

## **Supplemental Survey to the 1999/2000 Post Harvest Survey - Zambia Data Documentation – Revised June 2010**

Conducted by the Central Statistical Office (CSO) April – May 2001, in conjunction with the Food Security Research Project from Michigan State University. The sample was all households who completed the 1999/2000 Post Harvest Survey conducted in August 2000.

### **Recommended file structure:**

C:\.....\Zambia\  
hh01\

**anal** - data files and syntax created from analysis

**data** - working data files and syntax used to convert from archive format to SPSS

**docs** - documentation for this survey

**lookup** – lookup files are stored in this directory

**newvars** –

**demog** – calculation of adult equivalents and household size

**income** – standardized computed variables for crop income, off-farm income, livestock income, productive assets income, landholdings, retained caloric values, fertilizer costs

**tmp** - the directory where temporary files can be stored - to be erased when finished

### **Data files descriptions (c:\...\Zambia\hh01\data):**

Household level variables - prov, dist, csa, sea, hh

### **Conventions:**

If a value in a variable was missing and the question was suppose to have been asked, a user missing value of -2 with a value label of \*Do not know\* was entered. If a variable was not to have been asked, then the sysmis value was recoded to -1 \*Not applicable\*. Both values of -1 and -2 were declared as ■user missing•.

Extremes were run on all continuous variables. If it was determined that a value was an extreme, that case was verified by examining the questionnaire. If the value in the data file was the same as the value on the questionnaire, the data were not changed, even though the values may not have been reasonable.

**Panel household identification:**

This dataset is part of a panel study. In the file, “id.sav” a variable has been added, called “paneltype”. 5 different types of households have been identified:

**paneltype Panel classification of household**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 matches all 3 years	4288	61.9	61.9	61.9
	1 2000 only	1273	18.4	18.4	80.3
	2 2000 & 2004 only	1070	15.5	15.5	95.8
	3 2004 & 2008 only	52	.8	.8	96.5
	4 2000 & 2008 only	230	3.3	3.3	99.9
	5 none of years match	9	.1	.1	100.0
	Total	6922	100.0	100.0	

A value of 0 identifies the household as having been interviewed all three years of the panel study. A value of 1 indicates the household was only interviewed in this survey conducted in 2001. A value of 2 indicates the household was interviewed only in 2000 and 2004. A value of 3 indicates that the household interviewed in 2001 did not match the 2004 household. A value of 4 indicates the household was not found in 2004 but was found again in 2008, a value of 5 indicates the household was interviewed in one other year but is not the same household as was interviewed in 2001.

**Description of files: (c:\...\Zambia\hh01\data):**

Survey Section	File Name	Key Variables	Variable names start with ...	Number of cases
Front Page	<b>allsurid.sav</b> – all surveyed households	dist csa sea hh	mnemonics	7,699
	<b>id.sav</b> – only households that completed the survey. Weight variables are in this file. wgt_pan should be used if the researcher is working with the panel subset of data. Wgt_pop should be used if the complete dataset is being used for analysis. Wgt_orig should not be used.	dist csa sea hh	mnemonics	6,922
	<b>hh.sav</b> – data from many different sections	dist csa sea hh	hh...	6,922
	<p><u>hh.sav notes:</u> Number of children under 11 (boys and girls) (hh01, hh02) is less under 5 (hh03): There are approximately 35 cases than number where adding hh01 + hh02 is less than hh03. Questionnaires were checked and data were entered correctly from the questionnaires.</p> <p>Fertilizer transport modes: The data were not collected consistently throughout. There were households where no fertilizer was purchased and “10” entered into HH39, Hh40, HH41. There are cases where fertilizer was purchased and “10” is entered into hh39 hh40 and hh41. There are cases where the questions were left as sysmis – these were recoded to –1 “not applicable”. Some of the enumerators did not understand the question so asked the mode of transportation for each type of fertilizer input – where mode of transport for basal was entered into hh39 and mode of transport for top dressing was entered into hh40 and hh41 was keyed as “10” Not applicable.</p>			
Section 1 – <b>Demography</b>	<b>adults.sav</b> – adult family member	dist csa sea hh mem	d..	24,284
	<u>adults.sav notes:</u> There are some cases where the head of household is no longer a member of the household (D02 = 2) but months away from household (D08) equal 0 or some other greater value. That is what was recorded on the survey instrument.			
Section 2 – <b>Previous Deaths</b>	<b>PrevDeaths.sav</b> – deaths within hh since 1996	dist csa sea hh pdmem	pd...	2,176
Section 3 – <b>Off-farm income remittances</b>	<b>SalWage.sav</b> – salaried employment or informal wage labour activities	dist csa sea hh wact w01	w...	1,864
	<u>SalWage.sav Notes:</u> There are 9 households appearing in SalWage where no amounts were entered for wages earned. In Adults.sav the value of D10 = 1 (yes – earned salary – wages). However, there are no data on how much was earned. You will see –2 “do not know” entered for all the variables for these households.			

