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.*

ANALYSIS OF THE ZATPID I RURAL HOUSEHOLD SURVEY: WITH EMPHASIS ON DATA QUALITY AND A COMPARISON OF CHARACTERISTICS OF SMALLHOLDER MAIZE PRODUCERS OF ZAMBIA

By

A. Kapena Sumbye

A PLAN B PAPER

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Agricultural Economics

1989

ACKNOWLEDGEMENTS

I wish to thank the members of my committee, Dr. Dan Clay, Dr. Eric Crawford and Dr. Michael Weber (my major professor), for their guidance. Their experience with Africa in general and Southern Africa in particular was a source of constructive and practical counsel.

I also wish to thank Mrs. Wendy Peters, a member of staff in the computer programming office of the Department of Agricultural Economics, for showing me everything I needed to know about SPSSPC to be able to analyze my data. I also wish to acknowledge the assistance rendered to me by the other members of staff in the programming office whenever Mrs. Peters was not available.

I would not forget to thank Miss Kathryn Baird (then student assistant) for sharing with me the task of data cleaning. Miss Baird did most of the "Bench mark" data cleaning while I concentrated on cleaning the "Flow" data.

My thanks also go to the following people (who were students at Michigan State University then): Ms. Pia Chuzu and Mr. Sipula Kapola, for their comments on my paper.

My thanks would not be complete without acknowledging Mrs. Elda Keaton's advice and assistance that made it possible for me to survive the final weeks of my research work when I had no source of funds for maintenance.

Finally, I wish to thank Mrs. Melba Lacey for managing to finish typing the text within an incredibly short period. Without her cooperation, financial circumstances would have forced me to leave the paper unfinished.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ANNEX TABLES

CHAPTER I

INTRODUCTION

Between December 1984 and July 1986, the Planning Division of the Ministry of Agriculture and Water Development, in conjunction with the Rural Development Studies Bureau of the University of Zambia and Iowa State University conducted a rural household survey. This is referred to as the ZATPID I Rural Household Survey, because it was conducted during phase I of ZATPID, a five year USAID/Zambia project. ZATPID stands for Zambia Agricultural Training, Planning and Institutional Development Project.

The aim of the survey was to provide data for use in analysis to inform policy making in Zambia. It was also supposed to provide a data bank for use by Zambian graduate students to meet the research requirements of their degree programs.

To date, very little comprehensive analysis of these data has been done to determine their accuracy and usefulness. The present study was, therefore, undertaken with the aim of completing a general review of the data set before selecting from it a subsample to further analyze.

To distinguish between the rural household survey (RHS) and this paper, the term "survey" is used to refer to the rural household survey, and the term "study" or "the paper" to refer to this paper.

A. STUDY OBJECTIVES

The objectives of the study are:

1. to review the methods and survey procedures used in the RHS;
2. to complete internal and logical consistency checks on the data base resulting from the 477 households surveyed in the flow portion of the RHS in order to identify the number of likely cases of valid information;
3. to select a subsample of households producing maize and to conduct a comparison of selected smallholder maize production characteristics;
4. to discuss the implications of the findings of both the consistency checks and the comparative analysis, and to recommend further cleaning and/or analysis of the data set.

B. RATIONALE

Since this survey was completed in the 1984-86 period, there has been very little done with the resulting data, especially on the part of Zambian analysts at the Ministry or the University. There was one piece of research conducted on this data by a Ph.D. student at Iowa State University in 1987, but this thesis was only defended as a preliminary version and so far has not been revised for final acceptance and publication. The opportunity to try to utilize this data is important because it provides opportunities for Zambian analysts to gain the computer and methodological skills necessary to utilize this and similar household data. Even if a judgement is made that little, if any, of the data is potentially useful, the process of systematically cleaning and analyzing the data will be a valuable learning experience. Thus, the following are important reasons for carrying out this study:

1. It is important to review the survey methods and procedures used in the RHS because the accuracy of any final analytical results will depend on the reliability of the survey methods and procedures used. This, in turn, will affect the interpretation and recommendations to be drawn from the survey.
2. The logical and internal consistency checks across households multiple data files will reveal the data which are inconsistent or do not agree. This will provide a useful cross check and help ensure that illogical data values and cases are not included for further analysis.
3. Data on maize growers have been chosen for further analysis in a subsample because maize is the crop grown by most people in the country and is also the principal food crop. If the analyses indicates that meaningful conclusions can likely be drawn on the basis of this subsample, then there may well be good reasons to conduct additional analysis of the data.
4. The discussion of the implications of the findings together with the recommendations will be useful both for the understanding of the maize growers in Zambia and for the future use of the RHS data.

C. RESEARCH QUESTIONS

The following are the principal research questions that will guide this study:

1. Are the methods and procedures followed in the RHS survey consistent with the likely reality of farmer practices and information recall capabilities?
2. Is there internal consistency in information among the various files (portions) of each household/data base?
3. Do the data indicate adequate implementation of the RHS methods and procedures?
4. After eliminating obvious cases of problem data from the sample, does analysis of a maize producing subsample suggest different household characteristics that are: a) statistically significant, b) valid in the sense of plausibility in the reality of rural Zambia?

D. OUTLINE OF THE REMAINING CHAPTERS

Chapter II reviews the methods and survey procedures used in the RHS. This will include describing the overall survey plan, the units and levels of observation, the sampling procedure and data collection instrument, and the resulting data files. Chapter II also reviews the logic of the RHS design and its phasing with the cropping calendar in Zambia.

Chapter III presents the steps taken in completing the logical and the internal consistency checks. The results of the consistency checks are also presented and discussed. Then a subsample is obtained consisting of maize growers, and results of the comparative analysis of these maize growers are presented in this chapter. Chapter III also discusses the implications of the findings and Chapter IV presents conclusions and recommendations regarding the future use of the data set.

CHAPTER II

REVIEW OF THE ZATPID I RURAL HOUSEHOLD SURVEY

A. INTRODUCTION

This chapter reviews the methods and survey procedures used in the RHS and describes the overall survey plan, the units of observation, the sampling procedure and data collection instruments, and also gives a brief description of the files. The chapter also reviews the logic of the RHS design in light of the cropping calendar. A detailed description of the aims, methods and survey procedures of the RHS are given in a report compiled at the end of the ZATDIDI survey entitled the "Rural Household Survey Volume I." The present chapter gives only a brief overview of essential material covered in this report.

B. OVERALL SURVEY PLAN

The Rural Household Survey was conducted by the Ministry of Agriculture and Water Development (Planning Division) in conjunction with the Rural Development Studies Bureau of the University of Zambia and Iowa State University.

The Planning Division team consisted of Iowa State University and Zambian personnel (including this author up to the end of the phase of training of the enumerators).

The study was aimed at collecting data on the subsistence and emergent farmers for use in agricultural policy making. Commercial farmers were excluded from the survey. Subsistence farmers were defined as those farmers with less than 2 hectares of land, who seldom produce a surplus above their household needs and, if they do produce a surplus, are likely to dispose of it through barter in their own communities rather than through any formal marketing network. Emergent farmers were defined as those farmers with between 2 and 40 hectares of land, who sell some production on the cash market. For the purposes of this paper, subsistence farmers and emergent farmers are collectively called smallholders and are treated as one group. Hence the terms small scale farmer, subsistence farmers and smallholder are used interchangeably in the subsequent sections of the paper.

The survey comprised five phases. The first phase was the "Quick listing" of household, the results of which were later used in selecting sample households. The quick listing was done in August and September, 1985.

The second phase was the "Bench Mark." The aim of the bench mark was to provide descriptive data, such as demographic data, data on attitudes of respondents, data on practices of households in the prior production season and market facilities available. This survey was done during October and December, 1985.

The third phase was the "flow." The purpose of the flow survey was to provide a continuous accounting of household cropping practices, crop inputs and outputs and other features such as use of household labor, food consumption, household incomes and expenditures throughout the cropping season. Thus, the flow phase ran from December 1985 to June 1986.

The fourth phase was the "Extended bench mark." It was conducted during the flow period (i.e., the cropping period). This survey was carried out because the bench mark survey was not complete. Data were collected on heights and weights of children, which were supposed to be collected during the bench mark period but were not obtained because scales were not available at the time. The extended bench mark survey also collected data on market services.

The fifth phase was the "Post harvest." The aim of this survey was to collect, during the post harvest period (July 1986-December 1986), much of the same data as those collected during the flow. The main difference between these two surveys is that the post harvest one was done only 2 or 3 times (and was not designed to be inclusive) over this six-month period.

The relationships among the above mentioned surveys can be summed up as follows:

(a) the bench mark and the extended bench mark are complementary and are both concerned about what happened "last year," where "last year" refers to the 1984-85 season;

(b) both the flow and the post harvest refer to the 1985-86 season. The flow runs from December 1985 to June 1986 (land clearing, planting and harvest periods). The Post harvest runs from July 1986 to December 1986. There was no separate questionnaire for the post harvest survey. Rather, certain parts of the flow questionnaire were repeated during the post harvest period.

C. LEVELS OF DATA AND UNITS OF MEASUREMENT

The data collected from each household were at the household, household-crop/animal, household-transaction, household-individual, household-implement, and household-visit level.

The basic units of measurement were hectares, acres and limas (1 lima = 1/4 hectare) for the area; and bags, kilograms, litres, bales, bags-unshelled and oxcart or sled for weight and volume.

For the purposes of the RHS, the household was defined as including all those individuals who live in close proximity to each other and who form one work-team under the guidance or direction of one leader, the head of the household. The household may include one or more young married couples and children if for their subsistence and other economic activities these young family members depend on the larger unit, headed by the leader who decides where to direct resources.

D. SAMPLE STRATIFICATION

The sample was stratified on the basis of farm size and ecological zone.

1. Farm size

The size strata used were:

(a) less than 2.0 hectares

(b) 2.0 - 9.99 hectares

(c) 10 - 39.99 hectares.

2. Ecological zones

The ecological zone was another principal level of stratification of the sample. There are four ecological zones in Zambia, namely the Northern High Rainfall area; the Western Semi-Arid Plains; the Central, Southern and Eastern Plateaux; and the Luangua-Zambezi Rift Valley.

The survey included only three of the four zones. The Luangua-Zambezi Rift Valley was omitted because the population density and agricultural production were too low for the accuracy required in the survey.

The provincial composition of the three zones was:

Zone I (Northern High Rainfall): Luapula, Northern, Copperbelt, part of Central and part of Northwest Provinces.

Zone II (Western Semi-Arid): part of Northwest and Western Provinces.

Zone III (Central, Southern, and Eastern Plateaux): Eastern, part of Central, Lusaka and Southern Provinces.

The Provinces were coded as:

1 = Central
2 = Copperbelt
3 = Eastern
4 = Luapula
5 = Lusaka (rural)
6 = Northern
7 = North Western
8 = Southern
9 = Western

E. SAMPLING PROCEDURES

The sampling procedures were mainly determined by the needs of the flow survey which required that enumerators visit each household every seven to ten days. The bicycle was used as the means of transport for the enumerators.

A cluster sample procedure was chosen so that distances between households could be kept to a minimum.

1. Sample sizes

The sample size, like other components of the sampling procedures, was determined primarily by the availability of funds and trained enumerators with respect to the flow survey.

There were enough resources for 40 enumerators. It was assumed that each of these enumerators would cover an average of 15 households per cluster or sample area during the flow. Thus there were 40 clusters sampled during the bench mark phase. The same clusters were used for the flow.

Since the bench mark consisted of only one visit per household, it was estimated that 25 households could be interviewed in a day by each crew of 8 interviewers and a supervisor that was moving from area to area. Thus the initial bench mark sample was 25 x 40 = 1000. The sample after dropouts was 911 households.

For the flow, 15 households were surveyed in each cluster. The 15 households in each cluster were a subset of the 25 households surveyed in the same cluster in the bench mark. The initial flow sample was 15 x 40 = 600. The sample after dropouts was 477 households.

2. Use of the Central Statistics Office Sampling Frame

Because of the desire to extrapolate sample results to the national levels, it was necessary to do accurate weighting of the households and sample areas. To get the data needed for the weighting, the sampling framework utilized by the Central Statistics Office in conducting the national census in 1980 was used.

The Central Statistics Office sample framework consists of a series of Census Supervisory Areas (CSAs) which are subdivided into Standard Enumeration Areas (SEAs). The CSAs in urban areas, national parks and forests, and those along the line of rail were excluded. The resulting sampling frame had approximately 1900 CSAs. These were divided according to the ecological zones in which they fell.

3. Weights

Both the bench mark and the flow data were weighted. The weights were designed to reflect the distribution of the farmers among farm size strata in the universe these weights were determined in the "quick listing" phase of the survey. The weights indicate how many households are represented by each household that was interviewed. Multiplying data values obtained in the survey by weights gives totals for the whole nation or a province.

The weights of any given cell in the flow are larger than in the bench mark because the sample size for the flow is smaller and this means that the rate of expansion must be greater to give estimates of the same nationwide totals.

This present paper will experiment with the weights developed, but will make only minor use of these weights because a) the analysis does not propose any aggregation to national level and b) the accuracy of the weights is questioned in RHS Volume 1, and this issue becomes even more difficult as the sample size is reduced in the present analysis when selected households are dropped due to bad data.

4. Household selection

The steps involved in selecting the households were:

a) Selection of the CSAs

The selection of the sample CSAs from within known ecological zones was done using the principle of "probability proportional to population." A listing of all CSAs was made along with the number of households in each CSA. Then selection was done using a starting point chosen at random between 1 and the sampling interval. The sampling interval was the total number of households in the zone divided by the number of sample areas or clusters in that zone. In this way the CSAs with few households had a smaller chance of being selected than those with a larger number of households.

b) Selection of the Households

As discussed above, the major determinant of the sampling procedure (and the sample size) for the households was the availability of resources, both in terms of finance and personnel.

Once the sample size was determined for each CSA and each size class, the specific households were selected. The selection was done by listing the households in each CSA by number, and then selecting a random starting number and incremented by the sampling interval. Thus the procedure for selecting specific households was similar to that used in selecting the sample CSAs. Note also that the households in the flow sample were drawn from the 911 households (the bench mark sample).

Below is the summary of the household sample composition:

(i)	bench mark	- original total households	=	1000
		households in each cluster or sample area	=	25
		total clusters	=	40 (i.e., 1000/25)
		sample after dropouts	=	911
(ii)	flow	- original total households	=	600
		households in each sample area	=	15
		sample after dropouts	=	477

F. DATA COLLECTION INSTRUMENT

The questionnaire was used as the instrument of data collection, and all answers were recorded on the questionnaire. The questionnaire for the bench mark and flow surveys were developed jointly between Iowa State University analysts and a committee of Zambian analysts from various agencies. The basic design was to assign as many as possible of the questions to the bench mark survey, and leave only those requiring careful measurement to the recurrent flow questionnaire. This did save on redundant questions, although the bench mark information applies to a different cropping season than the flow data, and for certain kinds of analysis this introduces problems. An important part of the flow survey design is carefully described in the survey documentation, and is repeated here because it is central to the logic tests that will be undertaken later in this study.

