AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Egerton University

Tegemeo Institute Of Agricultural Policy And Development

DEVELOPING INCOME PROXY MODELS FOR USE BY TITLE II-FUNDED NGOS IN KENYA: A TECHNICAL REPORT FOR NGOS AND USAID/KENYA

David Tschirley and Mary Mathenge

Tegemeo Working Paper 6

October 2003

Tegemeo Institute of Agricultural Policy and Development
P.O. Box 20498
00200 City Square
NAIROBI, KENYA

tel: 254-2-2717818
fax: 254-2-2717819
email: egerton@tegemeo-org

TABLE OF CONTENTS

I. Introduction 1

II. Income Proxy Models: What Are They and How Can They Be Useful? 1
A. Background 1
B. Monitoring or Impact Evaluation? 3
C. What Steps Are Needed to Develop an Income Proxy Model? 3
D. What Steps Are Needed to Use the Income Proxy Models? 5

III. The Income Survey 6
A. Sample Design 6
B. Questionnaire Design 8
C. Training and Field Work 9
D. Data Entry and Cleaning 10

IV. Model Development 10
A. Definition of Income Components 10
B. Types of Proxy Variables Tested 10
C. Results 12

V. Using the Models 13

VI. An Overview of the Rural Economy in the Study Zones 15
A. Crop Production and Marketing Behavior 15
B. Livestock 19
C. Structure of Household Incomes 20
D. Staple Purchase and Sales Behavior by Income Class 24

ANNEXES

ANNEX **PAGE**

A. Sampling Guidelines for Proxy Surveys 30

B. Proxy Questionnaire 39

C. Sampling Procedures Used in NGO Full Income Surveys 50

D. Enumerator Manual for NGO Full Income Surveys 58

E. Selected Prices Used in Valuation of Agricultural Production and Sales 74

F. Model Results 75

G. Caloric Needs Used for Calculation of Consumption Adult Equivalents 102

H. Procedures for Generating Income and Income Component Estimates Using SPSS/windows Syntax File 103

LIST OF TABLES AND FIGURES

TABLE **PAGE**

1. Target and Actual Sample Sizes for Each NGO, Full Income Survey, 2001/02 6
2. Predictive Power by Zone 12
3. Predicting Means by Income Component by Zone 14
4. Prediction of Households into Respective Income Groups 14
5. Top 10 Crops Grown in Each Zone: Percent Growing and Mean Value Produced . 16
6. Top 10 Crops Grown in Each Zone: Percent Selling and Mean Value Sold 18
7. Commercialization Indicators by Zone 19
8. Livestock Ownership 19
9. Livestock Products: Production by Zone 20
10. Detailed Income Levels and Shares, by Zone 21
11. Aggregated Income Levels and Shares, by Zone 22
12. Incomes by Per Capita Income Tercile 23
13. Income Shares by Per Capita Income Tercile, by Zone 24
14. Purchase Behavior on Main Staples, by Income Tercile and Zone 26
15 Maize Market Position of Households by Zone 27
16. Selected Household Indicators by Maize Market Position 28

FIGURE **PAGE**

1. Overview of Process for Developing and Utilizing an Income Proxy Model 4

Developing Income Proxy Models for use by Title II-funded NGOs in Kenya: A Technical Report for NGOs and USAID/Kenya

By

David Tschirley and Mary Mathenge

I. Introduction

Title II-funded NGOs in Kenya administer a range of interventions designed to enhance the welfare of rural households in vulnerable areas of the country. Many of these interventions are oriented towards USAID/Kenya's Strategic Objective of increased household income in target areas. Yet measuring income in rural areas of Africa is difficult, due to the many different sources of income and the methodological challenges of quantifying each. The expense in human and other resources of collecting, cleaning, and processing good quality income data is beyond the capacity of all but dedicated research projects. To facilitate such reporting, Tegemeo Institute and MSU Department of Agricultural Economics worked with NGOs to develop income prediction models which the NGOs could use in future years to report on these outcomes. The main objective of this work was to develop an integrated package that would allow USAID-funded NGOs working in Kenya to monitor rural household income and income components using easy-to-collect proxy variables. The package was to include 1) sampling guidelines for the periodic income proxy surveys, 2) model questionnaires for these surveys, 3) econometric models relating the proxy variables to household income and income components, 4) SPSS/Windows syntax files based on these models that generate the quantitative estimates of household income and income components, and 5) a manual for operating the package.

This paper details the specific procedures utilized to develop the income proxy method for Kenya NGOs, reports on the performance of the method, and brings together in one place each part of the package needed to implement the method. The next section provides general background on income proxy methods; section III reports on the full income survey that formed the basis for development of the proxy method; section IV provides details on model development, including definition of income components, the types of proxy variables tested, and the performance of the models, while section V briefly explains how the models are to be used. Finally, section VI provides an overview of the rural economy in the study zone based on the detailed income data set collected by the NGOs.

II. Income Proxy Models: What Are They and How Can They Be Useful?

A. Background

An income proxy model is one part of a package of procedures that NGOs, donors, governments, or research institutions can use to monitor rural household income and income

components using easy-to-collect proxy variables. The model is a set of algebraic equations that relate these proxy variables to components of income:

$$\hat{Y}_i = a_i + b_{i1}X_{i1} + b_{i2}X_{i2} + \ldots + b_{in}X_{in} + e_i$$

where,

$\hat{}_i$	is estimated income from component i,
a_i	is a constant (or intercept) term for income component i,
$b_{i1} \ldots b_{in}$	are the coefficients (fixed numbers) that quantify the relationship of each proxy variable to income component i,
$X_{i1} \ldots X_{in}$	are the selected proxy variables for income component i, and
e_i	is a random error term.

Taken together, the various components in the model sum to total household income:[1]

$$\hat{Y} = \sum_{i=1}^{C} \hat{Y}_i$$

where,

$\hat{}$	is estimated total income,
$\hat{}_i$	is estimated income from component i, and
C	is the number of income components.

These algebraic relationships are developed using standard "ordinary least squares" econometric techniques applied to a household data set which contains detailed data on household incomes and the proxy variables. Once this detailed data set is collected and the model is estimated, one needs only to collect the proxy variables to obtain estimates of income components and total household income. These simple ***proxy surveys*** will typically be conducted once a year, or however often the institution wishes to track household income. The much more detailed and time consuming ***income survey*** needs to be done once at the beginning of the project cycle and preferably again at the end of the cycle for validation purposes. The complete package which defines the income proxy methodology includes 1) sampling guidelines for the periodic proxy surveys, 2) a model questionnaire for these surveys, 3) the set of econometric models relating the proxy variables to household income and income components, 4) SPSS/Windows syntax files based on these models that use the proxy data to generate the quantitative income estimates, and 5) a manual for operating the package.

---

[1] If desired, the models could be developed to return *per capita* household income, as opposed to total household income.

The usefulness of an income proxy methodology derives from the importance of household income as an objective of development activities: an important overall development goal in nearly every developing country is the reduction of poverty and improvement in the incomes and well-being of rural households. Thus, measurement of household income is one logical choice for monitoring the effects of policies and programs oriented towards accomplishing this goal.

B. Monitoring or Impact Evaluation?

The econometric models in the income proxy methodology are designed to capture the *association* between income and the proxy variables, and to return as accurate a prediction as possible. As such, they can be used directly to *monitor* the types of economic activities that households engage in, and the incomes they derive from these activities. The models themselves are not designed to allow conclusions regarding cause and effect; to use these models for *impact evaluation* (for example, to measure the impact of an NGO's agricultural production and marketing assistance on agricultural and overall household income), they need to be integrated into an overall approach which includes the following elements:

- A sampling design that distinguishes between participants (the target population for the intervention being evaluated) and non-participants (the non-target population),
- A baseline survey conducted prior to the beginning of the intervention, distinguishing between likely participants and likely non-participants,
- The collection of complementary data regarding the physical, economic, and social environment of the participating and non-participating households.

It is beyond the scope of this paper to go into detail on impact evaluation;[2] suffice it to say that, within such an integrated approach, use of income proxy models can allow more frequent monitoring (because it will be less costly and less time consuming), provide a richer set of monitoring results covering the range of the households' economic activities, and reduce the cost of the impact evaluation.

C. What Steps Are Needed to Develop an Income Proxy Model?

Figure 1 provides an overview of the process for developing and utilizing an income proxy model. Once the original, detailed data are collected and the prediction model is developed (Phases I and II), one need apply only Phases III & IV for the remaining years of the program before collecting a new full data set to re-estimate the prediction model and perform a full evaluation of the program.

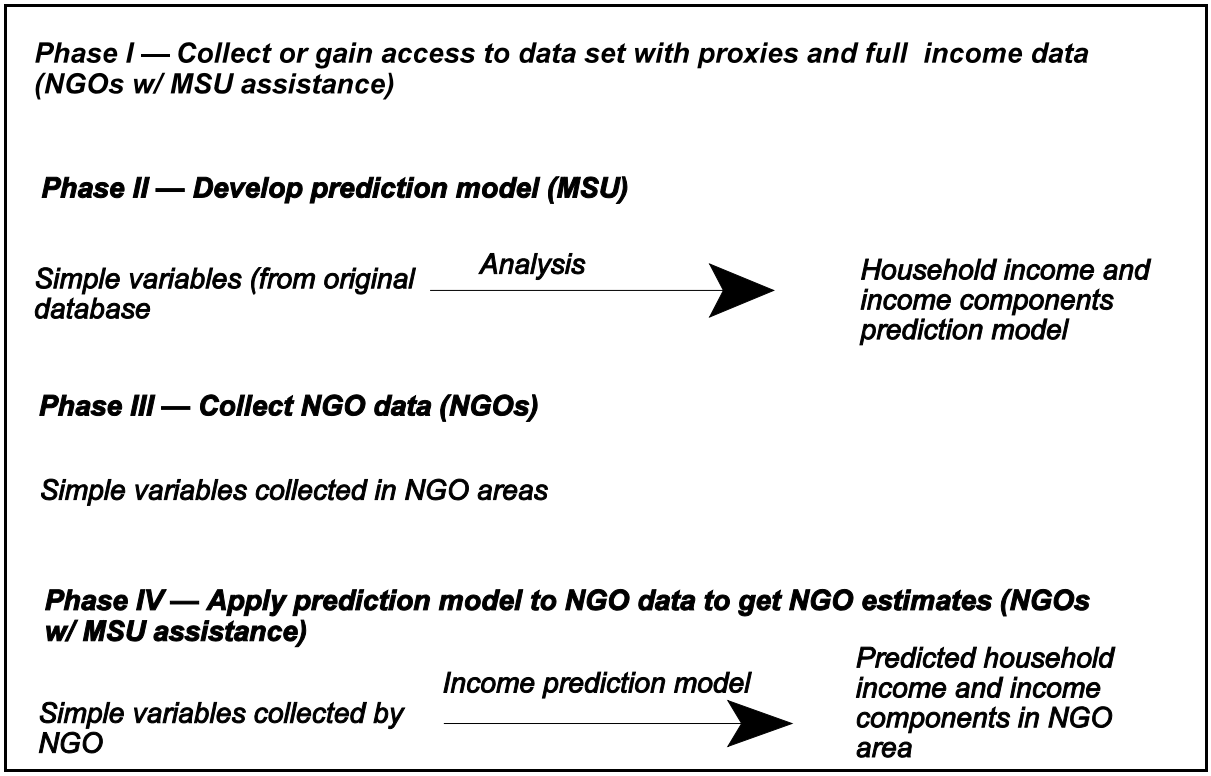
---

[2] For a good introduction to this topic, see Ravallion, Martin (1999). "The Mystery of the Vanishing Benefits: Ms. Speedy Analyst's Introduction to Evaluation". Policy Research Working Paper ..., Washington, D.C., World Bank.. This can be downloaded from the web by going to www.worldbank.org/research/, choosing "poverty", then searching for "Ravallion" under "Policy Research Working Papers".

To develop the model, the analyst must work closely with NGOs to:

1. **Understand the design and operation of the interventions that are being monitored, and the economic environment where they are being implemented**. The analysts developing the model need this type of information to define a set of econometric models that are meaningful for the NGO, and that can be estimated with acceptable accuracy with proxy variables.

Figure 1. Overview of Process for Developing and Utilizing an Income Proxy Model



2. **Define a relevant and feasible breakdown of income components to be modeled**. The preferred definition will depend primarily on the types of economic activities which are most important in the area where the intervention is taking place. For example, in a pastoral area with little crop production, the latter may be grouped into a single component, while livestock activities might be broken into several components. In an area of heavy cropping activities where livestock is less important, the reverse might hold.

3. **As much as possible, anticipate the proxy variables that will be used to model each component**. While not every proxy variable can be defined prior to the data analysis, many can be, and identifying a comprehensive list of probable and possible proxies ahead of time will improve the modeling results. As in the definition of income components, there will be substantial similarities in the definition of these

variables across NGOs, but if the income components are not identical, neither will the proxy variables be.

4. **Define target and non-target populations consistently across NGOs, and develop a sample frame to allow stratification on this basis**. If several NGOs will be using the income proxy methodology to report on their programs, and if their success will be judged in part on these results, then it is best that they follow a common definition of target and non-target populations. If a common approach is not followed, then it will be impossible to determine whether differences between NGOs in reported changes in income are due to differing definitions, or to differing effectiveness of the interventions. Whatever definition is chosen needs to be *workable* in terms of available time and resources, and *meaningful* in a reporting context. See Annex A (Sampling Guidelines for Proxy Surveys) for more detail on this issue.

5. **Design and conduct a detailed *income survey*** that will provide the data to estimate the models. This should be a common effort among all NGOs.

6. **Estimate the models**. The data must be entered, cleaned, organized, and then analyzed to develop the prediction models.

7. **Develop a model questionnaire for the proxy surveys**. Defining the models involves defining the most efficient set of proxy variables for each income component. Once this is done, a questionnaire is designed to collect just these proxy variables in future years. These questionnaires consist almost entirely of yes/no questions, with quantification of a limited number of variables. Thus, these questionnaires are much shorter, the interviews are shorter and easier to conduct, and the data are much easier to enter and clean than a full income survey. See Annex B for the proxy questionnaire designed for Kenya NGOs.

8. **Develop a data processing routine** to convert the proxy variables into estimates of income components and total income. Tegemeo/MSU have developed a SPSS/Windows syntax file that performs this function. It is available in electronic version upon request.

D. What Steps Are Needed to Use the Income Proxy Models?

Once the eight steps above are completed, NGOs have the tools they need to generate estimates of household income and income components with greatly reduced data collection and processing time. With the frequency the NGOs desire (for example, yearly), they need to conduct the proxy survey, enter and clean the proxy data, and run the processing routine. See Chapter V for more detail.

III. The Income Survey

A. Sample Design

The main purpose of the NGO household income survey was to generate sufficient data of good quality on household incomes and income sources to allow estimation of geographically disaggregated models of these income components. Some level of geographical disaggregation of the models – rather than single models for each component over all NGOs – was important due to factors that tend to vary systematically over space. These include prices, crop mix, productivity levels, relative importance of different income sources, and type of NGO interventions, among others. Though the modeling process accounts for some of these differences by capturing them in proxy variables, previous experience in Kenya and Mozambique showed clearly that regional models would perform better than national models.

Zones were thus defined to provide sufficient geographical disaggregation while keeping the modeling job feasible; since models would be developed for each income component in each zone, the number of models needed increases rapidly with the number of zones defined. Following discussion with NGOs and USAID personnel regarding the characteristics of each NGO intervention zone, it was decided to define four zones for the survey, as detailed in Table 1.

Table 1. Target and Actual Sample Sizes for Each NGO, Full Income Survey, 2001/02

Zone	NGOs included	Target Sample Size	Actual Sample Size
1	CARE	200	195
	Total for zone	200	195
2	WVI	100	100
	FHI	100	100
	Total for zone	200	200
3	ADRA	67	66
	CRS	67	71
	TS-HPI (Taita Taveta only)	66	49
	Total for zone	200	186
4	TS-HPI (all districts except Taveta)	200	193
	Total for zone	200	193
	Overall total	800	774

CARE's USAID-funded activities all take place in Nyanza province of western Kenya. This area is quite distant from all other NGO intervention areas, and has agro-ecological and socio-economic conditions that distinguish it from those other areas. Thus, it was a natural choice as a single zone. TS-HPI's USAID-activities are concentrated in a relatively large area of the coast, but also include the inland district of Taita Taveta, which is substantially different both agro-ecologically and socio-economically from the coastal areas. Thus, all TS-HPI areas except for Taita Taveta were grouped in a single zone. WVI and FHI both work in the Arid North, WVI in an irrigated zone in Turkana and FHI in mountainous areas around Marsabit which receive more rain than the surrounding low-lying areas. Households in each area practice a mix of pastoral livestock and settled agriculture, which distinguishes them from all other zones. WVI and FHI were thus grouped together in a single zone for sampling and modeling purposes. CRS and ADRA both work in low potential areas east and southeast of Nairobi. The areas have similar income levels, beekeeping is more important in each of these areas than in other NGO areas, and crop mix tends to be similar. These two NGO were therefore grouped in one zone; because of Taita Taveta's substantial differences from TS-HPI's other coastal areas, and a crop mix more similar to that in the CRS and ADRA areas, Taita was also grouped with these two NGOs into a single zone.

To ensure sufficient observations for modeling activities in each zone while keeping the data collection and cleaning process manageable, it was decided to sample 200 households in each zone, 100 target and 100 control. Control and target households were defined based on the type of program the NGO ran. CARE and ADRA were judged to have programs with significant spillover effects on households within entire villages, whether or not each households was directly reached by an NGO intervention. For this reason, entire villages in which CARE or ADRA operated were defined as target villages, and all households in those villages were considered target households for sampling purposes. Villages in the same general area but in which the NGO did not operate were defined as control villages, and sampled separately. All other NGOs were judged to have more focused interventions with few if any spillover effects on households not directly participating in the program. In these cases, control and target households were selected from the same set of villages, based on whether or not they directly participated in the NGO program.

Each NGO was provided with forms to collect and organize the information needed to design a sample in which villages would be selected "probability proportional to size", and equal number of households would then be selected through simple random sampling within each village[3]. To develop a PPS sample, information is needed on the size (population or number of households) of every village in which the NGO operates, and every village in a defined control area. In practice, CARE and ADRA were able to provide this information, and their samples were designed PPS. For other NGOs, villages were selected with simple random sampling, and weights were developed based on estimates of the size of sampled villages relative to the total population or number of households throughout target and control areas. See Annex C for the SPSS syntax and related notes that were used to generate final sample weights for each NGO.

[3] See Annex A for a discussion of PPS sampling and the reasons why it is typically preferred in sample design. See Annex C for the sampling forms each NGO was to use.

B. Questionnaire Design

Data collection for the Income Proxy Work was done in two rounds, each covering a period of 6 months, so as to reduce the recall period and improve the quality of data collected. The first round of data collection was done in June/July 2001, covering the period between October 2000–March 2001 for cropping activities and January 2001 to June 2001 for all the other activities as indicated in section III of the questionnaires. Round two was done in February 2002, covering April –Sept 2001 for cropping activities (thus completing a full year of cropping activities) and July to December 2002 for all other activities (also completing a full year for non-cropping activities). Note that the recall period for cropping activities was meant to coincide with the cropping seasons of each region; harvest from these seasons each occurred during the recall period of the surveys, and costs incurred in obtaining that harvest were included in the data collection to calculate net values of production. Production seasons are as shown below:

East Kenya:	Main harvest	Jan-March
	Short harvest	Jul-September
West Kenya:	Main harvest	Jul-October
	Short harvest	Nov-January

Separate but similar structured survey instruments were designed and used for each of the two data collection rounds.[4] These were designed to collect detailed income data and likely proxy variables for model building.

The survey instruments covered the following topical areas: identifying variables; crop inventory; cropping activities for last harvest; inputs (seeds, fertilizer, chemicals, hired labor) for last harvest; map of fields for current season; livestock investments; demography; businesses run by members of the family, informal labor activities; salaried/permanent wage employment; purchases for home consumption; household assets, and perception questions on various economic indicators. Both income and cost data were collected in order to compute net incomes. The crops part as shown in section I of the questionnaires was the most complex, involving detailed crop production and sales data based on respective cropping seasons. To facilitate this recall and minimize the probability of missing any crops, enumerators were trained to develop a map of fields before every crop table, to list crops that were grown on each filed, and only then to collect data on each crop in each field.

The livestock section collected data on current stocks of each animal, purchases and sales of live animals, production and sales of livestock products, and costs associated with livestock production such as labor, vaccinations, tick control, and animal feed.

Demographic information as it relates to household members, their ages, gender, levels of schooling, and involvement in the household's economic activities, was also collected. The key issue in this section involved the definition of a household and its implication for

[4] Questionnaires from both rounds are available upon request, in Word format. Software incompatibility with WordPerfect precludes integrating them into this paper as an Annex.

sampling and interpretation of results, since this was found to differ across different cultural settings. For the purposes of this survey, a common definition was adopted which considered a household to be composed of individuals living in the farm and sharing resources (e.g, labor contributed to farming activities, and food). Details of this definition and its refinement in the case of polygamous households are given on page 2 of the Enumerator manual in Annex D. This definition was used in defining what part of informal/business and salaried labor earnings was accounted for as being generated by the household. With this, any income from any of the other family members and friends not included in the family definition, but which was remitted in some way to the family, was accounted for as remittance.

