

Technical Note on the 2011 Haiti National Sampling Frame¹

1. Methodology

Given the diverse program of data collection that the new master sampling frame must support, a multi-tiered design has been employed. The master sampling frame will be based on a multi-stage stratified cluster sampling format. This design will include 1500 total Enumeration Areas (SDEs) that can be subdivided into statistically identical replicate samples. Possible combinations of replicates are two replicates of 750 of SDEs, 3 sub-samples of 500 SDEs, 4 sub-samples of 375 SDEs, and 5 sub-samples of 300 SDEs. The selection of the Enumeration Areas (SDEs) will form the first stage sampling and the subsequent stages will depend on the sample design of the desired survey. The selection of SDEs to form clusters for the frame will utilize Population Proportional to Size (PPS) sampling methodology based on the updated SDE population. Specialized surveys requiring complex or uncommon stratification, surveys with high levels of geographic disaggregation, and surveys with strict confidentiality requirements will not be based on the master frame but must select clusters directly from updated census frame.

1.1 Domains of Estimation

The master sampling frame is designed to be sufficiently large to accommodate estimates at varying domains, while also taking into consideration the resources, both in terms of funding and staff, to maintain a large master frame. The following are the domains of estimation that the frame will be able to accommodate.

- National
- Aire Metropolitan/Other Urban/Rural
- Earthquake Affected/Non-Earthquake Affected Communes
- Aire Metropolitan / 10 Departements

1.2 Size of the Master Sample

The size of the master sample was determined based on the need to be able to do multiple large household level surveys without over-burdening communities with multiple surveys. It was estimated that the average large household survey would be 500 SDEs, and that at least three of surveys of this size are planned for the coming months. Therefore the size of the master sample was decided to be 1500 SDEs.

1.3 Updating of Primary Sampling Units

The design of the frame is multi-stage cluster sampling format with the first stage sampling involving the selection of Primary Sampling Units (PSUs). The subsequent stages will be survey-based. The PSUs for the survey will be Enumeration Areas (EAs) as created for the 2003 census subject to one measure of

¹ The design of the 2011 national sampling frame was constructed as a collaboration between the IHSI, the World Bank, and Marco International. Selection of the master sample was done by Ruilin Ren (Macro International). Technical documentation was prepared by Kristen Himelein (World Bank) and Felipe Jacome (World Bank). Questions can be addressed to the IHSI or to khimelein@worldbank.org.

size (MOS). One MOS for the clusters will be taken as approximately 700 individuals (150 households) but vary between 7 individuals (4 households) and 9758 individuals (1370 households). These totals are based on the update of the 2003 using the 'quick counts.'

The 'quick count' methodology was jointly implemented by the IHSI and Center for Disease Control and Prevention. 500 SDEs were selected for inclusion into the 'quick count' based on the following three stratifications: rapidly expanding urban zones, earthquake affected zones, and the rest of the country. Within the 'rapidly expanding urban' stratum, two further stratifications were made into 'expanding urban periphery' and 'dense urban core.' Within the earthquake affected stratum, there were further divisions based on the availability of GIS information into 'Camp et MTPTC' (indicating the SDE contained a camp and was also mapped by the MTPTC project), 'Camp et UNOSAT' (indicating the SDE contained a camp and was also mapped by the UNOSAT project), 'Camp à proximité /MTPTC' (indicating the SDE was near a camp and mapped by the MTPTC project), 'Camp à proximité /UNOSAT' (indicating the SDE was near a camp and mapped by the UNOSAT project), 'MTPTC alone' (indicating the SDE was mapped by the MTPTC project alone), 'UNOSAT alone' (indicating the SDE was mapped by the UNOSAT project alone), 'Camp alone' (indicating the SDE contained a camp but was not mapped by either MTPTC or UNOSAT), 'Camp à proximité alone' (indicating the SDE was situated near a camp but was not mapped by either MTPTC or UNOSAT), and 'Autres communes affectées non classées' (indicating that the SDE was in an earthquake affected zone but that it was not mapped using the MTPTC, UNOSAT or camp GIS²).

The percent changes for each of the above quick-count strata are given in the table below.