> "The flow questionnaire was split into two parts. Pages 1-10 remained with the enumerator throughout the enumeration period. Entries were made into pages 1-10 throughout the field work with the visit number being recorded for most data entries so that the time of use of an input was known. Pages 11-16 contained questions on labor use, food consumption, income receipts and expenditures which were completed for each household on each visit or interview. These 6 pages were picked up by the field supervisor and returned to Lusaka to allow editing, coding and data entry to proceed while the field work was still proceeding. The reader should understand that one set of pages 11-16 was completed on each visit. Each set was given a visit number and the days since last visit were recorded. This same data was recorded on page 1 of the base book as a cross check."

An important observation is needed on the above description. Page 7 of the flow questionnaire was designed to collect information on production of each crop and there was no place on this page to record the visit number, unlike all other pages from 1 to 10 as described above. The result of this oversight is that it is impossible to know when the information on crop production was recorded. Hence, it is impossible to know during what time of the year the harvest actually took place. Such information is an important check as to whether the enumerators were following instructions and making the entries on pages 1-10 throughout the field work with the visit number being recorded as an indicator of the time of use of an input. Clearly this was not possible for the timing of outputs. An important logic check regarding output measurement is whether the enumerators were actually visiting the household after the harvest period was completed, and obtaining good estimates of production of each crop. Since it is unlikely that all crops are harvested at once, there should generally be more than one visit required to measure output.

The only other way to determine the likely timing of visits during the harvest phase is to check the dates on flow questions 11-16 to see when the enumerators were visiting the households for collecting data. This analysis will be done in the present study with the assumption that if the household was not visited to complete the labor use, consumption, and expenditure information on pages 11-16, then it is likely that the data on pages 1-10 were not completed (especially the output estimates) after the fact as the survey called for. To the extent that there are output data, they may represent an estimate by the enumerator during one of the early visits to the household during the planting and weeding periods of the year. If this is the case the output information becomes very suspect for large measurement errors.

G. DATA FILE CONTENT AND FILE LEVEL

1. Bench Mark Data Files

As shown in Annex Table 1, the bench mark questionnaire resulted in nine data files, running from Card01 to Card09. Most of these files contain data measured at the household level, with household-individual, household-crop, and household-implement also being important.

The contents of the files in the order Card01 to Card09 are: particulars of the household, crop production last year, livestock and poultry, nuts and fruit production, marketing and rural services, attitudes, and inventory of farm equipment and tools.

Some files were recorded in subparts. For instance Card03 was recorded in two subfiles called Card31 and Card32. To get one file for Card03 requires JOIN ADDING (in SPSS PC+) the two subfiles.

Files at levels other than the household need aggregating to the household level when comparisons are being made with other data files at the household level. For example, Fcard07 has 1373 cases at household-crop level. This aggregates to 446 households, and can be analyzed at the household level.

2. <u>The Flow Survey Data Files</u>

The flow questionnaire resulted in sixteen data files, running from Fcard01 to Fcard16, where F stands for "flow" (see Annex Table 2). None of the flow files contain data at the household level. The levels used include household-visit, household-crop, household-livestock, and household-transaction levels. The data for Fcard11 through Fcard16 were recorded in subparts, where each subpart stood for each of the 9 provinces. This means, for example, that to have only one Fcard11, all the subparts of Fcard11 would have to be join added.

The contents of the files in the order Fcard01 to Fcard 16 are: dates of visits; crop production; use of fertilizer, lime and manure; crop production-pesticide use; use of animal power in crop production; use of tractor power for crops; other crop expense and crop production; crop dates; household livestock; demography of household and labor force; household labor record; household labor record, short-term workers; household food consumption; household income since last visit; and household expenses since last visit.

H. THE RHS DESIGN AND THE CROPPING CALENDAR

The flow portion of the RHS was designed so that it would capture all the activities that take place on the farm during a full planting season, namely land preparation, planting, weeding, and harvesting. The visit interval was to be 7 days to 10 days in order to reduce memory recall problems. The basic survey design for all the questions in the flow phase required that the farmer report activities since the previous visit. The only exception was for consumption, where the farmer was asked only about the "last three days." As discussed in the document RHS Volume 1, there was considerable confusion on this issue. Apparently many enumerators visited a household and if no one was present to answer questions, this was counted as a visit although no data were collected. When the enumerator visited the household the next time, many times the information recall period was from the last time the enumerator visited, not from the last time specific questions were asked. This resulted in many gaps in recall, and means that some unknown number of cases may contain underestimated values.

Figure 1 is a general cropping calendar showing likely harvest periods for selected crops. Actual timing of activities may differ slightly depending on zones and weather conditions.

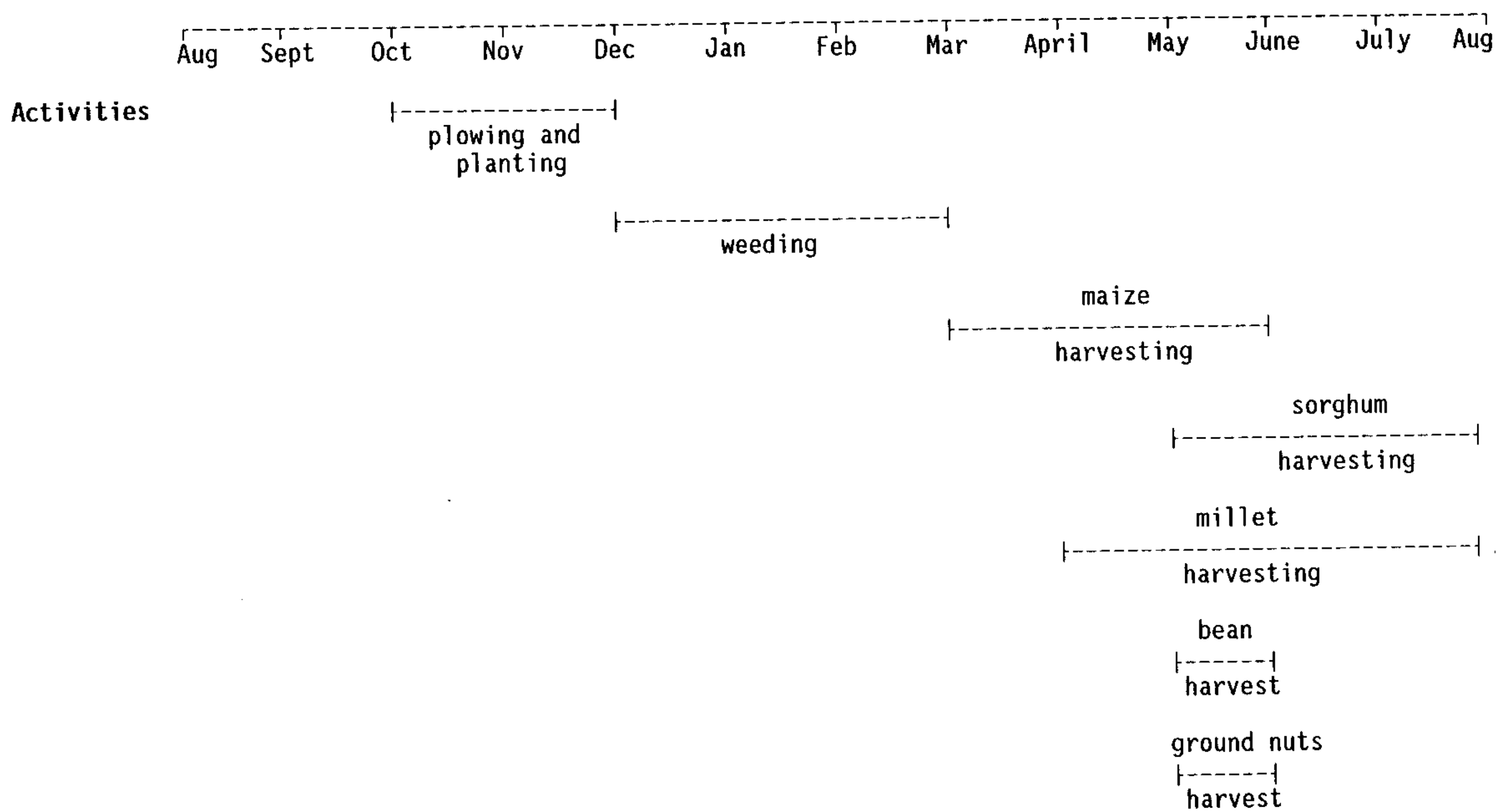

Figure 1. General Cropping Calendar with Selected Crop Harvest Periods.

The important issue that Figure 1 helps illustrate is that the instructions for the survey as to when to stop collecting data were also quite inadequate. The general instruction was to continue to visit each sample household until all crops grown by the household had been harvested, and this was to be determined by the enumerator. As will be shown in the consistency checks, it appears that many households were not visited after April or May. This would appear plausible only if a household planted only maize and harvested very early in the season. Furthermore, since the survey was to also collect consumption, expenditure, and income data, it does not seem adequate to stop collecting such data just because the harvest is completed. This decision rule will result in consumption and expenditure data for very unequal periods, especially if consumption and/or expenditures are affected by the size of harvest (which is often an important issue to inform in household surveys). Likewise, some of the most interesting periods of household behavior may be missed since it is after harvest that households may have the greatest potential to vary both consumption and expenditure behavior. The issue of frequency of visit and cropping season reality will be examined again during the logical data tests.

I. CROP INFORMATION

The following is basic background information on the principal crops grown in Zambia. This information comes from previous studies and from the author's general knowledge of farming conditions. This information will be helpful as a general check against the results to be obtained from the analysis of selected data results from the flow survey.

1. <u>Maize</u>

Maize is the principal food crop in Zambia. It is produced by all the types of farmers found in the country (smallholder farmers, emergent farmers, commercial farmers and state farms).

Although maize is grown throughout the country, the majority of production comes from Central, Southern and Eastern Provinces. Small-scale producers, estimated at 600,000 households, are thought to account for about 60% of the official maize sales (World Bank, Zambia: Policy Options and Strategies for Agricultural Growth 1984). Commercial farmers, estimated at 500 households, account for the remaining 40%.

The maize yield for the country during the 1984-85 season was 22 bags/ha or 1980 kg/ha. The yields for the individual provinces were 2430 kg/ha for Eastern, 1800 kg/ha for Luapula, 2070 kg/ha for Northern, 1890 kg/ha for North Western, 2070 kg/ha for Southern and 1080 for Western (Quarterly Agricultural Statistics Bulletin 1985: MAWD Planning Division).

The range of maize yield betwen 1982 and 1985 was 1620 to 1980 kg/ha for the whole country, 1890 to 2430 kg/ha for Central Province, 1350 to 1890 for Cooperbelt Province, 1170 to 1620 kg/ha in Eastern Province, 1800 to 1890 for Luapula Province, 1530 to 3780 kg/ha for Lusaka, 2070 to 3060 kg/ha for Northern Province, 1440 to 1890 for North Western, 1620 to 2430 for Southern and 630 to 1080 kg/ha for Western Province.

2. Sorghum and Millet

Sorghum and millet are grown almost entirely by small farmers. They are usually grown under shifting cultivation and are frequently intercropped with other crops, especially beans, sweet potatoes, ground nuts and pumpkins. Both crops can grow in many parts of the country. They are, however, common where rainfall is uneven, such as in the Gwembe valley of Southern Province. There is little official sales of both the crops.

3. Cassava

Cassava is the principal food in most areas where maize is not the principal staple food. Cassava production is dominant in the north.

Cassava would grow in most climatic conditions. Therefore, the distribution of cassava growing areas is more due to historical factors than to climatic ones. Where it is grown, cassava is grown on scattered fields and harvested over a period of almost two years. Hence it is difficult to estimate yields. Little cassava is marketed.

4. Cotton

Cotton is mainly grown in Southern, Central and Eastern provinces. It requires light to moderate rainfall, a short wet season, and a lot of sunshine. It is produced by both commercial farmers and smallholders, with the latter dominating. It is almost entirely grown for sale.

5. Sunflower

Sunflower is grown throughout the country by all the groups of farmers. Small-scale farmers dominate in its production. The crop requires a short growing period compared to hybrid maize SR52. Its attractiveness to smallholders is that it is comparatively free of pest and disease.

6. Soybean

Soybean is mostly grown by commercial farmers. It is grown on both dry land and under irrigation, and it is grown in rotation with maize or wheat.

It has not been attractive to smallholders in that it requires use of herbicides, is difficult to harvest, and is not considered tasty. Soybean is mainly grown for sale to oil product dealers.

7. Ground Nuts

Groundnuts are mainly grown in the Eastern Province where they are grown on relatively big plots. Elsewhere, groundnuts are mostly grown on small plots by both medium and small farmers, and used mostly for household consumption.

CHAPTER III

LOGICAL AND CONSISTENCY CHECKS ON THE RHS AND COMPARATIVE ANALYSIS OF MAIZE GROWERS

A. INTRODUCTION

This chapter presents the steps taken in doing the logical and consistency checks and discusses the results and implications thereof. It also presents the subsample of maize growers selected after the completion of the consistency checks and reviews the results of the comparative analyses done on the subsample.

B. DATA CLEANING STEP I (Data Preparation)

This step involves data from both the bench mark and the flow surveys.

1. Action

SPSS/PC+ was chosen as the statistical package to be used for the data analysis. This was partly due to the fact that, along with the raw ASCII data that the Zambia project office provided, they also sent SPSS program files (include files). These program files included the SPSS commands to create SPSS system files from the ASCII files, which are needed to do the data analysis.

To create the SPSS system files from the ASCII data files each of the 25 program files (9 for the bench mark, 16 for the flow) had to be checked against the ASCII data files to be sure the data values were actually in the columns specified by the program files. The programs also included "most" of the coding labels (value and variable labels, see RHS Vol. 1 annex VII and annex III). Some of the code labels were missing and had to be added.

2. Logic

This was the most basic visual check of the data.

3. Procedure Used

The editor included in SPSS (Review) was not appropriate for checking the program files against the ASCII files because it has a limitation of 80 columns and the ASCII data files had a width of more than 80. Therefore, a different text editor had to be used for this check. The IBM Personal Editor was used.

Each of the 25 programs had to be executed within SPSS, and the SPSS system files that were created, saved to disk.

C. DATA CLEANING STEP II (Consistency: Consumption and Production)

This step involves the use of data from the flow survey only.

1. Action

At this stage of the cleaning process, data from Fcard01 and Fcard14 were used. Fcard01 contains visit days, while Fcard14 contains information about household expenditures and food consumption. The two files were join matched by household IDs and those households which appeared only in one of the two cards were excluded from the analysis.

2. Logic

One of the purposes of this cleaning step was to check whether each visit number (i.e., a number given each time the household was visited) recorded in Fcard01 had a corresponding visit number record in Fcard14. It was possible that a visit number in Fcard01 had no corresponding visit number in Fcard14, implying that the enumerator recorded in Fcard01 that he had visited the household but did not in fact interview the farmer in question, or he just forgot to record the visit or vice versa.