Minor corrections and improvements were made in the second round questionnaire as a result of the first round experience. For example, during data cleaning for the first round data, it was discovered that several crops had been indicated in the crop inventory table but left out under the crop table. The second round questionnaire, therefore, included a section for the missed crop during the first round as shown in page 2.

C. Training and Field Work

In preparation for each round of field work, two training sessions were held. The first session involved training of trainers with the aim of going over the questionnaire with supervisors from the NGOs to prepare them for enumerator training and field supervision. This was done in Nairobi by Tegemeo/MSU researchers. The training involved understanding the key issues in the questionnaire plus the general approach to questionnaire administration as explained in the Enumerator manual given in Annex D.

Each NGO did their own recruitment of field enumerators from their project sites. The training of trainers was then followed by the training of enumerators in all the respective NGO sites. This was done by the NGO supervisors who went through the training of trainers, with assistance from Tegemeo staff. It involved going over the same process as with the training of trainers.

The training of enumerators was followed immediately by field work, whereby each NGO conducted the process as was appropriate for them. There was however frequent communication between the NGO's and Tegemeo staff.

There were a few difficulties encountered in the field.. For example, field personnel for the FHI project encountered cattle rustlers in one of the sites and had to suspend data collection until the danger subsided. In general, insufficient supervision during round 1 resulted in some households indicating that they had produced certain crops which were then not detailed in the proper crop production section. As reported in the previous section, a special page was therefore designed in round 2 to obtain data on these missed crops. Tegemeo also provided staff to supervise data collection in each NGO area during round 2, to minimize the possibility of further error.

D. Data Entry and Cleaning

Data entry for each round was done at the TS-HPI offices in Nairobi immediately after field data collection was complete using data entry forms designed in SPSS/Windows. Tegemeo/MSU researchers prepared the DE templates, while each of the NGOs provided at least one person to enter their own data, under the supervision of Tegemeo staff.

After the data entry process was finalized, data cleaning proceeded with Tegemeo/MSU at the forefront and every NGO providing at least one person in one of the weeks to assist in the cleaning process. The data cleaning process involved checking and correcting for data entry errors and data collection mistakes, whenever possible.

IV. Model Development

A. Definition of Income Components

Previous income proxy work in Kenya defined eight income components for estimation: retained cereals and tubers, sold cereals and tubers, retained fruits and vegetables, sold fruits and vegetables, industrial crops, livestock, informal off-farm, and salaries and remittances. Because the NGO data set had so few industrial crops (such as coffee, sugar, and tea), this component was dropped, and the few observations of those crops were integrated with cereals and tubers into "field crops". Thus, the income components defined for the NGO income proxy models were:

1. Retained field crops
2. Sold field crops
3 Retained fruits and vegetables
4. Sold fruits and vegetables
5. Livestock (sales of animals plus production and sales of animal products)
6. Informal off-farm
7. Salaries and remittances.

Retained and sold agricultural production components (1-4) were valued in the same way, using median district level sales prices for each survey period in districts that had at least 10 price observations. Districts that had less than 10 price observations during one survey period were merged with nearby districts to obtain at least 10, and production and sales in those districts were then valued at the same median price. See Annex E for an indicative list of prices used in the analysis

B. Types of Proxy Variables Tested

In attempting to estimate each of these components, emphasis was placed on identifying proxy variables that would be straightforward to collect and process, and which had strong logical and empirical links to the level of income from the component. Seven general types of variables were used in the models:

- Measures of the intensity of involvement in the activity. Measures of intensity varied by component, but for the agricultural components typically included the number of items within the category that the household produced (for example, the number of food crops that the household cultivated), and the number of items that it sold (or whether it sold any, or not). For off-farm components, this set of variables generally included the number of people involved in the activity (informal off-farm or salaried labor & remittances), and the number of months in the year in which someone was involved. This set of variables also included indicators of the specific nature of involvement in the activity (e.g., what general type of wage labor, or what type of informal business activity)

- Production function variables. These were the same for all cropping activities: total acres owned (rather than the more difficult to collect acres in specific crops), use of fertilizers (yes/no), and hiring of labor (yes/no).

- Selected quantitative variables. Quantitative variables are more complex to collect and process than typical proxy variables, but are needed because production levels can fluctuate substantially from year-to-year based on rainfall and other factors. By quantifying the production of the most important food crop and cash crop, these quantities can themselves proxy for yield levels of other crops within their category. This should substantially improve the performance of the method over time. We used five quantitative variables in the models: the quantity produced of the "most important" food crop for home consumption, the quantity produced of the food crop that gave most sales income, the quantity produced of the industrial crop that gave most sales income, the quantity produced of the "most important" fruit or vegetable for home consumption, and the quantity produced of the fruit or vegetable that gave most sales income. By allowing the households to specify their "most important" crop in these various categories and quantifying that, the models should do a good job capturing the effect of changing cropping patterns in rural areas o the country.

- Farmer assessment of the crop harvest. This set of variables includes adverse event variables for the crop production components, such as damage from several sources (yes/no), the number of crops that were completely lost due to any problem, and the farmer's overall assessment of the quality of the year's harvest. These variables will help the models capture year-to-year changes in weather and pest problems.

- Household characteristics, such as schooling of the head of household, whether the household is female-headed, and the estimated value of non-land assets held by the household.[5]

- Household ranking of the relative importance of the income source compared to other sources.

[5] This proxy variable is generated from a regression using simple yes/no responses to the ownership of a set of 15 assets. Thus, it is not necessary to collect number owned and value of a large set of assets to obtain this variable.

- Interaction terms. We made very liberal use of interaction terms to get maximum value out of the variables used. For example, by interacting the number of months that anyone in the household earned income from any informal off-farm activity (a simple yes/no question) with indicators of the type of activities that the household Was involved in (also yes/no questions), we obtained a proxy for the number of months worked in that specific activity; this variable, and others like it, was quite useful in several of the models.

These variables were generated in spss/windows syntax for all income components in all zones. They were then tested in a stepwise regression framework, which admitted only those variables whose explanatory power surpassed predetermined minimum levels. In this way, the most efficient set of proxy variables was defined for each model.

C. Results

In order to test the performance of developed models, it is important to assess how well they explain variations/changes in household incomes and how well they predict income and income sources. Full model results, including goodness of fit, proxy variables in each regression, their coefficient values and level of statistical significance, can be found in Annex F. Table 2 gives the Coefficient of Determination (R^2) for each model in each zone. The Coefficient of Determination measures the proportion of variation in the dependent variable that is explained by the model. Overall, most of the models explain the changes in total household income quite well (over 85% in all zones). The R^2 for the component incomes in

Table 2. Predictive Power by Zone

Component	Zone 1	Zone 2	Zone 3	Zone 4
	---------- R^2 ----------			
Retained field crops	0.768	0.824	0.707	0.598
Sold field crops	0.861	0.996	0.815	0.617
Retained f&v	0.804	0.827	0.899	0.677
Sold f&v	0.975	0.703	0.948	0.947
Livestock	0.632	0.742	0.856	0.798
Informal off-farm	0.958	0.711	0.836	0.865
Salary & Remittance	0.774	0.961	0.888	0.819
Total Income	0.878	0.951	0.888	0.874

Note that:

- R^2 value of 1.0 means the model perfectly predicts each and every value of the dependent variable, while
- R^2 value of zero means that you would do just as well saying that everyone is equal to the mean.

each zone are also fairly high, indicating that most of the variation in component incomes can be explained by the proxy variables. The sold crop components however, perform slightly better (explains above 87%) except for sold field crops in Zone 4 and sold fruit & vegetables in Zone 2. Generally, livestock models seem more difficult to explain, as evidenced by the relatively low R^2 given in the table. [6]

In terms of comparing the actual and predicted values, Table 3 shows that the mean total household income is generally predicted with higher accuracy than most of the component incomes i.e. less than 1%. Zone 4 models are predicted with greatest accuracy while Zone 2 has relatively larger errors. The models for Zones 1 and 3, which have apparently shown a lot of similarities, are fairly well and almost equally predicted. As for income sources, the retained fruits and vegetables and salaries and remittances seem to perform better than the others.

One potential use of these models is to classify households into broad income groups. Table 4 compares how well the models classify households into three income groups of equal size, or income terciles. The results show that over 90% of those households actually in the top income tercile were correctly predicted to be in that tercile by these models. Only 1% of the best-off households were predicted to be in the bottom tercile -- all other errors of classification were of only one tercile. Of those households actually in the bottom tercile -- the poorest households -- 71% were correctly classified by these models. Of the 29% that were misclassified, nearly all (26%) were misplaced by only one tercile group. Thus, the models perform better in identifying the better-off households than they do in identifying the poorest households, but still correctly identify nearly three-quarters of these poor households.

V. Using the Models

Using the models developed in this work to generate estimates of income and our seven income components involves first collecting the simplified proxy data, entering it into a specific data structure, and then running the SPSS/Windows syntax file which converts the proxy data into estimates of household incomes and income components. In practice, the results generated by the syntax file then need to be critically reviewed to be sure they are reasonable, and underlying proxy variables need to be examined for implausible cases. See Figure 1 in section II.C. for a graphic presentation of this process.

[6] The R^2 values for the component incomes are for those households that had incomes from that source -- households with no incomes from the source were left out. In calculating R^2 values for total household income, households with no income from a given source were assigned values of zero, since the proxy variables will also indicate unambiguously whether or not a household had such income. For this reason, the R^2 values for total household income are higher than the weighted average values over all the components.

Table 3. Predicting Means by Income Component by Zone

Component	Zone 1			Zone 2			Zone 3			Zone 4		
	Actual Value	Predicted Value	% Error	Actual Value	Predicted Value	% Error	Actual Value	Predicted Value	% Error	Actual Value	Predicted Value	% Error
Retained field crops	10,928	10,815	-1.0%	3,667	3,882	5.9%	15,065	14,922	-0.9%	19,087	19,051	-0.2%
Sold field crops	4,831	4,909	1.6%	5,045	4,938	-2.1%	5,598	5,428	-3.0%	1,609	1,609	0.0%
Retained f&v	2,109	2,135	1.2%	997	995	-0.2%	4,996	5,028	0.6%	13,759	13,801	0.3%
Sold f&v	5,596	5,593	-0.1%	3,323	2,939	-11.6%	1,983	1,963	-1.0%	19,685	19,658	-0.1%
Livestock	10,430	10,419	-0.1%	4,897	5,297	8.2%	9,184	9,202	0.2%	13,514	13,516	0.0%
Informal off-farm	22,019	21,633	-1.8%	8,227	7,819	-5.0%	8,433	8,570	1.6%	29,011	29,031	0.1%
Salary & Remit.	34,525	34,159	-1.1%	16,235	16,397	1.0%	40,147	39,626	-1.3%	117,082	116,349	-0.6%
Total Income	**90,438**	**89,663**	**-0.9%**	**42,390**	**42,268**	**-0.3%**	**85,406**	**84,739**	**-0.8%**	**213,747**	**213,014**	**-0.3%**

Table 4: Prediction of Households into Respective Income Groups

		Predicted Income		
		Tercile 1	Tercile 2	Tercile 3
Actual Income	Tercile 1	71%	26%	3%
	Tercile 2	10%	85%	5%
	Tercile 3	1%	7%	92%

Annex A is a sampling guide to be used in designing the sample for the proxy surveys. Annex B contains the model questionnaire that can be used to collect the needed proxy data. During actual NGO data collection in 2003, additional sections were added to this questionnaire at the request of NGOs. This can be done -- modules or sections can be added -- as long as a) nothing is removed from the model questionnaire and b) the basic structure of the model questionnaire is not altered. If any sections are removed, it will not be possible to run all the prediction models accurately. If the structure of the questionnaire is altered, the syntax file which generates results will have to be modified to run properly, and these modifications can become complex if substantial changes are made in the questionnaire.

Annex H provides ste-by-step instructions for entering the proxy data, structuring and saving the files, and running the SPSS syntax file to generate results. ***It is imperative that these procedures be followed closely to avoid substantially increasing the complexity of generating these income proxy results.***

VI. An Overview of the Rural Economy in the Study Zones

This section provides a descriptive overview of the rural household economy in the study zones, based on results of the full income survey.

A. Crop Production and Marketing Behavior

Crop production in most of the NGO sites entails the usual dryland rainfed agriculture, as most of the intervention areas are in the marginal areas of the country where food security is an issue.

As shown in Table 5, maize is most widely grown in all zones with at least 95% of the population involved. It also turns out to be either the most or second-most valuable in mean value of production among those who produce the crop. Mangoes in Zone 4 and Miraa in Zone 2 have the highest mean production value among those households producing those crops. Miraa in Zone 2 is produced entirely in the FHI region, but only grown by 21% of the sample which is approximately 42% of the FHI sample alone.

There is a general similarity across zones, in that most of the crops in the top ten cut across other zones and no one zone has more than two crops in the top ten that do not appear in at least one other zone as shown in the table.

Note that no traditional cash crops (coffee, tea, industrial sugar cane, cotton) appear in the top 10 of any zone. This is due to the fact that most NGO intervention areas are in the marginal agricultural areas which do not favor the production of such crops. Most of the crops in the top ten are cereals, pulses and some horticultural crops. Fruits and vegetables are extremely important in Zone 4, followed by Zone 2. They comprise seven out of the top ten crops in Zone 4 and four out of the top ten in Zone 2. Zone 3 has only bananas and pawpaws.

Table 5. Top 10 Crops Grown in Each Zone: Percent Growing and Mean Value Produced

Zone 1 (CARE)			Zone 2 (WV, FHI)			Zone 3 (Adra, CRS, TT)			Zone 4 (TS-HPI)		
Crop	**% Growing**	**Mean gross value prod. (Ksh)**	**Crop**	**% Growing**	**Mean gross value prod. (Ksh)**	**Crop**	**% Growing**	**Mean gross value prod. (Ksh)**	**Crop**	**% Growing**	**Mean gross value prod. (Ksh)**
dry maize	99	7,104	dry maize	95	3,083	dry maize	98	13,010	dry maize	99	10,222
dry beans	84	3,054	sorghum	95	2,087	cowpeas	83	2,387	cassava	89	5,271
sorghum	80	2,458	cowpeas	79	329	**pig. peas**	67	1,835	cowpeas	85	2,705
sukuma wiki	65	4,234	g. grams	53	471	sorghum	63	2,555	bananas	82	4,752
tomatoes	59	2,868	dry beans	50	1,400	g. grams	62	3,841	**coconuts**	81	9,018
sweet pot.	56	1,149	tomatoes	37	1,388	**pumpkin**	48	4,764	**cashew**	78	7,640
groundnuts	46	4,549	pawpaws	36	1,100	dry beans	48	3,218	mangoes	75	16,413
cowpeas	43	2,046	suk. wiki	35	2,075	cassava	44	884	pawpaws	73	1,123
cowpea leaves	42	579	**onions**	23	436	bananas	41	2,368	indig. veg.	68	857
indig. veg.	37	1,714	**miraa**	21	16,449	pawpaws	38	1,377	tomatoes	58	5,445
ALL CROPS		26,219	ALL CROPS		12,742	ALL CROPS		32,941	ALL CROPS		66,283

Note: Bolded crops are in top ten only in that zone; mean gross value of production is among those households producing the crop.

Although almost the entire sample in the four zones engages in maize production, only a small proportion of them sell maize as shown in Table 6. Less than 10% of the sample in Zones 2 and 4 sell maize. This implies that maize is mainly grown for home consumption in these zones. In zones 1 and 3, 41% and 34% of households, respectively, sell maize. Later in the section we will examine whether there is evidence that these households need to purchase maize back later in the year.

Miraa in Zone 2 gives the highest mean income among those growing and those selling, but contrary to expectations, only 49% of those who grow made any sales. It would appear that the lack of sales among more than half the growers of miraa is due to very low production: sellers produced an average of 136 kg of miraa, while non-sellers averaged only 3 kg of production.

Note that the mean value of both production and sales in Zone 4 are more than double that in any other zone. This could be a result of higher yields in this area due to the relatively favorable agro-climatic conditions. As expected, Zone 2 has the lowest mean value of production and sales.

Table 7 gives some commercialization indicators by zone. Commercialization generally defines the proportion of agricultural production that is marketed. The table shows the relatively low levels of commercialization in almost all NGO sites. In three of the four zones, less than one-third of produced crops are sold in any quantity. Zone 2 has the lowest proportion of value sold and has a median value of zero and a mean number of crops sold of only 1.4. These results are partly determined by the drought that affected that region during the survey period; yet drought is a common feature in this zone, so these results are not considered atypical. On the other hand, Zone 4 has the highest proportion of value sold (followed very closely by Zone 1) and by far the highest median value sold, which could be due to the presence of crops like cashew, coconuts and mangoes which are mainly grown for cash as shown in Table 6. The median value sold represents the middle value such that half of households have values above, and one-half have values below, the median. A median value sold of zero for Zone 2 implies that at least 50% of the sample did not sell any of their crops.

It is important to note that median is preferred over the mean as an "average" measure when extreme values are present, because the mean is heavily affected by these values (Ramanathan, 1993). In other words, the median, unlike the mean, is not affected by extreme values (outliers), and hence may give a better reflection of the situation of a "typical" household.

Table 6. Top 10 Crops Grown in Each Zone: Percent Selling and Mean Value Sold

Zone 1 (CARE)			Zone 2 (WV, FHI)			Zone 3 (Adra, CRS, Taita Taveta)			Zone 4 (TS-HPI)		
Crop	**% Selling**	**Mean gross value Sold (Ksh)**	**Crop**	**% Selling**	**Mean gross value Sold (Ksh)**	**Mean Value Sold**	**% Selling**	**Mean gross value Sold (Ksh)**	**Crop**	**% Selling**	**Mean gross value Sold (Ksh)**
dry maize	41	4,613	dry maize	9	2,367	dry maize	34	6,299	dry maize	3	3,605
dry beans	45	2,287	sorghum	8	980	cowpeas	29	1,740	cassava	34	4,075
sorghum	22	1,868	cowpeas	9	301	**pig. peas**	19	2,737	cowpeas	15	2,052
sukuma wiki	75	4,511	g. grams	25	168	sorghum	16	820	**bananas**	19	6,869
tomatoes	63	3,483	dry beans	5	4,813	g. grams	72	3,622	**coconuts**	75	6,213
sweet pot.	33	905	tomatoes	28	3,342	**pumpkin**	1	3,665	**cashew**	89	8,123
groundnuts	57	3,012	pawpaws	35	817	dry beans	19	4,944	mangoes	59	19,656
cowpeas	33	3,016	suk. wiki	33	5,646	cassava	23	771	pawpaws	18	967
cow. leaves	46	564	**onions**	36	620	bananas	34	1,679	indig. veg.	30	721
indig. veg.	46	3,243	**miraa**	49	32,235	pawpaw	28	1,834	tomatoes	40	6,950
ALL CROPS		11,816	ALL CROPS		5,798	ALL CROPS		7,038	ALL CROPS		24,883

Notes: 1) Percent selling is among those that grew the crop, 2) Mean value sold is among those selling

Table 7. Commercialization Indicators by Zone

Zone	Mean # of Crops Produced	Mean # of Crops Sold	Mean % of Total Value of Crop Production that is Sold	Median Value Sold (Ksh) (Mean)
Zone 1 (CARE)	10	4.4	27	3,839 (11,816)
Zone 2 (WV and FHI)	8	1.4	13	0 (5,798)
Zone 3 (Adra, CRS, TT)	10	2.6	16	2,640 (7,038)
Zone 4 (TS-HPI)	15	4.3	28	9,810 (24,883)

B. Livestock

Livestock production is a major activity in most NGO sites, with cows, bulls, goats and chicken rearing being the main activities. Table 8 gives the percentage of households owning different livestock types by zone and the median number owned among those owning. Nearly all the cows are local or indigenous, this being due to the fact that most of the NGO intervention areas are dry and not conducive to the raising of grade animals. As expected, Zone 2 has the highest ownership of cows, bulls and sheep, but very low ownership of chickens. The relatively low ownership of chickens can be attributed in part to the continued presence of pastoral dimensions to these households' livelihood strategies, which makes it difficult to rear chicken because of the need to occasionally move them from place to place.

Table 8. Livestock Ownership

Zone	Cows, Bulls, Calves		Sheep		Goats		Chickens		Bees (Hives)	
	% Owning	Median # Owned	% Owning	Median # Owned	% Owning	Median # Owned	% Owning	Median # Owned	% Owning	Median # Owned
Zone 1 (CARE)	74	6	22	2	57	4	83	10	1	2
Zone 2 (WV and FHI)	76	5	44	5	61	7	42	5	3	2
Zone 3 (Adra, CRS, TS-HPI/Taita)	61	2	14	2	62	5	89	10	41	5
Zone 4 (TS-HPI/Coast)	65	3	2	3	67	6	85	17	4	1

Note: # owned is among those owning

Zone 3 has the highest ownership of bee hives, this being the major off-farm activity in the Ikutha division of Kitui district where ADRA is based. Consequently, this zone has the highest median gross value of honey as shown in Table 9.

