2012 Enquête sur les Conditions de Vie de Ménage Après le Séisme Sample Selection and Weighting

This document outlines the procedures used to select the sample for the *2012 Enquête sur les Conditions de Vie de Ménage Après le Séisme* (ECVMAS). The items covered include a brief description of the sample size calculations, a discussion of the selection, the outcome of the fieldwork, and the calculation of the expansion factors.

Sample Size Calculations

Though the ECVMAS 2012 was designed as a cross-cutting multi-topic household survey, it was necessary to limit the number of variables of interest to do the sample size calculations. For these purpose, the primary variables were participation in the labor force and unemployment.[1] The mean, standard deviation, and design effects were calculated from the most recent household survey data available, the 2007 *Encuesta de Empleo y Economia Informal* (EEEI), and various sample distributions were considered, including equal, proportional, and optimal. Given the need to take into account the relative importance of both variables, as well as maintain minimum sample sizes with geographic domains of interest, a practical allocation was used for the final distribution.

Table 1 : Selected Design Effects

	mean	se	deff	deft
Unemployment Rate				
overall	0.1684	0.0085	5.34	2.31
rural	0.2782	0.0096	3.45	1.86
urban	0.0936	0.0124	5.18	2.28
Net Activity Rate				
overall	0.4684	0.0066	4.08	2.02
rural	0.4482	0.0071	3.50	1.87
urban	0.4832	0.0104	2.57	1.60

Source: Labor Market Survey 2007

The sample is a two-stage stratified cluster sample with a total of 500 clusters or *sections d'énumération* (SDE), and is designed to be representative at the level of the ten départements as well as nationally. In addition, a separate stratum was created to represent the population living in internally displaced person camps. Each SDE is expected to contain eight households in rural areas and 16 households in urban areas. This decision was made to limit the size of the rural clusters keep the workload of the rural teams with the constraints of the field calendar. Since the calculations from the previous dataset indicated very high intracluster correlation coefficients for the variables of interest in rural areas, the cluster size could be reduced without a substantial loss in statistical power. To ensure representativeness at the level of the département, a minimum of 30 SDEs within a stratum (department) was maintained to achieve a minimum level of precision[2], while the remained of the sample was apportioned based upon the variation of the principle variables of interest.

Table 2: Final Distribution of SDEs for ECVMAS

department	rural	urban	Total
Non-Camp Sample			
Aire Metro	0	90	90
Artibonite	32	18	50
Centre	32	7	39
Grand'Anse	26	7	33
Nippes	27	3	30
Nord	26	18	44
Nord-Est	19	12	31
Nord-Ouest	27	9	36
Ouest	41	7	48
Sud	29	7	36
Sud-Est	28	5	33
Total	287	183	470
Camp Sample			30
Total			500

[1] Consumption and poverty are also key indicators to be produced from the survey, but reliable data was not available on these measures. It is likely, however, that poverty and unemployment are highly correlated, and therefore the two employment variables should be sufficient.

[2] This is based on achieving a 95% certainty for a ± 5 percentage point confidence interval for a proportion requires a sample size of 384. Since the design effects are unknown, the sample size is adjusted by a factor of 1.2 as compensation.

Therefore, the ECVMAS sample includes 470 SDEs from the échantillon-maître to represent the non-camp population, and 30 camp segments to represent the camp population. Of these 183 are urban SDEs, and 287 are rural. The total sample size is therefore expected to be 2,664 urban households and 2,296 rural households, for a total sample size of 4,960 (8 households per SDE except for SDEs in Aire metro and camps where this number will be 16).

Sample Selection

The sample for the ECVMAS survey was selected from two separate sampling frames – the Haiti master sampling frame or *Échantillon-maître*, which covers all non-camp areas, and the information from the Camp Coordination and Camp Management Cluster[3], which covers populations currently living in camps and therefore excluded from the échantillon-maître.

