



# STATISTICS

*SIERRA LEONE*

Sierra Leone Integrated Household Survey

Process Manual

*DRAFT*

## *Summary*

This manual contains details on the processes involved in treating and transforming the data for use in computation of the household consumption aggregate of the Sierra Leone Integrated Household Living Condition Survey (SLIHS). It covers the following areas:

- A. Preparation of the data
  - 1) ID Changes
  - 2) Structural Edits
  - 3) Treatment of outliers
  - 4) Program procedures in CPro
- B. Methodology for computing the aggregate
  - 1) Directory Structure
  - 2) SPSS programs
  - 3) File Specifications
- C. Observations and considerations
  - 1) Prices
  - 2) Checklist of improvements
  - 3) Sample issues
  - 4) Milestone dates
  - 5) Reference Documents
  - 6) Tabulations
  - 7) Archive structure
  - 8) Product Recodes
  - 9) Other links
  - 10) Contacts

The data entry screen were designed using IMPS (CENTRY and CONCOR). Most of the programs designed to process and prepare the data for computation of the aggregate were written in CPro (Census and Survey Processing Software). IMPS and CPro are developed by the U.S. Census Bureau. The programs for computing the household aggregate were designed in SPSS (Statistical Package for the Social Sciences). World Bank standards and template were used to classify the items in the aggregate.

1. ⇒ ID Changes

There are two id systems used for tracking the households: Field IDs and System IDs. The Field IDs are based on the actual household ID as it was listed in the EA. For the data entry, it was decided to recode the enumeration areas into sequential System ids. This created several problems. Because automated control systems were not sufficiently established prior to data entry, the changes in enumeration ids from Field Id to System Id created some confusion in tracking. Furthermore, the insertion (as opposed to appending) of 26 EAs to complete the sample during the second six months of data entry caused duplication of ids between the first six months and second six months. For this reason, a program was developed which changes all the IDS for the first six months to the appropriate final system ids.

{clicking the hyperlinks will run the programs. If you do not wish to run the program but only want to find its location, then go to the file as indicated by the hyperlink}

A listing of the IDs is provided in an excel document. These contain the three methods used for tracking the questionnaire. The original selected EA (or field code) a sequential system ID and a third set of modified IDs. These were changed due to insertion of 20 Eas during the six month of field work.

<..\Other\IDChanges.xls>

In order to replace the old ids since the insertion caused the inadvertent duplication of many household ids, two programs were designed to modify the ids.

Part A: IDCHANGEA.PFF

Part B: IDCHANGE.B.PFF

The program changes those ideas for the first six months only and should be run for both Part A & B. The output containing the corrected IDs should then be combined with the second six months of data to complete the data set with the correct IDs.

## 2. ⇒ Structural Edits

The files entitled Part\_A1 and Part\_B1 now contain household level ids that do not have duplicates. However, there are cases where the individual ID still requires some editing for structure (i.e. consistency across records) and basic consistency. The series of edits contained in the following programs deal with these edit:

Part A: [AEdits.bch](#)

Part B: [BEdits.bch](#)

Part C: [CEdits.bch](#)

These edits corrected for line drift, duplicates etc.

## 3. ⇒ Treatment of Outliers

A process of checking and correcting for extreme values was required before producing the final data files for computation of the aggregate or use in National Accounts. The method used in dealing with outliers was through an analysis of a correlated ratio the median of which was used to base the limit of acceptability. In some cases the values were dynamically imputed and in others statically imputed. Dynamic imputation was used in consumption items from Section 8H, 9A and 9B. This was done in order to not prejudice the results with a constant replacement value. When dynamic imputation was used, the outlier was defined by defining resistant fences to establish limits of acceptability based the inter-quartile ranges of the median values.<sup>1</sup> CSPro software was used to automate the process. The following provides a section by section review of the treatment.<sup>2</sup> For details on the imputation frequencies, listings are provided in the folder:

C:\Sierra Leone\_Household Survey\_2003\Doc\Reports\ImputeListings

### **Part A**

The following are specific ways in which the sections outliers were dealt with.

#### *Section 2: Education*

All components of reported education expenditures were compared with the median values and statically imputed if they exceeded the median value.

---

<sup>1</sup> For more information on these specifications please refer to the article: Statistical Methods for Developing Ratio Edit Tolerances for Economic Data, Journal of Official Statistics, vol. 15, no. 14, 1999 (pgs. 517-535) by Jenny Thompson and Richard S. Sigman.

<sup>2</sup> No treatment of outliers was done if the number of observations of any given product was less than or equal to 5.

### *Section 3: Health*

These were done manually since there was not that many observations that seemed unreasonable.

### *Section 4: Occupation Income*

Static imputed using the following table and parameters.

Industry Type	Upper	Lower	Replacement	
Agriculture	Daily	26,000	50	5,000
	Weekly	80,000	100	10,000
	Two weeks	160,000	200	20,000
	Monthly	1,066,502	600	55,583
	Quarterly	1,834,000	1,750	175,000
	Yearly	1,570,000	2,050	205,000
Retail	Daily	75,000	50	5,000
	Weekly	120,000	150	15,000
	Two weeks	390,000	400	40,000
	Monthly	800,000	1,000	100,000
	Quarterly	1,215,000	1,475	147,500
	Yearly	4,200,000	3,500	350,000
All other	Daily	42,000	70	7,000
	Weekly	210,000	350	35,000
	Two weeks	875,000	875	87,500
	Monthly	665,000	1,400	140,000
	Quarterly	1,800,000	4,000	400,000
	Yearly	9,800,000	7,000	700,000

Secondary Occupation was a faster technique given that there were fewer observations. The mean per week income received was used and three standard deviations were used to impute or 1,255,759 and replaced with the median weekly wage.

#### Report

#### WEEK

How many weeks did you work	Mean	N	Std. Deviation	Median
1	160845.602	533	638573.52468	50000.0000
2	93449.9114	175	330637.05560	25000.0000
3	35108.7500	80	56143.83947	11166.6667
4	40349.4000	125	74740.65621	15000.0000
5	28322.9896	230	56924.20766	8000.0000
6	23848.7040	125	53245.25480	8000.0000
7	13921.6931	81	16427.84647	7142.8571
8	21815.0069	182	62672.20243	6125.0000
9	17639.6794	61	20487.06256	8888.8889
Total	<b>78898.3668</b>	<b>1592</b>	<b>392287.59719</b>	<b>15000.0000</b>

#### Section 6: Anthropometrics

Ranges were determined by Height and Weight limits as established and used by the Editor's Manual in DHS Surveys (OCR Macro). This height and weights are disaggregated by gender and age in months.

Age	Male				Female			
	Weight (kg)		Height (cm)		Weight (kg)		Height (cm)	
	Min	Max	Min	Max	Min	Max	Min	Max
0-2	0.5	10	36	74	0.5	9.0	36	72
3-5	1	13	45	83	1	12	44	80
6-8	2	15	51	87	2	14	50	86
9-11	3	16.5	56	91	2.5	15.5	54	90
12-14	4	17.5	59	96	3	16.5	57	95
15-17	4	18.5	62	100	3.5	17.5	60	99
18-20	4	19.5	64	104	3.5	18.5	62	102
21-23	4.5	20.5	65	107	4	19.5	64	106
24-26	4.5	23	67	108	4.5	21.5	66	107
27-29	5	24	68	112	5	23	68	111
30-32	5	24.5	70	115	5	24.5	69	114
33-35	5	25.5	71	118	5	25.5	71	117
36-38	5	26	73	121	5	27	72	120
39-41	5	27	74	124	5	28	74	122
42-44	5	28	75	127	5	29	75	124
45-47	5	29	77	129	5.5	30	77	126
48-50	5	30	78	132	5.5	31	78	129
51-53	5	31	79	134	5.5	32	79	131
54-56	5.5	32	80	136	6.0	33	81	133
57-60	5.5	33	82	139	6.0	34.5	81	136

Because some of the values reported were on the order of magnitudes greater (i.e. missing decimal points), the imputation would divide by 10 and then rerun again.

There were high amounts of anthropometric data imputed. It is likely that this data is not useable. In the end the information was not imputed since it was determined that the values unworkable.

### *Section 7: Rent*

Problem Noted: There were some unreported rents as well as misreported (i.e. owner estimated rents reported as actual rents etc.).

Method: An index derived from adding the primary characteristics of the house (i.e wall, floor and ceiling). This index is used as the primary parameter or dimension of the hotdeck. The median was examined using a normalized (logarithmic distribution) and the inter-quartile range examined.

Note: An attempt was made to build a model using regression techniques based on the same index but the model overestimated the rent and was not useful in predicting a value. Due to the random and arbitrary nature of assigning rent, it was determined that dynamic imputation based on the quality index would not preserve the integrity of the system used.