In this step a household would be excluded from further analysis if it appeared only once in one of the two cards. In other words, at this point the concern was to check whether the farmer actually participated in the survey even if for one visit.

Fcard14 consumption data was chosen to be matched with Fcard01 because as part of the consumption portion of the questionnaire it represents the most complete and most frequently recorded data set of all the files that were to be created from the flow survey. This is so because Fcard14 contains information pertaining to consumption and the survey design was for each household to be visited every 7 to 10 days with the consumption portion of the questionnaire being completed each time.

3. <u>SPSS Procedure and Results</u>

a) Fcard01 contains household IDs and visit numbers (each time an enumerator visited a household, a number was assigned to that visit). This is just a record of each visit made to a household, therefore this file is at the household-visit level. This file contains 8819 cases.

b) Fcard14 contains household expenditures and food consumption information. This file was also recorded at the household-visit level. It contains 8848 cases.

c) Join match these two files by household ID and visit number. Exclude those cases which did not have corresponding entries in both files. This yields a file with 8263 cases.

d) Aggregate the 8263 case file to the household level. This is to check how many households are represented in the 8263 file. The resulting file contains 477 cases.

4. Implications of the Results

a) The 8263 cases are the cases that had corresponding entries in both Fcard01 and Fcard14. Fcard01 and Fcard14 are supposed to have the same number of cases. Therefore, the result of 8263 cases indicated problems in the data.

b) The 477 households obtained after aggregating represent households found in both Fcard01 and Fcard14 that had received at least one visit each. This confirms that of the original 600 households included in the flow survey, only 477 households were left in the final sample.

c) More rigorous tests at a later cleaning stage involved taking visit numbers greater than one and considering their distribution over the survey period as the criterion for qualification for further analysis.

D. DATA CLEANING STEP III (Complete Production Data)

This cleaning report involves the flow survey only.

1. Action

The cleaning process here related results from the planting and harvesting parts of the flow survey. The two files involved were Fcard02 and Fcard07, containing information about planting and harvesting, respectively. Households reporting contradictory information, such as a quantity harvested but nothing planted, were excluded from the analysis.

Also households that reported planting no crop and/or harvesting no crops were excluded. All households that reported planting and harvesting at least one crop were included.

2. Logic

This cleaning step is complex because of the possibility of confusing actual zero values (which might result from total crop failure) with poor survey methods and lack of supervision of the enumerators. There are at least two clear possibilities as discussed below:

a) Concerning the contradiction between planting and harvesting information for the same crop, it is not possible that a farmer who did not plant a particular crop could harvest the same crop. These cases are considered as bad data.

b) The households that reported planting no crops at all and/or harvesting no crops were excluded because, according to the RHS report, either the zero indicated cases where there was total crop failure, or it represented an error when the enumerator should have put "missing" or some other value. Because it is no longer possible to tell which was which, the only logical thing is to exclude the cases containing zero production of a crop that was indeed planted. Given that the rains were generally good throughout Zambia during this survey period, there was no widespread crop failure. A few households may indeed have experienced total crop failure for various reasons, but the number of cases is likely to be small.

3. <u>SPSS Procedure and Results</u>

a) Join add the two subfiles of Fcard02 (the file with planting information) to get one file. The resulting file had 1668 cases, at the household-crop level.

b) Aggregate the file containing 1668 cases to household level. The aggregated file had 474 households.

c) Drop from the file containing 474 households those households with total area planted equal to zero. The resulting file contained 462 households. In other words 12 households were dropped.

d) Aggregate Fcard07 (the file with information on harvesting) to the household level. The aggregated file had 446 cases at the household level, from 1373 cases at the household-crop level.

e) Exclude from the file containing 446 households those households that reported harvesting zero. The resulting file had 430 households.

f) Join match, by household IDs, the adjusted file containing planting information (462 households) with the adjusted file containing information on harvesting (446 households). The resulting file had 434 households.

4. Implications of the Results

a) The 474 households obtained after aggregating Fcard02 are households that reported planting at least one crop. Of the three households dropped, two came from Nchelenge district in Luapula and the third came from Kalabo in Western Province. None of the three households appeared in the file on livestock (Fcard09) hence it is not possible that they were specialized livestock farmers. However, all the three households appeared in the file on food consumption (Fcard14). The two households from Luapula had 23 cases each reported and the household from Western province had 11 cases reported. The appearance of cases on consumption data indicate that these households actually participated in the flow survey. But since there is no output data for them, they must be excluded from the current analysis.

b) The 462 households obtained in procedure item (c) above are farmers who reported planting something, supported by the fact that they reported the name of at least one crop that they planted. It is likely that these 12 households omitted contain bad data, so they were excluded from further analysis.

c) The 446 households obtained after aggregating Fcard07 are households that reported harvesting something. This means only that a crop code was recorded for these households. This says nothing about how much of the crop was produced, only that the enumerator had written down a crop name on the questionnaire, and that this information was entered into the data file.

Finding that only 446 of the original 477 households surveyed even mentioned crop output indicates the likelihood of poor attention to measuring output in the survey and hence poor data. The alternative implication is that 31/477 or 6.5% of the sample of farmers did not produce any crops, or specialized just in livestock, or perhaps were recorded as farmers, even though in reality they only lived on a farmstead, while really working at some other occupation. While these results are possible, given the author's knowledge of small farmers in Zambia, and of the survey design, such results are unlikely.

d) The 430 households obtained in procedure item (e) above and obtained by excluding from the 446 those cases where a crop code appeared but the quantity reported actually equalled zero. Thus 16 of the 446 households who reported production of at least one crop showed a zero amount produced. The zeroes could be cases of crop failure, or incorrect data. Again, the most likely case is of bad data. Given the diversified cropping patterns of small farms in Zambia, it is difficult to imagine that 16 households in the sample produced only one crop, and also experienced total production failure of this crop.

e) The 434 households obtained after join matching the adjusted files containing, respectively, planting and harvesting information and adjusting for the unmatching households, are households that harvested something and reported planting at least one crop. The end result of this cleaning step is that about 10% of the cases (45 of 479) are excluded from analysis mainly because of fundamental data quality problems.

E. DATA CLEANING STEP IV (Consistency: Production and Visits)

This cleaning report involves the data from the flow questionnaire only.

1. Action

The consistency check at this stage was to join match by household ID the file containing the 477 households (the households that had received at least one visit during the flow survey) with the file containing 434 households (the households that had both planted and harvested something) and discard those households which did not match. This would leave only the households that qualify with respect to the criteria of minimal visit and basic input/output information.

2. Logic

It is only logical that a household would be considered as having participated in the survey if it meets at least all the criteria looked at so far, namely planting something, harvesting something and being visited at least once during the survey period.

3. SPSS Procedure and Results

The join match command was used to match, by household IDs, the file containing 477 households (the households that had been visited at least once during the survey period) with the file containing 434 households (households that had planted and harvested something) and retained only the households that matched. The resulting file contained 434 households.

4. Implications of the Results

The 434 households in are those that met all the requirements considered. In other words, for careful analysis of basic household input and output issues, 434 households should be the sample to start with, instead of the original 477 in the data base.

F. DATA CLEANING STEP V (Valid Maize Production Data)

The analysis in this step and subsequent ones goes beyond the basic requirements considered so far. More rigorous considerations are introduced, such as going beyond one visit only and how these visits were distributed over the survey period.

Maize production practices will also be analyzed. Conclusions reached from the analysis will be used to gain an understanding of the general characteristics of the farmers in the survey.

1. Action

In this section a subsample of maize growers is drawn from the 434 households identified who meet the minimum requirement considered in the previous sections. To define a maize grower, emphasis is put on analysis of maize production and area planted data.

2. Logic

Maize is selected for further analysis because maize is the most important crop in Zambia, both in terms of number of households involved in growing it and of the relative amount grown. Of the 434 households who harvested something, 394 reported harvesting maize among other crops. If analysis of the data obtained in the survey does not yield statistically valid and realistic estimates of farmer behavior with maize, then it is doubtful whether analysis of data obtained for other crops identified in the survey is worthwhile.

3. <u>SPSS Procedures Used in Selecting the Basic Maize Subsample</u>

a) Convert all maize production information to a standard weight unit of Kg (the conversion factors used are shown in Appendix Table 4). This resulted in 394 households. The number of cases of maize at the household-crop level (that is before aggregating to 394) was 463.

b) Convert all area of maize planted information into standard units of hectares. The number of households that reported planting maize was 410. The number of cases reported of maize hectarage before aggregating were 451 (household-crop level). The conversion factors for hectarage are presented in Appendix Table 3.

c) The next step was to join match each of these two files with the basic sample of 434 households obtained in Step IV so that only those maize growers who meet the requirements of the basic subsample remain. The resulting subsample has 382 households.

d) Compute maize yield on the remaining 382 households. The results show that some households had maize yield that was zero; others had yield values as dot (·) which stands for missing data in SPSS. In the case where a zero value appeared for yield there was maize hectarage data missing (·) A yield missing value (·) appeared for households where both maize hectarage and maize kilograms produced were missing. Both cases result in unusable information, and these households must be removed from the subsample. The resulting subsample had 361 households.

The 361 households remaining are the maize subsample with the minimum or basic qualification considered so far. Quite a number of households (73 or 17%) were dropped from the overall sample of 434 households in this step. It is likely that only a small percentage of those dropped did not in reality grow maize, with the remainder being dropped as a result of measurement and/or data processing errors. We do know that of the 434 households who reported growing any crops at all, some 92 percent (394) reported growing maize. This suggests that perhaps as many as 40-50 percent of the cases dropped in this step result from bad data.

G. DATA CLEANING STEP VI (Consistency: Maize Production and Visits)

This step involves data from the flow survey only.

1. Action

In this step the basic subsample of maize (361 households) is exposed to more stringent logic and internal consistency criteria. These include using the criteria of 10 seperate visits during the six months of the flow survey as the minimum number of visits acceptable in order to obtain good data. A second criterion is whether the household participated in the survey up to the month of May 1986 as the minimum number of months of participation.

2. Logic

The flow survey officially was to run from December 1985 to June 1986, although many households were still being visited in July and some even in August. The expected number of visits was approximately 4 per month or a total of 24. It seems logical that any set of household data based only on 10 or fewer visits (i.e., less than 40 percent of anticipated visits) would not give accurate enough information about what happened during the six-month flow period. An important dimension of this issue is when these visits were made. If two visits were made each month over the period, 10 visits might be enough.

To examine the issue of the distribution of visits, it was decided to use the consumption questionnaire as an indicator of whether the enumerator was still visiting the household in the later months of the survey. (Since there was no visit number on the output data (card07), it is impossible to know in which month this information was obtained.) The following logic was applied to develop this test. In order to get a full set of both crop input and output data on a relatively short (1-2 week) recall basis, the household must be visited after planting, weeding, and harvesting activities have taken place. The general harvest period for maize begins as early as April, but extends well into June and July. Based on this information, it was decided that a compromise decision was required in order to not reject the possibility of an early harvest and hence that it was considered valid for data collection to have stopped in May. It was therefore decided that if at least one visit was made to the household in May, then the case is acceptable. It was considered extreme to require at least one visit in June. Thus May was chosen as the tentative cutoff month.

3. SPSS Procedure and Results

a) Leave in the sample only the households that received at least 11 visits during the flow survey period. The variable maximum number of visits was obtained by aggregating the number of visits as measured on Fcard01. The value of maximum number of visits for each household has to equal the number of count or the total number of cases that were collapsed into one value for each household in Fcard01 during aggregation. The resulting file had 336 households.

b) Leave in the sample only the households that received at least 11 visits and also participated in the survey with at least one visit in May 1986. The resulting file had 319 households. The 319 households are the maize subsample composed of households that met all the requirements considered so far.

4. Implications of the Results

Over 10 percent (42 households) of the subsample of maize growers from step (v) above (361) are removed if we apply this more restrictive test of the timing of household visits. If the month of June were used, only 264 households would remain in the subsample, and only 154 would remain if July were used. This presents an informative picture of the cumulative effect of the survey design decision (stop collecting data when the harvest is over) and the survey management/supervision problems. The existing documentation indicates that as the flow survey period progressed, it became more and more difficult for the project to sustain the financial and human resources to adequately supervise the survey. One of the most critical concerns this raises is whether the output (production) data collected really represents what the farmers reported after their harvest was completed, or whether the data obtained was an estimate either by the farmer or the enumerator at some point during the growing season. For the data from the consumption and expenditure part of the survey it is clear that the months of May, June and July are significantly undersampled.

H. DATA CLEANING STEP VII (Consistency: Bench Mark and Flow)

This step involves data from both the bench mark and the flow survey.

1. Action

This analysis aimed at finding out whether all the 477 households in the flow survey were indeed selected from the bench mark sample (911 households). A second question is how many of the 319 maize subsample from the flow survey can be matched with valid cases of data from the bench mark survey.

In this analysis a bench mark file which had all the 911 households was join matched by household IDs with a flow file which had all the 477 households. If any of the 477 households did not match, it means that that household was not in the bench mark.

2. Logic

It is important to ensure that all the households in the flow sample also appear in the bench mark. Otherwise it would not be possible to make comparisons on certain variables between the flow and the bench mark. Furthermore, once work begins on the maize subsample from the flow, it may be important to compare household behavior from the bench mark period.

3. SPSS Procedure Used

The join match command was used to match the households in the flow file containing 477 households with the households in the bench mark file containing the 911 households. Then those households which do not match, that is, whose identifying variables do not appear in both the bench mark and the flow were dropped. The result was that all 477 households in the flow also had identifying variables in the bench mark. In other words, all the 477 households matched.

4. Implication of the Results

The results show that all the 477 households in the flow were drawn from the bench mark. Therefore it is analytically possible to compare variables between the flow and the bench mark samples. However, when the 319 households in the maize subsample from the flow were matched with their counterparts from the bench mark, there were only 251 valid data (yield) cases. The remaining 68 households had either missing data (yield) or had zeros as values for the data in the bench mark. This limits the amount of actual comparison of input and output information in the two surveys.

I. DATA ANALYSIS STEP 1 (Production)

This step begins the comparative analysis of selected data in the subsample of households producing maize.

1. Action

Here we begin to examine basic production characteristics of households growing maize.

2. Logic

Given the concern we have about the accuracy and survey implementation problems in measuring output in the flow survey, it is important to begin this analysis by examining the basic production information that has been obtained. One useful step is to compare these results with what is generally known about smallholder agriculture in Zambia. One important issue that has to be discussed in conjunction with our review here of the survey results is that of sample weights. Recall that a "quick listing" of farm size of households in the selected standard enumeration areas (SEAs) was completed. This was part of the process of developing a basis for selection of the sample households according to the three farm size categories used in the study. The results of the "quick listing" reported in the report, RHS Volume I, indicated the following general distribution of farms in the study areas.

Category		Percent of farms
I	< 2 hectares	76.2%
II	2.0-9.99 hectares	22%
III	10.0-39.99 hectares	1.8%

In the actual sample drawn for the flow survey a different sampling rate was applied to each of these farm size categories. However, when results for the three categories are aggregated to obtain sample-wide results, different weights must be used. Part of the logic of this first data analysis is reviewing the effects of using the weights developed for this survey.

