Sample for the Enquête sur les Conditions de Vie de Ménage Après le Séisme (ECVMAS)

Summary

The sample for the ECVMAS survey was selected from two separate sampling frames – the Haiti Échantillon-maître, which covers all non-camp areas, and the information from the Camp Coordination and Camp Management Cluster[1], which covers populations currently living in camps and therefore excluded from the échantillon-maître. 500 SDEs/camp segments have been selected into the survey sample, from which 15 households will be drawn, for a total sample size of 7,500 households.

Sample Composition

In situations where limited data is available on which to base sample size calculations, a common guideline is to require a minimum of 30 SDEs within a stratum to achieve a minimum level of precision[2]. This guideline was used in the construction of the échantillon-maître in order to ensure minimum representativeness in each of the ten départements. This guideline was also applied to the camp population to ensure the same minimum level of representativeness, effectively treating the camp population as a separate département for the purposes of sample selection. 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. The échantillon-maître is divisible into three statistically identical replicates of 500 SDEs. Since only 470 of the 500 SDEs are required for the non-camp sample, the replicate was sub-sampled.

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 sections d'énumération (SDE) from the approximately 12,000 in the country. The master sampling frame can be subdivided into three 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

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.

---

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

[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.

Table 1 : 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
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.

Table 2: Final Distribution of SDEs for ECVMAS

department	rural	urbain	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

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 May 2011, with the next update scheduled from August 2011. The most recent update shows 1,001 active camps with a population ranging from two people to 34,000 people. The total camp population is estimated to be approximately 634,000 individuals, a decrease of seven percent from the previous estimate from March 2011, and representing roughly six 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 500 households, then these segments will enter individually into proportional to size selection.

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 3500 households, the camp would be divided into seven equally sized segments of 500 households. These seven will enter into the PPS selection model. If Camp A segment 3 and Camp A segment 7 are chosen, it will then be necessary to draw on a GIS map the seven segments, randomly number them, and then select segments 3 and 7. It is extremely important that the segments be numbered randomly, otherwise it is possible to introduce substantial bias into the relatively small camp sample.

Table 3: Population Living in Camps (by size)

Number of Households	Number of Sites	Households	Individuals
1-9	131	673	2,495
10-19	153	2,188	7,834
20-99	429	20,829	76,520
100-499	218	45,415	175,028
500-999	45	30,245	116,614
1000+	25	59,294	255,803
Total	1,001	158,644	634,294

Source: Haiti CCCM Cluster (DTM v2.0 – May 2011)

In addition, five replacement camp segments will be selected to be used if one of the selected camps is too small at the time of the survey or the security situation does not permit surveying in these areas.

Number of Households Per SDE

The calculations for the number of households per SDE are very difficult due to the lack of quality data on the two main variables of interest for the survey – poverty and unemployment. Using available data from the 2007 Labor Market Survey, the design effects are calculated for two main employment indicators, the unemployment rate and the net activity rate. In both cases, they are extremely high, indicating a high degree of intra-cluster correlation. Additionally, while it is not possible to calculate the design effects specifically for the poverty rate, a deff of about 2.0 is generally expected. In this case the marginal benefit of increasing the number of households surveyed per cluster is low, and therefore a design with a larger number of clusters and a smaller number of households per cluster is preferable. This, however, is offset by budget and logistics limitations. Therefore it was decided in discussions with the IHSI that the minimum feasible cluster size was 15 households.

Table 4 : 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