Master Sample

The sample for the ECVMAS is selected from the Haiti échantillon-maître (constructed in July 2011 with further documentation available from the *Institut Haitien de Statistique et d'Informatique* [IHSI]). The échantillon-maître is based on the selection of 1500 SDEs from the approximately 12,000 in the country. The master sampling frame can be subdivided into four statistically identical replicates, each of which is representative at the national level, at the level of Aire Metropolitan/Other Urban/Rural, Zones Affected by the Earthquake/Zone Not Earthquake Affected, and at the level of each of the 10 départements. The number of SDEs per département was determined in the échantillon-maître using probability proportional to the population raised to the 0.35 power at the département level, and the distribution within each département between urban and rural was proportional to the population. This procedure was used to guarantee minimum sample sizes within each of the 10 départements.

Sub-sampling into ECVMAS sample

The sub-sampling process starts with one of the three replicate sets from the échantillon-maître. Since the procedure for sampling the SDEs into the échantillon-maître oversamples rural areas (as the départements with the lowest population which require oversampling are predominantly rural), it was decided to maintain all of the selected urban SDEs in the ECVMAS. Therefore only rural areas are sub-sampled and steps were taken to continue to maintain a minimum of 30 SDEs per département. The final outcome continues to reflect the oversampling of rural areas in the échantillon-maître, but the distribution of percentages is now closer to the population distribution.

Table 3 : Distribution of the Percentages of Estimated Population and Sample Population

	Estimated Population		échantillon-maître		ECVMAS		probability of selection from échantillon-maître into ECVMAS	
département	rural	urbain	rural	urbain	rural	urbain	rural	urbain
Aire Metro	0.0	21.2	0.0	18.0	0.0	19.1		1.00
Artibonite	5.2	5.0	7.2	3.6	6.4	3.8	0.89	1.00
Centre	5.8	1.4	6.8	1.4	7.2	1.5	0.94	1.00
Grand'Anse	5.6	1.7	5.6	1.4	5.5	1.5	0.93	1.00
Nippes	3.9	0.5	5.6	0.6	5.7	0.6	0.96	1.00
Nord	5.6	3.5	5.6	3.6	5.5	3.8	0.93	1.00
Nord-Est	4.1	2.5	4.2	2.4	3.8	2.6	0.90	1.00

[3] Further information and documentation is available at www.cccmhaiti.info.

Nord-Ouest	5.7	2.3	6.0	1.8	5.3	1.9	0.90	1.00
Reste-Ouest	10.6	1.3	9.0	1.4	9.1	1.5	0.91	1.00
Sud	6.0	1.7	7.0	1.4	6.6	1.5	0.83	1.00
Sud-Est	5.7	0.7	6.4	1.0	5.7	1.1	0.88	1.00
Total	58.3	41.7	63.4	36.6	61.1	38.9		

The probabilities of selection are 1 for urban areas, since all SDEs selected into the échantillon-maître are then selected into the ECVMAS. The other probabilities of selection range from 0.83 in Sud rural to 0.96 in Nippes rural.

Camp Selection

The selection procedures used to select the camp sample are essentially the same as those to be used in the non-camp sample, though they will differ in implementation. A current list of camps and their populations is maintained by the International Organization for Migration (IOM) and this list is updated every three months. The latest camp list was completed in July 2012. The most recent update shows 447 active camps with a population ranging from 30 people to 36,000 people. The total camp population is estimated to be approximately 386,749 individuals, representing roughly 3.5 percent of Haiti's total population.

The majority of the camp population lives in large camps (more than 500 families). Since the selection of the first stage of the sample will be done with probability proportional to size, it was necessary to segment the large camps into enumerable pieces. This was done because to costs of listing a camp of 8,500 families would be prohibitive both in terms of money and logistical support. Therefore large camps will be segmented into equal-sized pieces of no more than 200 households, then these segments were randomly selected.

The physical segmentation of the selections will be done following the selection in coordination with IOM. For example, if Camp A has a population of 1400 households, and was selected twice by the proportional to size selection program, the camp would be divided into seven equally sized segments of 200 households. These seven segments should have clearly defined boundaries and be defined on a map prior to the random selection. Two of these seven segments would be selected into the survey.