Categorie de SDE	SDE	ni (sde)	2003 Census Population	2011 Estimated Population	
1.- Zones urbaines en expansion					
Périphérie urbaine en expansion	405	44	696,368	925,450	32.9%
Centre urbain très dense	394	44	368,417	367,245	-0.3%
2.- Zones affectées					
Camp et MTPTC	351	38	260,159	424,424	63.1%
Camp et UNOSAT	278	30	305,462	298,318	-2.3%
Camp à proximité /MTPTC	637	70	529,300	533,126	0.7%
Camp à proximité / UNOSAT	780	85	1,659,252	1,313,309	-20.8%
MTPTC alone	169	18	196,048	264,324	34.8%
UNOSAT alone	52	6	106,536	89,803	-15.7%
Camp alone	97	11	89,972	70,020	-22.2%
Camp à proximité alone	144	16	112,243	149,442	33.1%
Autres communes affectées non classées	802	88	519,043	479,734	-7.6%
3.- Communes stables	7858	50			
Total SDE	11967	500			

² Further information on the stratification is available through Ministry of Public Works, Transport and Communications (MTPTC) at www.mtptc.gouv.ht and UNOSAT at www.unitar.org/unosat/maps/hti.

Counts were then conducted in these SDEs to record the number of buildings, households and residents. This information was then used to then derive mean expansion or contraction factors for the 12 quick count update strata, and these factors were applied to the remaining SDEs within those strata to arrive at the estimated 2011 population. Note that no other demographic or auxiliary information was used for the updates.

1.4 Stratification

Currently, there are 10 administrative departements and 104 communes. In addition, though Port-au-Prince is part of the Ouest departement, the population is completely urbanized and difference in many important characteristics from the 'Reste-Ouest.' Therefore the Aire Metropolitan was considered to be a separate departement for the purposes of sample selection.

The approximately 12,000 SDEs de-limited by the 2003 census were explicitly stratified into 26 strata (urban/rural in each of the ten departements, and at the level of the six communes constituting the Aire Metropolitan), though it is expected that the majority of surveys will not have sufficient sample size to be able to disaggregate to this level. The size of each stratum within the master sample follows a modified proportional allocation in order to match the sample distribution of the sample to that of the population as closely as possible while still maintaining minimum sample sizes within the domains of estimation listed above.

With the exception of the Aire Metropolitan, SDEs were allocated into the 10 remaining domains proportional to the population raised to the power of 0.35. This was done to maintain the balance between equal (which would require a power of 0) and proportional to the population (which would require a power of 1), and to ensure that at least 90 SDEs were allocated into each domain of estimation. This was based on the guideline that 30 SDEs is the approximate minimum size necessary to have domain level estimates, and that this level of precision would be required for each of the three upcoming 500-replicate SDE surveys.

The number of SDEs for the Aire Metropolitan was chosen to be 270 SDEs in the master sample. This value represents a compromise between the proportional with power 0.35 allocation used in the majority of the sample selection (186 SDEs) and a pure proportional allocation (299 SDEs). While it is important to ensure the minimum sample sizes in the smaller departments, because of the economic importance of the Port-au-Prince and because of the high concentration of earthquake affected zones in this area, it was deemed important to preserve the majority of the precision for the Aire Metropolitan. Therefore 270 SDEs were chosen in the Aire Metropolitan, and the remaining 1230 SDEs allocated proportional to the 0.35 power as described above.

The remaining level stratification below the domain of estimation level (urban/rural in the 10 departements and the six communes in the Aire Metropolitan) was done proportional to the population (with a power of 1).

