

Sample Design
Somaliland Household Survey 2013

July 2013

Submitted by:
Kimetrica, LLC
2017 10th Street
Boulder, CO 80302

Submitted to:
World Bank
Economic Policy and Poverty Reduction

Contents

Background	3
Sample households per cluster	4
Sample Allocation	4
Frame Changes.....	5
Listing	7
Replacement Clusters	7
Error in UNFPA Listing.....	8
Cluster not Found	8
Security	8
Response Rates.....	9
Weighting.....	10
Annex A:	12
Listing Operation.....	12
Orientation.....	12
Initial Count.....	13
Segmenting	13
Listing	13
Selecting Households.....	17

Background

Somaliland does not have a census-based frame to use for designing a representative sample survey. Most surveys in Somaliland rely on the UNDP population survey of villages that date from 1990. During the preparatory phases for the SHS 2013, it was felt that this survey of villages was not the most accurate given the elapsed time that had passed and the likelihood that population centers may have shifted (given climate, political and cultural realities).

Other frames were reviewed for possible use instead of the UNDP village list, including a water point survey and the most recent voter registration database.

The voter registrations lists were reviewed and compiled and though there were apparent inaccuracies in some of the counts of eligible voters in the polling stations as well as GPS points which were inconsistent or missing, it was decided that the list of polling stations with the number of registered voters would at least be more accurate in indicating settlements for the PSU than the survey of villages. In addition, excepting the regions of Sool and Sanaag where there were security concerns, most villages where the polling stations were located did have GPS locations and so could be geo-referenced. The map below provides the distribution of the polling station GPS points.

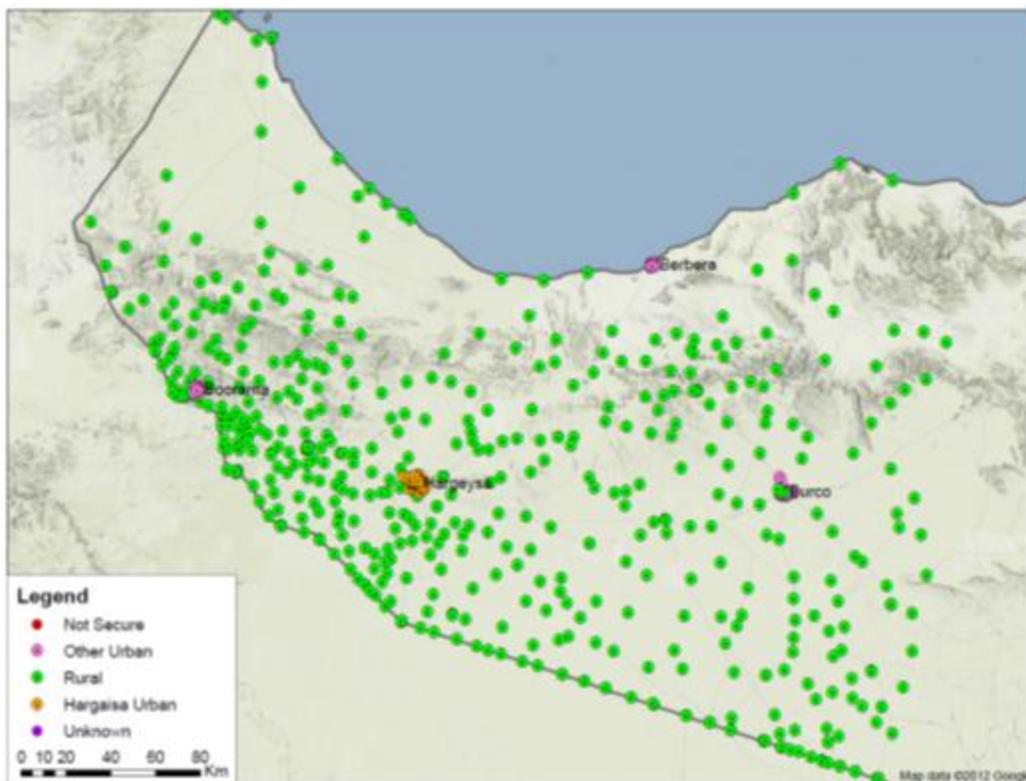


Figure 1: Map of polling stations with geo-referenced points

Sample households per cluster

The survey is designed to provide statistics at the national level and a disaggregation between the capital, Hergaisa and the rest of the country (combining the rural and other urban cluster). During the design phase it was estimated that a sample size of about 1700 households might provide the required statistical power to compute the consumption aggregate and subsequent poverty line. Initially the selection of 12 households per sample cluster (PSU¹) was considered. However, it was decided to reduce the number of sample households per cluster (PSU) to 9, and increase the number of sample clusters in order to reduce the corresponding design effects and increase the level of precision. Since the survey was not an explicit longitudinal baseline study (a panel type survey), attrition of the sample was not a primary consideration.

Kimetrica contracted the services of a sampling expert, Mr. David Megill, to further review the frames and provide guidance in terms of the stratification.

Sample Allocation

One issue that had to be addressed is the allocation of the sample to the different strata. Three primary strata were defined as: Rural, Hergaisa and Other Urban. The population proportion varied by stratum, and the general agreement in informal discussions was that about 50% of the population was urban and 50% were rural. The UNDP Somaliland MDG report cites a population in Somaliland of about 3.5 million (including pastoralists) and estimates that two-thirds of the population is rural or nomadic². IDPs were not included in these counts. For this reason, the SHS 2013 design assumed a population distribution during the SHS 2013 sample design phase was 60% rural and 40% urban; this was also validated by examining the distribution of the electoral registry. The number of registered voters indicates that 60% are rural; 22% in other urban settlements and 18% are in Hergaisa. Voter registries include people in settlements that are 18 and over.

