Sampling design for Cambodia child labour survey 2001

4.2.1 Number of households sampled in 2001 and sampling fraction

The Cambodia Child Labor Survey (CCLS) conducts in the nationwide representative sample of 12000 households within 600 sampling units (villages) and target to the children who aged between 5 and 17 years old. It is sample to provide information on child labor forces and research on condition of child labor in the various field of social and economics. Therefore the purpose to be selected more observation in urban than rural area. Survey estimates will be produced for three super strata,viz.,1-Phnom Penh, 2-Other Urban with 11 sub-strata and 3-Rural area with11sub-strata.

Table 1: Number of villages and Households and Number of households sampled in 2001

Domain	Number of village	Total Number of households	No.of household in Sampled villages	No. villages Sampled CSSPPS	No. households Sampled CSSEQP	
					Per village	Total
1- Phnom Penh	637	173,678	69,554	128	20	2,560
2- Other Urban	907	224,950	115,961	280	20	5,600
3- Rural	11,862	1,790,035	48,064	192	20	3,840
Total	**13,406**	**2,188,663**	**233,579**	**600**	**20**	**12,000**

Table 2: Sampling Fraction of Village and Households

Domain	Villages (%)	Households in Selected Village (%)	Household Overall (%)
1- Phnom Penh	20.1	3.68	1.47
2- Other Urban	30.8	4.83	2.49
3- Rural	1.6	7.99	0.21
TOTAL	4.5	5.13	0.55

4.2.2 Sampling frame

The result of General Population Census of Cambodia 1998 was taken for using as based of sampling design of this survey.

4.2.3 Characteristics of the survey

The sample for CCLS is stratified sample selected in two stages:

A- The first stage selection

The primary sampling units PSUs, villages will be selected from the list of village for every stratum within the domain is to be listed in order of: province/city, district, commune, and village code. The method of Circular Systematic Sampling with Probability of inclusion of village Proportional to its Size (CSS-PPS) will be use to select of villages.

The frame would contain in addition to the code for the above identification particulars, the name of the villages as well as the number of household in the village as know at that time. The number of household in the village as in the sampling frame will be as it size. The actual number of households in selected village will be obtain later by direct listing

in this will generally be different from the number recorded in the sampling frame. To avoid ambiguity, the number of households in a village as recorded in the frame will always be referred to as its size. The total of village in Phnom Penh will be denoted by N and the size of i-th village in the domain by Si, for i= 1, 2, ...,N. The number of village to selected from the domain by the (CSS-PPS) selection of sub-sample is explained bellow:

Step 1: Prepare a tabular layout using seven columns and N row - one for each in the domain. The seven columns are: 1- Serial number of village (i), 2- Identification code of the village consisting of province, district, commune, and village codes, 3- Name of village, 4-Size of village(Si), 5- Lower limit of selection interval (Li): Lo=1 and Li=S1+S2+.....+S(i-1) +1. 6-Upper limit of selection interval (Ui): Ui=S1+S2+.....+Si, the cumulative size, for i=1,2,.....,N, (note that column 6 has to be calculate before column 5, 7- Order of selection.

Step 2: calculate the sampling interval as I=U_N/n, round off to the nearest integer. Here U_N is the last entry in column 6, the total of the size of village.

Step 3: Chose a random integer R in the range 1 to U_N from the supplied table random number.

Step 4: Take R1=R, and generate a sequence of n selector number R1, R2, R3,.....,Rn in the following way. Get the next R from previous R by adding (I) in to it, and if the total exceed U_N, by subtracting U_N from the total.

Thus for j = 2,3,.....,n, Rj=Rj-1 +I, If this does not exceed U_N ; =(Rj-1+I)-U_N, otherwise.

Step 5: If the j-th Selection number Rj fall in the i-th selection interval (Li,Ui),that is , if Li<=Rj<=Ui, selected village no.i, as the j-th village to be sampled and write (j) in the column for (order of selection against village no. i.

In other words, the number of households in the village will be used as the measure of size. Sample village selection will be done through the use of a computer program. The Sample allocation is given below.