Table 1: Distribution of Master Sample within 11 Domains of Analysis

no	Name / ID	Percent of Total Population	Percentage Urban	Urban	Rural	Domaine
1	Aire Metropolitaine	0.199	1.000	270	0	270
2	Artibonite	0.162	0.320	52	110	162
3	Centre	0.073	0.163	20	103	123
4	Grand'Anse	0.045	0.205	21	82	103
5	Nippes	0.033	0.106	10	83	93
6	Nord	0.103	0.395	55	84	139
7	Nord-Est	0.039	0.377	37	61	98
8	Nord-Ouest	0.067	0.228	27	92	119
9	Reste-Ouest	0.143	0.130	20	135	155
10	Sud	0.078	0.177	22	103	125
11	Sud-Est	0.058	0.135	15	98	113
Total		1.000	0.386	549	951	1500

It should be noted that previous discussions regarding the master sample included stratification along socio-economic zones within the Aire Metropolitan. It was decided not to include this stratification because the socio-economic calculations were outdated – having been based on the 2003 census. Since there has been significant change in both the distribution of wealth in Port-au-Prince during this period, there were concerns that this stratification could actually decrease the efficiency of the sample. It was decided that socio-economic status was better approximated by the commune residence within the Aire Metropolitan, and therefore this variable was used instead for the selection.

1.5 Division of Master Sampling Frame into Replicates

The master sampling frame can be subdivided into statistically identical replicates depending on the needs of a particular survey. Possible combinations of replicates are two replicates of 750 of SDEs, 3 sub-samples of 500 SDEs, 4 sub-samples of 375 SDEs, and 5 sub-samples of 300 SDEs. The number of SDEs in the domains of estimation under each of the replicates is presented in the appendix to this document, but in general

Split	Number of SDEs per Replicate	Domains of Estimation
1	1500	<ul style="list-style-type: none"> ○ National ○ Aire Metro/Urban/Rural ○ Affected/Non-affected ○ Departement ○ Within Departement: Urban/Rural* ○ Within Aire Metropolitan: Commune*
2	750	<ul style="list-style-type: none"> ○ National ○ Aire Metro/Urban/Rural ○ Affected/Non-affected ○ Departement
3	500	<ul style="list-style-type: none"> ○ National ○ Aire Metro/Urban/Rural

		<ul style="list-style-type: none"> ○ Affected/Non-affected ○ Departement
4	375	<ul style="list-style-type: none"> ○ National ○ Aire Metro/Urban/Rural ○ Affected/Non-affected
5	300	<ul style="list-style-type: none"> ○ National ○ Aire Metro/Urban/Rural ○ Affected/Non-affected
6	250	<ul style="list-style-type: none"> ○ National ○ Aire Metro/Urban/Rural ○ Affected/Non-affected
<p>* not complete – some areas with low population will not have sufficient size – such as urban areas in the Nippes department, or the population of Tabarre within the Aire Metropolitan.</p>		

1.6 Estimation Procedures

The estimates from surveys using the master frame will not have self-weighting designs owing to varying allocation to different strata. The selection of SDEs to form clusters for master frame will generate design weights per strata which will be combined together with household and individual weights from the surveys to provide final weights for the data.

All estimates from the surveys will require weighting to account for the complex variance calculations due to multi-stage cluster sampling methodology. The weights are calculated using the probabilities of selection from the multiple stages of selection. In terms of the first round of selection, the probabilities were recorded during the PPS sampling procedures. In the second stage, the probability of selection for the individual households depend on the total number of households listed in the SDE, any segmentation that may have been performed in the field (mainly relevant for camps), and limited replacement selections (as required).

Appendix 1: Replicate Samples within the Master Sampling Frame

Achieved sample allocation for tow splits

Domain	Split 1			Split 2		
	Urban	Rural	Total	Urban	Rural	Total
Aire Metropolitaine	135		135	135		135
Artibonite	26	55	81	26	55	81
Centre	10	52	62	10	51	61
Grand'Anse	10	41	51	11	41	52
Nippes	5	42	47	5	41	46
Nord	27	42	69	28	42	70
Nord-Est	19	30	49	18	31	49
Nord-Ouest	14	46	60	13	46	59
Reste Ouest	10	67	77	10	68	78
Sud	11	52	63	11	51	62
Sud-Est	7	49	56	8	49	57
Haiti	274	476	750	275	475	750

