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 **completed** the interview – **use this file to merge in location variables**
Household level questions	+hh10.sav	hhid	1,309		General household level questions.
Notes on hh10 file: 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.					
Household	allhhid10.sav	hhid	1,342		All households that were to be interviewed – **use only if want to know how many households were not interviewed**
Inventory of crops	incrop10.sav	hhid, crop	15,406		Crop inventory- field crops, fruit trees & vegetables (tc = tissue culture)
Field level information	field10.sav	hhid, harvest, field	8,735		Field level data - acreage, tenure, land preparation types and costs
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	Fertotal – 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.
Notes on labour: 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. *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 livestinput10 file, but not in the livescost10 to avoid double counting*
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.
Notes on demog10.sav	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 is 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. The reason for returning to the household was not collected. 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.				
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	

Lookup tables

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

Type of data	File name	File to be used with	Key variables	Number of cases	Comments
Crop quantity conversion to kgs	Cropconv.sav	croplev10.sav	crop, unit	806	Use this file to convert all harvested/sold crop units to kgs.
Fertilizer quantity conversion to kgs	fertconv.sav	fert10.sav	ferttype, fertunit	155	File used to convert fertilizer units into kgs

Type of data	File name	File to be used with	Key variables	Number of cases	Comments
Crop prices	pricecrop.sav	croplev10.sav	crop, dist	955	Created with PriceCrop.sps. Developed using the following approach: district median if >=10 observations, otherwise zonal median if >=10 observations, otherwise provincial median, then national median.
Fertilizer prices	pricefert.sav	fert10.sav	ferttype, 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
Subdirectory "demog"					
Adult equivalents and size of household	ae_hhsize_10.sav	hhid	1,309	aehh10 – adult equivalents hhsize10 – household size	ae_hhsize_10.sps – see note at end of documentation regarding method used to compute adult equivalents
Subdirectory "income"					
All income variables in one file	income10.sav	hhid	1,309	Main variables are: income10 (sum of crpinc10, offrin10, vnetlv10)	merge_income.SPS
Income notes: 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, vlivecost10 includes only those costs that match previous years. The income total uses vlivecost10 .					
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
Livestock income notes: 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

Documentation files

C:\...\Kenya\Kenyahh2010\docs

File name	Contents
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

Data files pertaining to TAMPA surveys conducted in 1997, 2000, 2004, 2007 and 2010. All files that can be used with these survey years are stored in the subdirectory C:\....\kenya\KenyaGen.

Purpose	File name	Number of Cases	Comments
C:\...\Kenya\KenyaGen\data			
Consumer Price Index	CPI_allyears.sav	5	The consumer price index is based on the year 2003/2004, using raw CPI data from the Ministry of Finance, Government of Kenya. The period is from June xxxx to May xxxx (xxxx refers to the various years). To reflate all years to 2003/04, divide by these CPIs for their respective years. The years are: 1995/96, 1996/97, 1999/2000 and 2003/2004.
Rain information for the villages covered in the TAMPA surveys	tampa_rain.sav tampa_rain.dta	107	File contains data at the prov, dist, div. village level. Altitude, latitude, longitude, rainfall for the long and short harvests as well as fraction of 20 day periods with <40mm rain for each season – NOT YET UPDATED
Panel participation	Panel_participation.sav	1243	Households that have participated in 1997,l 2000, 2004, 2007 and 2010 surveys.
C:\...\Kenya\KenyaGen\docs			
Documentation of rainfall data	Kenya Rainfall Data.pdf		
Main and short season rain periods defined	Rainfall Periods for Tegemeo Sample Villages.pdf		

Miscellaneous Notes on the Rural Household Survey 2010
Egerton University - Tegemeo Institute / MSU
Updated – January 2011

Household Numbers

The total number of households that completed the interview was 1309. Of those, 1243 were interviewed in all 4 panel survey years (shhpanel).

There were gaps in numbering in both the TAPRA sample.

Brief Documentation for all files

Most of the files contain a variable **'comment'**. This variable consist any issues that were noted during cleaning that are specific to the particular case or set of cases. If no comments were added during cleaning, the variable was removed.

1. **allhhid10:** It is preferred that analyst use the hhidfinal10 file which contains only the households that were interviewed. The allhhid10 file contains all the households that were supposed to be interviewed. No major issues were noted in this file

2. **hhidfinal10:** This is a generated file. It contains all the households that completed the survey. It is at household level and contains the identifying variables for the household. The total number of cases is 1309.