Survey Section	File Name	Key Variables	Variable names start with ...	Number of cases																								
	<b>Bus_Activ.sav</b> – formal or informal business activities	dist csa sea hh bact b01	b...	3,133																								
	<b>Bus_Activ.sav Notes:</b> This file has duplicate cases if you use the key variables DIST CSA SEA HH MEM BACT B01 These are valid duplicate cases – the data in the rest of the variables are different – different incomes-expenditures were received for the same activity in the same location.																											
	<b>Remit.sav</b> – remittances given and received	dist csa sea hh rmem	r...	7,632																								
<b>Section 4 – Farm Land and Use</b>	<b>Field.sav</b> – details about each field	dist csa sea hh field	f...	28,289																								
	<p>Field.sav Notes: There are 5 cases where F08 (year field first acquired) does not seem valid (1689, 985,998) – the dates entered are what were on the questionnaire. There are 5 households where crop data exists, but no field data were collected. They are</p> <table border="1"> <thead> <tr> <th>dist</th> <th>csa</th> <th>sea</th> <th>hh</th> </tr> </thead> <tbody> <tr> <td>302</td> <td>23</td> <td>2</td> <td>98</td> </tr> <tr> <td>402</td> <td>44</td> <td>3</td> <td>111</td> </tr> <tr> <td>703</td> <td>8</td> <td>1</td> <td>35</td> </tr> <tr> <td>704</td> <td>4</td> <td>2</td> <td>5</td> </tr> <tr> <td>901</td> <td>33</td> <td>23</td> <td>7</td> </tr> </tbody> </table> <p>There are many households (about 50) where field information was collected, but there are no crop data. Some of these fields were fallow; others had crops but no crop data were collected by the enumerators.</p>				dist	csa	sea	hh	302	23	2	98	402	44	3	111	703	8	1	35	704	4	2	5	901	33	23	7
dist	csa	sea	hh																									
302	23	2	98																									
402	44	3	111																									
703	8	1	35																									
704	4	2	5																									
901	33	23	7																									
<b>Section 5 – Crop Stocks and Sales</b>	<b>crop.sav</b> – details about each crop harvested stored sold or exchanged and price (note – cassava, vegetables, fruits and green maize are not listed). Crops are not associated with specific fields.	dist csa sea hh crop	s...	14,923																								