The upper and lower limits require a power transformation of the related rent value.

QUALITY INDEX	Mean	N	Std. Deviation	Median	LL	UL
3	106813	465	159426	60000	12,143	474,342
4	115013	660	143406	72000	9,238	748,246
5	108960	75	105424	72000	12,143	474,342
6	186590	240	220445	120000	7,500	1,920,000
7	168706	51	139146	120000	7,500	1,920,000
8	201984	128	184014	159000	42,426	678,823
9	350213	254	490994	180000	9,468	3,194,018
10	767800	20	791694	432000	14,299	17,120,162

## **Part B**

### *Section 8 B Equipment*

These were imputed through manual observation. Any value exceeding 1,000,000 for any item was imputed with the median for that asset.

*Section 8C:* Imputed with the mean times three standard deviations for the product declared.

### *Section 8F: Inputs*

Keying errors were identified and 9 digit numbers were also replaced with the median value for the category. One major problem is noted in treating outliers is that many of them were clearly due to keying errors. These took the appearance of a leading digit followed by an entire field of "0".

### *Section 8H: Own produce (two components):*

A. Problem noted: Value of non-standard units

Method: The median unit value for a product was examined. A tolerance of 4 times the ratio was established. The median value of that unit was imputed if the limit was exceeded.

B. Problem 2 noted: Annual consumption was high due to a poorly reported amount, a keying error or problem with some enumerators cumulating the amount of product consumed per visit.

Method: The ratio used was consumption per equivalent adults by region and product. The lower limit was established at 0 or missing if the product was declared as consumed during the period of the interview. The upper limit was established as 1.5 times (inner fence) the inter-quartile. The distribution was left non-normalized.<sup>3</sup>

---

<sup>3</sup> For details on the limits established for Section 8H, 9A and 9B by individual products, and quintile or region refer to the excel spreadsheet ConsumptionLimits.xls. These tables are extensive. They can also be examined by referring to the ASCII file entitled: PartB\_Limits.lup

*Section 9A1 and 9A2: Non food*

Problem noted: Extreme values in reported amounts of consumption.

Method: Once the consumption quintiles were established the median value per product based on quintile was examined and the inter quartile ranges were established using the inner fence rule. The values examined were logarithmically transformed to lessen the effects of outlying values. This was done for two primary reasons. Since the values imputed were based on quintile medians, it was found that too many values were being imputed within the quintile without a power transformation. The curve was normalized, increasing the upper limit tolerance. Likewise, low values needed to be identified although they were only imputed if 0 or missing.

*Section 9B: Food Purchases*

Problem noted: Extreme values in reported amounts of food purchases. This could have been due to keying errors however the pattern suggests that the reporting problem was more extensive and not attributable to random keying errors. The problem is more likely due to the manner in which the diaries were kept and transferred to the questionnaire (the questionnaire required the reporting of food purchases on a quarterly or annual basis. This entailed a field aggregation from the diary to the questionnaire.

Method: Median values by region and product were examined using the inner fence rules. There was no power transformation of the items since it was felt that the extensive nature of the reporting problem required an inclusion of more values. Likewise, the only low values imputed were blank or missing.

*Section 10 Non Farm enterprises.*

Most were left untreated since it was difficult to determine correlations for imputation.

The factors in annualizing the expenditures were based on the median observations. Since a annualizing seemed to over state expenditures the following factors were used.

<b>Inpute code</b>	<b>Amount spent in last 2 weeks</b>	<b>Amount spent in last 3 months</b>	<b>Amount spent in 12 months</b>	<b>Bi Week</b>	<b>Month</b>	<b>Annual</b>
Hired Labour	14000	40000	120000	8.57	3.00	1.00
Raw Materials	40000	90000	180000	4.50	2.00	1.00
Spare Parts	42000	67500	45000	1.07	0.67	1.00
Article of resal	84000	150000	200000	2.38	1.33	1.00
Rental of Land	14000	90000	120000	8.57	1.33	1.00
Rent. of Machine	35000	30000	160000	4.57	5.33	1.00
Maintenance	15000	60000	40000	2.67	0.67	1.00
Rent. of Vehicle	40000	70000	2150000	2.00	1.00	1.00
Fuel/Lubricant	45000	50000	25000	0.56	0.50	1.00
Other Transport	10000	29400	40000	4.00	1.36	1.00
Electricity	8000	45000	120000	15.00	2.67	1.00
Water	7000	45000	66000	9.43	1.47	1.00
Taxes	2800	16500	25000	8.93	1.52	1.00
Other Expenses	10000	45000	37000	3.70	0.82	1.00
<b>Total</b>	<b>20000</b>	<b>60000</b>	<b>60000</b>	<b>3.00</b>	<b>1.00</b>	<b>1.00</b>

<b>Time unit</b>	<b>How much money goes to household?</b>	
Daily	3000	<b>100</b>
Weekly	10000	<b>30</b>
2 weeks	15000	<b>20</b>
Monthly	60000	<b>5</b>
Quarterly	100000	<b>3</b>
Yearly	300000	
Total	20000	

*Section 10D: Non-farm own consumption*

Problem noted: Outliers in reported values of consumption of non-farm items.

Method: For the purposes of the aggregate only question 5 was dealt with and used .

A static imputation of the median was used only for blanks and missing or if the amount consumed by the household exceeded the mount reported as received as payment. The mean of the amount consumed by region was used to impute.

Region	Mean	N	Std. Deviation	Median
Southern	.2086	129	.16261	.1667
Eastern	.3109	85	.21860	.2857
Northern	.2485	176	.19004	.2000
Western	.1291	86	.16083	.0500
Total	.2272	476	.19175	.1703

*Section 12B. Durable items:*

Problem noted: Extreme reported values noted for certain products.

Method: Same as the non food although the normalized lower limit was used since durable items are discrete and generally similar in valuation across regions.

Product Code	Mean	N	Std. Deviation	Median	LL	UL
301	129902.9	652	278799.6	50000	1,663	1,263,072
302	195017.9	28	151865.6	200000	14,482	1,381,068
303	42370	130	49300.92	30000	2,465	304,290
304	831224.5	49	2104345	400000	25,000	6,400,000
306	37451.46	103	20709.06	35000	12,353	80,954
307	29541.78	656	39223.98	20000	5,303	84,853
308	76803.45	435	86580.01	50000	6,889	348,372
309	135000	19	146714	75000	7,394	912,871
310	369428.6	35	751692.5	250000	9,375	2,400,000
311	313934.4	61	325014.2	250000	42,085	1,247,479
313	469007.1	70	505493.9	350000	76,614	1,794,720
314	107000	17	139153.2	70000	4,930	608,581
315	32065.22	69	17664.34	30000	8,639	81,026
316	170391.9	74	119824.2	150000	25,298	988,212
317	829375	8	871437.4	500000	4,743	47,434,165
318	3418333	12	2387189	3500000	78,678	60,372,562
322	541428.6	7	1086499	150000	15,654	849,644
323	204000	5	95289.03	200000	19,245	1,558,846

Durable sell was examined at the median. If the values were blank or 0 then half the high quartile was imputed for the value for a given year. If 1.5 the interquartile range was encountered as the upper limit, the same was used. The value for the upper limit which was used for imputing was kept in UL of the data look up file (LIMITS for Part B).

Note: A final run for treating all other outliers was done. The [following table](#) shows the ranges established. The outliers were computed as the Mean plus three standard deviations except in the case of income for non-farm enterprises and savings. 4 times the standard deviation was used for the net income range of non-farm enterprises. (Section 10D).

The [following link](#) contains the limits as established using the interquartile ranges for other items of consumption.

