EGERTON UNIVERSITY
TEGEMEO INSTITUTE OF AGRICULTURAL POLICY AND DEVELOPMENT

AND

MICHIGAN STATE UNIVERSITY

TEGEMEO AGRICULTURAL MONITORING AND POLICY ANALYSIS PROJECT (TAMPA 2)

HOUSEHOLD SURVEY 2004 DATA DOCUMENTATION

Support for this study was provided under the Tegemeo Agricultural Monitoring and Policy Analysis (TAMPA) 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.

2005

SAMPLING METHOD

TAMPA Sample

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. 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 possible.

SUMMARY OF HOUSEHOLDS SURVEYED

Out of the 1997 Tampa survey sample of 1540 households, there were 1397 households that were interviewed. 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 Tampa sample for this survey. In the USAID sample 788 households were interviewed. In the Tegemeo Sample 110 households were interviewed. Two exemplary farmers were interviewed. The TAMPA sample interviewed 1397. The total number of households interviewed was 2297. The data for page one of the survey instruments are contained in two files: allhhid04.sav and hhidfinal04.sav.The first file (allhhid04.sav) contains all the original selected households to be interviewed. The second file (hhidfinal04.sav) contains only those households that completed the interview for this 2004 survey (1397 hhids) of the TAMPA sample. This file should be used to merge the identifying characteristics to the other files as needed.

DATA FILE DESCRIPTIONS FOR RURAL HOUSEHOLD SURVEY

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)

DATA FILES

Directory: C:\Docs\Kenya\Kenyahh2004\augdata

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
Household identification	hhidfinal04.sav	hhid	1,397		All households that **completed** the interview – **use this file to merge in location variables**
Household level questions	hh04.sav	hhid	1,397		General household level questions
Household	allhhid04	hhid	1,483		All households that were to be interviewed – **use only if want to know how many households were not interviewed**
Inventory of crops	incrop04.sav	hhid, crop	21,160		Crop inventory- field crops, fruit trees & vegetables
Field level information	field04.sav	hhid, harvest, field	11,471		Field level data - acreage, tenure, land preparation types and costs
Cropping patterns	croplev04.sav	hhid, harvest, field, crop	26,480		Crop level data - crops grown, seed information, harvest, sales & buyers, amount spoiled for fruits and vegetables
Fertilizer used	fert04.sav	hhid, harvest, field	9,997	Fertotal – amount used was standardized to kgs	Field level file - types and amounts of fertilizer used
Labour inputs	labour04.sav	hhid, sizefild, crop, activity	25,502		Labour inputs for largest and second largest maize fields
Maize seed	MaizeSeed04.sav	hhid, season, field, sdtype	3,035		Types of seed varieties used and their sources
Fertilizer - inputs purchased with own cash	Tfert04.sav	hhid, inputype	2,457		Fertilizer and other inputs purchased/hired.
Home consumption purchases	purch04.sav	hhid, purch	13,364	kg1, kg2, kg3, kg12, totkgpch = kgs purchased	Purchases for home consumption by 4-month periods

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
Farm inputs on credit	input04.sav	hhid, input	614		Inputs received or bought on credit
Fruits and vegetables purchased for home consumption	consumpt04.sav	hhid, fooditem	6,812	totkgs = Total kgs purchased – 30 days kgsjan = total kgs purchased in January	Expenditure on food items over the past 30 days and in January
Livestock	lstsld04.sav	hhid, livecode	5,272		Livestock inventory and sales
Livestock products	lstprd04.sav	hhid, liveprod	2,257		Livestock production and sales
Adult household members from previous surveys	demogA04.sav	hhid, mem	7,276		Adult household members listed in 2000 or 2002. Data are: sex, age, relation to head, currently in school, years of schooling, months living at home, why left, engage in business/informal labor or salaried employment, if chronically ill for 3 or more months.
Additional adult members	demogA_A04.sav	hhid, mem	645		Adult household members not listed in 2000 or 2002, same questions plus why joined household and if had other income before joining
Children - < 15 years old	demogC04.sav	hhid, mem	3,770		Age, sex, relationship to head, level of education, information about biological mother and father, if chronically ill for 3 or more months
Mortality since 2001	mort04.sav	hhid, pdmem	203		Previous deaths, cause, symptoms, sex, year and month died, relation to head, level of education
Business / informal labour	hhinc04.sav	hhid, mem, activity	1,194	Low, medium, high=# of low,	Business and informal labour activities

Type of data	File name	Key variables	Number of cases	Computed variables	Comments
				medium and high income months	
Salaries and pensions	salwag04.sav	hhid, mem, activity	1,156		Salaries / permanent employment-pensions and remittances
Household assets	asset04.sav	hhid, asset	8,572		Household agricultural assets

Lookup tables

C:\Docs\Kenya\Kenyahh2004\lookup

Type of data	File name	File to be used with	Key variables	Number of cases	Comments
Crop quantity conversion to kgs	Cropconv.sav	croplev04.sav	crop, unit	555	Use this file to convert all harvested/sold crop units to kgs.
Fruits and vegetables conversion to kgs	Consumptconv.sav	consumpt04.sav	fooditem, unitpur	53	Convert consumption units into kgs
Fertilizer quantity conversion to kgs	fertconv.sav	fert04.sav, Tfert04.sav	ferttype, fertunit	85	File used to convert fertilizer units into kgs
Purchases conversion to kgs	purchconv.sav	purch04.sav	purch, unit	85	Conversion of purchase units into kgs.