Survey Section	File Name	Key Variables	Variable names start with ...	Number of cases
	<p><u>crop.sav notes</u>: The data have been verified. There are 322 cases where sold and bartered are more than harvested. There are 8 cases where kg stored is greater than kg harvested. There are 81 cases where a quantity was sold but no price was collected.</p> <p>If the analyst matches these crops with the 99/00 PHS data, there will be major differences. There could be fewer crops for the household in the SS than were in the PHS or there could be more in the SS than were mentioned in the PHS survey. The crops listed in the PHS were listed as starters for the SS. The household could add more crops or indicate that they really didn't raise that particular crop.</p> <p>No cleaning checks were made between the two files (PHS and SS) to decide if a crop had not been entered in the SS, that should have, or that a crop was not entered into the PHS that should have. This would have been extremely time-consuming exercise.</p>			
	<b>veg_fruit.sav</b> – vegetables, fruits and green maize	dist csa sea hh crop	vf...	3,239
	<p><u>veg_fruit.sav notes</u>: Key variables are dist csa sea hh crop however there are households that had cabbage and Chinese cabbage which was coded as 31, these cases had different values in vf02 and vf03 so could not be combined. Other households had specified ■other vegetables■ which were different, so there can be more than one case with the code of 47 for the household.</p> <p>The cases where there were no sales of vegetables/fruit were not keyed into the data file. A decision was made to refer to the field file and if there was land allocated to ■garden■ or ■orchard■, it can be assumed there were vegetables / fruit grown, but not sold. Should this data be needed for some future analysis, then it can be entered at that point.</p>			
Section 6 – Crop purchases	<b>purch.sav</b> – since last harvest	dist csa sea hh purch	p...	5,957
	<p><u>purch.sav notes</u>: There are several cases where the price appears to be high for the unit. It is possible that the enumerator may have collected the price for the total amount purchased rather than the price per unit. Cases that appeared to be extremes were verified with the questionnaire.</p> <p>For purch = 8, 9, &amp; 10, P03 = 19 and P05 = 19 may be correct codes, but no conversion factors have been decided because of the complexity of the issue. If 'home consumption' was interpreted as 'for household means', this unit would not have applied, possibly.</p> <p>For purch = 8, P05 = 15, and purch = 10, P05 = 10, they are data collection errors. They were not changed in the database, the analyst must decide what to do.</p> <p>Comparing the quantities purchased with the number of adults in the household, there are 15 households where the quantities seem to be extreme. These data were verified with the questionnaires. There may be data collection recording errors in the unit or quantity. The analyst must decide what to do.</p>			
Section 7 –	<b>service.sav</b> – 99/00 agricultural season	dist csa sea hh srcode	sr...	6,226

Survey Section	File Name	Key Variables	Variable names start with ...	Number of cases
<b>Service Provision</b>				
Section 8 – <b>Conservation Farming</b>	<b>soilcons.sav</b> – 99/2000 and 2000/2001 agricultural season	dist csa sea hh seprac	sc...	41,526
Section 9 – <b>Cotton roducer – Ginning Firms</b>	<b>cotton.sav</b> – experiences with cotton growing	dist csa sea hh year	ct...	3,284
Section 10 – <b>Fertilizer acquisition</b>	<b>fertacq.sav</b> – transaction channel questions	dist csa sea hh fert	fr...	1,742
	<u>Fertacq.sav notes:</u> FR04 - some enumerators misinterpreted the ■cost of transport• for a 50 kg bag and instead recorded the ■cost of a 50 kg bag•. So you will see small distances in FR03 and large values in FR04 for about 13 cases. There are 20 some cases where the distance from the source of fertilizer acquisition (FR03) seems rather large for the source being “within the village” (FR02 = 1). However, the data in FR03 for these cases are what was on the questionnaires.			
Section 11 – <b>Milk &amp; Egg Production</b>	<b>milk_egg.sav</b> – production over last 12 months and price	dist csa sea hh lv	lv...	3,048
	<u>Milk_Egg.sav notes:</u> Not all of the enumerators drew a line in the egg part of the table to create another column for the month. Sometimes, the number of months would be recorded as only 11 months, other times the number of months was not recorded, but 12 months of information were recorded. In the cases where there were only 11 months recorded, we coded a -2 for the 12 <sup>th</sup> month. This may or may not be accurate. We could not tell which month had been left out. Therefore, doing analysis by month will not be reliable for the egg section of the table.  If you add fresh milk sold (lv=2) and sour milk sold (lv=3) and subtract that total from fresh milk produced (lv=1), you will find several cases where the household sold more than was produced. These household data were verified with the questionnaires and reflect what the questionnaire said.			
Section 12 – <b>General Information</b>	Data for this section can be found in the hh.sav file.			
Section 13 – <b>Crop Forecasting</b>	This information is documented in the \Zambia\cfs\0001\docs directory as part of the Post Harvest survey done by the Central Statistical Office.			