Listing of variables treated for outliers:

Livestock	Non Food Infrequent	Section B: Expenditure
S8AQ23_HOW_MUCH	S9AQ3_HM_SPENT	Enterprise 2
S8AQ26_VALU_SALE	S9AQ4_SPENT_3MNT	S10BQ5B_SPEND_12
S8AQ29_TOT_VALUE		S10BQ6B_SPEND_3M
	Non Food Frequent	S10BQ7B_SPEND_2W
Equipment	S9A2Q3_SPENT_12M	S10BQ12B_SPENDON
S8AQ35_VALUE_NOW	S9A2Q4_SPENT_3MO	S10BQ13B_SPENDON
S8AQ37_RENTALVAL		S10BQ14B_SPENDON
	Food Purchases	
Plots	S9BQ3_AMOUNT_TOT	Section C: Assets
S8BQ4_FARM_SIZE	S9BQ4_AMOUNT_3MT	Enterprise 1
S8BQ7_WORTH_NOW		S10C1Q2A_HM_SELL
	Non Farm Enterprises	S10C1Q2B_HM_SELL
Harvest C1	Section A	S10C1Q2C_HM_SELL
S8CQ8_VALU_SALES	S10AQ21_BORROWED	
S8CQ11_VALU_SALE	S10AQ22_MAIN_SOU	Section C: Assets
S8CQ13_TOT_VALUE		Enterprise 2
	Section B: Expenditure	S10C2Q2A_HM_SELL
Harvest C2	Enterprise 1	S10C2Q2B_HM_SELL
S8CQ22_VALU_SOLD	S10BQ5A_AMOUNT	S10C2Q2C_HM_SELL
S8CQ26_VALU_SALE	S10BQ6A_AMOUNT	
	S10BQ7A_AMOUNT	Section D: Revenue
Ag Inputs	S10BQ12A_SPENDON	S10DQ5_VAL_CONSM
S8FQ2_HOW_MUCH	S10BQ13A_SPENDON	S10DQ6_AMT_SPENT
	S10BQ14A_SPENDON	S10DQ10_VAL_CONS
Own Consumption		
S8HQ10_HM_SELL		

Section E: Net Income

S10EQ1A\_AMT\_HH  
S10EQ2A\_AMT\_PERS  
S10EQ4A\_AMOUNT  
S10EQ6A\_AMOUNT  
S10EQ8A\_TOT\_AMT

Outward Transfers

S11BAQ8\_TOT\_AMT  
S11BAQ9\_TOT\_VAL  
S11BAQ10\_VAL\_OTH

Inward Transfers

S11BBQ8\_CASH\_TOT  
S11BBQ9\_FOOD\_VAL  
S11BBQ10\_OTHER

Section 12: Credit

S12AQ6\_TOTAL\_AMT

Section 12: Assets

S12BQ3A\_PUR\_PRIC  
S12BQ3B\_PUR\_PRIC  
S12BQ4A\_SELL\_NOW  
S12BQ4B\_SELL\_NOW

Section 12: Savings

S12CQ4\_CURR\_VALU  
S12CQ5\_VALUE\_ADD  
S12CQ6\_VAL\_DRAWN

Note: All miscellaneous  
income and expenditure  
were treated for apparent  
error in keying (i.e. 9  
digit items)

4. ⇒ Programs Procedures in CSPro

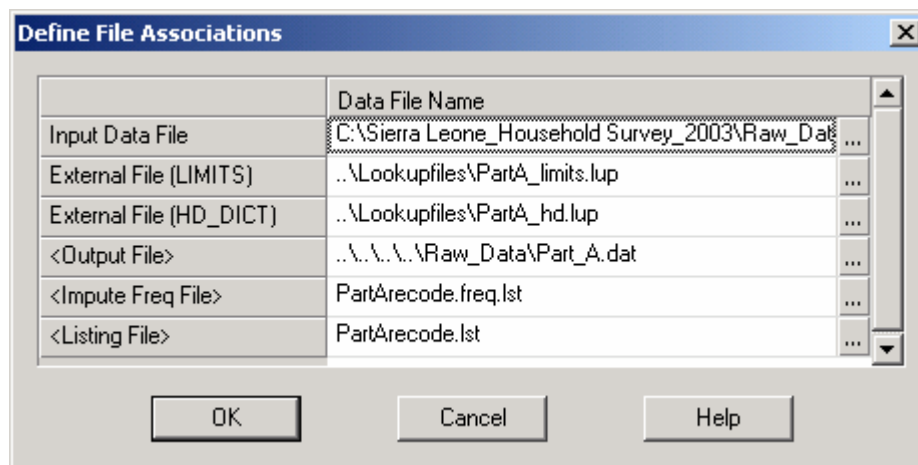
**IMPORTANT: Run PREP.BAT (DOS BAT FILE) to configure the directory structure on your hard drive.**

After running the two edit programs (Aedits and Bedits) two primary CSPro programs are required in order to recode and treat the data. These in turn have a sequential process built into them.

### **Part A**

Program: PrepA.bch (see Appendix for Program and file specifications)  
Input: C:\Sierra Leone\_Household  
Survey\_2003\Raw\_Data\Intermediate\PartA\Part\_A\_DDMMYY.dat  
Output: C:\Sierra Leone\_Household Survey\_2003\Raw\_Data\Part\_A.dat

Screen Shot:



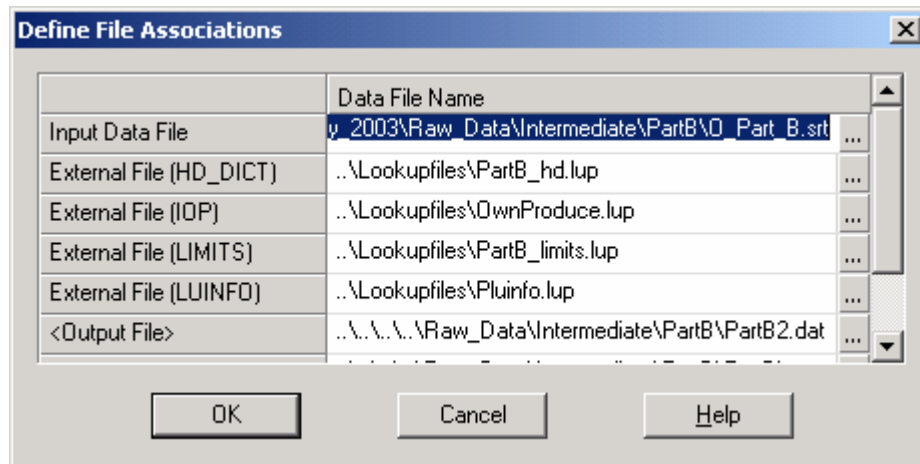
External Dictionaries: Limits.DCF      PartA\_limits.dat  
HotDecks.dCF      PartA\_HD.dat

1. *The name and path of the output file is critical as it is hard coded into the SPSS files.*
2. This program will do the following:
  - A. Recode of district 5 (Sherbro) to district 2 (Bonthe), standardize distance to water into meters, deal with reported rent for outlying values.

## Part B

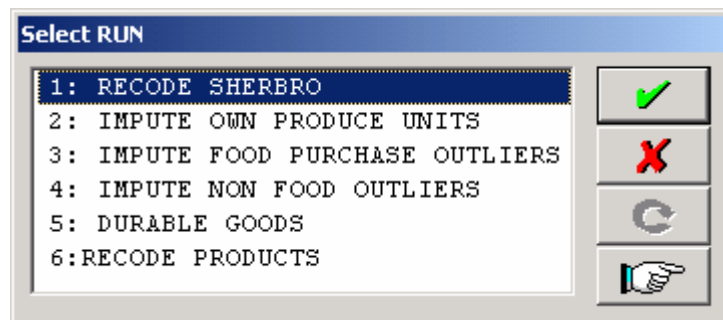
Program:	PrepB.bch (see Appendix for Program and file specifications)
Input:	Various iterations
Output:	<b>C:\Sierra Leone_Household Survey_2003\Raw_Data\Part_B.dat</b>

Screen Shot:



External Dictionaries: LuInfo.dcf	Pluinfo.lup (Household information)
Limits.dcf	PartB_limits.lup
Hotdecks.dcf	PartB_hd.lup
ImpOwnProd.dcf	Ownproduce.lup

1. *The name and path of the output file is critical as it is hard coded into the SPSS files.*
2. First Sort the data, eliminate a fishing record
3. The following is the series of imputations to be followed.



4. The first step requires a recode of district 5 (Sherbro) to district 2 (Bonthe). This output file is named Part\_B1.dat.
5. The subsequent steps should be followed in sequence.

### *Recoding*

The following recoding was done:

1. Imputation of Sherbro district to Bonthe urban.
2. Food products were recoded to a standard. Appendix A shows the recode scheme.

A new variable was defined in the data dictionary. The recodes were done into this variable.

1. ⇒ Directory Structure <sup>4</sup>

<b>Main directory:</b>	<b>Storage Purposes</b>
“\COUNTRYCODE_SURVEYNAME_YEAR”	Parent directory of all sub-directories
<i>Sub-Directory</i>	
“\RAWDATA”	All survey data sets before aggregation
“\PROGRAMS”	All computer program files
“\AGGDATA”	All aggregated data files
“\DOC”	All survey and aggregation documentation, and survey reports

There are 8 files as outlined by the World Bank template for computing the aggregate.

There are as follows:

Household composition HHSIZE.xx,	Table 1	File	1
Purchased food expenditures EXFDBUY.xx,	Table 2	File	2
Self-produced food consumption EXFDPRD.xx,	Table 3	File	3
Education expenditures EXEDU.xx	Table 4	File	4
Health care expenditures EXHLTH.xx	Table 5	File	5
Other frequent non-food expenditures EXNFDFQ.xx	Table 6	File	6
Total infrequent non-food expenditures EXNFDINFQ.xx	Table 7	File	7
Total household expenditures HHEXP.xx	Table 8	File	8

Each section below outlines the variables and the assumption used in calculating the specific item. Omissions and annualization schemes are outlined.