From Table 9, the main livestock products are milk and eggs, with most households being involved. Milk and milk products give the greatest income compared to other products, with Zone 4 having the highest median/mean value of milk production. This is not surprising since Zone 4 represents the TS-HPI project area whose intervention is in milk production and marketing. Note that, although Zone 2 has the least percentage of households producing eggs or rearing chicken, it has the highest median gross value (and the second highest mean value) among those producing.

Table 9. Livestock Products: Production by Zone

Zone	Milk		Eggs		Honey		Hides & Skins	
	% producing	Median gross value produced, Ksh (mean)	% producing	Median gross value produced, Ksh (mean)	% producing	Median gross value produced, Ksh (mean)	% producing	Median gross value produced, Ksh (mean)
Zone 1 (CARE)	62	7,200 (13,455)	78	265 (1,011)	1	NA	13	200 (319)
Zone 2 (WV and FHI)	66	4,060 (6,660)	38	706 (1,515)	2.4	NA	28	90 (239)
Zone 3 (Adra, CRS, TS-HPI/Taita Taveta)	52	4,800 (9,900)	82	475 (1,110)	28	1,278 (6,863)	11	90 (135)
Zone 4 (TS-HPI/Coast)	60	32,400 (34,790)	75	450 (1,538)	2.5	NA	10	60 (62)

Note: Median and mean value produced are among those producing

C. Structure of Household Incomes

Table 10 presents detailed per capita income levels and shares by zone. Household incomes comprise the total net value of all productive and income earning activities to the household, both cash and in-kind. As expected, the highest per capita incomes are found in Zone 4 and lowest in Zone 2. This pattern is partly attributable to different agricultural endowments of the two regions, and partly to the dramatically higher income shares from salaries and remittances in Zone 4.

The largest share of income for Zones 1 and 3 is from retained field crops which is mainly from the usual dry land cultivation of cereals and pulses for home consumption. As expected, the largest income share for Zone 2 is from livestock (21%), with most of the crop components except retained field crops and informal having no contribution. On the other hand, salaried activities and remittances contribute the largest share of income for Zone 4 due to job availability in the tourism industry, these being coastal areas.

Table 10. Detailed Income Levels and Shares, by Zone

Zone	Median *per capita* Income Level, Ksh (Mean)	Median (Mean) Income Share from ...						
		Retained Field Crops	Sold Field Crops	Retained F&V	Sold F&V	Livestock & livestock products	Informal off-farm	Salaried off-farm + remittances
Zone 1 (Care)	10,200	**18.8**	2.7	2.9	2.2	6.3	8.5	3.8
	(20,061)	**(29.4)**	(8.5)	(5.0)	(8.8)	(3.5)	(20.9)	(23.8)
Zone 2 (WVI and FHI)	3,036	8.7	0.0	0.0	0.0	**21.4**	0.0	1.3
	(7,002)	(18.3)	(2.3)	(2.7)	(4.6)	**(31.6)**	(14.0)	(26.5)
Zone 3 (CRS, ADRA, TS-HPI/Taita)	9,022	**21.7**	1.8	4.5	0.0	2.6	9.4	9.1
	(15,907)	**(24.4)**	(6.6)	(8.8)	(2.5)	(11.6)	(18.9)	(27.3)
Zone 4 (TS-HPI/Coast)	15,443	10.2	0.0	6.8	5.6	2.0	9.0	**28.9**
	(24,655)	(15.3)	(1.1)	(10.5)	(9.4)	(7.2)	(19.0)	**(37.6)**

Note: Median income shares do not sum to 100%; mean shares do sum to 100%

Consistent with the low degree of commercialization as shown in Table 6, the sold components contribute negligibly to household income.

In Table 11 we aggregate these seven income sources into three (cropping agriculture, livestock, and off-farm) to facilitate broad comparison of income strategies in the zones. The table shows that off-farm income is the most important activity in every zone, though it is in a virtual tie with cropping agriculture in Zone 1. Over 40% of the median per capita income in all the zones come from off-farm activities and remittances, the exact contribution differing across zones. Households in Zone 4 have on average close to 2/3 of their income from off-farm activities, owing to its location in the coastal areas. Crop incomes contribute much of the rest as livestock contributes negligibly except in Zone 2. This is not unexpected given that the study sites are all in the marginal areas with unfavorable weather conditions for crop production, resulting in higher concentration in off-farm activities.

It is not only important to understand household income levels and shares, but also how this income is distributed among participating households. Table 12 breaks households in each zones into three groups of equal size (terciles) based on income, and presents mean and median incomes for each of these groups. The table shows a skewed income distribution towards the highest income groups especially for the poorer zones. The top 33% of the sample earn about nine times more than the lowest group in Zones 1 and 3. This gap increases to 20 times in Zone 2 (the poorest zone) and decreases to 6 times in Zone 4 (the zone with highest mean and median incomes). This implies lower income disparities for the well endowed zones than for the 'poor' ones. Overall, about 21% of households in this sample have nominal per capita incomes above the national GDP per capita for 2000 (ksh.22,943) as given by the Economic Survey, 2001 which is very similar to the results of the Tampa Household Survey in 2000 (ksh. 22,112). This share is highest in Zone 4 (34% above the national mean) and lowest in Zone 2 (7%).

Table 11. Aggregated Income Levels and Shares, by Zone

Zone	Median *per capita* Income Level, Ksh (Mean)	Median (mean) income share		
		Ag	Livestock	Off-farm + remittances
Zone 1 (Care)	10,200	**39.8**	6.3	40.7
	(20,061)	**(51.7)**	(3.5)	(44.7)
Zone 2 (WVI and FHI)	3,036	17.7	21.4	**25.9**
	(7,002)	(23.1)	(34.9)	**(41.9)**
Zone 3 (CRS, ADRA, TS-HPI/Taita Taveta	9,022	38.7	2.6	**46.3**
	(15,907)	(42.3)	(11.6)	**(46.2)**
Zone 4 (TS-HPI/Coast)	15,443	29.9	2.0	**61.0**
	(24,655)	(36.3)	(7.2)	**(56.6)**

Note: Median income shares do not sum to 100%; mean shares do sum to 100%

Note that Zones 1 and 3 are nearly identical in income distribution as well as income shares (see table 10) due to similarities in the agro-climatic conditions of the two regions. Zone 1, however, has more skewing of income at the top end, as shown by the higher mean income, but similar median, in tercile three as compared to Zone 3.

Our final income table (Table 13) provides a further breakdown of income shares into income terciles by zone. The relationship between income levels and the share of income coming from off-farm activities has received a great deal of study in Africa and other developing areas of the world. The interest in this topic stems from the desire to identify the most effective ways out of poverty for low income rural households. Results have generally shown that, while in Asia the lowest income households are most reliant on off-farm incomes, in Africa the pattern is typically reversed: higher income households tend to have greater shares of income from off the farm, while lower income households do not have access to the more remunerative types of off-farm activities, and thus must rely primarily on agriculture. This pattern would be expected to be strongest in areas where the density of population and infrastructure are low, and where agro-ecological potential is low.

Table 12. Incomes by Per Capita Income Tercile

Zone	Median *per capita* Income Level, Ksh	Median Income by Income Tercile		
	(Mean)	Tercile 1 (lowest income)	Tercile 2	Tercile 3 (highest income)
Zone 1 (Care)	10,200	3,422	10,200	27,903
	(20,061)	(3,759)	(10,800)	(45,512)
Zone 2 (WVI and FHI)	3,036	486	3,036	9,494
	(7,002)	(240)	(3,022)	(17,789)
Zone 3 (CRS, ADRA, TS-HPI/Taita Taveta	9,022	3,802	9,022	27,605
	(15,907)	(3,844)	(9,477)	(34,430)
Zone 4 (TS-HPI/Coast)	15,443	6,800	15,443	41,760
	(24,655)	(7,035)	(15,962)	(50,968)

It is thus not surprising that the table shows agriculture's share of income falling sharply while off-farm share rises sharply, as income rises in zones 1, 3, and 4. Livestock shares are low in each of these three zones, and change little across the different income groups. Note however, that absolute income levels from cropping agriculture are higher among the highest income groups than they are among the lowest income groups. Thus, it is clear that the better-off households are earning higher incomes from all broad sectors of economic activity, whether cropping agriculture, or livestock or off-farm, but that off-farm incomes are the primary reason that they have risen to the highest income tercile. This is an important finding that bears further research to elucidate potential programmatic implications for NGOs.

Patterns in Zone 2 are slightly different. First, the lowest income households were badly affected by drought, and thus earned almost no income from cropping agriculture. Aside from this, the clearest pattern which emerges in Zone 2 is that households move out of livestock in a relative sense and into off-farm activities as incomes rise. From the bottom to the top income terciles in Zone 2, livestock shares fall by 33 percentage points while off-farm shares rise by 39 percentage point.

Table 13. Income Shares by Per Capita Income Tercile, by Zone

Zone	Income Tercile	Median Income, Ksh (Mean)	Median Income Share from ...		
			Agriculture	Livestock	Off-farm
Zone 1 (Care)	1 (lowest)	3,422	0.66	0.03	0.21
	2	10,200	0.40	0.12	0.37
	3 (highest)	27,903	0.12	0.04	0.79
Zone 2 (WVI and FHI)	1 (lowest)	486	0.01	0.40	0.19
	2	3,036	0.22	0.25	0.13
	3 (highest)	9,494	0.20	0.07	0.58
Zone 3 (Adra, CRS, TS-HPI/Taita Taveta)	1 (lowest)	3,802	0.47	0.04	0.35
	2	9,022	0.44	0.03	0.39
	3 (highest)	27,605	0.29	0.02	0.66
Zone 4 (TS-HPI/Coast)	1 (lowest)	6,800	0.47	0.01	0.41
	2	15,443	0.33	0.03	0.58
	3 (highest)	41,760	0.18	0.02	0.77

Note: Income shares are medians, and thus do not sum to 100%

D. Staple Purchase and Sales Behavior by Income Class

Earlier sections of this chapter showed that off-farm incomes are important contributors to household overall income in the survey areas. We also know that most of these areas have relatively low agro-ecological potential. Under these circumstances, once would expect purchases to be an important source of food staples for some households. We therefore examine staple food purchase behavior in this section, focusing on two groups of staples: maize grains and flours, and wheat flour, bread, and rice.[7]

Table 14 examines the proportion of households purchasing each of these staple groups, and quantities purchased, by income level in the four zones. The first pattern which emerges is that 9 out of every 10 households in Zones 1, 3, and 4, and 8 out of 10 in Zone 2, purchase wheat flour, bread, or rice. Smaller but still substantial percentages purchase maize grain and meal. The second key pattern is that, with the exception of Zone 2, maize grain and flours show evidence of being "inferior goods" in economic terms[8]. Wheat flour, bread, and rice

[7] Wheat flour, bread, and rice were grouped together because purchase behavior for each was very similar, and all were substantially different from maize grain and maize meal.

[8] "Inferior" and "normal" goods are economic terms based on observed household behavior, and are not meant to imply any judgements about the suitability of such foods for human consumption.

are "normal goods" in every zone. In other words, as income rises, households the Zones 1, 3, and 4 are less likely to purchase maize grain and flours, while in every zone households are more likely to purchase wheat or rice as income rises. The quantities of wheat and rice purchased also tend to rise with income. In Zone 2, maize purchases tend to follow a more "normal" pattern, with both the percent of households purchasing and the quantities purchased rising with income. This difference between Zone 2 and the other zones is consistent with the very low incomes in Zone 2.

Tables 15 and 16 are potentially useful in addressing the question of whether substantial numbers of households in these low-potential areas sell maize during the harvest and then find themselves obligated to purchase it at higher prices during the short season to meet consumption needs. This is a frequent concern in rural areas of Africa, and is often conceived as "distress sales" at very low prices during harvest to meet pressing needs, followed by purchases later in the season at high prices to meet basic consumption needs. How common this pattern is, and what type of household might fall into it, is typically not informed by empirical information about actual household behavior. To begin addressing the issue, in Table 15 we classify households into four mutually exclusive and exhaustive groups with respect to their behavior in the maize market: households that neither sold nor purchased maize, those that only sold, those that only purchased (grain or flours), and those that both purchased and sold. By design, all households in the sample fall into one of these categories.

In Table 16, we present the mean per capita incomes of each of these group, along with quantitative information on their maize production, purchase, and sales behavior.

Several patterns emerge. First, households that both purchased and sold maize – the only ones which might be engaging in distress sales followed by high cost purchases -- is the smallest group in every zone. Second, in Zones 2 and 4, the proportion in this category is so low – about 1% – that in these zones we can unambiguously reject the hypothesis that a significant share of the rural population engages in distress sales of maize in the harvest season only to repurchase later in the year at high prices. It also appears unlikely in these two zones that many households sold maize and then were unable to purchase it later in the year when they needed it – only 6.1% and 2.6% of households in Zone 2 and Zone 4, respectively, sold grain without purchasing it later in the year.

Table 14. Purchase Behavior on Main Staples, by Income Tercile and Zone

Zone and Income Tercile	Maize grain or flours		Wheat flour, bread, or rice	
	% Purchasing	Median (mean) quantities purchased among those purchasing (kg)	% Purchasing	Median (mean) quantities purchased among those purchasing (kg)
Zone 1 (Care)				
Tercile 1 (lowest)	63.0	54 (81)	88.0	12 (17)
Tercile 2	52.9	88 (118)	92.2	12 (21)
Tercile 3 (highest)	44.0	90 (126)	98.2	20 (27)
Zone 2 (WVI and FHI)				
Tercile 1 (lowest)	27.7	29 (70)	76.6	6 (9)
Tercile 2	25.9	31 (45)	74.7	6 (7)
Tercile 3 (highest)	44.9	42 (115)	91.5	20 (24)
Zone 3 (CRS, ADRA, TS-HPI/TT)				
Tercile 1 (lowest)	65.2	100 (157)	90.4	9 (14)
Tercile 2	44.0	40 (99)	96.2	14 (25)
Tercile 3 (highest)	51.7	90 (121)	93.2	14 (22)
Zone 4 (TS-HPI/Coast)				
Tercile 1 (lowest)	49.2	54 (112)	96.8	18 (28)
Tercile 2	49.2	58 (102)	95.2	20 (28)
Tercile 3 (highest)	35.5	44 (103)	98.4	23 (40)

Table 15. Maize Market Position of Households by Zone

Position in maize market	Zone			
	1 (Care)	2 (WVI, FHI)	3 (CRS, ADRA, TS-HPI/Taita Taveta)	4 (TS-HPI/Coast)
	% of households			
Neither sold nor purchased	22.4	61.6	27.9	53.4
Sold only	24.3	6.1	18.7	2.6
Purchased only	36.6	31.2	40	43.4
Both purchased and sold	16.7	1.2	13.4	0.5

Third, the evidence in Table 16 suggests that households that only sold maize (who are concentrated in Zones 1 and 3) were unlikely to need to repurchase maize for basic consumption needs later in the year. Households in this group had the largest mean and median maize production, highest mean and median net maize availability, and also the highest mean and median per capita income of any group. The median household in this group retained maize equivalent to about 3,000 kcal per adult equivalent in the household, which would have been enough to meet the calorie needs of the entire household only with the retained maize .[9]

Finally, households who both purchased and sold maize (these are also concentrated in Zones 1 and 3), despite having the lowest median per capita incomes, had the second-highest net maize availability of any group, only slightly below households that only sold. Taken together, this evidence suggests strongly that distress sales of maize were not a widespread problem in this survey area; if there were households who compromised their food security through such sales, they were a very small minority.

[9] Adult equivalents are based on the calorie needs of household members of differing ages and genders. 3,000 kcal/day is considered adequate for a normally active adult African male between the ages of 18 and 30. Caloric needs for different ages and females are lower. See Annex G for actual values used in this analysis.

Table 16. Selected Household Indicators by Maize Market Position

Position in maize market	% of hhs	Median (mean) per capita income (ksh)	Maize Behavior						
			% producing	median (mean) per capita kg produced among those producing	% purchasing	median (mean) per capita kg purchased among those purchasing	% selling	median (mean) per capita kg sold among those selling	Median (mean) maize availability in kcal/ae/day
Neither sold nor purchased	39.7	9,352 (15,719)	90.9	72 (124)	0		0		1,465 (808)
Sold only	13.1	10,370 (21,912)	100.0	225 (381)	0		100	51 (143)	3,045 (1,916)
Purchased only	38.2	9,391 (17,948)	92.3	47 (73)	100	10 (18)	0		1,103 (786)
Both purchased and sold	9	7,115 (16,546)	100.0	171 (222)	100	18 (31)	100	30 (61)	2,542 (1,853)

Note: maize availability in kcal calculated as production plus purchases minus sales, divided by number of consumption adult equivalents in household.

ANNEXES

Annex A

Sampling Guidelines for Proxy Surveys

Income Proxy Surveys:
Guidelines for PVO Sampling

By Rui Benfica

and

David L. Tschirley

June 2003

1. Introduction

To report results with greater accuracy and reliability across the different areas where PVOs operate, and to increase the comparability of reporting across PVOs, it would be appropriate that all organizations followed, to the extent possible, some basic steps in the design of their samples. The guidelines presented here are aimed at providing PVOs with some key principles to be applied and steps to be followed, in order to improve the quality of their data and reporting, given constraints on time, personnel, and money. These guidelines do not represent USAID "policy", but rather technical suggestions to be applied whenever possible. The closer these guidelines are followed the better the USAID Mission will be able to track performance and impact across the board. Some PVOs are already implementing their surveys using the approach suggested here or a version that is close to it.

This paper is in no way meant to be a comprehensive guide to survey sampling. Consult survey sampling texts for questions which may emerge from reading this paper. A helpful and relatively accessible guide to survey sampling is Graham Kalton, "Introduction to Survey Sampling", Quantitative Applications in the Social Sciences Paper No. 35, Sage Publications. 1985.

2. Basic Principles of the Sampling Approach

The basic principles suggested are:

- Besides the usual target group, include a control group in the sample;
- Draw samples of similar size in the control and target groups;
- Design samples that are probability proportional to size (PPS) in both target and control groups;
- Present results separately for target and control groups

Background and, where relevant, specific steps to follow in applying these principals are presented in the following sections.

2.1. Control and Target Groups

To compare households assisted and not assisted by PVO programs, the sample should include both a target and a control group. The question then is how to develop a definition of these two groups that is *workable* in terms of available time and resources, and *meaningful* in a reporting context. Given the various types of programs in place and the likely indirect impact over undefined areas, there is seldom a straightforward, "correct" definition of the two. Therefore, each PVO needs to develop a definition they consider workable and meaningful, according to their specific circumstances.

In doing so, be clear about the *level* at which you make the definition:

- Defining the two groups at the ***household level*** implies that you can have both target and control households in a single village. This may be most meaningful for interventions which are easily targeted to specific households and which have little spillover or

demonstration effect on other households. However, if the intervention does have significant spillover or demonstration effects, then a household level definition may not be the most meaningful. In any case, a household level definition will require lists of all households stratified (classified) as target and control. Developing such lists may require substantial additional work prior to fielding the survey. Thus, in general, a household level definition will typically require more time and resources - will be less workable - than a village level definition.

- Defining the two groups at the ***village level*** assumes that entire villages are affected by the interventions of the PVO, or not. Such a definition is most meaningful when an intervention has significant spillover or demonstration effects. Preparing the sample using a village level definition may require significantly less time and effort than using a household level definition, so in general the village level approach is the most workable.

Since many PVO interventions have spillover and demonstration effects, defining target and control groups using a village level approach will typically provide the best combination of workability and meaning for PVO impact surveys. If a PVO already has lists of target and control (participant and non-participant) households for its villages, and if it is confident that its interventions have few spillover or demonstration effects, then it might consider using a household level approach. The discussion in this paper is oriented towards a village level approach.

2.2. Sample Size

The size of the sample must be decided at three levels:

9. The total sample size in each group - target and control. We will refer to this number as n.
10. The distribution of that sample over villages i.e., the number of villages in each group (v).
11. The number of households to interview in each village (h).

Total sample size in each group: The primary purpose of defining control and target groups is to compare the means of selected variables across those groups. For example, you may want to know whether the maize yield in the target group is significantly higher than in the control group. This comparison of means across groups is most statistically efficient when the samples in the two groups are of equal size. Allowing the sample size in the groups to differ, for example by allowing each sample to be proportional to the size of its group, reduces the efficiency of the comparisons to be made. Thus, your design should call for total samples of equal size in the target and control groups. Given the practical problems of fielding surveys, actual sample sizes might differ slightly, but these differences should be minimized.