Table 4: Population Living in Camps (by size)

Number of Households	Number of Sites	Households	Individuals
1-49	159	4,737	17,774
50-99	109	7,602	26,613
100-199	83	11,257	40,690
200-499	56	18,366	67,815
500-999	24	17,386	69,388
1000+	16	37,583	164,469
Total	447	96,931	386,749

Source: Haiti CCCM Cluster (July 2012)

Outcome of Fieldwork

The fieldwork was conducted from August to December 2012 by mobile teams. In total, five SDEs needed to be replaced during the fieldwork due to insecurity. These replacements were randomly selected within the same section of the commune if possible, otherwise within the commune. Within selected SDEs, the required number of households (16 in urban areas and 8 in rural areas) were interviewed in all cases, with the exception of two SDEs. In one of those two, only 12 households were available, and therefore only 12 were interviewed. In the second SDE, issues with the cartography prevented the teams from locating and interviewing the complete sample. The weights were adjusted to compensate for these altered probabilities of selection.

Weight Calculations

Because two separate sampling frames and difference sampling methodologies were used, the weights were constructed separately for the household sampling frame and the camp sampling frame. The weights were then combined and a post-stratification population adjustment was applied to smooth any inconsistencies resulting from the time discontinuity of the construction of the two frames.

Household Weights

The households sample was selected in three stages. The first stage selection is the selection of one of three possible replicates of the master sampling frame.

$$p_1 = \frac{1}{3}$$

The second stage includes the probability proportional to size selection of 470 total clusters, stratified according to the distribution in table 2. The probability of selection at this stage is :

$$p_2 = \left(\frac{m \cdot c_h}{C_H}\right)$$

where *c* is the number of households in stratum *h*, and C_H is the total population of households living in stratum *H*. Finally, either 8 or 16 households were selected from the total from the list in the segment.[4] The probability of selection at this stage is :

$$p_3 = \frac{n}{C'_H}$$

where *c'* is the number of households recorded during the listing operation and *n* is the number to be selected (8 in rural areas and 16 in urban areas). The total weight for the household sample is therefore the inverse probability of selection or:

$$w = \frac{3}{p_2 \cdot p_3}$$

Camp Household Weights

The camp sample was selected in three stages, and therefore the weight calculations include three components. The first stage selection includes the probability proportional to size selection, as well as the subsampling of 30 of the selected 40 units into the primary sample (with the remaining as replacements). The probability of selection at this stage is :

$$p_1 = \left(\frac{40 c_i}{C}\right)\left(\frac{30}{40}\right)$$

[4] The instructions for the listing of the segments in the camp sample were given to the implementing partner as to list all households within the pre-defined and selected segment. During the implementation, however, eight segments were listed only up to 200 households. This means that the value of S_i for those segments is underestimated, and therefore the weight slightly overestimated. This difference though is likely to be negligible for most estimations.

where *c* is the number of households in camp *i*, and *C* is the total population of households living in camps according to the July 2012 camp list. In the second stage of selection, the selected camps are divided into segments of approximately 200 households, and segments are then selected using simple random sampling to be listed and enumerated. The probability of selection at this stage is :

$$p_2 = \left(\frac{s}{S_i}\right)$$

where *s* is the total number of segments to be selected (one in all cases except Canaan and Ancien Airport Militaire, which have three and two, respectively), and S_i is the total number of segments in camp *i.* Finally, 16 households were selected from the total from the list in the segment.[5] The probability of selection at this stage is :

$$p_3 = \frac{16}{c'_i}$$

where *c'* is the number of households recorded during the listing operation and *n* is the number selected. The total weight for the camp sample is therefore the inverse probability of selection or:

$$w = \frac{1}{p_1 \cdot p_2 \cdot p_3}$$

Post-stratification

To reduce the overall standard errors and weight the population totals up to the known population figure, a post-stratification correction (wps) is applied. While this correction does reduce overall standard errors (see Little et al, 1997), in this context was primarily used to adjust total population sizes to census projection estimates as the totals from the survey were almost uniformly low across strata (likely the result from uneven quality in the SDE listing operation).

The final panel weight is then:

$$w_{final} = w * w_{ps}$$

[5] The instructions for the listing of the segments in the camp sample were given to the implementing partner as to list all households within the pre-defined and selected segment. During the implementation, however, eight segments were listed only up to 200 households. This means that the value of S_i for those segments is underestimated, and therefore the weight slightly overestimated. This difference though is likely to be negligible for most estimations.