<sup>4</sup> Taken from the World Bank: Household Survey Expenditure Aggregation Guidelines

*Table 1 Household composition and weights summary table*

Table 1. Household composition and weights summary table																															
	Variable name	Variable definition	Aggregated from variables in X files	Syntax and data files																											
0	HHID	Unique household identification ID																													
1	HHSIZE	Total # of residents in the household		HHSIZE.xxx																											
2	FAO_ADQ	<p>Total sum of adult equivalent <b>FAO_ADQ</b> This is the aggregated sum of values of individual adult equivalent scales in a household. The FAO scale is based on the assumption that there are different consumption levels of individuals based on their ages and sex.</p> <table><tr><td></td><td>Male</td><td>Female</td></tr><tr><td>0-1 yrs</td><td>0.27</td><td>0.27</td></tr><tr><td>1-3 yrs</td><td>0.45</td><td>0.45</td></tr><tr><td>4-6 yrs</td><td>0.61</td><td>0.61</td></tr><tr><td>7-9 yrs</td><td>0.73</td><td>0.73</td></tr><tr><td>10-12 yrs</td><td>0.86</td><td>0.78</td></tr><tr><td>13-15 yrs</td><td>0.96</td><td>0.83</td></tr><tr><td>16-19 yrs</td><td>1.02</td><td>0.77</td></tr><tr><td>20 and above</td><td>1.00</td><td>0.73</td></tr></table>		Male	Female	0-1 yrs	0.27	0.27	1-3 yrs	0.45	0.45	4-6 yrs	0.61	0.61	7-9 yrs	0.73	0.73	10-12 yrs	0.86	0.78	13-15 yrs	0.96	0.83	16-19 yrs	1.02	0.77	20 and above	1.00	0.73		HHSIZE.xxx
	Male	Female																													
0-1 yrs	0.27	0.27																													
1-3 yrs	0.45	0.45																													
4-6 yrs	0.61	0.61																													
7-9 yrs	0.73	0.73																													
10-12 yrs	0.86	0.78																													
13-15 yrs	0.96	0.83																													
16-19 yrs	1.02	0.77																													
20 and above	1.00	0.73																													
3	CTRY_ADQ	Total sum of country-specific adult equivalent if a country prefers different weights as defined above		HHSIZE.xxx																											
4	WTA_HH	Household weight, HHSIZE multiplied by population weights. This is the weighting co-efficient to derive household-level estimates of indicators.		HHSIZE.xxx																											

Observations:

1. The household identification needs to be suppressed or changed to preserve confidentiality.
2. CTRY\_ADQ: The household composition using the Ghana measure was used for the survey.
3. WTA\_HH: Household weights have not been computed at the time of this report. The sample is self-weighting and based on the number of households in the sample, a weight of 251.6 is used but this is based on 1986 population statistics.

*Table 2 Purchased Food Expenditures*  
*All expenditures are annualized and in current local prices*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		
5	FDRECALL	Food purchase recall period, Two weeks=1, One month=2, Other=actual days of recall period		
6	FDBRDBY	Breads, rice in all forms, maize, wheat, cereal and cereal cassava, plantain, and potato, couscous, etc.		EXFDBUY.xxx
7	FDMEATBY	All meats and poultry, including game meats and their by products, excluding fats.		EXFDBUY.xxx
8	FDFISHBY	Fish, shell fish, land snails, land crabs, and frogs		EXFDBUY.xxx
9	FDDAIRBY	Milk, cheese, other dairy products and eggs		EXFDBUY.xxx
10	FDFATSBY	Vegetable oil, animal fats and oil-rich nuts		EXFDBUY.xxx
11	FDFRUTBY	Fruits		EXFDBUY.xxx
12	FDVEGBY	Vegetables		EXFDBUY.xxx
13	FDSWTBY	Sweets, desserts, cookies, candies, jams, honey		EXFDBUY.xxx
14	FDBEVBY	All beverages, including coffee, tea, cocoa, mineral water, soft drinks, and juices		EXFDBUY.xxx
15	FDRESTBY	Food consumed in restaurants and canteens		EXFDBUY.xxx
16	FDOTHBY	Items not mentioned elsewhere		EXFDBUY.xxx
17	FDTOTBY	Total annual expenditure of purchased foods, in current local prices, SUM( <b>FDBRDBY, FDMEATBY, FDFISHBY, FDDAIRBY, FDFATSBY, FDFRUTBY, FDVEGBY, FDSWTBY, FDBEVBY, FDRESTBY, FDOTHBY</b> )		EXFDBUY.xxx

1. The household identification needs to be suppressed or changed to preserve confidentiality.
2. FDRECALL: Need to determine given the status of the diaries.
3. Food items were aggregated accordingly and annualized depending upon the amounts reported as annual or quarterly.

*Table 3 Values of Self-Produced Foods, including gifts of food*  
*All values are annualized in current local prices*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		
18	FDBRDPR	Breads, rice in all forms, maize, wheat, cereal and cereal cassava, plantain, and potato, couscous, etc.		EXFDPRD.xxx
19	FDMEATPR	All meats and poultry, including game meats and their by products, excluding fats.		EXFDPRD.xxx
20	FDFISHPR	Fish, shell fish, land snails, land crabs, and frogs		EXFDPRD.xxx
21	FDDAIRPR	Milk, cheese, other dairy products and eggs		EXFDPRD.xxx
22	FDFATSPR	Vegetable oil, animal fats and oil-rich nuts		EXFDPRD.xxx
23	FDFRUTPR	Fruits		EXFDPRD.xxx
24	FDVEGPR	Vegetables		EXFDPRD.xxx
25	FDSWTPR	Sweets, desserts, cookies, candies, jams, honey		EXFDPRD.xxx
26	FDBEVPR	All beverages, including coffee, tea, cocoa, and juices		EXFDPRD.xxx
27	FDOTHPR	Items not mentioned elsewhere, including foods received as gifts if they cannot be categorized above.		EXFDPRD.xxx
28	FDTOTPR	Annual total monetary value of self-produced foods and foods received as gifts, in current local prices, <b>SUM(FDBRDPR, FDMEATPR, FDFSHPR, FDDIARPR, FDFATSPR, FDFRUTPR, FDVEGPR, FDSWTPR, FDBEVPR, FDOTHPR)</b>		EXFDPRD.xxx

1. The household identification needs to be suppressed or changed to preserve confidentiality.
2. Food items were aggregated accordingly and annualized depending upon the amounts reported. They were aggregated over the observation period and annualized only for observed consumption (and not expected).

*Table 4 Education Expenditures*  
*All variables are annualized and in current local prices*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		
29	EDTUITON	Tuition		EXEDU.xxx
30	EDBOOKS	Fees for text books and stationery		EXEDU.xxx
31	EDUNIFMS	Fees for uniforms		EXEDU.xxx
32	EDEXTRA	Fess for extracurricular activities		EXEDU.xxx
33	EDRMBRD	Fees for room and board, if boarding school. May include food for day scholars.		EXEDU.xxx
34	EDTRNSP	Fees for transportation to schools		EXEDU.xxx
35	EDMTNCE	Fees for school maintenance and other informally charged fees		EXEDU.xxx
36	EDOTHER	Expenditures for education not mentioned above		EXEDU.xxx
37	EDAGG	To be used only if educational expenditure is reported as a single item.		EXEDU.xxx
38	EDTOT	Annual total education expenditures in current local prices, SUM( <b>EDTUITON,EDBOOKS, EDUNIFMS,EDEXTRA, EDRMBRD,EDTRNSP, EDMTNCE,EDOTHER, EDAGG</b> )		EXEDU.xxx

1. The household identification needs to be suppressed or changed to preserve confidentiality.
2. All education expenses were taken from Part A, Section 2. Part A was disaggregated by student. Although the question was later asked in the consumption section, these reported amounts were not used.