Achieved sample allocation for three splits

Domain	Split 1			Split 2			Split 3		
	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Aire Metropolitaine	90		90	90		90	90		90
Artibonite	18	36	54	17	37	54	17	37	54
Centre	7	34	41	7	34	41	6	35	41
Grand'Anse	7	28	35	7	27	34	7	27	34
Nippes	3	28	31	4	27	31	3	28	31
Nord	18	28	46	19	28	47	18	28	46
Nord-Est	12	21	33	12	20	32	13	20	33
Nord-Ouest	9	30	39	9	31	40	9	31	40
Reste Ouest	7	45	52	7	45	52	6	45	51
Sud	7	35	42	7	34	41	8	34	42
Sud-Est	5	32	37	5	33	38	5	33	38
Haiti	183	317	500	184	316	500	182	318	500

Achieved sample allocation for four splits

Domain	Split 1			Split 2			Split 3			Split 4		
	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Aire Metropolitaine	68		68	68		68	67		67	67		67
Artibonite	13	27	40	13	27	40	13	28	41	13	28	41
Centre	5	26	31	5	26	31	5	26	31	5	25	30
Grand'Anse	5	21	26	5	21	26	5	20	25	6	20	26
Nippes	2	21	23	2	21	23	3	21	24	3	20	23
Nord	14	21	35	14	21	35	13	21	34	14	21	35
Nord-Est	9	15	24	9	15	24	10	15	25	9	16	25
Nord-Ouest	7	23	30	7	23	30	7	23	30	6	23	29
Reste Ouest	5	34	39	5	34	39	5	33	38	5	34	39
Sud	5	26	31	5	26	31	6	26	32	6	25	31
Sud-Est	4	24	28	4	24	28	3	25	28	4	25	29
Haiti	137	238	375	137	238	375	137	238	375	138	237	375

Achieved sample allocation for five splits

Domain	Split 1			Split 2			Split 3			Split 4			Split 5		
	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Aire Metropolitaine	54		54	54		54	54		54	54		54	54		54
Artibonite	11	22	33	11	22	33	10	22	32	10	22	32	10	22	32
Centre	4	20	24	4	20	24	4	21	25	4	21	25	4	21	25
Grand'Anse	5	16	21	4	17	21	4	17	21	4	16	20	4	16	20
Nippes	2	17	19	2	16	18	2	16	18	2	17	19	2	17	19
Nord	11	16	27	11	17	28	11	17	28	11	17	28	11	17	28
Nord-Est	8	12	20	8	12	20	7	13	20	7	12	19	7	12	19
Nord-Ouest	5	19	24	5	19	24	5	18	23	6	18	24	6	18	24
Reste Ouest	4	27	31	4	27	31	4	27	31	4	27	31	4	27	31
Sud	4	21	25	4	21	25	5	20	25	5	20	25	4	21	25
Sud-Est	3	19	22	3	19	22	3	20	23	3	20	23	3	20	23
Haiti	111	189	300	110	190	300	109	191	300	110	190	300	109	191	300

Achieved sample allocation for six splits

Domain	Split 1			Split 2			Split 3			Split 4			Split 5			Split 6		
	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Aire Metropolitaine	45		45	45		45	45		45	45		45	45		45	45		45
Artibonite	9	18	27	9	18	27	9	18	27	9	18	27	8	19	27	8	19	27
Centre	4	17	21	4	17	21	3	18	21	3	17	20	3	17	20	3	17	20
Grand'Anse	3	14	17	3	14	17	3	14	17	4	14	18	4	13	17	4	13	17
Nippes	2	14	16	2	13	15	1	14	15	1	14	15	2	14	16	2	14	16
Nord	9	14	23	10	14	24	9	14	23	9	14	23	9	14	23	9	14	23
Nord-Est	6	10	16	6	10	16	7	10	17	6	11	17	6	10	16	6	10	16
Nord-Ouest	5	15	20	4	16	20	4	16	20	4	15	19	5	15	20	5	15	20
Reste Ouest	3	23	26	3	23	26	3	22	25	4	22	26	4	22	26	3	23	26
Sud	3	18	21	3	17	20	4	17	21	4	17	21	4	17	21	4	17	21
Sud-Est	2	16	18	3	16	19	3	16	19	3	16	19	2	17	19	2	17	19
Haiti	91	159	250	92	158	250	91	159	250	92	158	250	92	158	250	91	159	250