Table 3: Distribution of villages and households in sampling frame

Sn.	Code and Name of Provinces	No. of Villages			No. of Households		
		Urban	Rural	**Total**	Urban	Rural	**Total**
1	01 Banteay Meanchey	59	544	**603**	18296	92559	**110855**
2	02 Battambang	62	507	**569**	25421	120137	**145558**
3	03 Kampong Cham	31	1717	**1748**	8112	303962	**312074**
4	04 Kampong Chhnang	26	520	**546**	7661	74648	**82309**
5	05 Kampong Speu	56	1263	**1319**	7432	107017	**114449**
6	06 Kampong Thum	55	682	**737**	12285	94427	**106712**
7	07 Kampot	15	462	**477**	6053	98920	**104973**
8	08 Kandal	20	1067	**1087**	10246	195509	**205755**
9	09 Kaoh Kong	11	116	**127**	5351	19179	**24530**
10	10 Kracheh	74	183	**257**	14738	34446	**49184**
11	11 Mondol Kiri	14	84	**98**	1260	4371	**5631**
12	13 Preah Vihear	31	173	**204**	4072	16547	**20619**
13	14 Prey Veng	42	1094	**1136**	10868	183205	**194073**
14	15 Pursat	63	412	**475**	10796	56962	**67758**
15	16 Rattanak Kiri	16	224	**240**	3165	13557	**16722**

16	17 Siem Reab	77	805	**882**	20418	105055	**125473**
17	18 Sihanouk Vile	85	0	**85**	27584	0	**27584**
18	19 Stueng Treng	17	111	**128**	4300	9846	**14146**
19	20 Svay Rieng	18	672	**690**	4100	94115	**98215**
20	21 Takeo	40	1076	**1116**	7233	147656	**154889**
21	22 Oddar Meanchey	21	83	**104**	3779	8388	**12167**
22	23 Krong Keb	16	0	**16**	5295	0	**5295**
23	24 Krong Pailin	58	0	**58**	4060	0	**4060**
SUB_TOTAL		**907**	**11795**	**12702**	**222525**	**1780506**	**2003031**
24	12 Phnom Penh	405	232	**637**	96407	75864	**172271**
GRAND_TOTAL		1312	12027	13339	318932	1856370	2175302

Table 4: Distribution Number of Villages and Households to be selected by provinces in Other Urban and Rural

Sn.	Code and Name of Provinces	No. of Villages			No. of Households		
		Urban	Rural	Total	Urban	Rural	Total
1	01 Banteay Meanchey	23	10	**33**	460	200	660
2	03 Kampong Cham	10	32	**42**	200	640	840
3	08 Kandal	13	21	**34**	260	420	680
4	21 Takeo	9	16	**25**	180	320	500
5	15 Pursat	14	6	**20**	280	120	400
6	12 Phnom Penh	73	55	**128**	1460	1100	2560
7	04 Kampong Chhnang	10	8	**18**	200	160	360
	05 Kampong Speu	9	12	**21**	180	240	420
8	02 Battambang	32	13	**45**	640	260	900
	24 Krong Pailin	5	0	**5**	100	0	100
9	14 Prey Veng	14	20	**34**	280	400	680
	20 Svay Rieng	5	10	**15**	100	200	300
10	07 Kampot	8	11	**19**	160	220	380
	09 Kaoh Kong	7	2	**9**	140	40	180
	18 Sihanouk Vile	35	0	**35**	700	0	700
	23 Krong Keb	6	0	**6**	120	0	120
11	10 Kracheh	18	4	**22**	360	80	440
	13 Preah Vihear	5	2	**7**	100	40	140
	19 Stueng Treng	5	1	**6**	100	20	120
	11 Mondol Kiri	2	1	**3**	40	20	60
	16 Rattanak Kiri	4	1	**5**	80	20	100
12	17 Siem Reab	26	11	**37**	520	220	740
	06 Kampong Thum	15	10	**25**	300	200	500
	22 Oddar Meanchey	5	1	**6**	100	20	120
	TOTAL	353	247	**600**	7060	4940	12000