*Table 5 Health Care Expenditures*  
*All variables are annualized and in current local prices*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		
39	HLCNSLT	Consultation and registration fees		EXHLTH.xxx
40	HLMEDC	Medication expenditures		EXHLTH.xxx
41	HLPROC	Medical procedure fees		EXHLTH.xxx
42	HLHOSPT	Hospitalization fees		EXHLTH.xxx
43	HLTRSP	Transportation fees		EXHLTH.xxx
44	HLINSURE	Health insurance		EXHLTH.xxx
45	HLOTHER	Expenditures on other health care items not mentioned above		EXHLTH.xxx
46	HLAGG	To be used only if health expenditure is reported as a single item.		EXHLTH.xxx
47	HLTOT	Annual total expenditures on health care in current local prices, SUM( <b>HLCNSLT, HLMEDC, HLPROC, HLTRSP, HLINSURE, HLOTHER, HLAGG</b> )		EXHLTH.xxx

2. The household identification needs to be suppressed or changed to preserve confidentiality.
2. Amounts reported from Part A were used in the following manner:
  - A. HLCNCNSLT: included consultations as reported in Section 3A, Part A except (omitted vaccinations, prenatal and postnatal):  
 Prenatal: taken from Section 3D and annualized  
 Postnatal: taken from Section 3C and annualized, vaccines omitted.  
 Vaccinations: taken from 3B and annualized by the number of vaccines.
  - B. HLMEDC: taken from Part A section 3A.
  - C. HLPROC: Not determined
  - D. HLHOSP: Taken from Part A, Section 3A.
  - E. HLTRSP: Taken from Part A, Section 3A.
  - F. HLINSURE: Not determined.
2. Although there are health related expenditures reported in section 9A and 9B, the recall period is less in Part A and disaggregated by individual. Also, these would have been the first questions asked and possibly more reliable. The amounts reported in Section 9A and 9B were not used to avoid overstating the aggregate.

*Table 6 Frequent Non-Food Expenditure Aggregation*  
*All expenditures are annualized and in current local prices*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		
48	NFDTBAC	Alcoholic beverages, tobacco and narcotics		EXNFDFQ.xxx
49	NFDUTIL	Expenditures paid for utilities, including water, electricity, fuel, gas, etc.		EXNFDFQ.xxx
50	NFDCLOTH	Clothing, shoes, fabrics, sheets, etc.		EXNFDFQ.xxx
51	NFDFMTN	Furnishing and routine household maintenance, tools and equipment for house and garden, and goods and services for routine household maintenance.		EXNFDFQ.xxx
52	NFDTRANS	Transportation, excluding those paid for education and health care if separable, otherwise, include the total transportation expenditures here only. Include expenditures for both public and private transportation utilization, include tickets, gasoline and maintaining fees. Purchase of private vehicles is counted as infrequent non-food expenditures in Table 7.		EXNFDFQ.xxx
53	NFDCOMM	Postal services; telephone and telefax services and internet services		EXNFDFQ.xxx
54	NFDRECRE	Recreational, cultural and sport services, newspapers, magazines, books and stationery (excluding books for schooling). Equipment, if not separable from the services, otherwise go to Table 7 for infrequent expenditures		EXNFDFQ.xxx
55	NFDTOTPR	Total monetary value of self-produced non-foods		EXNFDFQ.xxx
56	NFDINSUR	Other insurance, excluding health insurance if separable, otherwise, here only.		EXNFDFQ.xxx
57	RNTACTL	Actual rent paid or mortgage paid		EXNFDFQ.xxx
58	RNTIF	Household's estimate of what they would pay if they rented their dwelling.		EXNFDFQ.xxx
59	RNTIMP	Imputed rent based on regression or other methodologies for every household		EXNFDFQ.xxx
60	RNTHH	Household rent: actual if the household paid rent, imputed otherwise. Or imputed rent for every household.		EXNFDFQ.xxx
61	NFDFQOTH	Other frequent non-food expenditures not mentioned above		
62	NFDFQTOT	Annual total frequent non-food expenditures, in current local prices, excluding rent, SUM(NFDTBAC, NFDUTIL, NFDCLOTH, NFDFMTN, NFDTRANS, NFDCOMM, NFDRECRE, NFDTOTPR, NFDINSUR, NFDOTHER)		EXNFDFQ.xxx

*Table 6 Frequent Non-Food Expenditure Aggregation*  
*All expenditures are annualized and in current local prices*

3. The household identification needs to be suppressed or changed to preserve confidentiality.
3. NFDTABC: Taken from section 9B: Food Purchases.
3. NFDUTIL: Taken from section 7 (water, electric, rubbish) and added Section 9B Water Charges (NGUMA Valley Water Authority) and section 9B fuel and Power (278-282).
3. NFDCLOTH: although reported in Section 9A as an infrequent expenditure item, they were aggregated into the file as frequent (201-219)
3. NFDMTN: Section 9B (283-302).
3. NFDTRANS: Items reported in 9A and 9B appear to be inconsistent. Codes include what appear to be durable items and therefore subject to use and depreciation (313-324).
3. NFDCOMM: Used 325-326 section 9B.
3. NFDREDRE: Used 337-346 and 363 Section 9B.
3. NFDTOTPR: Used actual amounts declared in the 2 week period section 10D, Question 5.
3. NFDINSUR: Not determined.
3. RNTACTL: Used section 7 Question 12A. Annualized.
3. RNTIF: Used section 7 Question 12B. Note that only when the response to owner occupancy was rented was the amount in 12A used. All other occupancy status were assumed to be 12B.
3. RNTIMP: Imputed using dynamic imputation and quality of the household.
3. RNTHH: RNTACTL or RNTIF or RNTIMP
3. NFDFQOTH: Section 9B 347-360 and 362 and 364-365.

Omitted or questionable items:

- |         |  |
|---------|--|
| 303-307 | Medical  |
| 308-312 | Transport  |
| 327-332 | Education  |
| 333-336 | Small appliances   |
| 350-352 | Writing and Drawing equipment                                  |
| 361     | Jewelry  |
| 358     | Writing and Drawing Equipment                                  |
| 359     | <i>Included but questionable since restaurant is included.</i> |

*Table 7 Infrequent non-food expenditures*  
*All prices are in current local prices*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		
63	NFDSAPPL	Expenditure on small appliances and electronics, such as coffee maker, pots and pans, cell phones, cameras, DVD, etc.		EXNFDINFQ.xxx
64	NFDINVES	Large, lump sum expenditure such as purchase of vehicles, TV, computers, video cameras, etc.		EXNFDINFQ.xxx
65	NFDUSEVL	Use value of large items. if available. If survey asked for the purchasing date and the price of the large items, then annual use value could be calculated by assigning a life span for each item. Include also payments for debts, such as car payment, etc.		EXNFDINFQ.xxx
66	NFDCEREM	Ceremonial type of expenditures such as spending for weddings, funerals, coming out of age, etc.		EXNFDINFQ.xxx
67	NFDIQOTH	Other infrequent non-food expenditures not mentioned elsewhere		EXNFDINFQ.xxx
68	NFDIQTOT	Total annual infrequent non-food expenditures, SUM(NFDSELEC, NFDUSEVL, NFDIOTH, ....), users may choose whether include lumpy sums such as NFDINVES, and NFDCEREM		EXNFDINFQ.xxx

2. The household identification needs to be suppressed or changed to preserve confidentiality.
2. NFDSAPPL: Used the small appliances declared in 12B as purchased less than one year ago. These items were selected by the amount of purchase. Any item less than 100,000 Leones was expended. Any amount greater than 100,000 was depreciated using straight line five-year.
2. NFDINVES: All items aggregated as described above. Items greater than 100,000.
2. NFDUSEVL: Depreciation schedule as per Deaton and Zaidi accounting for inflation and real interest rate (guess?).
2. NFDCEREM: Taken from Section 11D: Weddings, Dowry and Funerals. However this component was excluded from the aggregate.
2. NFDIQOTH: Not determined.
2. NFDIQTOT: NFDSAPPL + NFDUSEVL+NFDIQOTH
2. Only included 275 and 276. Furniture to be determined.

Omitted or questionable items:

- 201-219 Counted in frequent items
- 230-233 Counted in Durables
- 244-247 Counted in Health items
- 234-274 Counted in either Durables or elsewhere
- 248-254 Counted in health
- 277 Included in Section 9B

## Calculated values for durable depreciation

Report

Median

Product Code	DEP
Furniture	.1607
Sewing Machine	.1303
Stove	.1686
Refrigerator	.0922
Fan	.1610
Radio	.3263
Radio Cassette	.1977
Record Player	.2403
3_In_One Radio	.1434
Video Equipment	.0893
T.V.	.0828
Camera	.1482
Iron(Electric)	.1545
Bicycle	.1975
Motor Cycle	.1363
Car	.1353
Boat	.1530
Canoes	.1318
Out Board	.1318
Total	.1692

Outboard was negative so just used 0.138 to depreciate.