But what size should the sample be? There is no easy answer to this question for various reasons. First, a theoretically recommended sample size is a function of the desired level of accuracy, which in turn depends on the variance in the variable to be estimated. In this case,

we have many variables to be estimated, each with different and unknown variances. Second, the sample size is a function of available time and resources, particularly human and financial. However, as a rule of thumb, having a sample size of at least 200 households, preferably more, in each group is desirable.[10]

Number of villages and number of households in each village: The determination of number of villages and number of households per village can proceed in two ways:

- If you first decide how many villages to work in, then the number of households to be interviewed in each village is determined by n/v, where n is the total sample size and v is the number of villages you have decided to visit. For example, if desired sample size in each group is 250 and you decide that you have the resources to work in 20 villages in each group, then the number of households to be interviewed in each village is 250/20 = 12.5. You would interview 13 households per village and achieve a sample size of n = 260.

- Alternatively, you can first decide how many households to interview in each village. In this case, the number of villages is determined by n/h, where h is the number of households you wish to interview in each village. If your desired sample size is again 250 and you decide to interview 15 households per village, you will need to work in 250/15 = 16.67 villages. Rounding, you would work in 17 villages, achieving a sample size of n = 255.

A common approach would be to decide that you want to spend one day conducting interviews in each selected village. You would then estimate how many interviews you can conduct in one day: that number becomes h. You then calculate v (number of villages in each group) as n/h.

It should be clear from this discussion that the determination of v and h is based primarily on pragmatic considerations. However, a statistical principle to keep in mind is that, for a given n (total sample size), the efficiency of your estimates will generally be greater if you have

[10] As an example of the results you can expect from a sample of 200, if you are estimating maize yield with a *simple random sample* of 200, and your sample mean is 1,200 kg/ha, with a sample standard deviation of 500 kg/ha (variance of 250,000; these would not be atypical numbers), then a 95% confidence interval for that mean is 1,200 +/- 1.96 * sqrt(250,000/200) = 1,200 +/-35. In other words, you have 95% confidence that the true mean is between 1,165 kg/ha and 1,235 kg/ha. Note again that this calculation is based on a simple random sample. The approach suggested here (called *cluster sampling*) results in wider confidence intervals for a given sample size (its use is nevertheless often justified because it is a much more workable design than a simple random sample). The increase in the confidence interval with cluster sampling depends principally on the number of households interviewed per village (for a given total sample size n, fewer households per village - and more villages - gives a better estimate) and the degree of homogeneity within villages. It would not be unusual for the confidence interval in a cluster sample design to be 2-3 times larger than the interval from a simple random sample. This means that if the same data were obtained from the procedures recommended here (same sample size, mean, and standard deviation), the 95% confidence interval on maize yield could be as large as 1,200 +/- 105 kg. Note also that this example ignores issues of non-normal distribution of yield data, a treatment of which is beyond the scope of this paper.

more villages and fewer households per village.[11] Thus, subject to your constraints ot time, money, and personnel, you should spread your sample over as many villages as possible.

2.3. Selection of Villages and Households

Once you have determined n, v, and h, you need to choose the actual villages in which to work, and the households to interview.

Selection of Villages: The sampling method recommended in this case is the selection of villages with **Probability Proportional to Size (PPS)**. This means that the probability of a village being selected is proportional to the size of that village. Thus, for example, a village with 400 households would have twice the probability of being selected of a village with only 200 households. Why use PPS and not another sampling method? First, PPS eliminates the need for weighting the results in the analysis by ensuring that each household has the same probability of being selected. Second, PPS allows one to draw equally sized samples in each village, regardless of its size. Having the same number of households to be surveyed in each village makes it easier to program the fieldwork – assuming that interviews take approximately the same time in each village.

With n, v, and h defined, the next step consists of classifying and listing by target and control group, all villages which could potentially be included in the survey. You must then obtain data on the population (or number of households) of each village. The selection of villages is done separately in the target and the control group, using the same procedures. PPS sampling is straightforward and described in the hypothetical example below.

The first step in this method is to list the villages and their total population. If population numbers are not available, you can use the total number of households in each village. You must then construct the cumulative ranges (cr) and probabilities (p) for each group. The example here is for the target area group of villages and assumes that the number of villages to be selected is 4. For the control group of villages, the same method is followed.

[11] This statement assumes that households are more similar to their neighbors in the same village than they are to households in other villages. This assumption is generally appropriate in rural Africa.

Table 1: Organization of village data for PPS sampling

Villages	# of HHs (*)	Cumulative Range (cr)	Probability (p)
Josina Machel	100	1-100	100/1500
1 de Maio	120	101-220	120/1500
3 de Fevereiro	220	221-440	220/1500
Agostinho Neto	80	441-520	80/1500
Lipilichi	160	521-680	160/1500
Napipine	240	681-920	240/1500
25 de Junho	90	921-1010	90/1500
Spartan	100	1011-1110	100/1500
Ujamaa	80	1111-1190	80/1500
Buckeye	310	1191-1500	310/1500

(*) Can also be in terms of total population.

There are 1500 households in the population to be sampled. The cumulative range (cr) keeps track of the interval of numbered households in each village. The order in which the villages appear in the list is not important. In this list, Josina Machel Village has the first 100 households, 1 de Maio has households 101-220, and so on. The probability (p) for each village is simply the number of households in that village divided by the total number of households in the survey area. The villages with greater numbers of households have larger probabilities of selection.

You may choose a sample of 4 villages in two ways: using a random number table, or using *systematic sampling*. Using a random number table, you select 4 random numbers between 1 and 1500 from the table. This can also be done using a computer application – simple spreadsheets have a statistical function for these purposes. Suppose that the numbers selected in this random selection are 20, 530, 1099 and 1420. These numbers should be located in the cr column and the villages corresponding to those cumulative range intervals will constitute the sample: Josina Machel, Lipilichi, Spartan and Buckeye. These villages have been selected with probabilities proportional to their numbers of households.

An alternative approach is to use systematic sampling. This consists in dividing the total number of households (1500) by the number of villages to be sampled (4) to get the sampling interval (375). A random number between 0 and 375 is chosen randomly from the random number table to determine the first village selection. If the random number selected is 150, then 1 de Maio is the first village. Then 375 is added to the random number to give 525, making Lipilichi the second selection, adding 375 again gives 900, making Napipine the third selection. Finally, adding another 375 gives 1,275 and makes Buckeye the last village selected.

Selection of Households: Once villages have been selected, for each of them the entire list of households is necessary – no detailed data on the household are needed, except for the name of the household head that identifies it. The actual selection of households is done

using **Systematic Sampling (SS)**. First, number all households in the village from 1 to n. The total number of households in each village j is THHj. Then, the actual selection process is made using lists for each village with the following steps for each village:

Definition of Sampling Intervals (SI). SI for Village j (SIj) is given by SIj= THHj/h. Note that, while h is the same across all villages sampled, SIj between villages varies because of the differences in their sizes. If h is 10 in each village, and THH for a given village j is 120, then the SIj is 120/10 = 12.

For each Village, the first household to be selected in its list is obtained by choosing a random number between 0 and its SIj (a simple scientific calculator or spreadsheet can be used to select random numbers). The corresponding household in the list of numbered households is picked. For example, with a selection interval of 12, the first random number between 0 and 12 might be 4: the fourth household on your list is selected.

Then the process continues by systematically picking up every "+ SIj" household in the list until the desired number of households for the Village is reached. This process allows for a selection of households uniformly distributed along the Village list. In our example, you would select households 4, 16, 28, 40, 52, 64, 76, 88, 100, and 112, for a total of the desired 10 households.

2.4 Summary of Sampling Procedures

In summary, we are suggesting that you engage in the following steps to design and execute your sample:

1. Define target and control groups. You should probably do this at the village level, rather than the household level. There is no single correct way to define these groups, so think through the issues and present your reasoning in the results report.

2. Define the total sample size in each group. Try to do at least 200 in each group, more if your resources permit. Design the sample to deliver equal sample sizes in each group, recognizing that final numbers may differ slightly.

3. Determine the number of villages (v) and the number of households per village (h) that you will interview. The final decision is based on pragmatic considerations (time, personnel, money), but remember that, for any given n, your statistical estimates will be more accurate if you spread your sample over more villages, implying fewer household interviews in each village; 200 interviews spread over 10 villages (20/village) are better than 200 spread over 5 villages (40/village). ***Conduct the survey in as many villages as your resources of time, personnel, and money will permit.***

4. Select v villages with ***probability proportional to size*** (PPS). See the discussion above on how to do this.

5. Select h households in each village using systematic sampling. See above.

2.4. Reporting of Results

In reporting your results, follow these principles:

1. Present clearly your definition of target and control groups. Recognize the limitations of your definition (none is ever perfect), but highlight the strengths and explain why you made the decision you did.

2. Present a clear but concise description of your sampling strategy in each group.

3. Whenever relevant, present results broken down by control and target groups.

4. In your breakdowns, indicate the number of observations that contributed to any given mean. This will assist the reader in assessing the numbers you present. For example, if you have a sample size of 200 in your target group but have a table reporting results for target households in one specific area, the number of observations for that table will be less than 200. Include this number in each of the cells of your tables.

5. Remember that most statistical packages assume simple random sampling when conducting statistical tests (e.g., for a difference in means). We have seen that the cluster sampling approach advocated here results in wider confidence intervals than does simple random sampling. As a result, for a given n it will be more difficult to conclude that there are statistically significant differences in means or proportions. Put another way, if you present the results of unadjusted statistical tests, you will sometimes be concluding that there are statistically significant differences when, in fact, there are not.

   If you want to present statistical tests, you need to adjust them to take into account the *sample design effect*. Consult a sampling text such as Kalton for how to do this. (Graham Kalton, "Introduction to Survey Sampling", Quantitative Applications in the Social Sciences Paper No. 35, Sage Publications. 1985.)

Annex B. Proxy Questionnaire

Egerton University - Tegemeo Institute/MSU
KenyaNGO Proxy Survey
June 2003

Identifying Variables:

	NAME (Please write)		**CODE**
NGO		NGO	
Province (Write name, then enter code at far right)		PROV	
District (Write name, then enter code at far right)		DIST	
Division (Write name, then enter code at far right)		DIV	
Location Sublocation (Write name, then enter code at far right)		SUBLOC	
Village (Write name, then enter code at far right)		VILL	
Household Number		HHID	
HH Name			
Target Is this a target household or a control household? (**1=Target, 2=Control**)		TARGET	
Respondent Name			
Date			
Enumerator (Write name, then enter code at far right)		ENUM	

PROV		**DIST**			**NGO**	
1 Coast	5 Western	Homabay=46.	Marsabit=38.	Taita taveta=13.	1=Adra	5= HPI/TS
2 North Eastern	6 Central	Kilifi=11.	Mbeere=36.	Turkana=82.	2=Care	6=WVI
3 Eastern	7 Rift	Kitui=31.	Rachuonyo=45.		3=CRS	
4 Nyanza	8 North Rift	Kwale=12.	Suba=44.		4=FHI	
		Malindi=14.	Tharaka=37.			

NGO _____ PROV _____ DIST _____ VILL _____

1. Household Members

Person Name	Person Number	Sex 1=male 2=female	Age	Relation to Head 0=head 1=spouse, 2=father/mother, 3=son/daughter, 4=other relative, 5=other non-relative	What is the last year of schooling that the household head completed? 0-12, or 13=some Univ.	Did this person engage in any **business or informal labor** activities during the past 12 months? (incl jua kali, farm kibaruas, farm other districts) **0=no , 1=yes**	Did this person have any **salaried employment** during any of the past 12 months? 0 = no 1 = yes
NAME	**PERNO**	**SEX**	**AGE**	**RELHEAD**	**YRSCHL**	**INFORMAL**	**SALARY**
	1			0			
	2						
	3						
	4						
	5						
	6						
	7						
	8						
	9						
	10						
	11						
	12						
	13						
	14						
	15						
	16						
	17						
	18						
	19						
	20						
	21						
	22						

NGO _____ PROV _____ DIST _____ VILL _____

AGRICULTURAL ACTIVITIES

Q2. How many **TOTAL ACRES** are you currently cultivating? **TACRES** _____

Q3. OTHER CROPS

Crop		Did you *plant* this crop during either main or short harvest? 0=no 1=yes	Did you apply any *fertilizer* to this crop during either harvest? 0=no 1=yes	Did this crop sustain any *damage* from pests, or weather, or disease, or any other problem? 0=no 1=yes	Did you *completely lose* this crop from any field during either harvest? 0=no 1=yes	Did you *sell* any of this crop over the past 12 months? 0=no 1=yes	Crop		Did you *plant* this crop during either main or short harvest? 0=no 1=yes	Did you apply any *fertilizer* to this crop during either harvest? 0=no 1=yes	Did this crop sustain any *damage* from pests, or weather, or disease, or any other problem? 0=no 1=yes	Did you *completely lose* this crop from any field during either harvest? 0=no 1=yes	Did you *sell* any of this crop over the past 12 months? 0=no 1=yes
CROP		**PROD**	**FERT**	**DAMAGE**	**LOSE**	**SELL**	**CROP**		**PROD**	**FERT**	**DAMAGE**	**LOSE**	**SELL**
Dry Maize	1						Greengrams	34					
Green maize	2						Sweet potato	43					
Coffee Mbuni	6						Arrowroots	44					
dry Beans	7						Barley	60					
Sorghum	8						Simsim	78					
Millet	9						Yams	81					
CoffeeCherries	11						Pigeon peas	141					
Wheat	13						Njahi	147					
Cotton	14						Miraa	148					
Sugar cane	15						Soyabeans	160					
Sisal	16						Green peas	167					
Foxtail millet	18						Bulrush millet	169					
Flowers	20						Yellowgrams	201					
Cowpeas	21						Teff	202					
Fodder	22						chick peas	203					
Irish potatoes	27						Safflower	206					
Cassava	28						Bixa	207					
Tobacco	29						B o y o - b l a c k grams	210					

NGO _____ PROV _____ DIST _____ VILL _____

Crop		Did you *plant* this crop during either main or short harvest? 0=no 1=yes	Did you apply any *fertilizer* to this crop during either harvest? 0=no 1=yes	Did this crop sustain any *damage* from pests, or weather, or disease, or any other problem? 0=no 1=yes	Did you *completely lose* this crop from any field during either harvest? 0=no 1=yes	Did you *sell* any of this crop over the past 12 months? 0=no 1=yes	Crop		Did you *plant* this crop during either main or short harvest? 0=no 1=yes	Did you apply any *fertilizer* to this crop during either harvest? 0=no 1=yes	Did this crop sustain any *damage* from pests, or weather, or disease, or any other problem? 0=no 1=yes	Did you *completely lose* this crop from any field during either harvest? 0=no 1=yes	Did you *sell* any of this crop over the past 12 months? 0=no 1=yes
CROP		**PROD**	**FERT**	**DAMAGE**	**LOSE**	**SELL**	**CROP**		**PROD**	**FERT**	**DAMAGE**	**LOSE**	**SELL**
Sunflower	30						Bururi	211					
Rice	31						Beans, fresh	700					
Groundnuts	33												

Q4. Considering both the short and main harvests, which of these crops gave you the *greatest amount of food for home consumption**?***
(WRITE the crop _____) **FOODOTH** _____

Q5. Again considering both the short and main harvests, what ***quantity*** of this crop (the one listed in the previous question) did you produce over the past year?

Quantity **QNTOTHF** _____

1=90 kg bag 11=50 kg bag 2=kgs 3=litre 4=crates 5=numbers
6=bunches 9=gorogoro 10=tonnes 12=debe 13=grams **Unit** **UNITOTHF** _____

Q6. Considering both the short and main harvests, which of these crops gave you the ***greatest cash income*** (from sales)? (**WRITE** the crop _____) **CASHOTH** _____

Q7. Again considering both the short and main harvests, what ***quantity*** of this crop (the one listed in the previous question) did you produce over the past year?

Quantity **QNTOTHC** _____

1=90 kg bag 11=50 kg bag 2=kgs 3=litre 4=crates 5=numbers
6=bunches 9=gorogoro 10=tonnes 12=debe 13=grams **Unit** **UNITOTHC** _____

NGO _____ PROV _____ DIST _____ VILL _____

Q8. FRUITS AND VEGETABLES

Crop		Did you ***plant*** this crop during either main or short harvest? **0=no 1=yes**	Did you apply any ***fertilizer*** to this crop during either harvest? **0=no 1=yes**	Did this crop sustain any ***damage*** from pests, or weather, or disease, or any other problem? **0=no 1=yes**	Did you ***completely lose*** this crop from any field during either harvest? **0=no 1=yes**	Did you ***sell*** any of this crop over the past 12 months? **0=no 1=yes**	Crop		Did you ***produce*** this crop during either main or short harvest? **0=no 1=yes**	Did you apply any ***fertilizer*** to this crop during either harvest? **0=no 1=yes**	Did this crop sustain any ***damage*** from pests, or weather, or disease, or any other problem? **0=no 1=yes**	Did you ***completely lose*** this crop from any field during either harvest? **0=no 1=yes**	Did you ***sell*** any of this crop over the past 12 months? **0=no 1=yes**
CROP		**PROD**	**FERT**	**DAMAGE**	**LOSE**	**SELL**	**CROP**		**PROD**	**FERT**	**DAMAGE**	**LOSE**	**SELL**
Tamarind	3						squash	124					
Banana stems	4						cucumber	125					
Banana	10						Brinjals	129					
Cowpea leaves	19						chilies	131					
Coconuts	23						Pineapples	133					
Cashew nuts	24						Pears	134					
French beans	25						Macadamia	135					
gourds	62						Tangerine	136					
tomatoes	63						Passion fruit	137					
sukuma wiki	64						garlic onion	138					
pepper	65						Indig. grains	139					
spinach	66						Indig. vegetables	140					
capsicum	67						Castor oil	146					
Watermelon	69						tree tomato	162					
Pawpaw	70						White suppoise	163					
Guava	72						Chewing sugarcane	170					
Mango	73						Pumpkin leaves	172					
Lemons	74						Zambarao	174					
Orange	75						Grapes	179					
pumpkin	76						Dhania	183					
Okra	77						Bean leaves	184					
Cabbage	93						Stefali	190					
carrot	94						Green coconuts	193					
Passion fruit-yellow(mero)	95						coconut-copra	194					
onions	96						Shalgeda	204					
Avocado	97						Mulberry	208					
Lugard	118						Grape fruit	209					
Matomoko	120						Sandra	212					

NGO _____ PROV _____ DIST _____ VILL _____

Q9. Considering both the short and main harvests, which of these crops gave you the ***greatest amount of food for home consumption?*** **FOODFV** _____
(**WRITE** the crop _____)

Q10. Again considering both the short and main harvests, what ***quantity*** of this crop (the one listed in the previous question) did you produce over the past year?

Quantity **QNTFVF** _____

1=90 kg bag 11=50 kg bag 2=kgs 3=litre 4=crates 5=numbers
6=bunches 9=gorogoro 10=tonnes 12=debe 13=grams **Unit** **UNITFVF** _____

Q11. Considering both the short and main harvests, which of these crops gave you the ***greatest cash income*** (from sales)? (**WRITE** the crop _____) **CASHFV** _____

Q12. Again considering both the short and main harvests, what ***quantity*** of this crop (the one listed in the previous question) did you produce over the past year?