### Lookup tables

C:\...\Zambia\hh01\lookup

Type of data	File name	File to be used with	Key variables	Number of cases	Comments
Crop quantity conversion to kgs	cropconv.sav	crop.sav	crop, unit	259	Converts all harvested/sold crop units to kgs.
Land conversion to hectares	ha_conver.sav	field.sav	f03	4	Converts area to hectares
Price per kg for a specific crop	crop_prices_9900.sav	crop.sav	prov, crop	153	Median price per kg for the crop at the provincial level – kwacha. Generated using 'crop_prices_9900.SPS'
Livestock median price	livestock_price_9900.sav	\postharv\9900\livestock.sav	prov, animal	44	Unit price for livestock – Kwacha. Generated using 'livestock_price_9900.SPS'
Livestock product prices	liveprod_price.sav	milk_egg.sav	prov, lv	22	Unit price for livestock products. Generated using 'livprod_price.SPS'
Conversion of purchases to kgs	purconv.sav	purch.sav	purch, unit	127	Converts all purchased units to kgs

### New Computed Variables

C:\...\Zambia\hh01\newvars

Type of data	File name	Key variables	Number of cases	Variables	Syntax File - comments
<b>Subdirectory “demog”</b>					
Adult equivalents and size of household	adult_equivalents_2001.sav	dist, csa, sea, hh	6,922	Aehh01 – Full time adult equivalents hysize01 – household size age15_59 – number of prime age members educhead – level of education of the head edu_max – maximum level of education in the household malehead – Male headed household femalehead – Female headed household	adult_equivalents_2001.sps – see note at end of documentation regarding method used to compute adult equivalents
<b>Subdirectory “income”</b>					
Crop income computation	crop_income_ss01.sav	dist, csa, sea, hh	6,922	gvharv_p: gross value of harvest for crops found in all years for panel, gvharv: gross values of all crops (same as gvharv_p for this year) gvsale_p: gross value of sales for crops found in all years for panel gvsale: gross value of sales cerealK, tubersK, com_crpK, otherK – categories of	crop_income_ss01.sps Using PHS data since we want to use the area planted and the data collected in the ss01 does not harmonize well.