3. SPSS Procedures Used

Compute basic counts and descriptive statistics as follows:

a) Percentage of farmers reporting growing maize = 394/434 x 100 = 90.8%; 434 and 394 were obtained in step (iv) and step (v) respectively in the cleaning process.

b) Percentage of households sole cropping maize = 267/319 x 100 = 83.7%.

c) Percentage of households intercropping maize = 51/319 x 100 = 16.3%.

d) Get household annual maize production summary statistics (see Appendix Table 5 for details).

The results were:

	unweighted	weighted
Mean	= 3114 kg	2618 kg
Median	= 1350 kg	1350 kg
Mode	= 450 kg	450 kg
Minimum	= 10 kg	10 kg
Maximum	= 54000 kg	54000 kg
Range	= 53990 kg	53999 kg
Skewness	= 4.57	4.58
Kurtosis	= 32.80	34.96
Standard deviation	= 5170	4232
Percentile (25 50 75)	= 450, 1350, 3600	450, 1350, 2880

Table 1 presents the unweighted maize results by a zone and farm size breakdown.

4. Implications of the Results

The percentage of households growing maize generally conforms to the widely held perception that a vast majority of smallholders grow maize. The proportion of the sample growing maize under sole crop conditions appears quite high for smallholders. Since the flow survey design did not collect data at the household-crop-field level, farmers were asked during the survey to summarize in their minds over all of the various different fields of maize they grew. Since it is common for smallholders to have more than one field of maize under cultivation on their farms, it seems possible that farmers may have reported data only about the field that was sole cropped. Further analysis of this issue is clearly important.

The sample estimates of household annual maize production are revealing of potential data problems. The sample unweighted (3114 kg) and weighted mean of 2618 kg per household suggests that the average smallholder produces more than enough maize for household consumption. But the sample median is only 1350 kg per household, and the sample mode, or most frequently occurring value, is 450 kg per household. The mean, median and mode are all types of averages. When sample data are normally distributed these three averages tend to be equal. When data are positively skewed, as in the case of our data, the mean will be larger than both the mode and the median.

Table 1. **Household Average Annual Maize Production by Zone and Farm Size** (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max	Percentile 25	Percentile 50	Percentile 75
Zone 1 (Northern High Rainfall)	117	2791	900	90	5954	10	54000	330	900	2700
Zone 2 (Western Semi-Arid)	64	1381	450	180	1919	60	7920	180	450	1755
Zone 3 (Central, Southern & Eastern Plateaux)	138	4192	2250	450	5262	15	29250	788	2250	5400
Farm Size 1 (<2 ha)	104	1468	540	450	2203	10	14760	203	540	1868
Farm Size 2 (2-9.99 ha)	187	3387	15530	900	4779	12	29250	486	1530	4140
Farm Size 3 (10-39.9 ha)	28	7406	4275	1350	10438	90	54000	2183	4275	8888

The skewness of 4.6 means that the data are positively skewed. This means that most of the farmers in the subsample had low to medium production values and that relatively few farmers had higher values.

The kurtosis (33) indicates that the data are leptokurtic or peaked. In other words, the data cluster around certain values. The value for both skewness and kurtosis is close to zero in a normal distribution. Thus, the distribution of household production is different from the normal distribution.

The standard deviation is a useful measure of likely dispersion of the estimates from our sample. In repeated sampling from a normal distribution, approximately 68 percent of the time values will fall within one standard deviation of the population mean; and approximately 95 percent of the time will fall within two standard deviations of the mean. This does not apply to skewed data like ours. Thus if this data were normally distributed, the standard deviation of 4231 kg would mean that if repeated samples were selected from the smallholder population under the same conditions 95 percent of the time (which is the confidence level employed) the production values would fall within 8462 kg from the mean of 2618.

The standard deviation (4231 kg) in this case is very large considering that the mean value is 2618 kg. An important question is whether this large degree of variation in the mean amount of production is indeed a reflection of the reality of farmers in Zambia, or whether the survey procedures and resulting data obtained are inadquate. The significance of such large standard deviations will be discussed further in later sections.

The percentile values mean that 25 percent of the farmers in the subsample had estimated annual household maize production below or equal to 450 kg, that 50 percent had production below or equal to 1350 kg and that 75 percent had production below or equal to 2800 kg.

The wide dispersion of the sample values is a serious concern when asking whether these are plausible results. For example, an important question is whether it is not likely that 25% of the population of smallholders produce 450 kgs or less of maize per year, or that 50% of the population produce less than 1350 kg per year. Depending on the number of household members, these levels of production may not even provide sufficient maize for the household's own consumption needs, let alone leave much for sale to the market.

Before going further in the discussion of these results, it is important to review the effect of using unweighted and weighted data. A second question is how did the data cleaning process of eliminating cases affect the distribution of the sample across the farm size categories.

Table 2 presents information on the number of cases (percentages) in the sample before and after data cleaning. It shows the number of cases (percentages) after the sample weights are applied (see Table 5 in the Annex which show the distribution of the 319 subsample breakdown by size strata and region). Clearly the process of cleaning the data by eliminating cases, but not changing the weights, affects the estimate of the total number of households in the nation. The weighted 477 sample expands with the sample weights to represent 592,852 households, whereas the 319 sample expands to represent only 390,434 households. However,

Table 2. Sample Breakdown by Farm Size Strata: Unweighted and Weighted

Farm Size Category	477 Sample Unweighted Number of Cases (Percentage)	477 Sample Weighted Number of Cases (Percentage)	319 Sample Unweighted Number of Cases (Percentage)	319 Sample Weighted Number of Cases (Percentage)
I. (<2 ha)	178 (37.3%)	336,693 (56.8%)	104 (32.6%)	202.9 (52%)
II. (2.0-9.9 ha)	260 (54.5%)	241,381 (40.7%)	187 (58.6%)	176,684 (45%)
III. (10-39.99 ha)	39 (8.2%)	14,778 (2.5%)	28 (8.8%)	10,862 (2.8%)
TOTAL	477 (100%)	592,852 (100%)	319 (100%)	390,434 (100%)

the basic percentage breakdown among the farm size strata is not altered significantly in the cleaning process. For example, the 319 sample has around 33 percent of the cases in strata I (the 477 sample has 37%). And the weighted 319 sample has roughly the same percentages in each size strata as the weighted 477 sample.

A major problem with the weighted results from both the 477 and the 319 samples is that they do not match the results from the quick listing, which was carried out to establish the weights in the first phase of the sruvey. For example, the quick listing found that 76 percent of the households were in farm size strata I, whereas both the 477 and the 319 weighted samples show some 51-57 percent in this category. Hence a major question mark must be placed beside the weights. Clearly, if work is to be done beyond the present study, the weights for the 319 subsample need to be reestimated, presumably to reflect more the results obtained in the quick listing, and to adjust for the reduced sample size.

Much of the analysis in the remainder of this report will not use the sample weights. Obviously sample-wide statistics from the unweighted data are misleading. Breakdowns by region and farm strata are, however, much more meaningful, especially if the relative proportions from the quick listing are kept in mind.

The results of the analysis of the household production data from a zone and farm size distribution perspective were presented in Table 1. These unweighted results reveal some interesting aspects, among these are:

a) the order of mean maize production (but not the absolute levels of production) with respect to zones is in line with what one would expect in Zambia. That is, the Central, Southern and Eastern Plateaux or zone 3 is likely to have the highest production level, followed by the Northern High Rainfall area or zone 1, and the Western Semi-Arid area zone 2 coming last with the lowest relative production level.

b) the order of average maize production with respect to farm size also seems to agree with what one would expect, namely that on a relative basis the smaller the farm, the lower the production levels.

c) the percentile distributions again reveal some relatively low levels of household maize production, especially for the smallest farm size strata. For example, 50 percent of the sample in strata I did not produce more than 540 kgs per year.

J. DATA ANALYSIS STEP 2 (Yield)

To continue the analysis of the maize data, a yield variable was created for each household. In general, maize yield should not vary nearly as much across households as did total production.

1. Action

Compute maize yield for each household and analyze by region and farm size, with the full subsample and a reduced or trimmed version of the subsample that eliminates the apparent extreme values. Also construct a graph of the sample yield values according to the area planted per household to maize.

2. Logic

It is useful to examine yield in a number of various dimensions:

a) Overall: A general look at yield is necessary in that it gives information concerning the efficiency of the use of resources, particularly land, that one would not get by looking at production alone. Production figures alone do not tell whether an adequate amount was harvested on a given land area, and therefore one would not be able to make even general comments about resource use.

b) Zone: Analysis of yield with respect to zone will show whether the zone factor has any likely influence on yield. In general, zone is an attempt to create more homogeneous climate and soil groups, and it is generally expected that these should influence yield.

c) Farm size: Analysis of yield over farm size should give a general indication of technology used, and its effect on yield. In general in Zambia the large small holders are thought to be heavier users of improved inputs, and most likely will have higher yields.

d) Sole crop and intercrop: A useful check on data quality is to look at sole crop and intercrop yields separately because it is easier to justify a very low yield level for intercrop since with intercropping the farmer may have had only a few stalks of maize per unit of area.

e) Graph of area planted by yield: This is a beginning step in analyzing variability in yield. The plot will show the range of farm size over which variability is high. This may help shape questions as to why variability levels differ across the area planted.

f) Exclusion of extreme yield values is a possible step to examine the effect of likely poor survey results. A large number of extreme values biases the results and hence the conclusions.

3. SPSS Steps and Results

a) For the summary statistics on yield, the same SPSS procedures used for production in section H are used. The only difference is that here we use "variable = maize yield" instead of "variable = maize production."

b) Graph maize yield with area planted. The results are shown in Figure 2.

c) The summary statistics on yield are presented below:

Maize yield (see Appendix Table 6 for details) 319 households.

	unweighted	weighted
Mean	= 1833 kg/ha	1752 kg/ha
Median	= 1111 kg/ha	1112 kg/ha
Mode	= 360 kg/ha	360 kg/ha
Minimum	= 9 kg/ha	9 kg/ha
Maximum	= 32400 kg/ha	32400 kg/ha
Range	= 32391 kg/ha	32391 kg/ha
Skewness	= 6.8	6.8
Kurtosis	= 62	63.5
Standard deviation	= 2835.38	2714
Percentile (25 50 75)	= 450, 1112, 2224	450, 1112, 2155

d) The results of the unweighted analysis of yields with respect to all the dimensions discussed above appear in Tables 3 to 8.

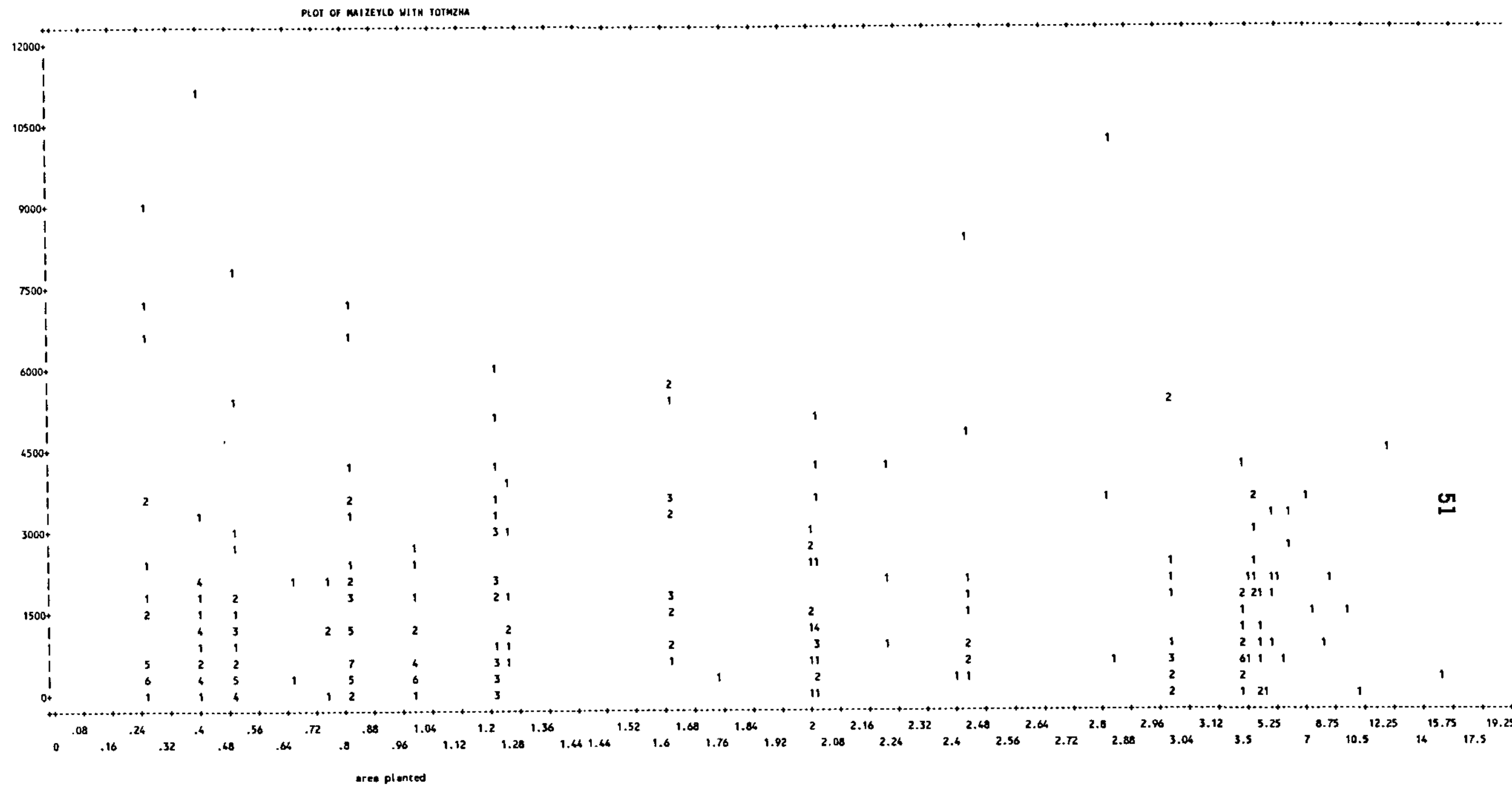

Figure 2. Maize Yield by Area Planted (Sole Cropped)

Table 3. **Household Annual Average Maize Yield (kg/ha) by Zone and Farm Size** (319 households) (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max	Percentile 25	50	75
Zone 1 (Northern High Rainfall)	117	1841	1112	360	3286	9	32400	360	1122	318
Zone 2 (Western Semi-Arid)	64	1131	720	360	1204	30	6480	360	720	1611
Zone 3 (Central, Southern & Eastern Plateaux)	138	2153	1390	556	2919	10	27242	653	1390	2565
Farm Size 1 (to <2 ha)	104	1368	774	360	1577	36	7920	360	774	1932
Farm Size 2 (2-9.99 ha)	187	2015	1112	1112	3387	10	32400	508	1112	2250
Farm Size 3 (10-39.9 ha)	28	2350	1899	3336	2215	9	1112	826	1899	3291