3. **hh10:** This file contains the household level questions. The file is at household level. Field observations:

   a) Variable "intview": (Megan) In our group, we used this code to mean anything related to the hh not being able to answer due to PEV. Household could have relocated, head/spouse could have died, etc.
   b) Variable "fallow": (Megan) We decided during training that this did not include land left for livestock grazing.
   c) Maize Market Access section (page 11): (Megan) These questions were really difficult for many of our households to answer. I'm not sure if this is because not everyone was a surplus producer or if households tend to have one buyer that they work with and therefore don't know much about the rest of the market. I don't think we should have too much confidence in these answers, at least from the areas I travelled. Also note that some areas were not only reliant on maize, so they are likely selling to buyers not captured here.
   d) (Milu) Variable "Tradenum". This question was general- when a number, say 10 is recorded, it may mean 10 bicycle traders, or 10 lorry traders, etc. It does not have buyer type connotation.
   e) Time allocation and decision making (page 12): (Megan) I would say we should use these data with caution. Not only is it difficult for the household/respondent to come up with a percentage, but also the enumerators were trained that this should only be asked with respect to time spent at the household. For example, even if the spouse lives in Nairobi half of the year, 75 percent of his/her time might be allocated to farming activities when at the

household. This method seems faulty to me, but the supervisors thought that was how it was asked in the past.
(Milu) Some respondents had difficulties coming up with proportions.

f) Land rental rates (page 14): (Megan and David) In some areas a respondent could also rent land already under productive tree (for example tea), meaning the value of the land is much higher. Also, land devoted to rice or sugarcane went for a very different rate than maize, for example. We tried to include "normal" ag land here, but I wonder how this came out in other groups. Also, a lot of people who don't rent land have no idea how much it costs. We probably should have also asked "is it possible to rent land in this area?" And also included "don't know" for these two questions on land rental rates. Some areas don't have two ag seasons, so there will either be the same value entered on both the year and season rate.here. Also, note that main and short season rates are very different. This question probably should include a qualifier for "Main" season.

g) Transport of fertilizer (page 30) Many households indicated that a bike and matatu were the same. Opiyo said they would handle this during data collection. May want to add motorbikes as an option - they are everywhere. People also relied a lot on donkeys for transport.

4. **incrop10:** This file is at crop level. It contains a question asking the season planted, 1 – main, 2 – short and 3 – both for the annual crops planted and the number of trees for the perennial crops produced or planted. This file was compared with the croplev10 file to verify data. During cleaning, more emphasis was directed to the crop file. Notice that commercial trees were transferred to the informal income section.

5. **field10:** This file is generated from the original crop file. It contains field level information. Some acreages were noted to be very small especially when related to the yields. The questionnaires were checked to confirm the data were entered correctly. It's possible there were enumerator errors in the calculation of the field size. No major issues noted. In the 2000 survey the variable "harvest" was called "season".

5. **croplev10:** This file is generated from the original crop file. The file contains details of the cropping pattern for the main and the short season. The file is at "hhid, harvest, field, crop" level. Duplicates were checked. More than one type of fodder can be in the same file. Fodder types were not distinguished by the type of crop (i.e. maize, grass, sweet potato leaves, etc.). There could be two cases for the same crop in the same field where the unit of sales is different, e.g. sales unit for mangoes. The file also contains information on amounts harvested and amounts sold from this harvest. There were 21 cases of volunteer crop which did not have seedtype and amount of seed. Those cases are noted in the sdtype variable. The seed cost for maize is repeated on the maizeseed file but sometime with some minor discrepancies. Analyst should work with the details on the maize seed file where applicable. Commercial trees were transferred to the informal income section. In the 2000 survey the "harvest" variable was called "season".

a) **Variable "tenure": (Megan)** There were some areas where households would farm in gov't owned swamps but considered it their land (in fact they may have included it in the first question on land owned). We tried to code that as 5,

but just something to note in case some things seem strange.

b)

7. **fert10:** This file is generated from the original crop file and contains information on types and amounts of fertilizer used on every field. No major issues were noted. In the 2000 survey the "harvest" variable was called "season".

8. **labour10:** The labour file contains details on labour for the largest monocrop maize field or the largest intercrop maize field (if maize was not grown as a monocrop).

   Some households did not have any labour input because the work was done by salaried labour. Some instances the person was supervising labor so the hours are lower than might be expected.

   Gang labor is included under family labor.

   Children are defined as < 15 years old.

9. **Maizeseed10:** The file contains details on maize seed type, purchase and prices. Note that seed information as also collected at the crop level in "croplev10.sav". However, the question asked in the crop table referred to the total quantity and did not ask for detailed information. The maizeseed10.sav file asked for expanded detail on maize only, allowing the respondent to indicate the different seed types used in the same field. In many instances the information in this file will be the same as in the croplev10.sav file. It is recommended that researchers use the information in this file for analysis of seed types used for maize.

10. **input10:** The file contains details of inputs that the household bought **on credit**. These inputs include fertilizers and other farming inputs. Inputs codes starting at 31 were thought to be capital expenses and should be removed for any "income" computations. The cash credit was quite difficult to capture. Some were specifying money (as the input type) while other gave the details of the input bought from the credit. In cases where the input was given in money form, the value was indicated in the InpValue and the InpUnit was given as number. No table lookup to standardize prices was created. The actual price quoted should be used.