Type of data	File name	Key variables	Number of cases	Variables	Syntax File - comments
				crops – value of production cerealS tuberS, comcrpS, otherS – categories of crops value of sales vegsales – no production information, only sales. Also included are variables of harvest and sales by each crop	
Off farm income	off_farm_income_ss01.sav	dist, csa, sea, hh	6,922	Net_off_farm: total net off farm income business costs subtracted (tot_work+ tot_bus+ remit_in) Gross_off_farm: total gross off farm income where business expenses are not subtracted (tot_work+gv_busincome + remit_in) Detailed variables are: tot_bus, gv_busincome, gv_buscost, remit_in, tot_work, non_ag, ag_wage, agric_sf (small farm ag wages) agric_lg (commercial farm ag wages)	off_farm_income_ss01.SPS
Live animal valuation	livestock_income_ss01_with_fishest.sav	dist, csa, sea, hh	6,922	“Livinc01_fish” – total income for livestock activities <u>including</u> estimated fish income livinc01 – total income for livestock activities <u>excluding</u> estimated fish income livprod – income from eggs and milk tot_livs – income from live sales and slaughtered animals	livestock_income_ss01.SPS Income from fish was estimated based on data from ss04 and ss08
Productive asset valuation	productive_assets_ss01.sav	dist, csa, sea, hh	6,922	prodast – draft animals and equipment value (Kwacha)	Productive_assets_ss01.SPS
Land holdings	land_variables_ss01.sav	Dist csa sea hh	6,922	Cropland – area under crops Landholdpanel – area under crops plus fallow land Includes variables for area planted for each crop. Garden and virgin land comes from the ss01 data.	Land_variables_ss01.sps The data was generated from the PHS 99/00 – the area planted to each crop was not captured in the supplemental survey. The area that was collected in the supplemental should not be used – it is not reflecting the cropped area correctly.
Fertilizer costs (these variables have not been incorporated into the income file)	hh_fertilizer_costs_kgs_ss01.sav	Dist csa sea hh	6,922	Tot_fert_kg – total kgs of fertilizer used for all crops Tot_fert_cost – total cost of fertilizer used for all crops Tot_fert_kg_p: total kgs of fertilizer used for crops asked about for panel analysis Tot_fert_cost_p: total cost of fertilizer used for crops asked about for panel analysis	hh_fert_costs_9900.sps Median price for basal and top dressing using the PHS 99/00 was calculated and used to value fertilizer. We are not able to distinguish between free fertilizer and that which was actually paid for, as

Type of data	File name	Key variables	Number of cases	Variables	Syntax File - comments
				Also variables for value of fertilizer used for each of the crops.	was done for the ss04 and ss08 data. There were approximately 250 observations in those years to received free fertilizer.
Calories retained	hh_level_crop_detail_calories_ss01.sav	Dist csa sea hh	6,922	Individual variables for each crop computing calories retained	hh_level_crop_detail_calories_ss01.sps
Income (gross and net)	income_gross_net_observed_ss01.sav	Dist csa sea hh	6,922	Gross_income_panel = gvharv+ vegsales+ tot_livs+ livprodprod+ gross_off_farm Net_income_panel = gvharv, vegsales, tot_livs, livprodprod, net_off_farm, -tot_fert_cost Gross_income_allhhs and net_income_allhhs are the same for this year.	income_gross_net_observed_ss01.sps

### Publications:

**SS01\_DataFileNotes.pdf.** Food Security Research Project, 2001. Supplemental Survey Data Documentation Notes. Michigan State University, September.

**SS01\_Enumerator\_manual.pdf.** Republic of Zambia, 2001. Supplemental Survey to the 1999/2000 Post Harvest Survey (for small and medium scale holdings) **Interviewers' Instruction Manual.** Food Security Research Project, Michigan State University and Republic of Zambia, March.

**SS01\_OrigQuest.pdf.** Republic of Zambia, 2001 Supplemental Survey to the 1999/2000 Post-Harvest Survey (for small and medium scale holdings) **Original Survey Instrument.** Republic of Zambia, April.

**SS01\_SynQuest.pdf.** Republic of Zambia, 2001 Supplemental Survey to the 1999/2000 Post-Harvest Survey (for small and medium scale holdings) **Modified Original Survey Instrument reflecting changes made during enumerator training.** Republic of Zambia, April.

**SS01\_TrainingNotes.pdf.** Republic of Zambia, 2001 Supplemental Survey to the 1999/2000 Post-Harvest Survey (for small and medium scale holdings) **Supplemental Training Notes.** Food Security Research Project, Michigan State University and Republic of Zambia, April.

### Questions regarding data validity:

If you find data that you wish to have verified with the original survey instruments, please follow these instructions: Send a copy of the syntax file that will produce a listing of data in question to Margaret Beaver at [beaverm@msu.edu](mailto:beaverm@msu.edu). She will either make corrections to the data and then issue new copies of the data to everyone or send a syntax file to everyone to run to make the corrections. It is important that you send a syntax file that will list the questionable data rather than just describing the problem.