Table 4. Average Household Maize Yield by Zone and Farm Size for Yield between 500 kg/ha and 5000 kg/ha (212 households) (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max
Zone 1 (Northern High Rainfall)	70	1922	1765	1112	1096	500	4500
Zone 2 (Western Semi-Arid)	38	1431	1334	556	812	500	3335
Zone 3 (Central, Southern & Eastern Plateaux)	104	1788	1612	556	1085	508	4744
Farm Size 1 (<2 ha)	61	1612	1440	556	915	500	3706
Farm Size 2 (2-9.99 ha)	128	1779	1483	1112	1080	500	4744
Farm Size 3 (10-39.9 ha)	23	2122	1927	3336	1204	532	4500

Table 5. Average Household Maize Yield by Zone and Farm Size for Sole Cropped Maize Only (267 households) (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max
Zone 1 (Northern High Rainfall)	97	1583	1112	360	1666	9	9000
Zone 2 (Western Semi-Arid)	47	1177	720	360	1290	30	6480
Zone 3 (Central, Southern & Eastern Plateaux)	123	2271	1440	556	3059	10	27242
Farm Size 1 (<2 ha)	84	1462	890	360	1668	36	7920
Farm Size 2 (2-9.99 ha)	157	1929	1170	2216	2716	10	27242
Farm Size 3 (10-39.9 ha)	26	2403	1899	2291	2271	9	11119

Table 6. Average Household Maize Yield by Zone and Farm Size for Intercropped Maize (52 households) (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max
Zone 1 (Northern High Rainfall)	20	3090	1223	90	7065	90	32400
Zone 2 (Western Semi-Arid)	17	1006	720	720	949	90	3336
Zone 3 (Central, Southern & Eastern Plateaux)	15	1185	845	222	821	222	2520
Farm Size 1 (<2 ha)	20	972	490	90	1008	90	2891
Farm Size 2 (2-9.99 ha)	30	2463	979	278	5799	120	32400
Farm Size 3 (10-39.9 ha)	2	1676	1676	1350	460	1350	2002

Table 7. Average Household Maize Yield by Zone and Farm Size for Sole Cropped Maize for Yield between 500 and 5000 kg/ha (181 households) (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max
Zone 1 (Northern High Rainfall)	61	1853	1620	1112	1099	500	4500
Zone 2 (Western Semi-Arid)	27	1437	1334	556	769	556	3336
Zone 3 (Central, Southern & Eastern Plateaux)	93	1823	1668	556	1117	508	4744
Farm Size 1 (<2 ha)	51	1600	1440	556	906	508	3706
Farm Size 2 (2-9.99 ha)	109	1783	1440	1112	1094	500	4744
Farm Size 3 (10-39.9 ha)	21	2164	1927	3336	1250	532	4500

Table 8. Average Household Maize Yield by Zone and Farm Size for Intercropped Maize for Yield between 500 and 5000 kg/ha (31 households) (unweighted)

Zone and farm size	# cases	mean yield	median yield	mode yield	std dev	min	max
Zone 1 (Northern High Rainfall)	9	2389	2046	1112	1011	1112	4500
Zone 2 (Western Semi-Arid)	11	1416	1350	720	950	500	3336
Zone 3 (Central, Southern & Eastern Plateaux)	11	1493	1350	556	741	556	2520
Farm Size 1 (<2 ha)	10	1674	1342	2880	1010	500	2891
Farm Size 2 (2-9.99 ha)	19	1758	1800	741	1024	5556	4500
Farm Size 3 (10-39.9 ha)	2	1676	1676	1350	461	2001	3352

4. Implications of the Results

The weighted and unweighted national average maize yield are relatively similar, and are within the range of the average maize yield for Zambia for the years 1982-85 (see pages 21-22). Hence these results seem plausible when compared to results of other studies.

Since the weighted and unweighted results are quite similar, and the weights are in doubt, no further analysis of sample weighted yield will be done.

The very large standard deviation is problematical, and is most likely related to measurement error rather than such large differences among farmers. The analysis of variance to be conducted later will indicate whether such large degrees of variance in the data will permit conclusions that yields across regions and farm size strata are statistically different.

To the extent that these are statistically different, the general pattern of maize yield across zones appears to agree with what one would expect in Zambia, namely highest yield level in Zone 3 (the Central, Southern, and Eastern Plateaux) followed by Zone 1 (the Northern High Rainfall) and the lowest yields in Zone 2 (Western Semi-Arid).

The pattern of maize yield across farm size also seems to be in line with what one would expect, namely that yield increases with increase in farm size. One reason for this would be that as the farm size grows, the farmer tends to employ some technology that would enhance yield, such as use of herbicides and pesticides, fertilizer, and animal traction.

Exclusion of the extreme yield values, that is considering only yield values between 500 kg/ha and 5000 kg/ha, improves the mean yield values for zones 1 and 2, but deteriorates it for zone 3, making it lower than that of Zone 1. A plausible explanation is that the extreme yield values that had appeared under zone 1 and zone 2 before adjustment were really outliers and possibly errors, while those which had appeared under zone 3 (the best production area of the country) were mostly valid values for this zone.

The results in Tables 5 and 6 are very inconclusive, in that yields are sometimes higher for intercropped maize than for sole cropped (see zone 1 and farm strata size 2). Most likely the intercropped data are not correct, and these estimates are also affected by fewer observations, as compared to the sole crop case.

The graph of yield by actual area planted reveals high variability in yield throughout, although it does appear to decrease with increases in area planted. In other words, there is more variability in yield levels for smaller farms than for bigger farms. Causes of overall high, but decreasing variability could be many and varied. The common one would be poor data in general, and proportionally larger error in measuring outputs on smaller farms, especially in the case of this survey where there were no data collected at the field level. This means that farmers would have their own criteria, different from each other, for choosing the fields to report to the enumerator and for summing production values. Obviously valid causes of variability would be differences in types of inputs used and/or differences in the efficiency of input use. Yet other valid causes would be geographical and climatic conditions.

K. DATA ANALYSIS STEP 3 (ANOVA)

1. Action

In this step, analysis of variance (ANOVA) is used to examine mean household yield of the maize sample, first for all the yield values reported and then for the yield values between 500 and 5000 kg/ha. In the second case, apparent extreme values (with respect to the likely Zambian situation) are dropped. To obtain maximum degrees of freedom, and to obtain more precise comparisons, the analysis focuses on sole crop maize only.

The critical significance level employed in hypothesis testing is 0.05 or 5% probability.

2. Logic

ANOVA provides a statistical tool for determining whether the mean maize yield values are significantly different (statistically speaking) given the large variance in the data.

In addition, ANOVA results may help show the relative importance of various factors in explaining the variation in maize yield, suggesting which factors to emphasize in more elaborate analysis.

3. SPSS Procedures Used

a) ANOVA for zone by farm size (all three zones, all three farm sizes), all reported yield values in the 319 household sample.

Summary of results:

Cell Means

Farm Size	Zone 1	Zone 2	Zone 3
1	1128.16 (27)	1036.38 (21)	1960.86 (36)
2	1780.68 (62)	1177.78 (24)	2313.16 (71)
3	1587.34 (8)	2635.40 (2)	2781.61 (16)

NB The numbers in parentheses refer to number of cases.

*** ANALYSIS OF VARIANCE ***

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig. of F
Main Effects	64746918	4	16186730	2.852	.024
Farm Size	14848407	2	7424204	1.308	.272
Zone	43293506	2	21646753	3.814	.023
2-way Interactions					
Size	5590598	4	1397650	.246	.912
Zone	5590598	4	1397650	.246	.912
Explained	70337516	8	8792190	1.549	.141
Residual	1464460101	258	5676202		
Total	1534797617	266	5769916		

b) Given the large variation in the data, the ANOVA for zone by farm size will be completed using only households with maize yield values between 500 and 5000 kg/ha for all zones and all farm sizes. Below are the results of this analysis.

Cell Means

		Zone 1	Zone 2	Zone 3
Farm Size	1	1737.93 (16)	1244.82 (10)	1653.90 (25)
	2	1828.38 (40)	1404.68 (15)	1854.20 (54)
	3	2414.24 (5)	2635.40 (2)	2007.49 (14)

NB The numbers in parentheses refer to number of cases.

*** ANALYSIS OF VARIANCE ***

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig. of F
Main Effects	7864217	4	1966054	1.723	.14
Farm Size	4186786	2	2093393	1.835	.16
Zone	3116547	2	1558273	3.814	.25
2-way Interactions					
Size	2125588	4	531397	.466	.76
Zone	2125588	4	531397	.466	.76
Explained	9989805	8	1248726	1.094	.36
Residual	196250457	172	1140991		
Total	206240262	180	1145779		

4. Implications of Results

The basic hypothesis of interest here is whether farm size affects yield, and also whether zone affects yield. Likewise the question is whether there is an interaction between the effects of zone and farm size on yield.

Using all cases of data (319) the calculated F ratio for the interaction effect is .246. The observed significance level is the probability of obtaining an F statistic (ratio) at least as large as the one calculated when all population means are equal. If this probability is small enough,the hypothesis that all means are equal is rejected. In this case the observed significance level is quite high (.912), meaning that we cannot reject and hence must accept the hypothesis that there is no interaction effect.

Given that there are no interaction effects (i.e. zone and farm size do not jointly affect yield), it is useful to continue hypothesis testing about the main effects. The calculated F for the zone effect is 3.814. The observed significance level is quite low (.023) meaning that we can reject the hypothesis of no zone affect on yield.

The calculated F for farm size is 1.308. The observed significance level is somewhat high (.272) hence we must accept the hypothesis that there is no statistically significant effect of farm size on yield.

Overall the large variance in the yield data appear to make it difficult to draw statistically significant conclusions from the data, except for the most obvious effects on yield of agroecological zone factors.

Working with households that have yield levels only between 500 and 5000 kg/ha, the calculated F for the interaction effect turns out to be 0.466 and the observed significance level 0.76. This level is again too high given our critical significance level of 0.05. Therefore we once again accept the hypothesis that there is no interaction effect.

Regarding the main effects, the calculated F for the zone effect is now 1.366 while the observed significance level turns out to be 0.25. The calculated F for farm size is 1.835 and the observed significance level is 0.16. Both 0.25 and 0.16 are relatively high levels and therefore we accept the hypothesis that there is no relation between maize yield and farm size or zones.

There are obviously factors other than zone and farm size that would be included in a more complete model to examine factors affecting yield. However this simple ANOVA test indicates that the large variance in the yield data may make it difficult to obtain high levels of statistical significance in more refined models.

L. DATA ANALYSIS STEP 4 (Household Food Production Levels)

1. Action

This step involved analyzing household per capita production levels for cereals and cassava (main staple foods), other food and cash crops. This includes an examination of the distribution of these results across the sample, and between two selected Provinces (Southern Province and North Western Province). Selected issues of food/cash crop mix are also examined.

2. Logic

The analysis of food production was prompted by the appearance of very low household annual maize production figures (10 kg - 500 kg) in the 319 household sample. To begin to examine if these levels of production were plausible, it is first necessary to calculate household annual per capita values in order to account for differences in family sizes (more complete analysis should use some measure of adult equilivents, but these were not possible in the present study).

The analysis of crop mix is done to shed some light on the potential of the households to supplement own-produced food. For instance, a cash crop would be a potential source of cash income that the household could use to buy food, if the data indicate a very low level of own household maize and/or other cereal production.

A look at two provinces, one of them considered to be one of the most developed provinces agriculturally and the other one as one of the least developed provinces would be another check on the validity of the food production. That is, it is generally expected that higher household per capita food production values would be found for households in the Southern Province than for those in the North Western Province.

3. SPSS Procedures Used and Results

a) Create a variable called total household staple food production by summing up all quantities produced of cereals (maize, millet, sorghum and rice) and cassava. The rest of the principal food crops (dry beans and groundnuts) produced by the household are then called other foods. Thus we potentially have three groups of crops produced by each household: (i) staple foods (ii) other foods, and (iii) cash crops. Only the staple food group (cereals plus cassava) is measured in total quantity produced (in kgs). The other crops are only identified as to whether they are grown or not by the household. As such these are only rough indicators of household production patterns. Likewise, the combining of cereals and cassave is recognized as a rough indicator, especially since the nutritional content of cereals and cassava is so different.

b) Compute household percapita staple food production by dividing the family size variable into the total household staple food production variable.

c) Compute household per capita maize production by dividing the family size variable into total household maize production.

d) Count other food crops and cash crops produced by each household.

e) Divide the sample into household per capita stapel food production quintiles and compute mean levels of production of staple foods for each quintile.

f) Compute household mean levels of per capita production of maize for each quintile defined above.

g) Compute average farm size for each quintile.

h) Count the percentage of households in each quintile that are from Northwestern Province and from Southern Province. Also break down the total number of households in each quintile by the 3 farm size strata.

i) Create a new variable that measures the number of households who grew maize and/or groundnuts, and cross tabulate these variables with cash crops grown, by the various staple crop quintiles.

j) Experiment with the effect of weighting the above variables with the sample weights.

k) Summary of results:

(i) Table 9 shows the unweighted and weighted percentage of households found to be growing the 10 principal crops that appeared in the survey results.

(ii) Table 10 shows the household per capita levels of staple food production, disaggregated into quintiles. A number of other indicators of household behavior are also shown in this table for the same staple food production quintiles. These are all unweighted sample statistics.

(iii) Table 11 presents the same information as Table 9, however estimates are weighted according to the sample weights.