    (Megan) May want to consider capturing large labour expenditures for commercial crops such as tea and coffee in future surveys in the input file.

11. **Fertsubsidy10:** (Megan) For the variable: "fsorc" - There was a surprising amount of confusion here. A lot of households knew that they picked up the fertilizer from the local church (for example) but didn't realize it was provided by the government.

12. **Decision10:** (Megan) This table was a long debated and apparently controversial addition to the survey. Some enumerators did not ask this activity by activity and crop by crop despite the training. There were many missing cases categories.

13. **Landmkt10:** (Megan) In retrospect, I think we forgot a major and important question here: "where is the land?" In some areas, people might own land in very different places, sometimes both in the village where the interview is taking place and

sometimes at a scheme elsewhere. This seemed particularly true of Rift Valley households and also in Kisii (where there is a lot of pressure on the land already). This is the reason you might find very different land values from households in the same village.

14. **purch10:** The file contains details of purchases on key items in 4 month groupings within the year and if the respondent could not answer in 4 month grouping, the response was given for the whole year.

    a) (Megan): Would like to see gift/relief split into two categories to see what areas receive food aid and which just rely on gifts from friends/family.

    b) The enumerators complained that this table took forever and that households had a very difficult time remembering purchases so far back in time. For that reason, a lot of them were probably lumped into the last few columns.

    c) (Milu): Personally and having listened to the way the respondents answer this question, I believe we just collect a CRUDE approximations. Unrecorded historical quantities and prices of highly frequent items is difficult to recall. A researcher interested in household consumption needs to revisit the same households a couple of times in the year to accurately capture data on consumption.

15. **livestock10:** Gives livestock inventory details. Purchases and sales were collected for cattle, not for any of the other livestock. No major issues noted.

    a) **(Megan);** Oxen proved to be a problem here. While they are technically an asset (per the table later in the survey), I imagine we might want to know how many were bought/sold/died/etc. during the year. I know that in one case, in particular, we added that info here and coded it as such. Something to think about for future livestock tables. Aadding a "born" column would have been useful. Normally the enumerators wrote the number born at the bottom of their surveys just so that we didn't have any questions about how the numbers worked out at the end of the day.

    b) Note that some households don't know the difference between local and cross. Also what happens when you cross a cross with a local? How should those cases be coded?

16. **Cowmilk10: (Megan)** This table requires a lot of enumerator quick-hand calculation to come up with these figures. Should try to work on an easier method to collect these data for the next round.

17. **liveprod10:** Gives details of production and sale of livestock products.

18. **Extension10:** (Megan and David) Variable "Others" - This was supposed to mean "what do you think other people in this village would be willing to pay (as compared to yourself)?" The questionnaire was not worded well. People had difficulty responding in a manner that would make sense relative to what they would have spent. The question was not understood based on some of the responses. There was a lot of confusion.

(Milu) Variable "Others" To avoid individual biases, in contingent valuation researcher we pose the question this way (Dr. C. Wolf). However, this didn't seem to elicit the required information. Most respondents would not say that they didn't know, they would say- 'go and ask them'

19. **Savings10**: (Milu) Two last questions: If the account is owned by head/spouse who works in town but still considered a member of the household- the respondent could not give all the details. Kms to the banking point may not make a lot of sense in some cases- a person living in Nakuru may be having an account in Nairobi, his bank may be having a branch in Nakuru/ATM, now, what is the distance in this case?

20. **demog10:** This file contains details of the demographic characteristics of the household. Adult household members listed in 2007 are in this file. Most of the household heads are in this file. However, some heads of household are in the additional adults file. 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 is 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>. 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.

    a) (Megan) In training, the enumerators were told that only those individuals residing in the house for one month of the year could be considered a member. To me, this seemed to create a lot of gray area between remittances and the unrecorded cost of maintaining the household member in another location. (MB - That instruction was not appropriate – the household could decide whether the person was a member or not, there was no time limit to be applied to that definition. Training should be modified for the next round.)

20 **demogA10:** This file contains details on additional adult members of the household not listed in 2007. Some original members returned and are listed here. The variable "mem" starts at 91 for new members. An adult is defined as 15 years or older.

21 **business10:** This file contains details on informal business household income. For hhid 367 the high average cost for livestock selling was higher than the high cost. The activity was accruing losses in the low and average months. In hhid 1488, the matatu business was incurring losses.

22 **salwag10:** Gives details on salaried income for the household. Remittance data were collected in this file. In a few cases the respondent did not know the salary earned by a member.

23 **asset10:** Gives details on the number and value of selected assets for the household. No major issues noted so far.

24 **Store10:** (Megan) For the variable "store", option 4 – room in main house - Note that this meant a specific room devoted to storage. For those households that stored grain in a bag that simply sat in the corner of a main household room, that was not recorded here.

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.

Gender	Age	AE
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