The selection probability for village i in stratum h will be computed as:

$$P_{hi} = (n_h / \sum_i MF_{hi}) \times MF_{hi} \qquad \text{(Form . 1)}$$

where :

P_{hi} = probability of selecting the i^{th} PSU in h stratum

n_h = number of sample villages to be drawn from h stratum

MF_{hi} = number of households in village i as recorded in the sample frame

ΣMF_{hi} = total number of households in stratum h as recorded in the sample frame

The design weight for the primary sampling unit is inversely proportional to its selection probability.

$$W_{hi} = 1 / P_{hi} = (1 / n_h * MF_{hi} / \sum_i MF_{hi})$$

$$= \sum_i MF_{hi} / n_h . MF_{hi}$$

If required to calculate at villages level of any characteristics (study variable) its denoted by y_{hi} , then the calculation y_{hi} in the stratum is $\hat{Y}_h$ based on sample villages

$$\hat{Y}_h = \sum_i w_{hi} . Y_{hi}$$

The estimate for total $\hat{Y}$ for all strata is the sum of estimate for each strata

$$\hat{Y} = \sum_i \hat{Y}_h = \sum_h \sum_i \left(\sum_i MF_{hi} / n_h . Mf_{hi} \right) . y_{hi}$$

B. Second Stage Selection

For each sample village (PSU), a field listing operation will be undertaken except for large villages. Large villages will be segmented first, comprising about 210 households based on the current household estimates by the village leaders. a segment will then be chosen randomly in which a complete listing of households will be done. This will entail carrying out a complete canvass of the PSU in order to make a current and complete listing of households contained within. The procedure will involve creating a sketch map for the PSU where physical boundaries in the village and the location of each household will be sketched. Canvassing, on the other hand, will entail a systematic covering of the entire village following a prescribed path of travel in order to make sure that all housing units in which the households reside will be accounted for.

After the listing operation is completed, a fixed sample size of 20 households will be selected in each PSU.The selection will be carried out using a circular systematic random sampling with a random start (CSS). The sampling interval will be equal to the current households in the PSU or segment divided by fixed 20, as the case maybe.

The selection probability for households j in village i will be computed as:

$$P_{hij} = (x_{hi} / MA_{hi}) \qquad \text{(Form. 2)}$$

where :

P_{hij} = probability of selecting the j^{th} households in village i

x_{hi} = number of households to be selected in village i^{th} in stratum h (20)

MA_{hi} = number of households actually residing in village i in stratum h at the time of the survey. This will be equal to the number of households listed in the village i if the village is not segmented. If only a segment of village i will be listed, this is equal to the number of households listed in the selected segment.

C. Basic Weights for Households information

The basic weight for a small villages (less than 211 households) computed as bellow:

$$W_{hij} = \sum_{I} M_{hi} * MA_{hij} / (n_h * x_{hi} * MF_{hij}) \quad \text{(Form 3)}$$

where :

$\sum MF_{hi}$, n_h are as defined in (form 1)
x_{hi} and MA_{hi} are as defined in (Form 2)

D. Weights for Households information for Large PSU's

Since large villages or those with an estimated number of households with 211 or more will be segmented, and only one randomly selected segment will be listed, and from the selected segment, the sample of 20 households will then be selected, thus, the weight for that PSU will be as follows:

$$W_{hij} = \left(\sum_{i} M_{hi} * MAL_{hi} * G \right) / \left(n_h * x_{hi} * MF_{hi} \right) \quad \text{(Form 4)}$$

where :

G = the number of segment in village i
MAL$_{hi}$ = the actually households to be list in selected segment
$\sum MF_{hi}$, **n$_h$** , **x$_{hi}$**, **MF$_{hi}$** = are as defined earlier

E. Estimation Procedure

E.1 Estimation Procedure for Household Information

The estimate of the stratum total is given by the following formula:

$$\hat{Y}_h = \sum_i \sum_j w_{hij}\, y_{hij} \quad \text{for} \quad \begin{matrix} j = 1, 2, ..., x_{hi} \\ i = 1, 2, \ , n_h \end{matrix} \quad \text{(Form. 5)}$$

where :