Table 9. Percentage of Households Growing Selected Crops; Using 319 Household Sample

Crop	(Unweighted) % Growing	(Weighted) % Growing
	-Percent of Households-	
Maize	100.0	100.0
Groundnuts	54.2	57.9
Millet	36.7	33.0
Beans	34.8	32.4
Sunflower	13.2	11.6
Sorghum	10.3	9.5
Rice	7.8	5.9
Cassava	7.5	4.5
Soybean	5.3	5.4
Cotton	5.0	4.5

Table 10. Household Per Capita Staple* Food Production, Food Crop Mix and Other Information by Quintiles (unweighted data)

Mean Per Capita Staple Food Production Quintiles	# Cases	Household Mean Per Capita Production (kg): Staple Food	Household Mean Per Capita Production (kg): Maize	Mean Farm Size (ha)	Percentage of Households from: North W. Province	Southern Province	Farm Size Strata 1	2	3
1	64	52	41	2.3	33	4	48	50	2
2	65	171	125	3.7	33	20	34	57	9
3	66	355	280	3.9	33	15	35	58	7
4	61	655	527	5.3	0	33	20	67	13
5	63	2383	1888	5.1	0	28	25	62	13

*Staple Food: Maize, millet, sorghum, rice and cassava.
Source: Computed from the data

Table 10. (continued)

Mean Per Capita Staple Food Production Quintile	Average Number of Food Crops Grown	Percentage of Households Growing: 1 Food Crop	2 Food Crops	3 Food Crops	4 Food Crops	5 Food Crops	Maize, No Cash Crops	Maize, G.Nut, No Cash Crops	Maize & Cash Crops	Maize, G. Nuts & Cash Crops
1	2.13	22	52	19	8	0	47	44	3	5
2	2.58	12	37	32	17	1	33	40	14	13
3	2.42	12	45	32	09	1	35	45	8	12
4	2.61	15	34	28	21	2	42	34	2	24
5	2.84	13	27	33	17	9	42	42	3	14

*Staple Food: Maize, millet, sorghum, rice and cassava.
Source: Computed from the data

Table 11. Household Per Capita Staple* Food Production, Food Crop Mix and Other Information by Quintiles (weighted)

Mean Per Capita Staple Food Production Quintiles	% Cases	Household Mean Per Capita Production (kg): Staple Foods	Household Mean Per Capita Production (kg): Maize	Mean Farm Size (ha)	Percentage of Households from: North W. Province	Southern Province	Farm Size Strata 1	2	3
1	21	44	34	1.7	26	5	67	32	1
2	20	168	137	2.7	26	19	54	43	3
3	19	328	272	3.1	48	12	49	48	3
4	20	568	446	3.3	0	27	43	52	4
5	20	2134	1591	3.3	0	36	45	51	3

*Staple Food: Maize, millet, sorghum, rice and cassava.
Source: Computed from the data

Table 11. (continued)

Mean Per Capita Staple Food Production Quintile	Average Number of Food Crops Grown	Percentage of Households Growing: 1 Food Crop	2 Food Crops	3 Food Crops	4 Food Crops	5 Food Crops	Maize, No Cash Crops	Maize, G.Nut, No Cash Crops	Maize & Cash Crops	Maize, G. Nuts & Cash Crops
1	2.15	20	52	20	8	0	42	52	2	4
2	2.32	13	52	24	10	0.003	26	47	16	11
3	2.45	17	39	28	14	0.02	40	47	17	11
4	2.47	14	43	27	14	0.02	44	37	1	18
5	2.80	07	35	35	14	8	35	46	2	17

*Staple Food: Maize, millet, sorghum, rice and cassava.
Source: Computed from the data

4. Implications of The Results

Results in Table 9 indicate relatively little effect of using the sample weights in the simple estimates of the percentage of households growing the 10 most important crops identified in the survey. After maize, groundnuts are grown by slightly over 50 percent of the smallholders. Next in importance is millet and beans (grown by about 35% of the smallholders), with the remainder of the food and cash crops being relatively unimportant (less than 12% of households grow these). This percentage (12%) seems to be too low considering the number of crops involved and the fact that cassava, which is an important food crop in some parts of the country is among them. It should be noted that households growing and marketing maize do consider this a cash crop.

Turning to results in Tables 10 and 11, a more comprehensive picture of the combination of crops grown on smallholders farms emerges. To begin this discussion, weighting the sample results does have a significant effect in almost all the variables estimated in the Table, so the majority of this section will concentrate on the weighted results in Table 11.

Using the standard FAO rule of thumb of annual cereals requirements of 200 kilos per person, some 40 percent (the bottom 2 quintiles) of the estimated smallholder population do not produce enough staple food to meet their own consumption requirements. Adding cassava to the definition of staple foods complicates this analysis somewhat, although from Table 9 it was shown that only about 4 percent of the households produce cassava. The importance of maize in meeting the basic food needs of the rural households can be seen by comparing the estimates of mean household per capita staple foods and maize. In all cases, maize constitutes more than 75 percent of the total. Households in the lowest quintile produce only 41 kgs per capita of maize, while those in the 2nd quintile produce an average 125 kgs per person. Smallholders in the upper two quintiles (upper 40%) do appear to produce enough maize and other food crops per person in their respective households to generate a marketable surplus. Households in the top quintile likely dominate smallholder maize sales.

An important question is whether these results are plausible and in agreement with other studies and well known characteristics of Zambian smallholders. A large portion of the smallholder population is estimated to not be producing enough food for their own requirements, and this is certainly possible. Recent studies in Malawi and Zimbabwe (Quinn, et al. and Rohrbach) show similar conditions. However, conventional wisdom in Zambia will likely question these relatively low household production estimates.

The mean farms size estimates for each quintile appear to be consistent with general logic, in that the relatively smaller farms, on average, appear to be the ones producing the least staple foods on a household per capita basis. There is also consistency in the fact that most of the households in the Northwest Province appear in the lower quintiles, while more of the households from the Southern Province appear in the upper quintiles. There is likewise some consistency in the distribution of the expanded results across the farm size strata categories. Some 67 percent of the households in the first quintile who are producing only an estimated 44 kgs per person of staple foods are in the smallest farm size strata.

What is somewhat questionable in these results is the fact that so many small and large farms are spread across the entire range of the mean per capita production quintiles. In fact, quintiles 2 through 5 look relatively similar in the distribution of farms across the farm size strata. Either there is a very wide range in output productivity across farms of varying size, including larger as well as smaller farms, or the measurement error was very high and somewhat randomly spread across the sample. Another way to see this problem is to analyze the household mean per capita staple and maize production data by farm size quintiles. These are shown in Table 12 and indicate the very wide range in estimates for a given farm size, since there is so much difference between the mean per capita estimates based on farm size, as compared to those based on household per capita quintiles (range of 34 to 1591 kgs of maize per person per household, when examined from the production quintile perspective, as compared to 305 to 750 kgs per person per household across the farm size quintiles).

In the bottom part of Table 11, a picture of the identified crop mix is given for the various households. The range of the number of crops per farm ranges from 2.15 to 2.8, with the highest production quintile farms growing the most food as well as cash crops. In general, the number of crops grown by smallholders appears relatively small, especially for the smaller of the group of smallholders. Maize and groundnuts are the most frequent crop combination, and only 4 to 18

Table 12. **Household Per Capita Mean Staple Food and Maize Production by Farm Size Quintiles** (weighted data)

Farm Size Quintiles	Household Mean Per Capita Staple Food Production (kgs)	Household Mean Per Capita Maize Production (kgs)
1	594	305
2	404	322
3	657	496
4	698	616
5	837	750

percent of the smallholders across the country combine at least one cash crop with the maize and groundnut cropping pattern. In the lowest per capita quintile of staple food production, only 6 percent of the households produce a cash crop. Hence, assuming the data are correct, food deficits for the majority of households in this quintile must be made up by reduced consumption, and/or by labor sales or other forms of non-farm income.

Overall these results suggest a picture of considerable food supply problems for 20 to 40 percent of the smallholders in Zambia. If accurate, such results have significant policy impolications. Alternatively, these relatively low household food production estimates are an indicator that the ZATPID I survey was not adequately designed and implemented to measure household food production. The only feasible way to resolve this uncertainty is to compare the internal consistency of these results with those of other surveys, and/or with portions of the ZATPID I data set that attempted to measure household consumption, expenditures and income.

CHAPTER IV

SUMMARY AND RECOMMENDATIONS

This chapter summarizes the results of the analysis and outlines recommendations for the future use of the Rural Household Survey (RHS) data.

A. SUMMARY OF THE FINDINGS

1. Survey Design Issues

The issues pertaining to the design of the survey include:

a) Crop output measurement. The absence of careful procedures to measure output, especially to measure output at the household-crop-field level has resulted in possible (likely) underestimation of household crop production. Because data were not collected at the household-crop-field level, the farmer had to sum in his/her own mind the production figures and other input variables related to the various crops in her/his fields. It is suspected that the farmers did not report values for all of the fields cultivated. It is most likely that some farmers would leave out some fields. In other words, omission of careful producers for measuring output at the field level leads to farmer information recall problems. It is hypothesized that the high variability of the production data has much to do with this survey design and related farmer information recall problem.

b) Related to the above measurement problem is the issue of vague instructions in the design about which crops to include in the survey (major crops only) and when to stop data collection (after harvest). It is very difficult to standardize on what is a major crop and a minor crop for different farmers, in various locations over the country. Likewise, letting the enumerators decide to stop data collection after crops were harvested removes significant standardization from the data set, and ignores the need to collect a complete set (over a given period of time) of consumption expenditures, income and labor data, in addition to crop production information.

c) The omission of information about the visit number in the file on production (harvests) makes it impossible to know the specific dates when harvesting took place and to check the procedures and implementation activities of the enumerators. This makes it difficult to verify the information on production, and to compare it to similar information from past studies or experience. This, in turn, makes it difficult to tell whether the information is plausible with respect to the Zambian smallholder farming situation in particular, and with the crop harvest calendar in general.

d) The absence of conversion factors for selected measurement units, especially for purchases and sales will make it difficult, if not impossible, for users of the RHS data to arrive at comparable results. For example, there are no standard conversion factors for going from unshelled to shelled crops. And different conclusions could be drawn by different users of the survey, depending on the conversions used.

e) The confusion and apparent inaccuracy of the data obtained from the enumerator's measurement of field areas suggest that survey design and implementation was not adequate. It should be expected that many smallholders would give imprecise information on the size of their fields and overall farm size. Yet the effort invested to have enumerators measure field and farm size produced as much inconclusive information as the data obtained from farmers directly. And imprecision in area measurements is one of the likely causes of such high variance of the yield and production data observed in the survey results.

f) The weights developed to extrapolate the sample results to the zone and national level do not appear to be correct. The cleaning process has reduced the sample size down to 319, and the existing weights appear to have been developed for the sample size of 477 households. These are obvious corrections required to update the weights. Yet the original weights for the 477 sample did not yield results similar to the "quick list." Thus more fundamental adjustments may be required in the weights, beyond simply correcting for the smaller sample size of 319.

Some analysts have also suggested that the omission of one ecological zone is problematical. However, the survey design was pragmatic in selecting this region of the country out of the survey. There was no intent to have survey results to represent the entire nation. This seems acceptable, given the resource constraints and objectives of the ZAPTID I survey.

2. Data Cleaning Issues

These issues include:

a) Instruction about "number of visits" were not followed in a relatively large number of cases; that is, few enumerators completed all the visits required. In some cases a farmer was visited only a few times during the whole survey period. In the data cleaning a substantial number of households were dropped from the sample when the criteria of a minimum of 11 visits was adopted. Due to inadequate supervision of the survey, and little timely data cleaning and consistency checking, the issue of a minimum number and timing of visits did not receive much, if any, attention.

b) Some data recording instructions were not followed by the enumerators, such as recording 999 where data were missing and zero where the actual value was supposed to be a zero. In most cases, where there was supposed to be 999 or a zero, a blank was left. This makes it difficult to discover the true situation. Because of this uncertainty, it seemed reasonable to drop out of the sample such cases of information. It is likely, however, that a number of valid cases, that is actual zero values, were eliminated in the process.

c) There was no documentation and no supervision follow-up during the survey of incredibly small and incredibly large production and other such values. For example, annual household production of maize of 10 kg or 50 kg. It is presently difficult to tell whether such values were actually so, or whether they were errors.

d) The data cleaning and analysis in the present study has focused on an important, but only a relatively small portion of the complete data set. It is highly likely that many additional data inconsistencies and other features will be discovered as the other portion of the data are cleaned and analyzed; for example, the crop input data, or the food consumption data. Also note in this regard that the still unpublished analysis completed by Nina Blid, a Ph.D. student at Iowa State University, only used 146 of the 477 households because the cleaning process indicated inconsistent and/or incomplete crop labor data. Unfortunately this Ph.D. dissertation has not been completed, nor has the draft material available documented the cleaning process that was followed. Yet the fact that such a large number of cases was dropped by an analyst trying to use specific portions of the data is not a good indicator. This is also consistent with results found in the present study.

3. Data Analysis Issues

The data analysis issues include:

a) The removal of 158 households from the original sample of 477 households, although indicative of poor data and implementation problems, did not seem to change significantly the distribution of the sample. Roughly the same percentage of the sample came from each zone and farm size strata. This suggests that a particular enumerator or supervisor was not the source of the poor data.

Rather, sample-wide design and implementation problems are suggested.

b) The analysis of the crop output data suggests that generally some 20% to 40% of the small scale farmers do not produce enough food crops for their own consumption, or that the output measurement bias is significant. The analysis also shows that the staple food deficient farmers apparently do not grow enough cash crops, which they could sell and use the proceeds to buy food to supplement their own-produced food. One method of analysis of this issue is to ask the question of whether the farmers, especially those whose total production of both food and cash crops appears inadequate for survival, have other sources of income and/or relatively low consumption. Some sections of the RHS contain such information as income, expenditure, gifts, and consumption which could be cleaned and explored to possibly gain insight into this question.

c) The analysis of the maize yield data is significantly affected by the large variance in average yields. The only statistically significant difference in yield in the ANOVA calculations resulted across zones. Farm size effects were not statistically significant. The plotting of the maize yield data seemed to indicate that yield measurement error is higher on smaller plots.

B. RECOMMENDATIONS

1. Additional analysis of the ZATPID I data. Although the present study raises a number of questions about the data, additional

analysis of the ZATPID I survey data is recommended in order to explore, among others, the issue of how households survive who do not produce what appear to be enough food crops and cash crops. For this purpose, Fcard14, Fcard15 and Fcard16 (the files with information on consumption, income, expenditure and gifts) would be of immediate use. To the maximum extent possible, this and other internal consistency analysis of the data should be done in Zambia in order to maximize the learning opportunities for Zambian analysts.

Some analysts may be tempted to conclude that the questions raised about the data in the present study are sufficient to conclude that more analysis is not warranted. However, this would ignore the potential learning opportunities that could be obtained from additional work with the data. Likewise the present study also has found a considerable number of positive features of the data, and related analysis. These seem to be internally consistent and may well represent the reality of smallholder conditions in Zambia. Such positive features of the data should not be overlooked.

A decision to stop analyzing the ZATPID I data at this point would also fail to recognize that the skills required to complete the analysis will be required in any future surveys and related analysis. Hence the ZATPID I data represent an important and timely opportunity to develop the skills and experience in Zambia that will be required to avoid design and implementation problems in future rural household surveys.

2. Efforts should be made to cross-check results of the analysis of the ZATPID I data with results from the analysis of data from similar surveys. For this purpose some new surveys might be conducted, the design of which should take into account the experience gained in the analysis of the ZATPID I Rural Household Survey. More importantly in the short-run, comparative analysis should be attempted using data from similar surveys already conducted, such as the Zambia Central Statistics Office "Comprehensive Agricultural Survey of 1987-88." This survey may in particular offer the opportunity to cross-check results from the ZATPID I survey.