C:\docs\Kenya\Kenyahh2000\docs

Documentation files

File name	Contents
2004_Synthetic_Questionnaire.pdf	Field questionnaire restructured to reflect the data file structure
2004_SurveyDocumentation.pdf	Documentation of data files, sampling methods, specific issues with the data set

Miscellaneous Notes on the Rural Household Survey 2004
Egerton University - Tegemeo Institute / MSU Updated – May 2005

Household Numbers

Original household numbers from the 1997 and 2000 survey range from 1 to 1838 for a total of 1397 households. There were gaps in numbering in both the Tampa sample.

Brief Documentation for all files

All the files except field04 and fert04 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.

1. **allhhid04:** It is preferred that analyst use the hhidfinal04 file which is contains only the households that were interviewed. This file contains all the households that were supposed to be interviewed. No major issues were noted in this file

2. **hhidfinal04:** 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.

3.hh04: This file contains the household level questions. The file is at household level. Question 3.2 there were two farmers with fodder maize and it was included in the comparison.
Question 18, some farmers indicated the month of August (month of interview). In these cases the flour was finished the same month as when the interviews were done.
Question 39 where there was only one season (rather than 2 seasons) the response was coded as NA.
Question 41, there appeared to be a problem with the perception of this question. Some farmers understood it as asking when the current head started being the head while some other farmers understood it as when the household settled there. In this case you will find some households indicating 2004 or 2003 for this question.

4. **incrop04:** This file is at crop level. It contains a Yes or no entry for the annual crops planted and the number of trees for the perennial crops produced or planted. This file was compared with the croplev04 file to verify data. No corrections were made to this file. During cleaning, more emphasis was directed to the crop file. Notice that commercial trees and sisal were transferred to the informal income section.

5. **field04:** This file is generated from the original crop file. It contains field level information. Some acreage was 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".

6.**croplev04:** 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. HHID 636 was not involved in any crop activities that year. Only mangoes were harvested and the relatives did that. The file also contains information on amounts harvested and amounts sold from this harvest. There were 18 cases of volunteer crop which did not have seed type and amount of seed. 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. Hhid 1288, there appear to be more avocados sold than harvested. This was as a result of differences in units. The unit for harvest was bags while sales were in numbers. Commercial trees and sisal were transferred to the informal income section. In the 2000 survey the "harvest" variable was called "season".

7 **fert04:** 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. **labour04:** Data were initially entered into two files, one for the largest field and one for the second largest field; each was cleaned separately and then merged into one labour file where further cleaning was done. The labour file contains details on labour for the largest maize field or the largest field (if maize was not grown) and the second largest maize field or the second largest field (if there was no second maize field). The variable "crop" was not a question asked on the original instrument. Most enumerators recorded the crop the activity was associated with on the questionnaires. Since the information was collected, it was added. In some fields, it was not possible to relate specific crops to specific activities. Usually those activities would have applied to all crops in the field. In these instances the crop variable was not filled in. Occasionally it was not possible to identify the crop that the activity was linked to.

Some households did not have any labour input because the work was done by salaried labour. There were a few issues of identification of the largest and second largest maize field. In a few instances, the largest field did not have maize and in other instances, the second largest field was not actually the second largest. The information involves inclusion of both intercrops and pure stands. Five households did not have any labour activity. Most of the largest field cases are maize – only 40 households do not grow maize in the main season. Many of the second largest fields are perennials and do not have the land preparation cost and labour costs. In some instances, the crop planted was maize but used as fodder. The code for fodder was
entered. Different types of fodder were not distinguished – there is only one code for fodder in this survey.

9. **Maizeseed04:** The file contains details on maize seed type, purchase and prices. Note that seed information as also collected at the crop level in "croplev04.sav". However, the question asked in the crop table referred to the total quantity and did not ask for detailed information. The maizeseed04.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 croplev04.sav file. It is recommended that researchers use the information in this file for analysis of seed types used for maize.

10. **Tfert04:** This file contains inputs that were bought **on a cash basis** by the households. The inputs contain mainly fertilizers but may also include other farming inputs.

11**. input04:** 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.

12**. purch04:** 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. Some fish was given in *gorogoros* (2kg tin) while others were given in numbers. These cases will appear as duplicate cases, but are not duplicates. The standard size of bread is 0.4 kgs; however, there were some households that bought larger loafs usually 1 kg. There were many cases of millet and sorghum in 1 or 2 kg packets. It appears that the enumerator and/or respondent were not able to differentiate between millet/sorghum flour and sorghum/millet grains which might account for the large variation. Where the milk price is quite high, milk was purchased in powder form. There are notes in the comment field for these cases.

13**. Consumpt04:** File gives details of purchase of selected horticultural crops over the last 30 days and also for January. If the respondent did not purchase the product in the last 30 days, the enumerator was to record the unit of measure for the January purchases in the section for the last 30 days. No major issues noted.

14**. Lstsld04:** Gives livestock inventory details. Purchases and sales were collected for cattle, not for any of the other livestock. No major issues noted.

15**. lstprd04:** Gives details of production and sale of livestock products. Under the "other (specify)" some people sold broilers in large numbers. We choose to let it remain here instead of transferring it to the informal income because we did not have all the cost involved in the production process.
16. demogA04: This file contains details of the demographic characteristics of the household. Adult household members listed in 1997,

2000 or 2002 are in this file. Most of the household heads are in this file. However, some heads of household are in the additional adults file.

17. demogA_A04: This file contains details on additional adult members of the household not listed in 2000 or 2002. There were 4 households where the head of the household is in this file. The variable "mem" starts at 31. An adult is defined as 15 years or older.

18. demogC04: The file contains details about the children in the household, who are younger than 15 years old. In the earlier surveys, children were not separated out into a separate table. Thus children already listed in earlier surveys will have a previously assigned member number. The member number for additional children starts at 51.

19. hhinc04: 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.

20. salwag04: 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.

21. asset04: Gives details on the number and value of selected assets for the household. No major issues noted so far.

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.

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