Quantity **QNTFVC** _____

1=90 kg bag 11=50 kg bag 2=kgs 3=litre 4=crates 5=numbers
6=bunches 9=gorogoro 10=tonnes 12=debe 13=grams **Unit** **UNITFVC** _____

Q13. **LIVESTOCK**

Animal		How many of this animal do you currently own?	Number sold of this type of animal over the past 12 months?	Animal		How many of this animal do you currently own?	Number sold of this type of animal over the past 12 months?
ANIMAL		**NANIM**	**SELLANIM**	**ANIMAL**		**NANIM**	**SELLANIM**
Grade bull	4			Cross calf	8		
Cross bull	5			Local calf	9		
Local bull	6			Goat	11		
Grade cow	1			Sheep	10		
Cross cow	2			Chicken	12		
Local cow	3			Duck	13		
Grade calf	7			Rabbit	16		

Q14. **LIVESTOCK PRODUCTS**

Livestock Product		Did you produce any of this product over the past 12 months? (0=no, 1=yes)	Did you sell any of this product over the past 12 months? (0=no, 1=yes)
ANIMPROD		**NPROD**	**SELLPROD**
Milk (fresh)	**1**		
Eggs	**2**		
Honey	**3**		
Hides & skin	**5**		
Other lvstk prod's	**6**		

Livestock Product		Production over past 12 months			
		Number of months of production	Quantity	Unit of Prod. 1 Kgs 2 litres	Frequency 1=daily 2=weekly 3=monthly 4=total for the period
liveprod			**Nmthpr**	**qprod**	**Unitpr**
Milk (fresh)	**1**				

Q15. In total, **over the past 12 months**, how much did you spend on **tick control?** (Ksh) **TICKCOST** _____

OF-FARM ACTIVITIES

Q.16. Participation in off-farm activities over the past 12 months

Month *Change starting and ending months as appropriate for timing of survey. Last month in list should be last month prior to survey.*		Did anyone in this household earn income from any kind of **business or informal labour** activities during the indicated months? (incl jua kali , farm kibaruas, farm other districts) (1=yes, 2=no)	Did anyone in this household earn income from any kind of **salaried employment** during any of the indicated months? (1=yes, 2=no)
MONTH		**INFMTH**	**SALMTH**
July 2002	207		
Aug	208		
Sep	209		
Oct	210		
Nov	211		
Dec	212		
Jan 2003	301		
Feb	302		
March	303		
April	304		
May	305		
June 2003	306		

Q17. Please indicate your net earnings from BUSINESS OR INFORMAL LABOUR during your **best SINGLE month** over the past 12 months (Ksh) **MAXEARN** _____

Q18. Did any of these BUSINESS OR INFORMAL LABOUR activities take place in an **urban area**? (1=yes, 2=no) **URBAN** _____

NGO _____ PROV _____ DIST _____ VILL _____

Q19. Business and informal off-farm activities, and salaried wage labour

Now please indicate if you or anyone in your family participated in any of the following off-farm activities over the past 12 months

Business and Informal Off-farm Activities			**Salaried Wage Labour**					
Activity		Over the past 12 months, did anyone in your household engage at any time in any of the following **business/informal** off-farm activities? (1=yes, 2=no)	Activity		Over the past 12 months, did anyone in your household engage at any time in any of the following **salaried wage labour** activities? (1=yes, 2=no)	Activity		Over the past 12 months, did anyone in your household engage at any time in any of the following **salaried wage labour** activities? (1=yes, 2=no)
ACTINF		**INFORMAL**	**ACTSAL**		**SALARIED**	**ACTSAL**		**SALARIED**
Informal/Business Activities			**Salaried Employment**			**Salaried Employment**		
Fish trading business	15		Surveyor	31		Civil servant	20	
Retail Shop /kiosk/ shopkeeping	30		Clerk	3		Manager	19	
Commercial tree selling	54		Lecturer	21		Social Worker	42	
Driver	12		Waiter	15		Watchman	17	
Wine tapper	65		Engineer	35		Accountant	37	
Vehicle Mechanic	51		Doctor	5		Driver	4	
Rental property	29		Veterinarian	25		Receives pension	10	
			Teacher	15		Industrial worker	8	

Q20. HOUSEHOLD ASSETS

AT PRESENT, how much/many of the following does this household own?

Agricultural asset	Quantity	Agricultural asset	Quantity	Agricultural asset	Quantity
ITEM	**QTY**	**ITEM**	**QTY**	**ITEM**	**QTY1**
14=animal traction plough		**19=truck**		**28=radio**	
15=cart		**22=water tank**		**29=zero-grazing units**	
18=car		**25=wheel barrow**		**46=telephone**	

Q21. IMPORTANCE OF INCOME SOURCES

Economic Activity		Please indicate the order of importance of each of these activities in the household's total income during the past 12 months -9=activity could not be ranked 0=did not give any income though produced 1=this activity gave the highest income of any activity, 2=this activity gave the second highest income ... ... -1=the household did not engage in this activity ***Enumerator***: **First place a -1 for all activities that the household did not engage in.** Then determine which of the remaining activities was the most important, second, etc.
ECONACT		**ORDER**
Crop production and sales (all crops)	1	
Livestock production and sales	2	
Farm kibarua	3	
Salaried labor	5	
Business activities	6	
Remittance	7	

Q22. Thinking about your **MAIN HARVEST**, would you consider your agricultural production to be reflective of a good production season, a normal production season or a poor production season? 1=Poor, 2=Normal, 3=Good **PRODYR** _____

Q23. **Enumerator:** Rate how you think the family's economic condition compares to most households in this area

1=better off 2= about the same 3= worse off **ECOMPARE** _____

Annex C

Sampling Procedures Used in NGO Full Income Surveys

This Annex includes a) Sample Design Sheets that each NGO was to use to organize their sampling information, b) the PPS Sample Definition Spreadsheet for ADRA target households (similar spreadsheets were developed for ADRA control households, and CARE control and target households), and c) the SPSS code used to generate final weights.

USAID/Cooperating Sponsor Income Proxy Survey

Sampling Design Sheet for NGOs with Targeted Interventions

This sheet should be used by NGOs implementing project interventions whose benefits acrue to specific participating households, with little or no spillover effect on non-participants. Based on previous discussions, we anticipate that **Technoserve/HPI, FHI, CRS, and WVI** will use this sheet.

Instructions:

1. Fill-out as many sheets as needed for each district that your USAID-funded project works in.
2. In the first column of each sheet, list each village that the project works in.
3. In the second column, indicate the number of beneficiary households in each village.
4. In the third column, indicate the number of non-beneficiary households in each village. Reasonable estimates are acceptable if you do not know the exact number.
5. In the fourth column, indicate the source for the number of non-beneficiary households.
6. Add any comments in the final column.

NGO: _____ **District:** _____

Village Name	**# of beneficiary households**	**# of non-beneficiary households**		**Comments**
		Number	**Source** (e.g., CBS data, NGO list, NGO estimate)	

USAID/Cooperating Sponsor Income Proxy Survey

Sampling Design Sheet for NGOs with Generalized Interventions

This sheet should be used by NGOs with broad interventions whose benefits to acrue to many or most households in a target village. Based on previous discussions, we anticipate that **CARE and ADRA** will use this sheet.

Instructions:

1. Fill-out as many sheets as needed to cover all *target villages*
2. For each target village, choose the nearest village where you have had *no interventions*. This will be the *control village.*
3. In the second column, indicate the total number of households in each village. Reasonable estimates are acceptable if you do not know the exact number.
4. In the third column, indicate the source for the number households.
5. Add any comments in the final column.

NGO: _____ **District:** _____

Village Name	**Total # of households in village**		**Comment**
	Number	**Source** (e.g., CBS data, NGO list, NGO estimate)	
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			
Target:			
Control:			

SAMPLE DEFINITION SPREADSHEET FOR TARGET HOUSEHOLDS, ADRA

	Size	Cum. Size
Target: Kwa Mbaki	73	73
Target: Kwa Ngindu	81	154
Target: Kisesini	104	258
Target: Kisou	99	357
Target: Kamuluni	43	400
Target: Kyambati	77	477
Target: Muambani East	63	540
Target: Muambani West	56	596
Target: Yaata	88	684
Target: Yamungu	150	834
Target: Nguumo	50	884
Target: Mukameni	56	940
Target: Tondomoni	56	996
Target: Ngaikini	52	1048
Target: Vendelani	85	1133
Target: Kanzokea	**62**	**1195**
Target: Nzunguni	**48**	**1243**
Target: Nguithyo	**38**	**1281**
Target: Mutuluni	**69**	**1350**
Target: Vutu	**46**	**1396**
Target: Ikisaya	**65**	**1461**
Target: Katumba	**32**	**1493**
Target: Lingithya	**12**	**1505**
Target: Tumbuni	**30**	**1535**
Target: Ngali	**27**	**1562**
Target: Kisuna	**48**	**1610**
Target: Kakindu	**69**	**1679**
Target: Methomaingi	**52**	**1731**
Target: Kimakimwe	**62**	**1793**
Target: Thunguta	**42**	**1835**
Target: Kakilai	**59**	**1894**
Target: Malumbani	**38**	**1932**
Target: Mwimbi	**137**	**2069**
Target: Nzeveni	**50**	**2119**
Target: Kwambuu	**152**	**2271**
Target: Yanzati	**124**	**2395**
Target: Kwekala	**103**	**2498**
Target: Matulu	**92**	**2590**
Target: Mutuni	**124**	**2714**
Target: Kyandula	**82**	**2796**
Target: Kisingo	**106**	**2902**
Target: Matinga	**206**	**3108**
Target: Ukamba	**241**	**3349**
Target: Mukundaa	**225**	**3574**
Target: Matemani	**155**	**3729**
Target: Timboni	**187**	**3916**
Target: Makueni	**88**	**4004**
Target: Mukameni	**59**	**4063**
Target: Ndiliu	**135**	**4198**
Target: Mulangoni	**247**	**4445**
Target: Kyangiu	**245**	**4690**

# of villages=	5
Selection interval=	2772
Starting point	1353

Note that the random starting point is generated anew each time the spreadsheet is recalculated. I took the first value of that random starting point and fixed it in cell H6 so that we could select villages.

	Selected Villages
366	Kamuluni
3138	Ukamba
5910	Tiva
8683	Syamatani
11455	Kilingoto

Replacement procedure

If a selected village is impossible to reach, then select the village immediately above it in the big list on the left. If that village is also impossible to reach, select the village immediately below the original village in the big list.
For example, if Kamuluni cannot be reached, then first select Kisou. If Kisou also cannot be reached, select Muambani East.

Target: Uthanga	**87**	**4777**
Target: Nthoiani	**47**	**4824**
Target: Mukonzo	**109**	**4933**
Target: Mwanyani	**142**	**5075**
Target: Kadengya	**148**	**5223**
Target: Ngai Ndethye	**82**	**5305**
Target: Kailembwa	**98**	**5403**
Target: Mwamba Isyuko	**45**	**5448**
Target: Kasoka	**142**	**5590**
Target: Memboo	**86**	**5676**
Target: Kyambusya	**80**	**5756**
Target: Tiva	**192**	**5948**
Target: Katokolo	**147**	**6095**
Target: Kivuti Center	**149**	**6244**
Target: Kisoji	**194**	**6438**
Target: Mutonya A	**150**	**6588**
Target: Mutonya B	**90**	**6678**
Target: Mbusyani	**85**	**6763**
Target: Mukuanima	**170**	**6933**
Target: Muathe	**60**	**6993**
Target: Kiisio	**82**	**7075**
Target: Imiwa	**50**	**7125**
Target: Makili	**30**	**7155**
Target: Muani	**30**	**7185**
Target: Ngwate	**40**	**7225**
Target: Kwa Musingi	**38**	**7263**
Target: Kamuvula	**30**	**7293**
Target: Kisou	**25**	**7318**
Target: Ndili	**41**	**7359**
Target: Ilaani	**48**	**7407**
Target: Matikoni	**22**	**7429**
Target: Imale	**19**	**7448**
Target: Kasula	**42**	**7490**
Target: Syaangwa	**61**	**7551**
Target: Kitambasye	**52**	**7603**
Target: Iiani	**96**	**7699**
Target: Ngawuni	**80**	**7779**
Target: Wiitu	**86**	**7865**
Target: Yumbu	**90**	**7955**
Target: Ngitini	**86**	**8041**
Target: Yauwa	**43**	**8084**
Target: Kyuasini	**38**	**8122**
Target: Wiitu	**38**	**8160**
Target: Kiangu	**30**	**8190**
Target: Kyakovi	**13**	**8203**
Target: Kivandeni	**137**	**8340**
Target: Muthue	**132**	**8472**
Target: Kyatulu	**126**	**8598**
Target: Syamatani	**122**	**8720**
Target: Mwangala	**104**	**8824**
Target: Itumo	**102**	**8926**
Target: Kikakaa	**91**	**9017**
Target: Mukuku	**84**	**9101**
Target: Maungu	**183**	**9284**
Target: Ngiluni	**165**	**9449**

Target: Kanyanzau	**157**	**9606**
Target: Kaluli	**135**	**9741**
Target: Mwanianga	**79**	**9820**
Target: Imale	**128**	**9948**
Target: Mang’’etheni	**115**	**10063**
Target: Imithumo	**90**	**10153**
Target: Kyanyaa	**87**	**10240**
Target: Kasikini	**157**	**10397**
Target: Mutalani	**118**	**10515**
Target: Kaaki original	**113**	**10628**
Target: Ingonzo	**110**	**10738**
Target: Kasivuni	**105**	**10843**
Target: Kinakoni	**71**	**10914**
Target: Ndulani	**68**	**10982**
Target: Kandongo	**66**	**11048**
Target: Kaaki	**40**	**11088**
Target: Masyondo	**183**	**11271**
Target: Kathima	**162**	**11433**
Target: Kilingoto	**154**	**11587**
Target: Kwaloti	**143**	**11730**
Target: Kituvwii	**121**	**11851**
Target: Kilingile	**105**	**11956**
Target: Kilungulu	**95**	**12051**
Target: Masoma	**54**	**12105**
Target: Kalivu	**121**	**12226**
Target: Minathini	**65**	**12291**
Target: Muteetu	**70**	**12361**
Target: Nzouni	**63**	**12424**
Target: Ndivuni	**93**	**12517**
Target: Mbuindune	**85**	**12602**
Target: Tangai	**66**	**12668**
Target: Mbukoni	**43**	**12711**
Target: Ivukuvuku	**47**	**12758**
Target: Maluma	**94**	**12852**
Target: Kakungula	**65**	**12917**
Target: Kaongoa	**85**	**13002**
Target: Ilusya	**70**	**13072**
Target: Masaini	**46**	**13118**
Target: Kangala	**64**	**13182**
Target: Imelu	**60**	**13242**
Target: Mbooni	**84**	**13326**
Target: Mitalani	**64**	**13390**
Target: Kamutei	**70**	**13460**
Target: Ndiithini	**54**	**13514**
Target: Mbakoni	**91**	**13605**
Target: Monguni	**74**	**13679**
Target: Yolomoni	**70**	**13749**
Target: Ndondoni	**62**	**13811**
Target: Ndithini	**50**	**13861**

SPSS SYNTAX AND NOTES TO GENERATE FINAL WEIGHTS

```
**  C:\KENYANGO\ARCH\LOOKUP\SYNTAX\W.SPS

** CREATES FILE C:\KENYANGO\ARCH\LOOKUP\W.SAV -- WEIGHTING FILE

*****
*****  NOTE THAT WE CAN NEVER MIX ZONES IN RESULTS, SINCE WEIGHTING IS NOT DESIGNED TO DO THIS.
*****  MUST GENERATE ALL RESULTS BY ZONE.
*****

get file='c:\kenyango\arch\dataround1\hhidngo1.sav'.
compute target=0.
if (hhtype=1) target=1.
agg out=*
 /break=ngo dist vill target
 /nhh=n(hhid).

**  ADRA SELECTED 5 TARGET VILLAGES PPS WITH SELECTION INTERVAL OF 2772, AND 5 CONTROL
**  VILLAGES PPS WITH SI OF 398.  SO WEIGHTS SHOULD BE SAME IN CONTROL VILLAGES, AND SAME IN
**  TARGET VILLAGES.  HOWEVER, THEY WERE  SUPPOSED TO SAMPLE 7 HHS IN EACH VILLAGE, BUT
**  ACTUAL NUMBERS DIFFERED SLIGHTLY.  SO WE SET WEIGHTS AS FOLLOWS:

**        FOR TARGET VILLAGES, W=2772/# OF HHS IN VILLAGE
**        FOR CONTROL VILLAGES, W=398/# OF HHS IN VILLAGE

if (ngo=1 and target=1) w=2772/nhh.
if (ngo=1 and target=0) w=398/nhh.

**  CARE SELECTED 25 CONTROL AND 25 TARGET VILLAGES PPS.  WERE TO SAMPLE 4 HHS IN EACH
**  VILLAGE, BUT ACTUAL NUMBERS DIFFERED SLIGHTLY IN SOME VILLAGES FROM THIS.  SO, AS FOR ADRA,
**  WEIGHTS WILL BE SI/NHH.

if (ngo=2 and target=1) w=204/nhh.
if (ngo=2 and target=0) w=305/nhh.

**  CRS DID not USE PPS.  ALSO HAD ONLY TARGET HHS.  THEY SAMPLED 10 VILLAGES TOTAL, 5 IN MWEA
**  DIVISION AND 5 IN GACHOKA DIVISION.  VILLAGES WERE SAMPLED WITH SRS AND EQUAL N IN EACH.
**  THUS IT NEEDS WEIGHTS AT VILLAGE LEVEL PLUS SECOND LEVEL WEIGHT  TO EXPAND THE GROUP OF
**  VILLAGES TO THE DIVISION

**        (From file c:\kenyango\sampleprep\CRS Sample.doc:
**        Total pop in all 3601 villages (Mwea) is 8602.  Number in sampled 3601 villages is 668.  ==> second  level
**        weight=12.9.
**        Total pop in all 3602 villates (Gachoka) is 12905.  Number in sampled 3602 villages is 577.  ==> second  level
**         weight=22.4.

if (ngo=3 and vill=360106) w=(154/nhh)*12.9.
if (ngo=3 and vill=360107) w=(118/nhh)*12.9.
if (ngo=3 and vill=360108) w=(79/nhh)*12.9.
if (ngo=3 and vill=360109) w=(123/nhh)*12.9.
```

```
if (ngo=3 and vill=360110) w=(194/nhh)*12.9.
if (ngo=3 and vill=360201) w=(82/nhh)*22.4.
if (ngo=3 and vill=360202) w=(165/nhh)*22.4.
if (ngo=3 and vill=360203) w=(147/nhh)*22.4.
if (ngo=3 and vill=360204) w=(92/nhh)*22.4.
if (ngo=3 and vill=360205) w=(91/nhh)*22.4.

** FHI AND WORLD VISION.  INSUFFICIENT INFORMATION. TREAT AS SRS WITH POP=6188 FOR FHI, 1528 FOR
** WV

if (ngo=4) w=6188/100.
if (ngo=6) w=1528/100.

** AS OF 26 NOVEMBER, DON'T HAVE SUFFICIENT INFORMATION FOR HPI.  SO:

** SET TAITA TAVETA OF HPI TO MEAN WEIGHT FOR ADRA AND CRS, SINCE TAITA TAVETA IS IN THEIR ZONE.
** SET REST OF HPI TO 1 (ALL IN SAME ZONE)

*temporary.
*select if (ngo=1 or ngo=3).
*des w.
if (dist=13) w=281.

if (sysmis(w) and ngo=5) w=1.

format target nhh w (f8.0).
execute.

save out='c:\kenyango\arch\lookup\w.sav'.
```

Annex D

Enumerator Manual for NGO Full Income Surveys

EGERTON UNIVERSITY

TEGEMEO INSTITUTE OF AGRICULTURAL POLICY AND DEVELOPMENT

DRAFT DATA COLLECTION MANUAL FOR

USAID-FUNDED NGOS

INCOME PROXY METHODOLOGY

June 2001

Introduction

This manual serves to guide field enumeration for the USAID-funded NGOs income proxy methodology development. However, it is a standard reference to all those involved in data collection and could be used as an enumerator-training handbook. The manual attempts to show how, specific questions are to be asked, how answers are to be filled and techniques of getting answers as desired by the instrument to meet income proxy methodology development data needs.

Things to do

- ☞ Introduce yourself on every visit and explain to the respondent the purpose of your visit to the household.
- ☞ Remember to request for consent and ensure confidentiality of the information in your introduction.
- ☞ Make sure you understand the questionnaire. Studying the questionnaire together with the manual will help you appreciate the use of this manual and give you a strong ground in later interviews; your objective should be to be able to ask the questions accurately and consistently without reading them.
- ☞ Internalize the questions to ensure good flow from question to question. This will allow the respondent to get involved.
- ☞ Ask questions in a similar way to each respondent. It is important that all your respondents get the same correct meaning from each question.
- ☞ Write legibly and make sure responses are entered to correct variables or cells. Keep the questionnaire neat and free from any damage.
- ☞ Remember to thank the respondent for the time and responses and allow him/her to ask any questions.
- ☞ Make sure you have completed filling the questionnaire before you conclude the interview.
- ☞ Submit the completed questionnaire to your supervisor promptly after checking for any pending calculations and coding.

Household definition

This survey examines the economic activities of a household, thus how to define a household is critical. For purposes of this survey, a household is considered to be composed of individuals living in the farm and sharing resources (e.g, labour contributed to farming activities, and food). This includes workers who live with a family. Utilizing this definition is relatively straightforward in the case of monogamous couples or female-headed households – the household is the husband (if relevant), the wife, all the wife's children still living at home and anyone else living in the household, including other family members and workers, contributing labour to the household, and eating with the household.

In the case of polygamous households, one should adhere to the following procedure:

- First, determine whether the various wives share resources – do they have common fields, and do they share the food from those fields?

    - If the answer is yes, then the household as enumerated in this survey must include the head, all wives, and all others living within the compound.
    - If the wives do not share resources, then the enumerator must randomly choose one of the wives, and define the household as that wife, her husband, and any children or others living with that wife and depending on or contributing to that wife's activities.

- In this latter case (the wives do not share resources), the husband will typically have resources that benefit all the households. Thus, it will not be correct to allocate all the husband's resources and income to the wife you have chosen to interview. In this case, you should have both the husband and the chosen wife present for the interview, and you should do your best to enumerate only the resources and income from the husband that accrue to that wife. You should, of course, enumerate all the chosen wife's resources and income, as well as those of the other members of her household.

The questions on the front page regarding polygamous households are designed to assist the enumerator in deciding how to define the household for those cases.