3. The design of future rural household surveys in Zambia. As a summary, some of the key aspects that future rural household surveys in Zambia should take into account are as follows:

   a) The levels of data collection, processing and analyses. It is important to clearly understand the issue of level of data collection and analysis (Crawford, et al.). This would include the household-crop-field level of data so as to minimize the problems that arise from farmer information recall problems. Whenever time and resources allows, it would also be important for enumerators to measure the exact size of fields.

   b) Visit numbers and data of visits. In multiple visit surveys it is important for the design to call for enumerators to visit the household each and every time period called for, even though little data may be actually obtained. This helps assure that enumerators are actually doing their work and building a relationship of trust with the farm household. This also helps assure that adequate data is obtained to show when events do not occur, as well as when they are present.

c) Adherence to carefully designed recording instructions. This would include ensuring that the difference is clear and consistent between missing values and zero values.

d) Follow-up on data obtained. Immediate and persistent follow-up and supervision is necessary, especially to check on any suspicious data. When enumerators realize that effective supervision is present, the quality of data can be improved as a survey goes forward. In contrast, what may have happened in the ZATPID I survey was a general decline in supervision and enumerator persistence, with relatively little concern for quality data by the 4th or 5th month of the flow survey.

e) Conversion factors. More careful indication of conversion factors among the various traditional units of measure of weight and quantity are necessary to ensure accurate and consistent analysis.

f) Gaps in data. To avoid gaps in data it is important that the survey design develop approaches to questions that are inclusive; in other words the questions should cover the complete time period since the last time the same questions was asked. This would ensure that even if, for some reason, the enumerator is not able to obtain information for a given visit, the information obtained in the next visit will cover the whole survey period.

BIBLIOGRAPHY

Blid, Nina. "The Determinants of Agricultural Productivity of Small-Scale Farm Households in Zambia." Draft Ph.D. diss., Iowa State University, 1988.

Chuzu, Pia and Mwita Rukandema. 1986. The Farming Systems of North-Western Province of Zambia I: The Sorghum-based traditional Subsistence Shifting Cultivation System, Solwezi (Zambia).

Crawford, Eric W., John S. Holtzman, John M. Staatz, Chris Wolf and Michael T. Weber. "MSU Experience in Research Design and Data Processing/Analysis." Paper presented at University of Zambia Food Security Research Network Workshop, Lusaka, Zambia, May 24-25, 1988.

Ministry of Agriculture and Water Development, Planning Division. 1986. Rural Household Survey, Volume 1.

_____. "Quarterly Agricultural Statistics Bulletin." July-December 1985. Lusaka, Zambia.

_____. "Annual Agricultural Statistical Bulletin." 1984.

Quinn, Victoria, Mable Chiligo and J. Price Gittenger. "Household Food and Nutritional Security in Malawi." Paper presented at Symposium on Agricultural Policies for Growth and Development, Lilongwe, Malawi, October 31-November 3, 1988.

Rohrbach, David D. "The Growth of Smallholder Maize Production in Zimbabwe: Causes and Implications for Food Security." Unpublished Ph.D. diss., Department of Agricultural Economics, Michigan State University, 1988.

World Bank. 1984. Zambia: Policy Options and Strategies for Agricultural Growth.

ANNEX TABLES

Annex Table 1. Data Files for Rural Household Benchmark Survey

File Name	Section of Questionnaire	No. of Records	Data Level
CARD01	A. Particulars of household	5834	Household-individual
CARD02	(Type 1) B 1.1-B 1.2 Land use	911	Household
	(Type 2) B 1.3 Crop production	1839	Household-crop
CARD03	(Type 1) B 1.4-B 1.12 Cropping practices	911	Household
	(Type 2) B 1.13-B 1.14 Irrigation and crop storage	911	Household
CARD04	B 2. Livestock production	911	Household
CARD05	B 3. Nuts and fruit production	911	Household
CARD06	C 1.-C 2. Purchases of inputs and sales of farm products	911	Household
CARD07	C 3.0-C 4.0 Purchases of consumption items and use of credit	911	Household
CARD08	C 5. Sources of extension information D 1. Attitudes towards farming D 2. Decision making D 3. Constraints to production D 4. Organizational involvement	911	Household
CARD09	E. Inventory of farm tools and implements	2677	Household-implement

Annex Table 2. Data Files for Rural Household Flow Survey

File Name	Section of Questionnaire	No. of Records	Data Level
FCARD01.A	P. 1 Dates of visits, provinces 1-5	3802	Household visit
FCARD01.B	P. 1 Dates of visits, provinces 6-9	4993	Household visit
FCARD02	P. 2 Crop production	1668	Household-crop
FCARD03	P. 3 Use of fertilizer and manure	582	Household-chemical
FCARD04	P. 4 Use of pesticides	47	Household-chemical
FCARD05	P. 5 Use of animal power	1144	Household-animal power
FCARD06	P. 6 Use of tractor power	24	Household-tractor
FCARD07	P. 7 Other crop expenses and crop production	1373	Household-crop
FCARD08	P. 8 Crop operation dates	1626	Household-crop date
FCARD09	P. 9 Household livestock	182	Household-livestock
FCARD10	P. 10 Household workers	2360	Household-workers
CENTRAL.F11	P. 11 Crop, livestock and other household work	4584	Household-crop, livestock and other household work
COPPERBE.F11	P. 11 " " " " " "	1758	" " " "
EASTERN.F11	P. 11 " " " " " "	6849	" " " "
LUAPULA.F11	P. 11 " " " " " "	5199	" " " "
LUSAKA.F11	P. 11 " " " " " "	2444	" " " "
CHINSALI.F11	P. 11 " " " " " "	2520	" " " "
ISOKA.F11	P. 11 " " " " " "	2363	" " " "
KASAMA.F11	P. 11 " " " " " "	1891	" " " "
MBALA.F11	P. 11 " " " " " "	4495	" " " "
MPOROKO.F11	P. 11 " " " " " "	1771	" " " "

Annex Table 2. (continued)

File Name	Section of Questionnaire	No. of Records	Data Level
NORTH-W.F11	P. 11 " " " " " "	1371	Household-crop, livestock and other household work
CHOMA.F11	P. 11 " " " " " "	3782	" " " "
KALOMA.F11	P. 11 " " " " " "	1853	" " " "
MAZABUKA.F11	P. 11 " " " " " "	1525	" " " "
MONZE.F11	P. 11 " " " " " "	3005	" " " "
SINAZON.F11	P. 11 " " " " " "	1760	" " " "
KALABO.F11	P. 11 " " " " " "	2627	" " " "
KAOMA.F11	P. 11 " " " " " "	4572	" " " "
MONGU.F11	P. 11 " " " " " "	1482	" " " "
SESHEKE.F11	P. 11 " " " " " "	854	" " " "
CENTRAL.F12	P. 12 Off farm, non-work, and total labor	4232	Household-labor
COPPERBE.F12	P. 12 " " " " " "	1785	" "
EASTERN.F12	P. 12 " " " " " "	4878	" "
LUAPULA.F12	P. 12 " " " " " "	3530	" "
LUSAKA.F12	P. 12 " " " " " "	1991	" "
CHINSALI.F12	P. 12 " " " " " "	1406	" "
ISOKA.F12	P. 12 " " " " " "	1998	" "
KASAMA.F12	P. 12 " " " " " "	2147	" "
MBALA.F12	P. 12 " " " " " "	3039	" "
MPOROKO.F12	P. 12 " " " " " "	1677	" "
NORTH-W.F12	P. 12 " " " " " "	1224	" "
CHOMA.F12	P. 12 " " " " " "	3314	" "
KALOMA.F12	P. 12 " " " " " "	884	" "
MAZABUKA.F12	P. 12 " " " " " "	1293	" "
MONZE.F12	P. 12 " " " " " "	2587	" "
SINAZON.F12	P. 12 " " " " " "	1260	" "
KALABO.F12	P. 12 " " " " " "	1638	" "
KAOMA.F12	P. 12 " " " " " "	3717	" "
MONGU.F12	P. 12 " " " " " "	1212	" "
SESHEKE.F12	P. 12 " " " " " "	858	" "

Annex Table 2. (continued)

File Name	Section of Questionnaire		No. of Records	Data Level
CENTRAL.F13	P. 13	Short-term workers	47	Household-st workers
COPPERBE.F13	P. 13	" " "	70	" " "
EASTERN.F13	P. 13	" " "	260	" " "
LUAPULA.F13	P. 13	" " "	102	" " "
LUSAKA.F13	P. 13	" " "	20	" " "
NORTHERN.F13	P. 13	" " "	654	" " "
NORTH-W.F13	P. 13	" " "	51	" " "
SOUTHERN.F13	P. 13	" " "	245	" " "
WESTERN.F13	P. 13	" " "	349	" " "
CENTRAL.F14	P. 14	Household food consumption	2441	Household-consumption
COPPERBE.F14	P. 14	" " "	1641	" "
EASTERN.F14	P. 14	" " "	5716	" "
LUAPULA.F14	P. 14	" " "	2642	" "
LUSAKA.F14	P. 14	" " "	2994	" "
CHINSALI.F14	P. 14	" " "	992	" "
ISOKA.F14	P. 14	" " "	1623	" "
KASAMA.F14	P. 14	" " "	1989	" "
MBALA.F14	P. 14	" " "	2645	" "
MPOROKO.F14	P. 14	" " "	1272	" "
NORTH-W.F14	P. 14	" " "	1053	" "
SOUTHERN.F14	P. 14	" " "	5559	" "
WESTERN.F14	P. 14	" " "	6779	" "
CENTRAL.F15	P. 15	Household income	268	Household-transaction
COPPERBE.F15	P. 15	" "	421	" "
EASTERN.F15	P. 15	" "	1245	" "
LUAPULA.F15	P. 15	" "	680	" "
LUSAKA.F15	P. 15	" "	532	" "
NORTHERN.F15	P. 15	" "	1831	" "
NORTH-W.F15	P. 15	" "	182	" "
SOUTHERN.F15	P. 15	" "	791	" "
WESTERN.F15	P. 15	" "	1944	" "

Annex Table 2. (continued)

File Name	Section of Questionnaire	No. of Records	Data Level
CENTRAL.F16	P. 16 Household expenditures	573	Household-transaction
COPPERBE.F16	P. 16 " "	994	" "
EASTERN.F16	P. 16 " "	2146	" "
LUAPULA.F16	P. 16 " "	1085	" "
LUSAKA.F16	P. 16 " "	924	" "
NORTHERN.F16	P. 16 " "	3526	" "
NORTH-W.F16	P. 16 " "	382	" "
SOUTHERN.F16	P. 16 " "	2307	" "
WESTERN.F16	P. 16 " "	4278	" "

Annex Table 3. Conversion Factors for Hectare

Conversion factors for hectare

1. Lima to hectare
   1 lima = 0.25 ha

2. Acre to hectare
   1 acre = 0.4047 ha

Annex Table 4. Conversion Factors for Kg

Conversion factors for Kg

1. Bags to kg

1 bag maize	= 90 kg	
1 bag sorghum	= 90 kg	
1 bag millett	= 90 kg	
1 bag beans	= 90 kg	
1 bag soybeans	= 90 kg	
1 bag cassava	= 90 kg--	(will be used by MAWD Planning Division for the purposes of the RHS)
1 bag ground nuts	= 80 kg	
1 bag rice	= 80 kg	
1 bag sunflower	= 50 kg	

2. Bale to kg

   1 bale of cotton = 65 kg

3. Unshelled bags to kg shelled

		No. of bags
1 unshelled bag of maize	= 18 kg shelled	(5:1)
1 unshelled bag of groundnuts	= 22.5 kg shelled	(4:1)
1 unshelled bag of millet	= 12 kg shelled	(7:1)

4. Litre to kg

   1 litre tin of grain ~ 3 kg of grain

5. Oxcart to kg

   1 oxcart full of bags = 540 kg (6 bags)

NB Most values were reported in kgs and shelled bags.

Annex Table 5. Number of Completed Questionnaires in Flow Survey

Province	District	CSA	Size Strata I	II	III	TOTAL
ZONE 1 - Northern High Rainfall						
Luapula	Kawambwa	36	6	4	1	11
"	Mansa	51	2	9	1	12
"	Nchelenge	12	4	7	0	11
"	Samfya	18	5	5	0	10
Northern	Chinsali	04	3	9	1	13
"	Isoka	20	3	20	2	25
"	Kasama	05	3	4	1	8
"	"	41	3	12	1	16
"	Mbala	01	1	10	2	13
"	"	24	5	4	1	10
"	Mporokosa	17	4	8	0	12
Copperbelt	Kitwe	01	6	2	1	9
"	Ndola Rural	30	8	4	0	12
"	Mkushi	28	4	9	0	13
Northwest	Mwinilunga	14	7	1	0	8
Total for Zone 1			64	108	11	183
ZONE 2 - Western Semi-Arid						
Northwest	Kabompo	34	5	3	0	8
"	Zambesi	25	4	3	0	7
Western	Kalabo	29	10	8	0	18
"	"	23	5	2	0	7
"	Kaoma	11	5	8	4	17
"	"	50	6	4	1	11
"	Mongu	31	11	2	0	13
"	"	61	5	8	0	13
"	Sesheke	14	4	3	2	9
"	"	21	1	6	0	7
Total for Zone 2			56	47	7	110
ZONE 3 - Southern Plateaux						
Eastern	Chama	10	6	6	0	12
"	Chipata	28	2	6	1	9
"	"	82	3	6	0	9
"	Katete	26	2	8	1	11
"	Lundazi	01	4	5	0	9
"	"	21	1	6	2	9
Central	Kabwe Rural	60	2	5	0	7
"	Mumbwa	18	0	12	4	16
Lusaka	Rural	27	7	8	0	15
Southern	Choma	23	4	8	11	23
"	Kalomo	14	5	7	1	13
"	Mazabuka	16	5	7	1	13
"	Monze	32	6	9	2	17
"	Binazongwe	03	5	6	0	11
Total for Zone 3			59	102	23	184
GRAND TOTALS			179	257	41	477

Annex Table 6. Number of Households for Flow Survey after Data Cleaning

Province	District	CSA	Size Strata I	II	III	TOTAL
ZONE 1 - Northern High Rainfall						
Luapula	Kawambwa	36	4	5	1	10
"	Mansa	51	0	0	1	1
"	Nchelenge	12	2	2	0	4
"	Samfya	18	3	3	0	6
Northern	Chinsali	04	3	6	1	10
"	Isoka	20	3	20	2	25
"	Kasama	05	3	4	0	7
"	"	41	2	10	1	13
"	Mbala	01	0	7	1	8
"	"	24	4	4	0	8
"	Mporokosa	17	3	7	0	10
Copperbelt	Kitwe	01	4	2	1	7
"	Ndola Rural	30	3	2	0	5
Central	Mkushi	28	1	2	0	3
Northwest	Mwinilunga	14	0	0	0	0
Total for Zone 1			35	74	8	117
ZONE 2 - Western Semi-Arid						
Northwest	Kabompo	34	1	2	0	3
"	Zambesi	25	0	0	0	0
Western	Kalabo	29	5	2	0	7
"	"	23	8	6	0	14
"	Kaoma	11	4	7	1	12
"	"	50	5	4	1	10
"	Mongu	31	2	1	0	3
"	"	61	1	4	0	5
"	Sesheke	14	2	2	0	4
"	"	21	1	5	0	6
Total for Zone 2			29	33	2	64
ZONE 3 - Southern Plateaux						
Eastern	Chama	10	4	6	0	10
"	Chipata	28	2	6	1	9
"	"	82	2	5	0	7
"	Katete	26	2	8	1	11
"	Lundazi	01	3	5	0	8
"	"	21	1	6	2	9
"	Petauke	60	1	3	0	4
Central	Kabwe Rural	16	7	3	1	11
"	Mumbwa	18	0	10	3	13
Lusaka	Rural	27	4	6	0	10
Southern	Choma	23	4	8	10	22
"	Kalomo	14	0	2	0	2
"	Mazabuka	16	4	6	0	10
"	Monze	32	3	1	0	4
"	Binazongwe	03	3	5	0	8
Total for Zone 3			40	80	18	138
GRAND TOTALS			104	187	28	319

Annex Table 7. Area Planted, Production and Yield of Maize by Household I.D.