*Table 8 Total annual household expenditures*

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
0	HHID	Unique household identification ID		HHEXP.xxx
1	HHSIZE	Total number of residents in the household		HHEXP.xxx
2	FAO_ADQ	See Table 1 for definition		HHEXP.xxx
3	CTRY_ADQ	Country-specific adult equivalent if differs from the above defined		HHEXP.xxx
4	WTA_HH	Household weight, HHSIZE multiplied by population weights. This is the weighting co-efficient to derive household-level estimates of indicators.		HHEXP.xxx
69	FPINDEX	Regional food price deflators, it could be similar to non-food price deflator		HHEXP.xxx
70	NFPINDEX	Regional non-food price deflator, it could be similar to food price deflator		HHEXP.xxx
71	CPIxx	Consumer price index for total expenditure from current year to xx year		HHEXP.xxx
17	FDTOTBY	Annual purchased food expenditures in local current prices, from Table 2	EXFDBUY.xxx	HHEXP.xxx
28	FDTOTPR	Annual expenditures of self-produced foods in local current prices, from Table 3	EXFDPRD.xxx	HHEXP.xxx
72	FDTOT	Total annual household food expenditure in current local prices	SUM (FDTOTBY, FDTOTPR)	HHEXP.xxx
73	FDTOTDR	Total annual hhd food expenditure in regionally deflated current prices	FDTOT/FPINDEX	HHEXP.xxx
38	EDTOT	Annual hhd education expenditure in local current prices, from Table 4	EXEDU.xxx	HHEXP.xxx
47	HLTOT	Annual hhd health expenditure in local current prices, from Table 5	EXHLTH.xxx	HHEXP.xxx
60	RNTHH	Annual hhd rent in local current prices, from Table 6	EXNFDFQ.xxx	HHEXP.xxx
62	NFDFQTOT	Annual hhd total frequent non-food expenditures, in local current prices, excluding rent, from Table 6	EXNFDFQ.xxx	HHEXP.xxx
68	NFDIQTOT	Annual hhd infrequent non-food expenditures, in local current prices, from Table 7	EXNFDINFQ.xxx	HHEXP.xxx
74	NFDTOT	Total annual hhd non-food expenditures in local current prices, including both frequent and infrequent non-food expenditures.	SUM(EDTOT, HLTOT, RNTHH, NFDFQTOT, NFDIQTOT)	HHEXP.xxx
75	NFDTOTDR	Total annual hhd non-food expenditures in regionally deflated current prices	NFDTOT/NFPINDEX	HHEXP.xxx

	<i>Variable name</i>	<i>Variable definition</i>	<i>Aggregated from variables in X files</i>	<i>Syntax and data files</i>
76	HHEXP	Total annual household expenditure in local current prices	SUM(FDTOT, NFDTOT)	HHEXP.xxx
77	HHEXPDR	Total annual household expenditure in regionally deflated current prices	SUM(FDTOTDR, NFDTOTDR)	HHEXP.xxx
78	HHEXPxx	Total annual household expenditure in regionally deflated, xx year's prices	HHEXPDR/CPIxx	HHEXP.xxx
79	SHAREFD	Share of food expenditure = $100 * (FDTOT / HHEXP)$ , this is an indicator that can be used to check the aggregation quality. Food share should decrease as household expenditure increases		HHEXP.xxx
80	PCEXP	Per capita expenditure in current local prices	HHEXP/HHSIZE	HHEXP.xxx
81	PCEXPDR	Per capita expenditure in regionally deflated current prices	HHEXPDR/HHSIZE	HHEXP.xxx
82	PCEXPxx	Per capita expenditure in regionally deflated local xx year prices	HHEXPxx/HHSIZE	HHEXP.xxx
83	QUINTILE	Quintile based on PCEXP, if not available, use PCEXP, by rural/urban areas First the sample should be divided into rural and urban sub-samples. Second, within each sub-sample, total population should be divided into 5 equally distributed quintiles weighted by "WTA_HH", based on PCEXP ranked from the lowest to the highest, The reason for creating Rural/urban quintiles is because in Africa countries, urban households typically spend much more than the rural households. If all the households are ranked based on PCEXP, the first three quintiles would be mainly rural population, while the last two quintiles would be mostly urban population. The rural/urban quintiles are designed to capture the relative differences in living standard within the rural and urban areas, respectively.		HHEXP.xxx
84	QUINTR	Quintile based on PCEXP, see above		HHEXP.xxx

In computing the aggregate non-farm own produce was omitted because it was felt that this was being double counted on the consumption side.

Although the sample size was 3720, 7 households were dropped in the aggregate:  
311502069 did not have any declared household members  
1208382035, 1309532005, 1311552181, 3103121002, 3107151118, 3203211064 had 0 food consumption. The final aggregate has 3713 households.

## 2. SPSS Programs

As per the World Bank recommendations a series of SPSS programs has been designed to facilitate the computation of the aggregate. These should be run in order. An SPSS file named “integrated.sps” is included which will execute the files in their proper order. All the programs are found in the directory:

*C:\Sierra Leone\_Household Survey\_2003\Programs\SPSS*

They follow the following order:

- A. Part\_A.SPS (Prepares PartA.dat for splitting).
- B. Part\_B.SPS (Prepares PartB.dat for exportation).
- C. Part\_A\_Section.sps (Splits and labels Part A files into their sections)
- D. Part\_B\_Section.sps (Splits and labels Part B files into their sections)
- E. Part\_C\_Section.sps
- F. [EXEDU.SPS](#)
- G. [EXHLTH.SPS](#)
- H. [HHSIZE.SPS](#)
- I. [EXFDPRD.sps](#)
- J. [EXFDBUY.sps](#)
- K. [EXNFDINFQ.sps](#)
- L. [EXNFDFO.sps](#)
- M. [HHEXP.sps](#)

All sectional files are stored in the following directory:

*C:\Sierra Leone\_Household Survey\_2003\Sectional*

All intermediary files are stored in the directory:

*C:\Sierra Leone\_Household Survey\_2003\InterFiles*

All final integrated data files are stored in the directory:

*C:\Sierra Leone\_Household Survey\_2003\Sector*

3. ⇒ File Specifications (recommended scheme for naming)

- 1. PrepA.bch** CSPro batch application designed to treat part A data for computing the consumption aggregate.

**Path:** C:\Sierra Leone\_Household Survey\_2003\Programs\CSPro\Batch\Prep

**Associated Files:**

PrepA.ord  
PrepA.mgf  
PrepA.pff  
PrepA.app  
PtA.dcf

- 2. PrepB.bch** CSPro batch application designed to treat part B data for computing the consumption aggregate.

**Path:** C:\Sierra Leone\_Household Survey\_2003\Programs\CSPro\Batch\Prep

**Associated Files:**

PrepB.ord  
PrepB.mgf  
PrepB.pff  
PrepB.app  
PtB.dcf

- 3. Data files** As the data files undergo several iterations a convention for naming these files and tracking the iterations and subsequent outputs was defined as follows.

**Path:** C:\Sierra Leone\_Household Survey\_2003\Raw\_Data\Intermediate\PartA

**Path:** C:\Sierra Leone\_Household Survey\_2003\Raw\_Data\Intermediate\PartB

**Associated files: DAT**

Part\_A\_DDMMYY.dat

Part\_B\_DDMMYY.dat

- 4. Output Files:** A series of outputs of the original data are required before arriving at the final data set.

**Path:** C:\Sierra Leone\_Household Survey\_2003\Raw\_Data\Intermediate

**Associated files : OUT**

Part\_A\_DDMMYY\_#.OUT

Part\_B\_DDMMYY\_#.OUT

- 5. Final Iteration:** A final iteration of the data files (after passing through the various stages) should be left for importation into SPSS.  
**NOTE:** This naming convention is important as it is hard wired into the SPSS coding.

**Path:** C:\Sierra Leone\_Household Survey\_2003\Raw\_Data

**Associated files: DAT**

Part\_A.dat

Part\_B.dat

- 6. Lookup files** CSPro made use of several lookup files to maintain auxiliary information. This auxiliary information is kept in data file

**Path:** C:\Sierra Leone\_Household Survey\_2003\Programs\CSPro\Batch\Lookupfiles

Associated files: **LUP**

Data Dictionaries	=>	Lookup file
HD_DICT	=>	Part_a_hd.lup OR Part_b_hd.lup
IOP	=>	OwnProduce.lup
LIMITS	=>	PartA_limits.lup OR PartB_limits.lup
LUINFO	=>	Pluinfo.lup

**7. ConsumptionLimits.xls** This is an important Excel file which contains all the limits which computes the external tables. It contains a macro activated using the ctrl-A key combination and computes the inter-quartile tolerance based on the inner fence rule.

**Path:** C:\Sierra Leone\_Household Survey\_2003\Programs\Other

**8. SPSS Files** The SPSS files to compute the aggregate are listed below. These follow a sequence and can be invoked from the file: integrated.sps.