Contributions made by the unmarried sons/daughters living away from home are captured as remittance to the recipient, hence an income to the recipient. Married sons/daughters not sharing farm resources with the household are excluded: their contributions being remittance, hence income to the recipient household member. "Heads of household" are evaluated the same as any other individual, based on whether or not they lived with the household during the period of reference. Non-relatives e.g. house helps, a shamba boy eating and sleeping in that house qualify as household members. Salaries or other compensation paid by the family to resident workers are not included as household income, as this would constitute double-counting.

Introducing the Interview

An interview is best done when introduced well. A confident and comfortable relationship between the interviewer and the respondent is the foundation to a good interview. If you seem bored, uninterested, tired or doing an unimportant task, the respondent will probably act in a similar manner.

Whom to Interview

Enumerators should make all efforts to have the male head of household and his wife present for the interview. In the case of female-headed households, the female head and another person

knowledgeable about the household's activities should be present. If this is not possible, then any adult member of the household who is knowledgeable about household activities including crops, livestock, and off-farm, qualifies to be a respondent. A respondent as well as the enumerator may consult any other member(s) of the household on different items of the questionnaire.

Using a structured Survey Questionnaire.

Three things are important when using a structured questionnaire

- Understand the questionnaire
- Know how to ask questions
- Follow instructions in the questionnaire

When asking questions

- Remain neutral. Nothing in your words or manner should imply criticism, disdain or approval to either the questions asked or the respondents answer.
- Ask questions to give them their correct meanings. If the respondent does not understand the question, repeat it in a simpler way without changing the meaning.
- Make the interview what it should be; **an enjoyable guided conversation**.

Instructions on the Questionnaire

- 'Skip' or 'Go to' instructions should be followed carefully so as to avoid skipping valid questions.

Probing

- Probing motivates the respondent to expand upon or clarify answers. It is a technique that works to eliminate unnecessary or irrelevant information. You should use neutral questioning or comment such as,

Repeat the question
Anything else?
Any other reason?
What do you mean?
Why do you feel that way?
Are there any other persons living in this household?
Do you have any other source of income?

***Except in the crop inventory table**, be sure to avoid leading questions while probing. They provide respondents with answers that may not apply to their circumstances, especially respondents who would want to give answers implied to be valid. In the crop inventory table, you*

do *want to mention specific crops, because the objective is to be certain that no crops are missed.*

If a respondent gives '*don't know'* answer because he/she thinks the answer is too personal; e.g. what does your husband do for a living? The wife might say that she does not know. Don't take this for an answer unless you are sure she means it. Instead remind the respondent that the information will be handled with complete confidentiality.

When asking questions seeking answers that may be confidential to some respondents, apply a good measure of tact not to harm the mood of your respondent. Such questions when asked well build on the interview, they make the respondent even more relaxed in later sessions.

Controlling the Interview

You are the interviewer, the one guiding the conversation, therefore you must control the interview in a courteous manner so that it is completed within a frame of time that does not throw in diminishing returns and perhaps more importantly to avoid irrelevant discourse. You can only do this if you think ahead of the respondent. It is very important that you always do so. If the respondent gives you a lengthy explanation politely interject to keep him/her focussed. If the respondent while answering a question seems to provide answers to questions in other sections of the questionnaire note these down. However, be sure to ask those questions again, you could find that they were not answers.

During the interview a person from a different household may come in to interrupt or contribute by answering questions on behalf of the household. Politely ask for his/her exit unless the respondent is comfortable with this, because you are only interested in circumstances of the target household.

The questionnaire is not short, the respondent may need to cook for a child from school, visit the bathroom or tend to a young one. Be sure to release them and continue from the point where you paused.

Ending the Interview

Thank the respondent for the time and co-operation even when it was scant. The introductory paragraph provides you with remarks to conclude the interview. Let the respondent know that we will go back to the same household for the second round to complete data collection.

Scope and Coverage

The survey covers 11 districts in the country. A total of 800 households will be interviewed in NGO intervention areas. The districts are Kilifi, Kwale, Taita taveta, Malindi, Kitui, Mbeere, Marsabit, Suba, Rachuonyo, Homabay, and Turkana. A second round data collection will be

done on same households interviewed in the first round, to complete data needs for developing the methodology.

The survey instrument covers the following question areas

K. Identifying variables
L. Crop inventory
M. Cropping activities for last harvest
N. Inputs (seeds, fertilizer, chemicals, hired labor) for last harvest
O. Map of fields for current season
P. Livestock investments
Q. Demography
R. Business, informal labor activities and salaried/permanent wage employment
S. Purchases for home consumption
T. Agricultural assets
U. Perception questions on various economic indicators.

Identifying Variables

NGO, Province, and District codes are given at the bottom of the first page.

DIV should be calculated as (DIST*100)+1, 2, etc., being sure not to repeat for any District. For example, the first division that CARE works in in Homa Bay should be (46*100)+1 = 4601.

VILL should be calculated as (DIV*100)+1, 2, etc., being sure not to repeat for any Division. Fro example, the first village that CARE works in in the Division whose code we just calculated above, should be (4601*100)+1 = 460101.

Households will be numbered 1-n by each NGO. Thus, identifying variables will be NGO and HHID. The supervisor for each NGO will assign the household numbers.

Date refers to the date the interview is carried out. The format is ***ddmmyy***

Q1 Crop Inventory

Discuss with the respondent to see that we have a common understanding of variations in seasons, though this may however differ by one or two months. Let the respondent understand that questions will be focussed to specific seasons and occasionally remind him/her in the course of the interview.

The table is divided into field crops, vegetables, and tree crops/fruits. These are further divided to crops to be prompted and others to be probed. Based on discussions with NGO personnel, the list of crops in each category will vary by area, depending on what crops are most common. The approximately 10 most common crops should be prompted, while the others should be simply probed.

Wild vegetables and other crops gathered from the bush for household consumption should not be included in crop inventory (Q1) or cropping activities (Q3/Q4). However, if the crops are sold, incomes should be captured in Q33 (Business and informal labor activities) as Agricultural Trading.

*Note that for field crops and vegetables the answer required is a **Yes** or a **No**, while for tree crops the* **number** *of trees are required.*

'Productive trees' refers to trees regardless of age, and it captures tree/fruit crops and any other tree earning income to the household. Income from timber, charcoal, or any products sold from such trees should be captured in the informal income table (Q33).

'Planted trees' excludes fruits and tree crops mentioned earlier.

When recording Crops with 2 outputs (e.g. coffee (coffee cherries and coffee mbuni), cowpeas (bean and leaves), sugarcane (consumption/commercial)) in the crop table, get the specific code for each product from the crop code table.

Section I: Agricultural Activities for Last Harvest

Field/Crop Worksheet

After filling the crop inventory, you should use the field/crop worksheet to organize information on each of the fields that the household had last harvest, and each crop on each of those fields. You can do this as a list of fields and crops:

Field 1	Crop 1, crop 2, crop 3 ...
Field 2	Crop 1, crop 2, ...

OR you can do it as a drawing, where each field is given a number, and you write into each field the crops that appeared on that field. These data will not be entered into the computer, but ***it is important that you go through this process so that you can organize the needed information to accurately complete the crop table.***

Q3/Q4 Cropping/harvesting/sales activities during last harvest

Questions in this section are to collect information on land preparation, crops, harvests and sales. Harvest seasons differ across regions hence the need for the **season** variable. These variations are as below.

Harvest	**Eastern**	**Western**
Main	Jan-March (planted in Oct of previous year)	July-Aug (planted in Mar/Apr of same year)
Short	July-Sept (planted in Mar/Apr of same year)	Dec-Jan (Planted in Oct last year)

A **field** is a portion of land with a uniform crop portfolio. It could be pure stand or more than one crop. Crops planted on one field have same acreage, land preparation and fertilizer information, these variables will be recorded only for the first crop in the field.

- To get acreage of fields, pacing is to be done and measurements converted to acres **(X*Y)/4800.** Give the respondent a chance to give acreage of each field, however, the enumerator should pace the field to establish or confirm the acreage. This is possible for small fields that are within or near the compound. For big pieces of land within the compound which cannot be paced the enumerator is to compare the respondent's estimates with the knowledge imparted during training on what the size of an acre should be. The same should be done for irregular fields. For land far away from home, the enumerator will have to rely on the respondent's answer. Probe further using the example of a nearby portion to approximate the size of the field.
- If a farmer has another farm in another district details of such farms will not be required in Q3/Q4 but should be noted in the questionnaire. In this work, sale of produce from such farms is considered a business and should be captured as such in Q33.
- Land preparation costs are the **actual costs incurred** in activities prior to planting**.** It excludes family labor, clearing shrubs/forests in virgin land and the cost of owned equipment. For perennial crops that were planted at an earlier season this cost is not required.
- For tree crops the number of seedlings could be taken to be equal to the number of trees if the respondent does not remember the exact numbers planted.
- Purchased maize seed is not necessarily hybrid. Retained hybrids are grains from purchased hybrid used in successive seasons.
- For crops harvested piecemeal or over some period e.g. tomatoes, cassava and French beans, harvest realized within the target period should be recorded.
- If crop is harvested in a 90kg bag or 50 kg bag that is not full, the enumerator should be sure to get equivalent smaller units like debe or gorogoro and record harvest quantity in these.
- In maize harvests, establish whether the quantity given is for shelled or unshelled and convert unshelled quantities to shelled. **The conversion ratio of unshelled to shelled is 2: 1.**
- Maize harvested and sold while green has a different crop code from dry maize.
- Page 6 is a continuation of the table on page 5. It should be filled when data from Q3/Q4 is more than what the table on page 5 can take.

- The crop inventory table (Q1) should be crosschecked against Q3/Q4 to ensure that all crop enterprises are captured in Q3/Q4.
- Use of fertilizer and manure, harvest, sales, buyer type, and rating yields are crop level questions that should be asked for each crop.

Q3/Q4 table should be filled in a systematic way. Establish all the fields by asking crop combinations and field level questions (acres, land preparation type and cost) a field at a time until you are through with all fields. Following this procedure will lead you to fill the table from the first variable **season** to the variable for land preparation cost (**Lpcost)** for all fields first. Then pick on the first crop in the first field and ask crop level questions row wise from the variable for fertilizer use **(Fert)** to the one for yield rating **(Yield)**. Proceed to ask these questions for all the crops.

Q6, Q9, Q11 Inputs

Instructions on seed type, variety, source

If input (seed, fertilizer or chemicals) is obtained in exchange for labor or any other product in-kind, the exchange cost of this input is the equivalent wage paid for similar labor hours, or the equivalent value of the commodity which was exchanged for the input. For input obtained on credit, cost of the input is the total amount the farmer will actually be charged and not the cost he/she has already paid. **Note that both cash and credit cost are recorded in the same cell.**

Fertilizer of one type could be obtained through cash purchases, on credit, in exchange for labor or free. Enter separate rows for each of these modes of obtainment.

- For example, if 1 90kg bag of DAP fertilizer was obtained through a cash purchase and then an additional 2 gorogoros of the same type was acquired through credit, then you should enter two lines for that type:

- Similarly, use separate lines to enter seed for the same crop and type but which was obtained in more than one way. The same should be done to chemical of one type obtained in multiple ways.

Q10. This question doesn't ask about retained chemicals because it assumes that all chemicals must come from outside the house

Q13 Expenditures on hired labor for cropping activities

The respondent would have told you about crops produced by the household. Use crops information and listed activities to probe for hired labor.

Question Q13 aims at capturing expenditures on casual labour (not salaried/regular) hired for the cropping activities which were enumerated in Q3/Q4. Activities that hired labor could be used for include

- Planting which includes ridging and application of basal fertilizer
- Stooking in maize harvesting
- Weeding including topdressing
- Stalking, mulching and pruning for tomatoes

This question ***excludes:***

- Family labour
- Gang labour are excluded
- Salaried labor (this will be captured in Q25)
- Payments to hired vehicles and animal driven carts

Terrace maintenance, de-silting, and water harvesting can only be included in the table if they benefit only one season. If they benefit more than a season they should be viewed as capital investments: cost outlays that **cannot** be charged to a particular season.

Hired labor could be paid with cash or in-kind payments or both hence the provisions in the table. Sometimes the respondent may not recall the actual amount paid for a given activity. When this happens, ask for the number of people hired, numbers of days worked and wage rate to calculate the costs.

Note.

A crop in which hired labor was not used in any of the activities does not appear on this table.

Q14 Salaried labour for cropping activities

This question asks for a single, total figure for salaried labour used on crops during the last season, i.e., during the season that we have been talking about until this point. Salaried labour includes workers paid on a regular interval, not a casual basis; remember to consider only the proportion of time spent specifically on cropping activities.

Q15, Q16 Other farmland

These questions first establish whether the farmer has any farmland, perhaps distant from the farm we have been discussing, that the enumerator has not yet talked about. If there is such farmland, then in Q15 we ascertain the production from that land. This table Q15

is much less detailed than the crop table (Q3, Q4), as it does not ask any questions about fields.

Section II: Agricultural Activities for Crops Currently in the Field

This section consists of a single page that is to be used to draw the configuration of fields and crops for crops that are currently in the field. We will obtain detailed information about these crops in the second visit. The section targets the crop that will be current at the date of the first round data collection i.e. crops planted in March/April 2001*(main harvest for western Kenya and short harvest for eastern Kenya)*

Section III: Other Activities since January 1 of this Year

Q17 Livestock revenue

Even when the household does not have livestock currently, make sure you capture any purchases and sales in the course of the period (since Jan 1, 2001).

Q18 Livestock costs

Ask for expenditures on each of these cost items i.e. tick control, vaccines and drugs, artificial insemination services and purchased feed one by one. Purchased feed includes commercial feed, Nappier grass, and fodder e.t.c. *Shoats are Sheep and goats*

Q19 Salaried farm workers for cattle, shoats and poultry

If the salaried worker does other jobs apart from livestock then consider only the proportion of time spent specifically for cattle. Note that **shoats** are sheep and goats.

Q20 Livestock output

The table requires a lot of concentration to fill. You will be required to do some calculations especially when for instance quantity of milk production varies within the six-month period due to changing numbers of cows on milk. In this case, if the farmer can give you total figures for production and sales since January 1, this will be much easier. If he or she cannot do this, you must do some calculations and enter the appropriate information.

If the household consumed an animal then it is likely that they got hide or skin; be sure to ask whether it was sold.

Litre conversions for various bottles

Tree top bottle - 0.75 liters
Soda bottle - 0.3 or 0.5 liters
Beer bottle - 0.5 liters

Q21 Demography

The table allows data to be collected on household member's names, age, gender, education and income earning activities. Ask D8 and D9 only if D7>=2; in other words, ask D8 andD9 only for persons in the table who have lived with the family for at least two months since January 1. This should be done with a good measure of tact and caution to avoid harming the mood of the respondent as this could interfere with the quality of the interview.

- *If D8 =1 ask Q31, if D9 = 1 ask Q32.*

Adopt this systematic approach to ask Q21. Get the names of all household members as per the above working definition and allocate a person number in ascending order to each: person no 1 is reserved for the head of the household. Then ask questions row wise from D2 to D9 for each member beginning with the one listed first (the household head).

Salaried employment refers to employment where a person receives a regular salary whether in the formal or informal sector.

Informal labor/business refers to any business enterprise or labor where the income is not consistent over the year.

Q22 Business and informal labour activities

Person name and number are to be adapted from Q21. A person could be engaged in more than one activity. Ask for estimated **gross** income and **cost** for each activity in each of the past 6 months. This includes incomes from share dividends and farms in other districts.

- *Make effort to interview the person who does the specified activity if the respondent cannot give the data and the person is available)*

- *If the respondent can only give negative values, indicate this in the gross and zero for cost*

If the household had farms away from home or commercial trees ask whether any sales were made and if so capture income received.

Q23 Salaried wage employment

The person number and names should correspond with those in demography table (Q21). If ***samewage*** variable is yes then all the other month variables should be skipped. Probe about income sources like Pension, the respondent may not remember to give this on their own.

Q23a,b,c Remittances

In this section you are to find out about any other individuals ***not listed in the demography table*** (Q21) that sent remittances to the household, or that received remittances from the household.

- *If a person is listed in the demography table, they must not appear in the remittance table*

Examples of persons who could enter into the remittance table include a male head of household who lived away from the family (and therefore was not listed in the demography table) and sent cash and/or in-kind remittances back to his family. Or it could include children who are away from the family studying, and to whom the family sent cash, or food, or any other remittance.

Keep in mind:

- Be sure to follow the skip rule after Q23b.

- A single person could both send remittances to the family and receive remittances from them. This could happen if, for example, someone not listed in the demography table visited home for a short time, bringing cash or other remittances with him; and when that person left, the family gave him food or some other item (including cash) to take with him. If this is the case, then fill-out both the **SENT TO** and the **RECEIVED FROM** sections of the table.

- The table allows for both in-kind and cash remittances.

- If the person sent or received more than one type of in-kind remittance, use one row for each kind.

Below is an example of a person who sent 10,000 Ksh in cash to the household, and who also received a 90 kg bag of maize from the household:

Q23c. We would like to know the total amount of remittances that these people have **SENT** or **RECEIVED** since January 1

Person		This person **SENT TO THE HOUSEHOLD** (the household received)...				This person **RECEIVED FROM THE HOUSEHOLD** (the household sent) ...			
		In-kind remittances			Cash (Ksh) Quantity	In-kind remittances			Cash (Ksh)
		Prod	Quantity	Prod		Unit	Quantity	Unit	
Perno		**Prod1**	**Quant1**	**Unit1**	**Cash1**	**Prod2**	**Quant2**	**Unit2**	**Cash2**
101	Joseph				10,000	1	1	1	0
102									
103									
104									
105									
106									
107									

Q24 Ranking economic activities

By the time you reach this part you would have known about the households income earning activities. First place a -1 for all activities that the household did not engage in. Then get the order of importance of the activities the household engaged in.

Q24a Food Aid Receipts

This section allows you to indicate any food aid (relief/gift) that the household received since 1 January. This will be most common in the FHI and WV areas, where large food aid distribution programs have been going on for some time. Simply indicate the total amount of each type of commodity that the household has received since 1 January.

Q25 Purchases for home consumption

Ask about the quantities of various items for home consumption that the household purchased in the past six months, and the expenditure on each item in the same period. The table is designed to allow the respondent to reply to the question in the manner easiest for him or her – the respondent can indicate daily, weekly, or monthly frequencies of purchase, or the total amount purchased since 1 January.

Remember:

- If a respondent says that they have purchased something daily or weekly, probe to be sure that they really purchased every day or every week without exception. For example, if they say they purchased daily, but in fact did not purchase every single day since 1 January, then try to calculate a total figure.

- The final item in the list is for "group membership fees". This includes fees for women's groups, or for any other type of farmer group that charges a membership fee.

The enumerator should be cautious since these questions could be sensitive.

Q26 Agricultural assets as at present

Get the number of specified assets the household owns.

Q28 Requires the enumerator to rate the economic condition of the household compared to others in the areas. The enumerator must be cautious how he/she frames this question

Q29 Is the same question as Q28, but is to be responded by the respondent, not the enumerator.

Remember:

- *The enumerator should make their own, independent assessment in Q28, then should allow the respondent to make their own assessment in Q29 without the enumerator influencing the response in any way!*

Q30 Comparing last harvest to previous production years

Ask the respondent whether the last harvest is reflective of a poor, good or bad harvest, compared to the same season in earlier years. This is a comparison he/she is aware of.