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
1	1	3	16	11	1.62	3060.00	1890.29
1	1	3	16	42	.40	900.00	2223.87
1	1	3	16	48	.81	270.00	333.58
1	1	3	16	53	.40	810.00	2001.48
1	1	3	16	64	.81	360.00	444.77
1	1	3	16	73	.50	720.00	1440.00
1	1	3	16	81	.81	900.00	1111.93
1	1	3	16	82	3.24	2250.00	694.96
1	1	3	16	94	1.21	360.00	296.52
1	1	3	16	97	2.02	1350.00	667.16
1	1	3	16	99	.81	450.00	555.97
1	2	1	28	29	2.43	810.00	333.58
1	2	1	28	32	.25	1800.00	7200.00
1	2	1	28	38	.25	900.00	3600.00
1	3	3	18	1	2.43	2250.00	926.61
1	3	3	18	6	2.02	2160.00	1067.46
1	3	3	18	11	.81	15.00	18.53
1	3	3	18	16	4.86	10530.00	2168.27
1	3	3	18	17	.40	4500.00	11119.35
1	3	3	18	45	2.02	20.00	9.88
1	3	3	18	49	2.43	1800.00	741.29
1	3	3	18	70	.81	1350.00	1667.90
1	3	3	18	71	3.24	2700.00	833.95
1	3	3	18	72	8.09	6660.00	822.83
1	3	3	18	101	4.05	2700.00	667.16
1	3	3	18	104	.75	900.00	1200.00
1	3	3	18	115	3.64	9000.00	2470.97
2	1	1	1	4	1.00	450.00	450.00
2	1	1	1	10	.65	270.00	412.40
2	1	1	1	49	3.00	720.00	240.00
2	1	1	1	126	3.25	1620.00	498.46
2	1	1	1	130	.25	90.00	360.00
2	1	1	1	148	15.00	2790.00	186.00
2	1	1	1	154	.50	450.00	900.00
2	2	1	30	49	.50	90.00	180.00
2	2	1	30	74	.40	180.00	444.77
2	2	1	30	86	.40	450.00	1111.93
2	2	1	30	94	.25	8100.00	32400.00
2	2	1	30	161	1.00	450.00	450.00
3	1	3	10	2	1.00	1170.00	1170.00
3	1	3	10	17	.50	900.00	1800.00
3	11	3	10	34	.50	2700.00	5400.00

Annex Table 7. (continued)

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
3	1	3	10	52	1.25	720.00	576.00
3	1	3	10	86	1.00	540.00	540.00
3	1	3	10	157	1.00	1800.00	1800.00
3	1	3	10	225	1.00	450.00	450.00
3	1	3	10	252	.50	1260.00	2520.00
3	1	3	10	288	.50	1350.00	2700.00
3	1	3	10	310	1.75	1350.00	771.43
3	2	3	28	9	2.21	2070.00	934.92
3	2	3	28	34	1.21	2700.00	2223.87
3	2	3	28	40	2.43	20250.00	8339.51
3	2	3	28	145	1.21	3510.00	2891.03
3	2	3	28	200	.40	630.00	1556.71
3	2	3	82	2	1.21	3780.00	3113.42
3	2	3	82	18	2.02	1710.00	845.07
3	2	3	82	31	2.43	11520.00	4744.26
3	2	3	82	41	3.24	4320.00	1334.32
3	2	3	82	140	.40	270.00	667.16
A001	A002	A003	A004	A005	TOTMZHA	TOTMZKG	MAIZEYLD
3	2	3	82	218	1.62	2250.00	1389.92
3	2	3	82	240	2.02	2700.00	1334.32
3	3	3	26	24	2.21	4320.00	1951.13
3	3	3	26	68	1.62	900.00	555.97
3	3	3	26	100	6.07	19170.00	3157.89
3	3	3	26	119	3.24	1800.00	555.97
3	3	3	26	157	2.02	4950.00	2446.26
3	3	3	26	187	8.50	18900.00	2223.87
3	3	3	26	245	3.64	2700.00	741.29
3	3	3	26	322	1.25	2250.00	1800.00
3	3	3	26	352	2.43	2520.00	1037.81
3	3	3	26	359	3.24	13950.00	4308.75
3	3	3	26	389	3.43	7200.00	2100.23
3	4	3	1	13	1.21	90.00	74.13
3	4	3	1	36	.40	450.00	1111.93
3	4	3	1	78	.40	450.00	1111.93
3	4	3	1	85	1.21	720.00	593.03
3	4	3	1	96	1.21	720.00	593.03
3	4	3	1	104	9.40	12780.00	1358.90
3	4	3	1	127	2.02	2250.00	1111.93
3	4	3	1	132	1.21	900.00	741.29

Annex Table 7. (continued)

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
3	4	3	21	24	4.05	3150.00	778.35
3	4	3	21	27	33.00	17550.00	531.82
3	4	3	21	37	2.02	10440.00	5159.38
3	4	3	21	51	3.24	270.00	83.40
3	4	3	21	52	1.62	5760.00	3558.19
3	4	3	21	71	7.21	10350.00	1434.69
3	4	3	21	94	1.62	5850.00	3613.79
3	4	3	21	103	5.00	4140.00	828.00
3	4	3	21	109	2.43	4140.00	1704.97
3	5	3	60	26	1.62	2700.00	1667.90
3	5	3	60	79	2.02	450.00	222.39
3	5	3	60	224	.81	2970.00	3669.38
3	5	3	60	256	1.00	2250.00	2250.00
4	1	1	36	6	1.21	270.00	222.39
4	1	1	36	14	.50	105.00	210.00
4	1	1	36	20	3.00	2160.00	720.00
4	1	1	36	73	.40	720.00	1779.10
4	1	1	36	80	.50	50.00	100.00
4	1	1	36	136	.25	10.00	40.00
4	1	1	36	142	.50	25.00	50.00
4	1	1	36	152	.25	45.00	180.00
A001	A002	A003	A004	A005	TOTMZHA	TOTMZKG	MAIZEYLD
4	1	1	36	178	.25	2250.00	9000.00
4	1	1	36	218	1.00	1800.00	1800.00
4	2	1	51	31	10.00	90.00	9.00
4	3	1	12	49	.25	630.00	2520.00
4	4	1	18	51	.50	900.00	1800.00
4	4	1	18	59	.25	450.00	1800.00
4	4	1	18	131	.25	900.00	3600.00
4	4	1	18	150	.50	540.00	1080.00
4	4	1	18	152	.50	1530.00	3060.00
5	1	3	27	9	.50	18.00	36.00
5	1	3	27	15	1.21	1080.00	889.55
5	1	3	27	26	3.24	1170.00	361.38
5	1	3	27	39	4.05	90.00	22.24
5	1	3	27	83	3.00	720.00	240.00
5	1	3	27	85	.50	3960.00	7920.00
5	1	3	27	88	2.00	270.00	135.00
5	1	3	27	93	.81	450.00	555.97

Annex Table 7. (continued)

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
5	1	3	27	96	.50	540.00	1080.00
5	1	3	27	102	.50	90.00	180.00
6	1	1	4	9	3.24	1620.00	500.37
6	1	1	4	14	2.02	2700.00	1334.32
6	1	1	4	15	3.24	5670.00	1751.30
6	1	1	4	93	2.02	1980.00	978.50
6	1	1	4	96	2.00	2880.00	1440.00
6	1	1	4	103	.81	1800.00	2223.87
6	1	1	4	215	4.00	14760.00	3690.00
6	1	1	4	224	2.43	1530.00	630.10
6	1	1	4	245	1.25	1530.00	1224.00
6	1	1	4	253	3.24	2160.00	667.16
6	2	1	20	3	4.00	14940.00	3735.00
6	2	1	20	6	1.62	9450.00	5837.66
6	2	1	20	17	12.00	54000.00	4500.00
6	2	1	20	21	1.21	2340.00	1927.35
6	2	1	20	25	4.00	11520.00	2880.00
6	2	1	20	39	3.00	6750.00	2250.00
6	2	1	20	43	2.00	3240.00	1620.00
6	2	1	20	51	2.81	9720.00	3459.81
6	2	1	20	54	.75	1530.00	2040.00
6	2	1	20	59	.81	2700.00	3335.80
6	2	1	20	81	1.25	5040.00	4032.00
6	2	1	20	92	.81	3420.00	4225.35
6	2	1	20	108	3.00	16200.00	5400.00
6	2	1	20	109	2.00	5760.00	2880.00
6	2	1	20	131	3.00	162000.00	5400.00
6	2	1	20	135	2.02	4140.00	2045.96
6	2	1	20	143	3.00	2250.00	750.00
6	2	1	20	144	.81	2340.00	2891.03
6	2	1	20	150	3.00	5220.00	1740.00
6	2	1	20	159	2.00	5400.00	2700.00
6	2	1	20	168	.25	1350.00	5400.00
6	2	1	20	183	1.00	2880.00	2880.00
6	2	1	20	193	.65	1350.00	2062.01
6	2	1	20	194	2.00	4770.00	2385.00
6	2	1	20	234	3.00	6120.00	2040.00
6	3	1	5	5	3.00	1800.00	600.00
6	3	1	5	62	.81	900.00	1111.93
6	3	1	5	65	2.00	900.00	450.00
6	3	1	5	215	1.00	360.00	360.00
6	3	1	5	242	1.00	1080.00	1080.00

Annex Table 7. (continued)

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
6	3	1	41	9	1.00	90.00	90.00
6	3	1	41	34	.50	12.00	24.00
6	3	1	41	36	3.00	360.00	120.00
6	3	1	41	66	.25	360.00	1440.00
6	3	1	41	70	.50	270.00	540.00
6	3	1	41	72	1.00	360.00	360.00
6	3	1	41	164	.50	270.00	540.00
6	3	1	41	173	1.00	270.00	270.00
6	3	1	41	230	1.00	630.00	630.00
6	3	1	41	237	1.00	270.00	270.00
6	3	1	41	250	.25	90.00	360.00
6	3	1	41	251	.50	90.00	180.00
6	3	1	41	260	3.00	720.00	240.00
6	4	1	1	14	1.25	1620.00	1296.00
6	4	1	1	18	2.83	1440.00	508.31
6	4	1	1	31	6.00	16200.00	2700.00
6	4	1	1	151	1.62	1800.00	1111.93
6	4	1	1	181	4.00	1350.00	337.50
6	4	1	1	183	.50	180.00	360.00
A001	A002	A003	A004	A005	TOTMZHA	TOTMZKG	MAIZEYLD
6	4	1	1	185	2.40	900.00	374.27
6	4	1	1	193	.40	450.00	1111.93
6	4	1	24	41	2.00	2250.00	1125.00
6	4	1	24	125	2.00	5400.00	2700.00
6	4	1	24	146	.40	450.00	1111.93
6	4	1	24	179	1.00	90.00	90.00
6	4	1	24	219	.81	900.00	1111.93
6	4	1	24	450	.40	40.00	98.84
6	4	1	24	540	.40	90.00	222.39
6	4	1	24	611	.40	810.00	2001.48
6	5	1	17	24	.75	900.00	1200.00
6	5	1	17	48	1.21	180.00	148.26
6	5	1	17	49	.50	180.00	360.00
6	5	1	17	53	.25	360.00	1440.00
6	5	1	17	71	1.00	300.00	300.00
6	5	1	17	96	.75	50.00	66.67
6	5	1	17	97	1.75	495.00	282.86
6	5	1	17	128	1.00	4500.00	4500.00
6	5	1	17	149	.81	1800.00	2223.87
6	5	1	17	150	.81	2880.00	3558.19
7	1	2	34	10	.75	90.00	120.00

Annex Table 7. (continued)

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
7	1	2	34	11	.50	180.00	360.00
7	1	2	34	172	.25	360.00	1440.00
8	1	3	23	2	.81	1890.00	2335.06
8	1	3	23	3	3.00	4050.00	1350.00
8	1	3	23	8	1.21	900.00	741.29
8	1	3	23	10	.40	1350.00	3335.80
8	1	3	23	12	1.21	5130.00	4225.35
8	1	3	23	14	1.21	486.00	400.30
8	1	3	23	32	1.62	1350.00	833.95
8	1	3	23	33	2.02	7200.00	3558.19
8	1	3	23	43	3.02	4500.00	1488.34
8	1	3	23	61	.81	22050.00	27242.40
8	1	3	23	65	.81	5400.00	6671.61
8	1	3	23	68	1.21	2430.00	2001.48
8	1	3	23	69	1.21	4050.00	3335.80
8	1	3	23	74	1.62	5850.00	3613.79
9	2	2	11	102	.40	900.00	2223.87
9	2	2	11	103	3.81	6300.00	1653.80
9	2	2	11	107	3.24	3150.00	972.94
9	2	2	11	123	1.21	6300.00	5189.03
9	2	2	50	6	.81	270.00	333.58
9	2	2	50	22	.25	150.00	600.00
9	2	2	50	41	.40	300.00	741.29
A001	A002	A003	A004	A005	TOTMZHA	TOTMZKG	MAIZEYLD
9	2	2	50	66	2.43	900.00	370.64
9	2	2	50	102	4.00	7740.00	1935.00
9	2	2	50	121	.25	180.00	720.00
9	2	2	50	123	.81	450.00	555.97
9	2	2	50	124	.50	360.00	720.00
9	2	2	50	129	.40	60.00	148.26
9	2	2	50	130	.25	120.00	480.00
9	3	2	31	1	.25	90.00	360.00
9	3	2	31	7	.50	630.00	1260.00
9	3	2	31	93	.25	180.00	720.00
9	3	2	61	50	.25	90.00	360.00
9	3	2	61	58	.81	1080.00	1334.32
9	3	2	61	76	.25	180.00	720.00
9	3	2	61	93	1.00	360.00	360.00
9	3	2	61	99	3.00	90.00	30.00
9	4	2	14	61	.81	2700.00	3335.80
9	4	2	14	82	1.00	1350.00	1350.00
9	4	2	14	121	1.21	2250.00	1853.22
9	4	2	14	144	1.00	180.00	180.00
9	4	2	21	7	1.21	1800.00	1482.58
9	4	2	21	8	2.02	1800.00	889.55

Annex Table 7. (continued)

Province A001	District A002	Zone A003	CSA A004	Household A005	Area Planted (Ha) TOTMZHA	Production (Kg) TOTMZKG	Yield (Kg/Ha) MAIZEYLD
9	4	2	21	29	4.05	5400.00	1334.32
9	4	2	21	57	2.43	3600.00	1482.58
9	4	2	21	81	3.24	900.00	277.98
9	4	2	21	149	3.24	1350.00	416.98

Number of cases read = 319 Number of cases listed = 319