number of cases	Comments
Crop quantity conversion to Kgs	Cropconv.sav	Croplev10.sav	Crop, unit	806	File used to convert harvested/sold crop units to kgs
Fertilizer quantity conversion to Kgs	Fertconv.sav	Fert10.sav	Fertype, fertunit	155	File used to convert fertilizer units in to kgs
Crop prices	pricecrop.sav	croplev10.sav	crop, dist	955	Created with PriceCrop.sps. Developed using the following approach: district median if $\geq 10$ observations, otherwise zonal median if $\geq 10$ observations, otherwise provincial median, then national median.
Fertilizer prices	pricefert.sav	fert10.sav	fertype, fertunit, dist	268	Created with PriceFert.sps. Followed standard approach as in PriceCrop.sps. Note that we also used a fertilizer price lookup file in the 2000 data set. Computation of Pfert is as with pricecrop.sps where we consider the district, zone, provincial and national prices in that order.
Prices for livestock products	priceLP.sav	lstprd10.sav	liveprod, dist	85	Created by PriceLP.sps. District price conversion for livestock products
Livestock selling prices	priceLS.sav	lstsld10.sav	livecode, dist	221	Created by PriceLS.sps. District price conversion for livestock sales
Prices of purchases	pricepurchase.sav	purch10.sav	purch, dist	206	Created by PricePurchase.sps. Price conversion for the prices of household purchases.
Livestock purchase price.	pricepurchLS.sav	lstsld10.sav	livecode, dist	221	Created by PricePurchLS.sps. Conversion for the purchase prices of livestock
Prices of seed	priceseed.sav	croplev10.sav	crop, sdtype, sunit, dist	1,910	Convert prices of seed into district prices Price of seed computed as in the other price lookup files. This file assigns a value to the seed used. Not all seeds were purchased.
Purchases conversion to kgs	Purchconv10.sav	purch10.sav	purch, unit	77	Conversion of purchase units into kgs.

## New Computed Variables

C:\...\Kenya\Kenyahh2010\NewVars  
     \demog  
     \income

Type of data	File name	Key variables	Number of cases	Variables	Syntax File
<b>Subdirectory “demog”</b>					
Adult equivalents and size of household	ae_hhsize_10.sav	hhid	1,309	aehh10 – adult equivalents hhsz10 – household size	ae_hhsize_10.sps – see note at end of documentation regarding method used to compute adult equivalents
<b>Subdirectory “income”</b>					
All income variables in one file	income10.sav	hhid	1,309	Main variables are: income10 (sum of crpinc10, offrin10, vnetlv10)	merge_income.SPS
<b>Income notes:</b> seed cost is included in this file but is NOT included in calculating expenses for crop income. Labour costs are also not included. Milk sales are included in the file but not used. Livestock costs - there are 2 variables, tlivescost10 includes all costs collected, vlivescost10 includes only those costs that match previous years. The income total uses vlivescost10.					
Crop income computation	cropinc10.sav	hhid	1,307	crpinc10, totcost (vprod, vsold, vret, lpcost, fertcost, seedcost)	cropinc10.sps
Off farm income	ofarminc10.sav	hhid	1,174	vsalrem, vinform, offrin10	offfarminc10.SPS
Live animal valuation	vlivestock_net10.sav	hhid	1,309	vcost_lv (Vetserv, sallvstk animfeed) – costs, vnet_ls, vprod_lp, vsold_ls, vpur_ls	livestock_income10.SPS
<b>Livestock income notes:</b> Two variables were computed – one calculating net income for cattle (vnet_lv_c - Net value cattle income 10 (live+animal prod) - cost) and another calculating income from the other animals (vnet_lv_o - Net value other livestock income 10 (live+animal prod) - cost). For panel analysis the net income from cattle would be used. Two expenditure costs were calculated: /tlvcost 'Total expenditures for all animal services' =sum(tlvcost) /vlvcost 'Expenditures matching previous years (animal feed and vet service)' = sum(vlvcost). The costs to match previous years was used “vlvcost” to calculate net income.					
Asset valuation	asset10.sav	hhid	1,309	Lsval_10 (panel), totlsval_10 (all 2010 hhs), eqval_10 (panel), toteqval_10 (all 2010 hhs), asval_10 (panel assets), totasval_10 (all 2010 hhs)	Asval10.sps

<b>File name</b>	<b>Contents</b>
2010_Original_Questionnaire.pdf	Questionnaire used in the field
2010_Synthetic_Questionnaire.pdf	Field questionnaire restructured to reflect the data file structure
2010_SurveyDocumentation.pdf	Documentation of data files, sampling methods, specific issues with the data set
2010_Enumerator_Manual.pdf	Instructions to enumerators

### Adult equivalence

The table shows the recommended conversion of different age categories and gender into adult equivalence. This table may be used together with the 3 demography tables for various computations.

The file called **ae\_hhsize\_10.sav** in the “c:\...\Kenya\kenyahh2010\NewVars\demog” subdirectory has already computed the adult equivalents using the breakdown outlined in the table below.

<b>Gender</b>	<b>Age</b>	<b>AE</b>
Both	<1 year	0.33
Both	1-2 years	0.46
Both	2-3 years	0.54
Both	3-5 years	0.62
Male	5-7 years	0.74
Male	7-10 years	0.84
Male	10-12 years	0.88
Male	12-14 years	0.96
Male	14 -16 years	1.06
Male	16 -18 years	1.14
Male	18-30 years	1.04
Male	30-60 years	1.00
Male	>60 years	0.84
Female	5-7 years	0.70
Female	7-10 years	0.72
Female	10-12 years	0.78
Female	12-14 years	0.84
Female	14 -16 years	0.86
Female	16 -18 years	0.86
Female	18-30 years	0.80
Female	30-60 years	0.82
Female	>60 years	0.74

**Document name: C:\...\Kenya\kenyahh2010\docs\2010\_SurveyDocumentation.doc**