$\hat{Y}_h$ = estimate of characteristic **y** for stratum **h**

y$_{hij}$ = any characteristic of household **j** in sample village **i** in stratum **h**

x$_{hi}$ = number of sample households in village i

n$_h$ = number of sample villages in stratum h

W$_{hij}$ = are as defined in (Form 3, 4)

The estimate for the total for all strata, $\hat{Y}$, was computed as the sum of the estimates for each stratum. That is:

$$\hat{Y} = \sum \hat{Y}_h \quad \text{(Form 6)}$$

h = 1 to 23

$$= \sum_h \sum_i \sum_j w_{hij}\, y_{hij}$$

$$= \sum_h \sum_i \sum_j \left[\sum_i M_{hi} * MA_{hij} / \left(n_h * x_{hi} * MF_{hij} \right) * y_{hij} \right] \text{(small village)}$$

or (village divided segments)

$$\left[\qquad \right]$$

$$= \sum_h \sum_i \sum_j \left(\sum_i M_{hi} * MAL_{hi} * G \right) / \left(n_h * x_{hi} * MF_{hi} \right) * y_{hij}$$

The estimated stratum mean is a ratio computed as:

$$r_h = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{\sum_i \sum_j w_{hij} y_{hij}}{\sum_i \sum_j w_{hij} x_{hij}} \qquad \text{(Form 7)}$$

where

y_{hij}, $n_{h,}$, x_{hi}, w_{hij} is as defined earlies

$$\sum_i \sum_j x_{hij} = 1 \quad \text{for} \quad j = 1, 2, ..., x_{hi}; \quad i = 1, 2, \ , n_h$$

The population mean is also a ratio, r, which was estimated using the following formula:

$$r = \frac{\sum_h \sum_i \sum_j w_{hij} \ y_{hij}}{\sum_h \sum_i \sum_j w_{hij} \ x_{hij}} \qquad \text{(Form 8)}$$

where

y_{hij}, $a_{h,}$, n_{hi}, w_{hij} is as defined earlies

$$\sum_i \sum_j x_{hij} = 1 \quad \text{for} \quad j = 1, 2, ..., x_{hi}; \quad i = 1, 2, \ , n_h; \quad h = 1, 2..... 23$$

E. 2 Estimation of Variances

In order to provide a basis for assessing the reliability or precision of CCLS estimates, the estimation of the magnitude of sampling error in the survey data shall be computed. Since most of the estimates from the survey are in the form of weighted ratios, thus variances for ratio estimates will thus be presented. The procedures in deriving the estimates for the variances are described below.

All variances of the ratio estimates will be given of the form:

$$\text{var}(r) = \frac{1}{\bar{x}^2} \sum (1 - f_h) \left(n_h / n_h - 1 \right) \sum \left(z_{hi}^2 - z_h^2 / n_h \right) \quad \text{(Form 9)}$$

where

$$r = \bar{y} / \bar{x} = \sum\sum\sum w_{hij} y_{hij} \Big/ \sum\sum\sum w_{hij} x_{hij}$$

$$\hat{x}^2 = \hat{X}^2 = \left(\sum\sum\sum w_{hij} X_{hij} \right)^2$$

$$z_{hi} = y_{hi} - r\, x_{hi}$$

$$y_{hi} = \sum_j w_{hij} y_{hij}$$

$$x_{hi} = \sum_j w_{hij} x_{hij}$$

$$Z_h = \sum_i z_{hi}$$

n_h = number of sample villages from stratum h

w_{hij} = weight for each individual in the sample household

The variance of the ratio estimate, r_h, in stratum h is given by the formula:

$$\text{var}(r_h) = \frac{1}{\bar{x}_h^2} (1 - f_h) \left[n_h / (n_h - 1) \right] \sum_i \left(z_{hi}^2 - z_h^2 / n_h \right) \quad \text{(Form 10)}$$

where

$$\bar{x}_h = \sum\sum\sum w_{hij} x_{hij}$$

f_h, n_h, z_{hi}, Z_h and are as defined earlier