Annex E:

Selected prices used in valuation of agricultural production and sales

Table C1. Prices of selected crops (Ksh/kg)

District	Maize		Cassava		Sukuma Wiki	
	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
Kilifi	11.11	8.88	3.03	5.05	10.00	10.00
Kwale	11.11	8.88	4.04	5.05	12.73	10.46
Taita Taveta	8.88	8.88	5.00	5.00	15.50	6.00
Malindi	11.11	8.88	4.77	5.00	12.73	10.46
Kitui	8.00	8.00	5.00	5.00	15.00	6.00
Mbeere	9.00	8.88	5.00	5.00	15.00	6.00
Marsabit	11.11	20.00	5.00	5.00	10.00	10.00
Suba	12.50	8.88	6.06	4.04	8.00	8.00
Rachuonyo	11.11	8.88	6.06	4.04	6.00	5.90
Homabay	11.11	8.88	6.06	4.04	10.00	12.00
Turkana	11.11	20.00	5.00	5.00	10.00	10.00

Table C2. Livestock Product Prices (Ksh/std unit)

Livestock Product	Price
milk (fresh)	20
eggs (for eating)	5
honey	60
hides/skins (cattle)	100
hides/skins (shoat)	30

Annex F: Model Results

CEREALS, TUBERS, PULSES, AND INDUSTRIAL CROPS

Zone 1, Retained

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.876	.768	.755	4441.82

a Predictors: (Constant), SORGRET, ZEROQSLD, IRPOTRET, QKEYRGNT qkeyret interacted with groundnuts, FERTAREA fertyes interacted w/ tacres, QKEYRDRM qkeyret interacted with dry maize, QKEYRET quant of prodn of key retained crop based on quantity retained approach, NOTHER, QKEYRSO qkeyret interacted with sorghum, FERTQRET fertyes interacted w/ qkeyret

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1460.856	913.148		1.600	.111
	QKEYRET quant of prodn of key retained crop based on quantity retained approach	2.421	.366	.701	6.611	.000
	QKEYRDRM qkeyret interacted with dry maize	4.359	.452	.425	9.652	.000
	NOTHER	467.798	97.616	.208	4.792	.000
	IRPOTRET	32125.680	4579.065	.259	7.016	.000
	QKEYRSO qkeyret interacted with sorghum	10.013	2.403	.230	4.168	.000
	QKEYRGNT qkeyret interacted with groundnuts	18.255	4.426	.149	4.125	.000
	FERTQRET fertyes interacted w/ qkeyret	-1.762	.381	-.492	-4.620	.000
	ZEROQSLD	2.826	.901	.129	3.137	.002
	FERTAREA fertyes interacted w/ tacres	460.102	147.426	.143	3.121	.002
	SORGRET	-2877.168	1392.180	-.115	-2.067	.040

a Dependent Variable: VRETNET

Zone 1, Sold

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.928(a)	.861	.853	4230.551	

a Predictors: (Constant), ORDCOMP order interacted with ecomapre, QKEYSCOW qkeysold interacted with cowpeas, QKEYSBE qkeysold interacted with beans, QKEYSCA qkeysold interacted with cassava, QKEYSGNT qkeysold interacted with groundnuts, TACRES Total acres cultivated - TACRES, NSLDQSLD nsoldoth interacted with qkeysold

b Dependent Variable: VSLDNET

Coefficients(a)

Model

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			1
(Constant)	-1288.521	928.783		-1.387	.168	
NSLDQSLD nsoldoth interacted with qkeysold	1.023	.059	.709	17.205	.000	
QKEYSCOW qkeysold interacted with cowpeas	12.712	1.259	.343	10.100	.000	
TACRES Total acres cultivated - TACRES	395.728	128.556	.125	3.078	.003	
QKEYSCA qkeysold interacted with cassava	-3.420	1.015	-.111	-3.368	.001	
QKEYSBE qkeysold interacted with beans	10.556	3.017	.119	3.498	.001	
QKEYSGNT qkeysold interacted with groundnuts	6.881	3.007	.075	2.288	.024	
ORDCOMP order interacted with ecomapre	229.386	102.436	.076	2.239	.027	

a Dependent Variable: VSLDNET

Zone 2, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.908(a)	.824	.812	3175.137	

a Predictors: (Constant), SORGRET sorghum was key retained crop, QKEYRET quant of prodn of key retained crop based on quantity retained approach, QKEYRMIR qkeyret interacted with miraa, QKEYRPIP qkeyret interacted with pigeon peas, NOTHER # of other crops, PRODYR, FERTQSLD fertyes interacted with qkeysold, TACRES Total acres cultivated - TACRES, QKEYRSO qkeyret interacted with sorghum, MIRAARET miraa was key retained crop, NZEROHRV # of crops with zero harvest, ZEROQRET nzerohrv interacted w/ qkeyret, QKEYRDRM qkeyret interacted with dry maize

b Dependent Variable: VRETNET net value of retained crops

Coefficients(a)

Model

	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			1
(Constant)	-1498.132	1014.895		-1.476	.142	
QKEYRET quant of prodn of key retained crop based on quantity retained approach	28.191	3.157	1.311	8.931	.000	
QKEYRDRM qkeyret interacted with dry maize	-21.097	2.855	-.933	-7.390	.000	
MIRAARET miraa was key retained crop	-19014.308	3636.471	-.261	-5.229	.000	
QKEYRSO qkeyret interacted with sorghum	-10.326	3.475	-.265	-2.971	.003	
PRODYR	1350.528	467.232	.131	2.890	.004	
ZEROQRET nzerohrv interacted w/ qkeyret	1.469	.294	.426	4.988	.000	
FERTQSLD fertyes interacted with qkeysold	-2.726	.844	-.134	-3.229	.001	
QKEYRMIR qkeyret interacted with miraa	63.670	21.260	.150	2.995	.003	
NOTHER # of other crops	371.826	65.041	.300	5.717	.000	
NZEROHRV # of crops with zero harvest	-443.104	104.744	-.300	-4.230	.000	
QKEYRPIP qkeyret interacted with pigeon peas	-68.026	34.617	-.066	-1.965	.051	
TACRES Total acres cultivated - TACRES	-285.369	109.686	-.143	-2.602	.010	
SORGRET sorghum was key retained crop	-1496.759	762.263	-.099	-1.964	.051	

a Dependent Variable: VRETNET net value of retained crops

Zone 2, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.998(a)	.996	.996	2698.393	

a Predictors: (Constant), ORDASSET order interacted with pvasset, TACRES Total acres cultivated - TACRES, QKEYSDRM qkeysold interacted with dry maize, QKEYSMIR qkeysold interacted with miraa, PVASSETS, WHTSLD wheat was key sold crop, BNSSLD bean was key sold crop, NSLDQSLD nsoldoth interacted with qkeysold, QKEYSBE qkeysold interacted with beans

b Dependent Variable: VSLDNET

Coefficients(a)

Model

	B	Std. Error	Beta	t	Sig.	1
(Constant)	1726.066	534.682		3.228	.002	
QKEYSMIR qkeysold interacted with miraa	96.545	1.705	.974	56.625	.000	
NSLDQSLD nsoldoth interacted with qkeysold	.631	.344	.033	1.835	.071	
PVASSETS	-.039	.007	-.044	-5.252	.000	
QKEYSBE qkeysold interacted with beans	120.207	20.970	.386	5.732	.000	
tTACRES Total acres cultivated - TACRES	-772.711	113.069	-.083	-6.834	.000	
BNSSLD bean was key sold crop	-74650.393	15801.002	-.301	-4.724	.000	
QKEYSDRM qkeysold interacted with dry maize	4.030	1.089	.038	3.700	.000	
WHTSLD wheat was key sold crop	7353.069	3095.120	.021	2.376	.021	
ORDASSET order interacted with pvasset	.007	.003	.019	2.204	.031	

a Dependent Variable: VSLDNET

Zone 3, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.841(a)	.707	.700	7459.143	

a Predictors: (Constant), NSLDQRET nsold interacted w/ qkeyret, ECOMPARE Enum - rate economic conditions of family compared to other households - ECOMPARE, NOTHER # of other crops, QKEYRET quant of prodn of key retained crop based on quantity retained approach

b Dependent Variable: VRETNET net value of retained crops

Coefficients(a)

Model

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			1
(Constant)	-4229.364	2492.792		-1.697	.091	
QKEYRET quant of prodn of key retained crop based on quantity retained approach	7.854	.687	.969	11.427	.000	
NOTHER # of other crops	433.100	119.262	.158	3.631	.000	
ECOMPARE Enum - rate economic conditions of family compared to other households - ECOMPARE	3053.164	1015.397	.127	3.007	.003	
NSLDQRET nsold interacted w/ qkeyret	-.280	.096	-.253	-2.916	.004	

a Dependent Variable: VRETNET net value of retained crops

Zone 3, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.909(a)	.826	.815	4408.729	

a Predictors: (Constant), NOTHER # of other crops, QKEYSGRG qkeysold interacted with greengrams, QKEYSCA qkeysold interacted with cassava, QKEYSMIR qkeysold interacted with miraa, QKEYSOLD quant of prodn of key sales crop based on value sold approach, NSOLDOTH # of other crops sold, NSLDQSLD nsoldoth interacted with qkeysold, QKEYSDRM qkeysold interacted with dry maize

b Dependent Variable: VSLDNET

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-1135.913	718.588		-1.581	.116	
NSLDQSLD nsoldoth interacted with qkeysold	.351	.099	.462	3.546	.001	
QKEYSGRG qkeysold interacted with greengrams	15.653	3.256	.189	4.808	.000	
QKEYSMIR qkeysold interacted with miraa	86.771	34.050	.095	2.548	.012	
QKEYSOLD quant of prodn of key sales crop based on value sold approach	6.875	1.163	1.204	5.910	.000	
QKEYSDRM qkeysold interacted with dry maize	-5.384	1.071	-.964	-5.027	.000	
QKEYSCA qkeysold interacted with cassava	-6.497	2.718	-.094	-2.391	.018	
NSOLDOTH # of other crops sold	835.084	201.909	.229	4.136	.000	

a Dependent Variable: VSLDNET

Zone 4, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.773(a)	.598	.589	12587.552	

a Predictors: (Constant), GGRAMRET green gram was key retained crop, QKEYRET quant of prodn of key retained crop based on quantity retained approach, QKEYRRI qkeyret interacted with rice, NOTHER # of other crops

b Dependent Variable: VRETNET net value of retained crops

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-2133.182	2600.445		-.820	.413	
QKEYRET quant of prodn of key retained crop based on quantity retained approach	7.067	.491	.689	14.386	.000	
NOTHER # of other crops	1433.963	295.677	.233	4.850	.000	
QKEYRRI qkeyret interacted with rice	16.276	8.554	.090	1.903	.059	
GGRAMRET green gram was key retained crop	21588.753	12670.284	.081	1.704	.090	

a Dependent Variable: VRETNET net value of retained crops

Zone 4, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.786(a)	.617	.602	2592.235	

a Predictors: (Constant), NSOLDOTH # of other crops sold, QKEYSCOW qkeysold interacted with cowpeas, QKEYSOLD quant of prodn of key sales crop based on value sold approach

b Dependent Variable: VSLDNET

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-1493.986	585.975		-2.550	.013	
QKEYSOLD quant of prodn of key sales crop based on value sold approach	1.567	.210	.535	7.453	.000	
QKEYSCOW qkeysold interacted with cowpeas	11.454	3.124	.263	3.667	.000	
NSOLDOTH # of other crops sold	1465.213	200.331	.520	7.314	.000	

a Dependent Variable: VSLDNET

FRUITS & VEGETABLES

Zone 1, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.897(a)	.804	.795	1618.845	

a Predictors: (Constant), ZEROQRET nzerohrv interacted w/ qkeyret, CBBGRET cabbage are key retained crop, QKEYRMG qkeyret interacted w/ mango, SUKRET sukuma are key retained crop, NFV # of f&v produced, NZEROHRV # of f&v crops with total prodn loss, NSLDQRET nsold interacted w/ qkeyret, QKEYRSU qkeyret interacted w/ sukuma, NFVQRET nfv interacted w/ qkeyret

b Dependent Variable: VRETNET net value retained over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	172.140	234.125		.735	.463	
NFVQRET nfv interacted w/ qkeyret	.426	.057	2.210	7.457	.000	
NFV # of f&v produced	196.349	33.443	.245	5.871	.000	
QKEYRSU qkeyret interacted w/ sukuma	1.579	.388	.406	4.065	.000	
NSLDQRET nsold interacted w/ qkeyret	-.408	.059	-1.964	-6.928	.000	
SUKRET sukuma are key retained crop	-851.616	331.500	-.102	-2.569	.011	
CBBGRET cabbage are key retained crop	3582.710	952.640	.126	3.761	.000	
QKEYRMG qkeyret interacted w/ mango	-1.425	.504	-.106	-2.829	.005	
NZEROHRV # of f&v crops with total prodn loss	-475.302	172.958	-.136	-2.748	.007	
ZEROQRET nzerohrv interacted w/ qkeyret	.524	.364	.141	1.439	.152	

a Dependent Variable: VRETNET net value retained over both rounds

Zone 1, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.987(a)	.975	.974	3543.171	

a Predictors: (Constant), SUKSLD sukuma wiki are key sold crop, NFVQSLD nfv interacted w/ qkeysold, NZEROHRV # of f&v crops with total prodn loss, WTRMSLD watermelon are key sold crop, BANSLD are key sold crop, QKEYSPP qkeysold interacted w/ pawpaw, QKEYSSU qkeysold interacted w/ sukuma, ZEROQSLD nzerohrv interacted w/ qkeysold, NSLDQSLD nsold interacted w/ qkeysold

b Dependent Variable: VSLDNET net value sold over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	1400.098	426.111		3.286	.001	
NSLDQSLD nsold interacted w/ qkeysold	1.181	.103	.821	11.473	.000	
QKEYSPP qkeysold interacted w/ pawpaw	7.009	.490	.224	14.297	.000	
QKEYSSU qkeysold interacted w/ sukuma	4.808	.732	.180	6.571	.000	
NFVQSLD nfv interacted w/ qkeysold	-.196	.078	-.157	-2.499	.014	
WTRMSLD watermelon are key sold crop	-10458.765	2587.859	-.055	-4.041	.000	
ZEROQSLD nzerohrv interacted w/ qkeysold	3.678	.800	.151	4.596	.000	
BANSLD are key sold crop	3294.046	1099.370	.041	2.996	.003	
NZEROHRV # of f&v crops with total prodn loss	-1282.732	409.967	-.065	-3.129	.002	
SUKSLD sukuma wiki are key sold crop	-1323.196	664.735	-.028	-1.991	.048	

a Dependent Variable: VSLDNET net value sold over both rounds

Zone 2, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.909(a)	.827	.809	1286.723	

a Predictors: (Constant), ORNGRET oranges are key retained crop, QKEYRPU qkeyret interacted w/ pumpkin, QKEYRLM qkeyret interacted w/ lemon, QKEYRTO qkeyret interacted w/ tomato, QKEYRPP qkeyret interacted w/ pawpaw, QKEYRWT qkeyret interacted w/ watermelon, ZEROQRET nzerohrv interacted w/ qkeyret, PAPARET pawpaw are key retained crop, TOMRET tomato are key retained crop, NSOLDFV # of f&v sold, NSLDQRET nsold interacted w/ qkeyret, QKEYRET quant of prodn of key retained crop based on quantity retained approach, NFVQRET nfv interacted w/ qkeyret

b Dependent Variable: VRETNET net value retained over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	12.703	164.988		.077	.939	
QKEYRPP qkeyret interacted w/ pawpaw	17.489	1.730	.580	10.112	.000	
QKEYRET quant of prodn of key retained crop based on quantity retained approach	7.337	1.376	.638	5.333	.000	
NSLDQRET nsold interacted w/ qkeyret	-2.876	.361	-2.702	-7.970	.000	
NSOLDFV # of f&v sold	233.465	54.799	.254	4.260	.000	
QKEYRPU qkeyret interacted w/ pumpkin	7.284	216.151	.001	.034	.973	
QKEYRWT qkeyret interacted w/ watermelon	-.031	1.953	-.001	-.016	.987	
NFVQRET nfv interacted w/ qkeyret	2.213	.334	2.813	6.623	.000	
ZEROQRET nzerohrv interacted w/ qkeyret	-2.504	.397	-.791	-6.314	.000	
QKEYRTO qkeyret interacted w/ tomato	-19.019	4.264	-.241	-4.460	.000	
TOMRET tomato are key retained crop	1436.525	712.059	.101	2.017	.046	
QKEYRLM qkeyret interacted w/ lemon	-4.078	1.591	-.129	-2.563	.012	
PAPARET pawpaw are key retained crop	-618.073	337.325	-.082	-1.832	.069	
ORNGRET oranges are key retained crop	-2285.028	1108.608	-.094	-2.061	.041	

a Dependent Variable: VRETNET net value retained over both rounds

Zone 2, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.838(a)	.703	.663	8797.162	

a Predictors: (Constant), NZEROHRV # of f&v crops with total prodn loss, QKEYSOR qkeysold interacted w/ orange, QKEYSLM qkeysold interacted w/ lemon, QKEYSMG qkeysold interacted w/ mango, QKEYSAV qkeysold interacted w/ avocado, BANSLD are key sold crop, NADULT number of adults in hh, NSOLDFV # of f&v sold

b Dependent Variable: VSLDNET net value sold over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-4987.376	2946.961		-1.692	.096	
NSOLDFV # of f&v sold	2314.738	448.411	.567	5.162	.000	
QKEYSMG qkeysold interacted w/ mango	12.990	3.358	.311	3.868	.000	
QKEYSOR qkeysold interacted w/ orange	172.846	125.210	.127	1.380	.173	
QKEYSLM qkeysold interacted w/ lemon	7.465	3.963	.138	1.884	.065	
QKEYSAV qkeysold interacted w/ avocado	17.670	22.051	.071	.801	.426	
BANSLD are key sold crop	2433.672	3359.532	.057	.724	.472	
NADULT number of adults in hh	491.439	692.279	.057	.710	.481	
NZEROHRV # of f&v crops with total prodn loss	-160.402	306.633	-.041	-.523	.603	

a Dependent Variable: VSLDNET net value sold over both rounds

Zone 3, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.948(a)	.899	.889	2916.066	

a Predictors: (Constant), QKEYRSU qkeyret interacted w/ sukuma, QKEYRCB qkeyret interacted w/ cabbage, QKEYRSC qkeyret interacted w/ chewcane, QKEYRLM qkeyret interacted w/ lemon, QKEYRMG qkeyret interacted w/ mango, QKEYRGD qkeyret interacted w/ gourds, QKEYRPU qkeyret interacted w/ pumpkin, QKEYRAV qkeyret interacted w/ avocado, QKEYRTO qkeyret interacted w/ tomato, NZEROHRV # of f&v crops with total prodn loss, NFV # of f&v produced, GRDRET gourds are key retained crop, CHEWRET chewing sugar cane is key retained crop, AVOCRET avocado are key retained crop, QKEYRET quant of prodn of key retained crop based on quantity retained approach

b Dependent Variable: VRETNET net value retained over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-411.338	407.634		-1.009	.314	
QKEYRET quant of prodn of key retained crop based on quantity retained approach	1.577	.480	.328	3.288	.001	
QKEYRPU qkeyret interacted w/ pumpkin	17.817	2.134	.215	8.348	.000	
NFV # of f&v produced	462.552	42.630	.321	10.850	.000	
QKEYRAV qkeyret interacted w/ avocado	25.822	3.773	.313	6.845	.000	
QKEYRGD qkeyret interacted w/ gourds	1.333	.557	.133	2.393	.018	
AVOCRET avocado are key retained crop	-5709.735	1846.592	-.143	-3.092	.002	
NZEROHRV # of f&v crops with total prodn loss	-784.745	199.896	-.108	-3.926	.000	
QKEYRSC qkeyret interacted w/ chewcane	-.749	1.229	-.021	-.610	.543	
QKEYRMG qkeyret interacted w/ mango	2.847	.492	.504	5.788	.000	
QKEYRLM qkeyret interacted w/ lemon	2.731	1.417	.051	1.927	.056	
CHEWRET chewing sugar cane is key retained crop	2541.855	1742.729	.048	1.459	.147	
GRDRET gourds are key retained crop	-118.268	890.281	-.004	-.133	.894	
QKEYRCB qkeyret interacted w/ cabbage	7.040	1.619	.114	4.348	.000	
QKEYRTO qkeyret interacted w/ tomato	3.259	2.672	.032	1.220	.224	
QKEYRSU qkeyret interacted w/ sukuma	4.659	1.459	.085	3.193	.002	

a Dependent Variable: VRETNET net value retained over both rounds

Zone 3, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.974(a)	.948	.944	1866.477	

a Predictors: (Constant), QKEYSAV qkeysold interacted w/ avocado, TOMSLD tomato are key sold crop, QKEYSCB qkeysold interacted w/ cabbage, QKEYSOLD quant of prodn of key sales crop based on value sold approach, BANSLD are key sold crop, FERTQSLD fertyes interacted w/ qkeysold, NSLDQSLD nsold interacted w/ qkeysold

b Dependent Variable: VSLDNET net value sold over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-102.561	279.778		-.367	.715	
FERTQSLD fertyes interacted w/ qkeysold	3.101	.571	.234	5.432	.000	
QKEYSCB qkeysold interacted w/ cabbage	12.447	.801	.434	15.540	.000	
NSLDQSLD nsold interacted w/ qkeysold	1.330	.112	.694	11.869	.000	
QKEYSOLD quant of prodn of key sales crop based on value sold approach	-.692	.160	-.195	-4.322	.000	
BANSLD are key sold crop	2964.181	602.788	.129	4.917	.000	
TOMSLD tomato are key sold crop	1704.840	554.293	.079	3.076	.003	
QKEYSAV qkeysold interacted w/ avocado	5.012	1.767	.073	2.836	.006	

a Dependent Variable: VSLDNET net value sold over both rounds

Zone 4, Retained

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.823(a)	.677	.654	8484.915	

a Predictors: (Constant), MANGORET mango are key retained crop, QKEYRCP qkeyret interacted w/ cowplvs, QKEYRPU qkeyret interacted w/ pumpkin, QKEYRCS qkeyret interacted w/ cashew, QKEYRSC qkeyret interacted w/ chewcane, QKEYRPP qkeyret interacted w/ pawpaw, NFV # of f&v produced, PRODYR farmer assessment of prodn year, NSLDQRET nsold interacted w/ qkeyret, FERTQRET fertyes interacted w/ qkeyret, COCORET coconut are key retained crop, QKEYRET quant of prodn of key retained crop based on quantity retained approach

b Dependent Variable: VRETNET net value retained over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-2291.613	2300.523		-.996	.321	
FERTQRET fertyes interacted w/ qkeyret	1.478	.237	.352	6.243	.000	
QKEYRPU qkeyret interacted w/ pumpkin	17.026	4.078	.186	4.175	.000	
QKEYRPP qkeyret interacted w/ pawpaw	6.872	1.331	.238	5.163	.000	
NFV # of f&v produced	539.323	112.186	.239	4.807	.000	
COCORET coconut are key retained crop	-4612.685	1644.502	-.160	-2.805	.006	
QKEYRCP qkeyret interacted w/ cowplvs	2.568	.758	.158	3.387	.001	
QKEYRSC qkeyret interacted w/ chewcane	11.512	3.579	.144	3.217	.002	
QKEYRET quant of prodn of key retained crop based on quantity retained approach	1.325	.304	.527	4.352	.000	
NSLDQRET nsold interacted w/ qkeyret	-.068	.019	-.392	-3.535	.001	
PRODYR farmer assessment of prodn year	1691.627	887.561	.087	1.906	.058	
QKEYRCS qkeyret interacted w/ cashew	18.578	5.964	.139	3.115	.002	
MANGORET mango are key retained crop	5444.015	2293.838	.130	2.373	.019	

a Dependent Variable: VRETNET net value retained over both rounds

Zone 4, Sold

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.973(a)	.947	.943	11396.120	

a Predictors: (Constant), ZEROQSLD nzerohrv interacted w/ qkeysold, FERTYES hh used fertilizer, BANSLD are key sold crop, QKEYSMG qkeysold interacted w/ mango, QKEYSSC qkeysold interacted w/ chewcane, COCOSLD coconut are key sold crop, SCHHEAD years of schooling for hh head, TARGET hh is target hh, NSOLDFV # of f&v sold, QKEYSTO qkeysold interacted w/ tomato, FERTQSLD fertyes interacted w/ qkeysold, QKEYSOLD quant of prodn of key sales crop based on value sold approach

b Dependent Variable: VSLDNET net value sold over both rounds

Coefficients(a)