File 1	HHSIZE.xx,	Household composition variables 1 to 4	Table	1
File 2	EXFDBUY.xx,	Purchased food expenditures variables 5 to 17	Table	2
File 3	EXFDPRD.xx,	Auto consumption variables 18 to 28	Table	3
File 4	EXEDU.xx	Education expenditures variables 29 to 38	Table	4
File 5	EXHLTH.xx	Health care expenditures variables 39 to 47	Table	5
File 6	EXNFDFQ.xx	Other frequent non-food expenditures variables 48 to 62	Table	6
File 7	EXNFDINFQ.xx variables 63 to 68	Total infrequent non-food expenditures	Table	7
File 8	HHEXP.xx	Total household expenditures variables 69 to 85	Table	8

1. ⇒ Prices

A survey of the markets was being conducted concurrent with the survey. The questionnaire was not clear regarding the necessity to convert non-standard units to standard units. Although an attempt was made to salvage this information (after six months of data collection) a review of the data showed inconsistent pricing and reporting. Some of the data is salvageable and should be brought together in a cohesive fashion.

Since there was a recode done to the final SPSS files and it may be that the product information requires the original detail, the data file does not show this recode.

Cookery was an average of all prices of cooked food from CPI data (87,88,89 Price Survey)

Dried Fish included (19,20):

Fresh fish was an average of all fish (16,17,18,23,24,25)

Smoked fish and dried fish were based on the same prices.

Oysters were shell fish

Other vegetables included cassava leaves and potato leaves.

The calculation of the price index can be reviewed in the following Excel data file for [food](#) and [non-food](#). The index was computed at the regional level due to CPI limitations.

2. ⇒ Checklist of Improvements

*There were problems associated with the questionnaire design. These are to be noted for future reference.*

The following listing is designed to provide documentation of the areas requiring improvement as they relate to data processing and the questionnaire.

The design of the questionnaire was inadequate in many ways. Issues such as poor coding, poor layout and faulty relationships within the questionnaire resulted in the reporting of either inconsistent or incoherent data.

The following are noted:

- Manner in which the diary was kept and entered into the questionnaire should be reviewed.
- Coding should follow the Classification of Individual Consumption according to Purpose (COICOP).
- Coding should be sequential for programming purposes. The sequence should be designed according to the required recode. This should be incorporated into the design.

- Coding should also keep in mind the need to define a recoding scheme for national accounts (i.e. the need to identify what are income, expenditure, asset and liability items).
- Coding should bear in mind the classifications required for the CPI.
- Coding should also bear in mind the need to assign calories to food items and should be as specific as possible.
- There should always be a consultation process in the design of the questionnaire.

## Questionnaire

- Lead questions should always be inserted to allow for a skip of an entire section. This also greatly facilitates the quality control when examining consistency of frequencies.
- Coding of “others” is extremely important. Some valuable information was lost because this was not coded. This could have been greatly facilitated using CSPro.
- Where possible units should be clearly stated (and carefully applied) and values stated as totals. Per unit values can be computed. But there needs to be consistency.
- A broader consultative process should have been worked into the questionnaire design.

## Section 1

1. Section 1 Q4: The use of “99” as a no response or not applicable code should be avoided in date fields, particularly in a YEAR field. Replaced BLANKS for all not applicable.
2. The lead in question has no field. It could be that there is a birth certificate but it is unavailable. There should be a field for the lead-in and then a skip of the field if there is no birth certificate.
3. The lines between the pages are not consistent.
4. There were too few lines. Some had to be added.
5. Why have the field for the ID of spouse if it is already given?
6. There should be a standard series of responses for “Other” throughout the questionnaire.
7. Not enough codes for ethnicity.

## Section 2: Education

1. Roster is small. Maybe the codes should be handles differently as opposed to in the column.
2. Code 000 for boarding or 9999 for couldn't pay. Under which column? Minute or hours.
3. The blank row to accommodate a heading (YR ad MT) should be grey.

#### Occupation 4B:

1. Look at activity codes. What is the difference between category I and J? What is the difference between 61 and 63 or 62 and 64? Where is teaching? Why ask 1-29 if these will be asked in section 8?
2. Activity code 44...why for building fences?

#### Occupation 4C:

1. Question 8 states 1,2 for Yes No and then asks for weeks and "00" if no. This is inconsistent.
2. Look at 4C12 and 4B9. The choices are the same but the value labels different for others. Also the value labels are not consistent throughout.

#### Occupation 4F

1. Q5. In the category defining why not looking for work, there is no student category. This should have been a code. They were placed as "Other". Also there is no retired or incapacitated.
2. What is the difference between 5 and 7?
3. What is the point of having Occupation code, industry code and activity codes? There was confusion as the enumerator switched the codes around.

#### Occupation 4G

1. 4GQ2 Occupation code substitutes 6 and 7 (which are valid codes) with Other and not working.

#### Section 6: Anthropometrics.

A. Out of 2940 about 874 were out of height and weight ranges as per DHS limits. This could have been due to a lack of decimal places in the data dictionary and possible field collection errors.

#### Section 8A: Agriculture

Land value owned frequently does not correspond to the sum of individual plots owned. Enumerator needs to check.

## Section 8B: Equipment.

- Everywhere else tries to get a unit value except this question. It is not known whether the value claimed is for the entire asset holding or per unit.
- Look at skip in equipment from Q33. Why is the question not the same as the others in that it asks..."In the last 12 months...did someone own?" Why skip to 38 and not to 36 given the way the question is structured.
- Equipment coding has some problems.

## Section 8 C

- Look at the unit of interview between Section C(1) and section B. The unit needs to be unique. Is it the individual, the farm or the crop?
- Why the skip to 13 from Part C Q10 (over the acreage)?
- In terms of men and women working on farms, it appears as there is some double counting going on. Perhaps the same people working the farm or household members. This should be stated clearly (exclude household members and only unique person where possible).
- Palm oil is a product harvested. This should be the palm nut and the processed part should be the oil. The sale of nuts is different than the sale of the oil.
- Extra page printed is confusing.

## Section 8E

- What is the point of Section 8E when there is another more detailed section on processing in Section 8G? Aggregating the two would be double counting.

## Section 8H

- The codes are different in own consumption from production.
- Non standard units were poorly recorded and frequently data entry assumed while numbers as decimals. There was a high amount of "other" rendering the non standard units virtually useless. Values declared were used as a proxy.

## Section 9 A and B

### Frequently purchase food and non-food

- The questionnaire design should report on each visit and not be grossed to annual or quarterly by the enumerator.
- Food items need to be coded with reference to their caloric value.
- Check food purchase item 187. What is this doing in here.
- Look at the double count possibility.

### Enterprise 10A

- Most of the responses were retail. It would have been good to know the specific activity of the retail sector. See 0113252351.
- Source of capital in several sections was diamonds. It should have been made clear.

### Enterprise 10B

- See the skip in Q2. If not purchased go to Q8? Why not go to next?
- Why ask for three enterprises and then stop with the expenditures of 2. Should have just done two enterprises.

### Enterprise Section 10D Revenue

- It leads by referring to response in Part B Q1. Look at where it is placed. Does this make sense? Does the question refer to the specific enterprise?
- What is meant by choice 4 in Q13, NOTAPPLICABLE? What is this category not included in all sections? Why selectively in only this out of all the questions.
- Review the logic of Part E: Net Income. Why is the distinction between personal and household? Why is the household net excluded from the question regarding where the money usually goes? Why not say then, “after deducting household and business expenditure is there money left over for X?” or something.
- Look at value labels for 10E and compare with other. Why fortnightly? Should always be the same.

## Section 11: Income

The coding method seems irrelevant. Why not record the actual numbers? Why not verify these with other section? The entire file was rendered useless because some of the entries were recorded as ranges and some were recorded as whole figures.

## Section 12 Assets:

- Assets should Account for electric availability. Sewing machine should distinguish between electric and manual. Asset differentiation for non-electrified rural areas would provide better indicators.
- Furniture should be specific. How many chairs or tables or armoires etc. Specific items of furniture may be more readily indicators of wealth in rural areas.

It is apparent that a consultation process did not take place when designing the questionnaire.

## Data Processing Issues

2. The use of sequential EAs is not necessary. The original IDs should have been kept. Inserting EAs into the master sample in the middle of the survey is not recommended particularly with sequential recoding.
2. Households should have been renumbered as sequential in order to preserve confidentiality and facilitate the count of the specific batch.
2. Data processing systems should have been designed in CSPro.
2. Data dictionaries should have been comprehensively labeled.
2. Variables names should have been consistent (i.e. ID sections).
2. Variable names should not be greater than 8 digits to facilitate the export functionality into SPSS since SPSS handle 8 character variable names.
2. Blank records should have been inserted even for no responses. This made it very difficult to evaluate data quality using frequencies.
2. In record sets where there are predetermined lists, even if there was no consumption it should have been included. Transposition is very difficult if this is not followed.
2. Space in the identifier of the data dictionary is not necessary and confusing.
2. Coding process for the “other” category not done. There is a large loss of important information. This can be dealt with easily in CSPro.
2. Values are frequently off by a factor one or two orders of magnitude. Particularly in non-farm enterprise. A check should have been integrated into the data entry system which would check against the value.
2. Consistent labeling for variables. And consistent names. For example...one variable for ID and the same naming conventions to facilitate cut and paste. Labels need to be clear and the names should never be over 8 characters long (i.e. SPSS limitations). See Section 10 Expenditure:  
S10BQ6A\_AMOUNT  
S10BQ6B\_SPEND\_3M (these are the same type of variable with same procedures).
2. Income, expenditure, capital accounts and credit accounts need to be given specific ids as opposed to sectional ids. This facilitates the identification of these variables and transfer of information for national accounts.
2. Person IDs tend to drift across the sections making matching difficult. May want to have a ghost age and sex column to track.
2. Tracking systems of the questionnaires did not exist.

2. A point of purchase is required for expenditures in order to provide the information for the CPI.

### 3. Sample Issues

The methodology used to calculate the sample can be found in the [following document](#).

*NOTE: The urban stratum of the sample was not explicitly stratified. The urban areas should be over-sampled.*

The sample is a self-weighting sample. The sample is based on 935,820 households or a grossing factor of 251.56.

### 4. ⇒ Milestone Dates:

- a. Start of survey:
- b. End of survey
- c. End of Data entry:
- d. Presentation of Aggregate: June 9, 2003
- e. End of edits: November 30, 2004
- f. Presentation of final files December 6, 2004

### 5. ⇒. Reference Documents:

For a list of reference documents go to the following folder:

C:\Sierra Leone\_Household Survey\_2003\Doc\Reference\Articles

### 6. ⇒ Tabulations

Programs for tabulating in CENTS were also done.

The quintiles have been defined in the data dictionary as additional fields. These can be found for Part A in Section 7 and for Part B in Section 9A0.

These required a simple batch program to insert the quintile. The programs are run after all the edits. However, because the CENTS programs use the original codes, these are not run through the recode program. It may be necessary to recode in SPSS.

7. ⇒ Archive Structure:

With permission from the World Bank the data was document and archived using the NESSTAR publisher.

The following are the final files:

The following are the primary files to be archived

**A. Households (Household ID)**

- EXEDU.sav
- EXFDBUY.sav
- EXFDPRD.sav
- EXHLTH.sav
- EXNFDFQ.sav
- EXNFDINFQ.sav
- HHEXP.sav
- HHSize.sav
- Section7\_Housing.sav
- Section9(8H)\_Food\_Consumption\_Purchase\_Own\_Con.sav
- Section9AB\_NonFood\_Consumption.sav
- Date\_Interview.sav

**B. Individual (Household ID + Individual)**

- Section1\_HouseholdRoster.sav
- Section2\_Education.sav
- Section3\_Health.sav
- Section4\_Occupation.sav
- Section5\_Migration.sav
- Section6\_Anthropometric.sav

**C. Enterprise Summary (Household ID + Enterprise)**

- Section10\_Non\_Farm\_Enterprise.sav
- Input/Asset (Household ID + Enterprise + Input)**
  - Section10B\_Enterprise1\_Expenditure.sav
  - Section10B\_Enterprise2\_Expenditure.sav
  - Section10C\_Enterprise2\_Assets.sav
  - Section10C\_Enterprise1\_Assets.sav

**D. Agricultural Summary (Household ID)**

- Section8\_Agricultural\_Household.sav
- Plot (Household ID + Plot)**
  - Section8B1\_PlotDetails.sav
- Crop (Household ID + Product code)**
  - Section8C1\_Harvest12months.sav
  - Section8C2\_Harvest2weeks.sav
  - Section8D\_Season.sav
  - Section8G\_Process.sav
- Input/Asset (Household ID + Input code)**
  - Section8A2\_Livestock.sav
  - Section8A3\_AgEquipment.sav

Section8F\_Inputs.sav

**E. Other Income and Expenditure (Household ID)**

Section11\_12\_Other\_items.sav

**Transfers (Household ID + Order)**

Section11BB\_TransfersIN.sav

Section11BA\_TransfersOUT.sav

**Credit (Household ID + Credit)**

Section12A\_HouseholdCredits.sav

**Assets (Household ID + Asset Code)**

Section12B\_HouseholdAssets.sav

Depreciation.sav

**Savings (Household ID and Account Code)**

Section12C\_HouseholdSavings.sav

**F. Community Questionnaire**

## 8. ⇒ Product Recodes

Recode	Description	8H	9B
001	Fresh Pepper	45	162.X
002	Salt		161
003	Pepper (Dried)		160
004	Spices and other condiments		162
005	Other miscellaneous food		176
006	Bitter Kola	29.X	122
007	Kola Nut	29.X	123
099	Cookry		167,168, 169,170, 171
100	Bread and Buns		105
101	Maize-Cob (fresh)	2	101
102	Maize-Flour/Dough	3	104
103	Millet Grain	5	102
104	Millet Flour	6	107.X
105	Rice (Paddy Grain) Local	1	103
106	Rice (Paddy Grain) Imported		
107	Guinea Corn	7	100.X
108	Sorghum	4	100.X
109	Bananas	30	135
110	Cassava-Roots	10	108
111	Cassava-Gari	11	114
112	Plantain	15	110
113	Sweet Potato	16	112.X
114	Coco Yams	14	109
115	Yams	13	111
116	Other Grains and flours	8,9	107.X
117	Cassava (Other Forms)	12	113,115
118	Other Roots and Tubers	17	112.X
119	Other starchy products		116
200	Bacon		193
201	Fresh Cow meat (no bone)	63.X	188
202	Chicken	60	142
203	Mutton	64	190
204	Pork	65	
205	Meat Sausage		195
206	Bush Fowl	62.X	145
207	Guinea Fowl	62.X	144
208	Other domestic poultry	61	143
209	Fresh Cow Meat (Bone)	63.X	189
210	Goat	66	191
211	Pig's Feet		192
212	Ham		194
213	Other domestic meat	67	197
214	Wild meat	68	
300	Dried Fish	69.x	155
301	Smoked Fish	69.x	152
302	Shell Fish	69.X	153
303	Fried Fish	69.x	156
304	Fresh Fish	69.x	154+158
305	Canned Fish		157
400	Eggs	70	146
401	Milk	71	147
402	Other Egg		146
403	Milk Powder		148

404	Baby Milk		149
405	Tinned Milk (unsweetened)		150
406	Other Milk Products		151
500	Palm Oil	21	130
501	Palm Kernel Oil		129
502	Groundnut oil		128
503	Coconut Oil	27	127
504	Other nuts and seeds	29.X	125
505	Margarine		132
506	Other vegetable oil/fats		133
507	Animal fats		126
508	Palm Nuts	25	124
509	Ground Nuts (roasted or raw)	23	120
510	Coconuts	26	125.X
511	Shea Butter	28	
600	Watermelon	31	139.X
601	Paw paw	34	139.X
602	Oranges, Tangerine	32	137
603	Mangoes	33	136
604	Pineapples	36	138
605	Other Fresh Fruit	37	139.X
606	Canned fruit		140
700	Bambara Beans	20	118
701	Cowpeas (Small)	22	117
702	Broad beans	24.X	119
703	Other pulses or legumes	24.X	121
704	Avocado pears	35	134
705	Tomatoes	40	162.X
706	Onions	41	162.X
707	Carrots	42	162.X
708	Okra	43	162.X
709	Gardens eggs, cucumbers	44	162.X
710	Cabbage/Lettuce	46	162.X
711	Spinach/Other Leafy Vegetables	47	162.X
712	Other vegetables	48	162.X
713	Cabbage		
800	Biscuits		106
801	Honey		173
802	Sugar		159
803	Ice cream, lollipop		175
804	Jams		172
805	Confectionary not frozen		174
900	Fruit juices		141
901	Coffee		163
902	Chocolate		164
903	Tea		165
904	Softdrinks		177
905	Local and imported beer		178
906	Palm Wine		179
907	Bamboo Wine		180

9. ⇒ Other links:

[Original EA Listing](#)

[Sample listing](#)

[Caloric Tables](#)

[Poverty Line Methodology](#)

10. ⇒ Contacts:

- A. Statistician General: Mr. Borbor Kandeh, SSL
- B. Deputy Director: Mr. Pessima, SSL
- C. Questionnaire and Sample Design: Mr. Adewoye McButscher, Mr. Sheik Rogers, SSL
- D. Field Work: Mr. Carter
- E. Data Entry: K.B. Danso Manu: CENTRY and CONCOR programs
- F. Data Management: Mrs. Davis and Mr. Tegeka Conteh
- G. Data editing, aggregate and archiving: Mr. Geoffrey Greenwell
- H. World Bank Advisor: Mr. John Ngwafon