Model

	B	Std. Error	Beta			1
(Constant)	-649.659	2252.843		-.288	.773	
QKEYSMG qkeysold interacted w/ mango	5.597	.459	.630	12.183	.000	
NSOLDFV # of f&v sold	1983.959	262.770	.161	7.550	.000	
FERTQSLD fertyes interacted w/ qkeysold	1.911	.376	.138	5.086	.000	
QKEYSTO qkeysold interacted w/ tomato	4.254	1.692	.054	2.514	.013	
QKEYSSC qkeysold interacted w/ chewcane	6.496	3.344	.037	1.943	.054	
BANSLD are key sold crop	19166.550	5329.922	.069	3.596	.000	
FERTYES hh used fertilizer	-7303.218	2474.127	-.073	-2.952	.004	
SCHHEAD years of schooling for hh head	428.101	153.833	.057	2.783	.006	
TARGET hh is target hh	4108.087	1927.818	.043	2.131	.035	
COCOSLD coconut are key sold crop	-14248.271	3075.559	-.129	-4.633	.000	
QKEYSOLD quant of prodn of key sales crop based on value sold approach	1.960	.472	.245	4.155	.000	
ZEROQSLD nzerohrv interacted w/ qkeysold	.862	.422	.046	2.044	.043	

a Dependent Variable: VSLDNET net value sold over both rounds

LIVESTOCK

Zone 1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.795	.632	.624	20970.07969

a Predictors: (Constant), MTHGOAT, NSLDLCOW, NSOLDLV, NSLDLBUL

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-3173.274	2289.923		-1.386	.167
	NSLDLBUL	10601.657	2670.178	.362	3.970	.000
	NSOLDLV	6460.564	1506.961	.216	4.287	.000
	NSLDLCOW	6770.424	1960.875	.311	3.453	.001
	MTHGOAT	73.431	26.492	.129	2.772	.006

a Dependent Variable: VNET_LV net livestock income, both rounds

Zone 2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	.861	.742	.725	7250.65418

a Predictors: (Constant), TARGET, NADULT number of adults in hh, MTHLCOW, SOLDGOAT, ECOMPARE Enum - rate economic conditions of family compared to other households - ECOMPARE, MTHGOAT, NLBULL, INT_LCOW, NSOLDLV, MKPROD total milk produced, NSLDGOAT, NGOAT

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1283.308	2173.067		.591	.556
	MTHGOAT	21.044	6.207	.454	3.390	.001
	NSOLDLV	7993.374	726.896	.542	10.997	.000
	INT_LCOW	1302.820	185.012	.337	7.042	.000
	NLBULL	-2126.572	418.149	-.229	-5.086	.000
	NGOAT	-167.918	50.993	-.342	-3.293	.001
	MKPROD total milk produced	17.041	2.539	.370	6.712	.000
	SOLDGOAT	-9393.479	1623.461	-.294	-5.786	.000
	NSLDGOAT	973.051	266.869	.270	3.646	.000
	ECOMPARE Enum - rate economic conditions of family compared to other households	-3121.406	950.306	-.132	-3.285	.001
	MTHLCOW	-96.487	31.566	-.160	-3.057	.003
	NADULT number of adults in hh	708.297	328.943	.084	2.153	.033
	TARGET	1306.518	1071.154	.047	1.220	.224

a Dependent Variable: VNET_LV net livestock income, both rounds

Zone 3

Model Summary

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.925(a)	.856	.843	7097.81673	

a Predictors: (Constant), SOLDMK sold milk, NSOLDLV, TARGET target hh?, NGCOW # of grade cows, INT_HNY soldhny interacted with nbees, SOLDCCOW sold cross cow, NSLDLBUL # of local bulls sold, NSLDLCOW # of local cows sold, NCCALF # of cross calves, NSOLDLP # of livestock products sold, INT_LCOW soldmk interacted with nlcow, NLBULL # of local bulls, SOLDHNY sold honey, MKPROD total milk produced

Coefficients

Model

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			1
(Constant)	836.607	1208.455		.692	.490	
MKPROD total milk produced	16.705	1.168	.599	14.299	.000	
INT_HNY soldhny interacted with nbees	677.165	47.856	.535	14.150	.000	
NSOLDLV	3171.811	680.210	.178	4.663	.000	
NSLDLBUL # of local bulls sold	10239.875	1826.465	.177	5.606	.000	
NSLDLCOW # of local cows sold	10182.196	2214.291	.154	4.598	.000	
SOLDCCOW sold cross cow	38961.187	7664.850	.162	5.083	.000	
NGCOW # of grade cows	-12617.770	3649.925	-.105	-3.457	.001	
NCCALF # of cross calves	-4401.940	1673.971	-.096	-2.630	.009	
NLBULL # of local bulls	-2109.811	541.400	-.146	-3.897	.000	
SOLDHNY sold honey	-4024.829	1949.086	-.079	-2.065	.040	
NSOLDLP # of livestock products sold	2435.866	1100.368	.077	2.214	.028	
INT_LCOW soldmk interacted with nlcow	1294.881	544.798	.097	2.377	.019	
TARGET target hh?	-2773.490	1204.861	-.070	-2.302	.023	
SOLDMK sold milk	-2823.468	1803.227	-.061	-1.566	.119	

a Dependent Variable: VNET_LV net livestock income, both rounds

Zone 4

Model Summary(b)

Model

R	R Square	Adjusted R Square	Std. Error of the Estimate	1
.893(a)	.798	.785	11138.16246	

a Predictors: (Constant), NSLDLCOW # of local cows sold, NCCALF # of cross calves, NRABBIT # of rabbits, NLBULL # of local bulls, NSLDGOAT # of goats sold, SCHHEAD years of schooling for hh head, SOLDCCOW sold cross cow, MTHGOAT mthmilk interacted with ngoat, TARGET target hh?, MKPROD total milk produced, SOLDMK sold milk

b Dependent Variable: VNET_LV net livestock income, both rounds

Coefficients(a)

Model

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			1
(Constant)	1798.363	1528.174		1.177	.241	
MKPROD total milk produced	20.931	.982	1.093	21.312	.000	
SOLDMK sold milk	-17286.140	3088.617	-.357	-5.597	.000	
SOLDCCOW sold cross cow	18083.076	4756.863	.142	3.801	.000	
NCCALF # of cross calves	3729.704	1120.761	.147	3.328	.001	
NLBULL # of local bulls	3851.030	1339.973	.102	2.874	.005	
SCHHEAD years of schooling for hh head	-286.424	131.349	-.077	-2.181	.031	
TARGET target hh?	-5487.632	2201.759	-.114	-2.492	.014	
NRABBIT # of rabbits	2045.655	599.627	.122	3.412	.001	
NSLDGOAT # of goats sold	2302.330	731.815	.111	3.146	.002	
MTHGOAT mthmilk interacted with ngoat	-42.258	18.718	-.092	-2.258	.025	
NSLDLCOW # of local cows sold	9782.560	4664.001	.072	2.097	.037	

a Dependent Variable: VNET_LV net livestock income, both rounds

INFORMAL OFF-FARM

Zone 1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	.979	.958	.957	19307.16000

a Predictors: (Constant), NMTHS, ORDCOMP, MAXNMTHS, PVASSETS, MTHFISH

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-25009.339	7319.665		-3.417	.001
	MAXNMTHS	.258	.005	.931	50.681	.000
	PVASSETS	9.119E-02	.020	.084	4.568	.000
	MTHFISH	1516.381	423.761	.069	3.578	.000
	ORDCOMP	1716.722	637.110	.048	2.695	.008
	NMTHS	1283.629	574.934	.043	2.233	.027

a Dependent Variable: VINFORM

Zone 2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	.843	.711	.698	16693.77810

a Predictors: (Constant), SCHHEAD years of schooling for hh head, URBAN, MTHRET, MAXNMTHS

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3071.936	2031.568		1.512	.134
	MAXNMTHS	.251	.025	.688	10.229	.000
	URBAN	19037.955	5502.407	.207	3.460	.001
	MTHRET	489.878	612.417	.050	.800	.426
	SCHHEAD years of schooling for hh head	1191.110	517.765	.146	2.300	.024

a Dependent Variable: VINFORM

Zone 3

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	.914	.836	.828	5817.25394

a Predictors: (Constant), ECOMPARE Enum - rate economic conditions of family compared to other households - ECOMPARE, URBAN, NPEOPLE, RETAIL, MAXNMTHS, MTHRET

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-4968.421	2008.665		-2.473	.015
	MAXNMTHS	.446	.023	.766	19.778	.000
	MTHRET	2217.162	587.585	.480	3.773	.000
	NPEOPLE	2448.310	586.838	.163	4.172	.000
	URBAN	5304.399	2159.220	.091	2.457	.015
	RETAIL	-14375.922	5495.240	-.328	-2.616	.010
	ECOMPARE Enum - rate economic conditions of family compared to other households	2063.413	893.985	.088	2.308	.023

a Dependent Variable: VINFORM

Zone 4

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
	.930	.865	.857	17106.85538

a Predictors: (Constant), ORDCOMP, MTHMECH, MTHRENT, NPEOPLE, TREESELL, FEMHEAD female headed hh, MTHWINE, DRIVER, MAXNMTHS

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-14218.705	4837.408		-2.939	.004
	MAXNMTHS	.505	.023	.847	22.375	.000
	TREESELL	-161306.694	19285.124	-.289	-8.364	.000
	NPEOPLE	7982.893	1464.095	.186	5.452	.000
	DRIVER	24724.935	8296.605	.098	2.980	.003
	MTHWINE	1493.126	520.973	.091	2.866	.005
	MTHMECH	2580.492	1077.863	.078	2.394	.018
	FEMHEAD female headed hh	-7585.053	3380.432	-.070	-2.244	.026
	MTHRENT	1760.923	852.893	.066	2.065	.041
	ORDCOMP	983.193	479.626	.064	2.050	.042

a Dependent Variable: VINFORM

SALARY AND REMITTANCE

Zone 1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.880	.774	.757	40747.6863

a Predictors: (Constant), MTHIND, MTHVET, MTHDOC, MTHWAIT, LECTURER, MTHENG, MTHSURV, SCHHEAD years of schooling for hh head, CLERK, NMTHS

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-3042.853	5365.774		-.567	.572
	NMTHS	7918.000	917.059	.424	8.634	.000
	MTHSURV	33296.416	4866.961	.285	6.841	.000
	CLERK	175811.652	21530.309	.392	8.166	.000
	LECTURER	240885.447	42354.414	.244	5.687	.000
	MTHWAIT	-19612.672	4152.112	-.218	-4.724	.000
	MTHENG	12958.602	3138.150	.175	4.129	.000
	MTHDOC	10659.088	3538.943	.129	3.012	.003
	MTHVET	21889.976	6837.142	.133	3.202	.002
	SCHHEAD years of schooling for hh head	1767.515	650.315	.124	2.718	.007
	MTHIND	5841.474	2387.504	.106	2.447	.016

a Dependent Variable: VSALREM value of sal and rem

Zone 2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980	.961	.955	13197.4903

a Predictors: (Constant), TARGET, ORDCOMP, MTHCIVIL, MANAGER, SOCWORK, TEACHER, MTHACCT, DRIVER, CLERK, WATCHMAN, ORDER Order of importance, SCHHEAD years of schooling for hh head, NMTHS, MTHCLERK, MTHWATCH, ECOMPARE Enum - rate economic conditions of family compared to other households - ECOMPARE, MTHTEACH

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
	(Constant)	41556.442	14277.025		2.911	.004
	NMTHS	2767.115	503.970	.178	5.491	.000
	MTHCLERK	61317.816	2949.021	1.296	20.793	.000
	CLERK	-378609.430	24382.153	-.942	-15.528	.000
	MTHTEACH	34143.784	2885.589	.883	11.833	.000
	SCHHEAD years of schooling for hh head	1648.748	479.623	.105	3.438	.001
	MTHCIVIL	9712.831	836.197	.290	11.615	.000
	TEACHER	-211751.789	28934.184	-.527	-7.318	.000
	ORDCOMP	9275.212	1623.780	.476	5.712	.000
	MANAGER	76045.342	15946.239	.110	4.769	.000
	SOCWORK	88968.783	15859.562	.129	5.610	.000
	MTHWATCH	6284.904	1574.625	.252	3.991	.000
	MTHACCT	9437.839	2308.080	.082	4.089	.000
	DRIVER	45710.681	14160.334	.066	3.228	.002
	WATCHMAN	-37442.260	14733.803	-.149	-2.541	.013
	ECOMPARE Enum - rate economic conditions of family compared to other households	-33848.789	7218.132	-.326	-4.689	.000
	ORDER Order of importance	-11862.844	3173.480	-.206	-3.738	.000
	TARGET	5989.632	2608.529	.048	2.296	.024

a Dependent Variable: VSALREM value of sal and rem

Zone 3

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.942	.888	.877	33792.9775

a Predictors: (Constant), SCHHEAD years of schooling for hh head, MTHCLERK, MTHPENS, MTHENG, NADULT number of adults in hh, ECOMPARE Enum - rate economic conditions of family compared to other households - ECOMPARE, MTHDOC, CIVIL, ORDASSET, NMTHS, MTHTEACH, PVASSETS

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-61127.327	11968.357		-5.107	.000
	ORDASSET	.159	.019	1.118	8.226	.000
	MTHTEACH	4129.640	1378.164	.132	2.996	.003
	MTHCLERK	7637.180	1746.692	.141	4.372	.000
	CIVIL	323133.716	37117.645	.289	8.706	.000
	NMTHS	6514.573	797.424	.332	8.170	.000
	PVASSETS	-.527	.086	-.831	-6.163	.000
	MTHENG	11578.664	2935.076	.124	3.945	.000
	NADULT number of adults in hh	7287.591	1745.295	.143	4.176	.000
	MTHDOC	12242.786	3014.102	.131	4.062	.000
	ECOMPARE Enum - rate economic conditions of family compared to other households	15706.014	5635.709	.090	2.787	.006
	MTHPENS	-3953.964	1416.149	-.093	-2.792	.006
	SCHHEAD years of schooling for hh head	1712.441	676.962	.097	2.530	.013

a Dependent Variable: VSALREM value of sal and rem

Zone 4

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.905	.819	.805	91903.4932

a Predictors: (Constant), PENSION, MTHSOC, SCHHEAD years of schooling for hh head, MANAGER, DOCTOR, NPEOPLE, MTHCLERK, MTHTEACH, NADULT number of adults in hh, ORDCOMP, VET

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-108805.828	24963.411		-4.359	.000
	NPEOPLE	58218.591	9591.179	.282	6.070	.000
	MTHTEACH	7402.026	1887.012	.169	3.923	.000
	MTHCLERK	11265.996	2683.583	.170	4.198	.000
	MTHSOC	35499.345	7804.914	.166	4.548	.000
	MANAGER	158130.844	47730.311	.122	3.313	.001
	NADULT number of adults in hh	10777.045	3081.708	.170	3.497	.001
	SCHHEAD years of schooling for hh head	4380.958	1273.226	.140	3.441	.001
	DOCTOR	443117.156	79524.678	.243	5.572	.000
	VET	511392.054	93948.042	.281	5.443	.000
	ORDCOMP	7304.919	2695.336	.116	2.710	.008
	PENSION	-66157.049	27771.417	-.089	-2.382	.019

a Dependent Variable: VSALREM value of sal and rem

Annex G:

Caloric Needs Used for Calculation of Consumption Adult Equivalents

Table B1. Kilocalorie requirements per day, by gender and age

Age	Female	Male
0-1 year	820	820
1-2 years	1150	1150
2-3	1350	1350
3-5	1550	1550
5-7	1750	1850
7-10	1800	2100
10-12	1950	2200
12-14	2100	2400
14-16	2150	2650
16-18	2150	2850
18-30	2100	3000
30-60	2150	2900
>60	1950	2450

Source: WHO, 1985

Annex H:

Procedures for Generating Income and Income Component Estimates Using Spss/windows Syntax File

To generate estimates of income and income components using SPSS/Windows, first clean the data. After cleaning, all conversion of questionnaire variables to proxy variables will be done by an SPSS syntax file developed by Tegemeo/MSU. The steps for using the SPSS for Windows package are as follows:

1. Enter the questionnaire data in the following files (note that "??" in each file name refers to the two-digit year in which the survey was conducted; for surveys done in 2003, "??" should be replaced with "03"):

File #	File Name	Relation to Questionnaire	Variables
1	HHIDFINAL??.SAV	First page – ID variables	key variables NGO ... ENUM
2	HH??.SAV	Household level file - all questions not in tables	key variables TACRES ... ECOMPARE
3	DEMOG??.SAV	Member level file - all data from Household Member table	key variables NAME PERNO SEX ... SALARY
4	OTHERCROP??.SAV	Crop level file -- all data from "Cereals, Tubers, Pulses, and Industrial Crops" table	key variables Crop Prod ... Sell
5	FV??.SAV	Crop level file -- all data from Fuit & Vegetable table	key variables Crop Prod ... Sell
6	LVSTK??.SAV	Animal level file -- Livestock table	key variables animal namim, ssellanim

File #	File Name	Relation to Questionnaire	Variables
7	LIVEPROD??.SAV	Animal products level file -- Livestock products table	key variables animprod nprod ... frqs1
8	OFFFARM??.SAV	Month level file – all data from "Participation in off-farm activities over the past 12 months" table	key variables month infmth, salmth
9	BUSLAB??.SAV	Activity level file – data from left hand portion of "Business and informal off-farm activities, and salaried wage labour" table	key variables actinf informal
10	SALWAGE??.SAV	Activity level file – data from right hand portion of "Business and informal off-farm activities, and salaried wage labour" table	key variables actsal salaried
11	ASSET??.SAV	All data from assets table	key variables item qty
12	ECACT??.SAV	All data from "Importance of Income Sources" Table	key variables econact order

2. Save these uncleaned files in a folder of your choice. This will be your copy of the original, uncleaned data, which should not be changed.

3. Create the folder c:\proxy??\incprox\data and copy all 13 uncleaned files to it. As in the file naming conventions in the table above, replace "??" with the two digit year of the survey, e.g., "03" if the survey was conducted in 2003.

4. Clean the files in c:\proxy??\incprox\data using procedures your NGO has developed with other surveys, and save the files to the same names. You will now have uncleaned, original data in a folder of your choice, and cleaned data in c:\proxy??\incprox\data.

5. Create the folder c:\proxy??\incprox\syntax and copy the file IncproxEstimateNGO.sps to it. Tegemeo/MSU will provide you with a copy of this file upon request.

6. Add ZONE1 through ZONE7 variables to IncproxVarsNGOs.sav, as instructed in IncproxEstimate.sps, and save to the same name.

7. Run IncproxEstimateNGO.sps. This file will create all required proxy variables and generate income results, saving them to the file IncomeNGO.sav. It will also deliver mean and median values for household income and income components in the SPSS Output Navigator.