Table 1: Distribution of registered voters across region.

Region	Total	% by region	HU	% by region	Other Urban	% by region	Rural	% by region
Awdal	178 980	15.0	0	0.0	67 577	26.4	110 403	15.3
M-Jeeh	415 050	34.9	213 812	100.0	18 421	7.2	182 817	25.4
Sahil	49 633	4.2	0	0.0	22 014	8.6	27 619	3.8
Sanaag	156 585	13.2	0	0.0	21 328	8.3	135 257	18.8
Sool	106 925	9.0	0	0.0	15 269	6.0	91 656	12.7
Togdheer	283 464	23.8	0	0.0	111 633	43.6	171 831	23.9
			213 812		256 242		719 583	
Total	1 189 637	100.0	(18%)	100.0	(22%)	100.0	(60%)	100.0

*Polling Stations 2009

¹ Note; the term cluster and PSU are used interchangeably.

² See pg. 11 of UNDP MDG Report 2010.

district, so although Gabiley is included as “other urban” it was selected from the polling station frame and so for selection purposes, was treated as a separate stratum.

Table 2 shows the distribution of the final sample of PSUs and also identifies the districts where the sample PSUs are located.

Table 2: Final sample distribution of regions, district, clusters and household by strata.

Strata	Regions	District	Clusters (PSU)	Households	
Hergaisa Urban	Marodijex	Hergaisa	64	576	
		Total	64	576	
Other Urban	Awdal	Boroma	7	63	
	Marodijex	Gabiley	1	9	
		Togdheer	Burco	14	126
	Sannag	Cerigabo	3	27	
		Sool	Lascanood	3	27
	Sahil	Berbera	3	27	
	Total		32	279	
Rural	Awdal	Boroma	8	72	
		Baki	3	27	
		Lughaya	1	9	
		Saylac	3	27	
	Marodijex	Hergaisa Rural	8	72	
		Gabiley	10	90	
		Balli-Gubadle	2	18	
		Salahley	4	36	
	Togdheer	Burco	14	126	
		Oodweyne	7	63	
		Buuhoodle	2	18	
	Sannag	Cerigabo	9	81	
		Ceel-Afwayne	4	36	
		Garadag	2	18	
		Badhan	3	27	
	Sool	Lascanood	5	45	
		Cynabo	7	63	
	Sahil	Berbera	3	27	
		Sheek	1	9	
		Total		96	873
	Total			192	1728

*Kimetrica, SHS 2013

Listing

In order to undertake the second stage selection of the actual sample households to be interviewed, a listing operation was undertaken. The listing operation was done concurrently with the actual data collection. The task of listing the households in sample urban EAs was facilitated greatly by the UNFPA mapping exercise that had been undertaken. As shown in Figure 2 previously, the EA boundaries were well defined and as it was still only recently conducted, the survey capitalized on using personnel from the UNFPA exercise to help guide the teams in the urban areas.

The listing operation in the rural areas was more challenging. The specific manual and instructions are provided in Annex A. Several problems were encountered during the listing operation, namely the inability to find or locate the specified settlement where the polling station was located. It is important to remember that a polling station is not a geographical entity. There could be one or several in any given settlement though often there is a one to one correspondence between the polling station and the settlement. A few cases arose where the settlements corresponding to a sample polling station could not be located and had to be replaced. GPS coordinates were available for most settlements on the polling station list except for those in the east where security problems prevented their collection. In any case, the lack of maps obliged a cartographic exercise on the part of the rural teams. Supervisors were trained in basic cartographic skills and were required to undertake the following exercise:

- Define the limit of the polling station and draw a rough boundary
- Drive around the settlement and divide it into roughly equal segments (between 80-140 households each)
- Randomly select a segment by casting a dice
- Proceed to list the selected segment
- Select nine (9) households using random tables provided to the supervisors
- Undertake the survey

Replacement Clusters

The problem of replacing clusters became evident particularly in the Sool and Sannag regions due to security issues. A practical approach was undertaken by using the “nearest secure neighbor”. The idea was to assure that the sample polling station in the same district had similar characteristics to those of the insecure sample PSU, in order to maintain the geographic representativeness of the sample and reduce the bias from the PSU nonresponse. This meant that in selecting the replacement we vetted the selection to assure that the same clan would be interviewed and the same livelihood. The distribution of replaced clusters is provided in the table 3 below.

Table 3: Replaced clusters by region. (Table of PSU replacements)

Region	Number replaced	Total	Reason	Percent of all replaced
Awdal	3	22	Not found	13.6
Marodijex	13	90	Error in listing	14.4
Togdheer	1	37	Not found	2.7
Sannag	4	21	Security	19.0
Sool	4	15	Security	26.7
Sahil	1	7	Security	14.3
Total	26	192		13.5

*Kimetrica, SHS 2013

As can be seen, 26 clusters had to be replaced. Except in the case of Hergaisa, the replacements were rural polling stations. The specific reasons are provided in the explanations below. A procedure was adopted whereby the Coordinator of the Household Survey would inform the Kimetrica Project Manager who would in turn work with David Megill, Sampling Consultant, to select the replacement. Each replacement PSU was selected with PPS from the remaining secure areas within the same district, in order to have consistent probabilities of selection and maintain the geographic representativeness of the sample. The PPS in rural polling stations would have been by the size of the registered voters and in the UNFPA lists, by number of households. The following reasons for replacement are provided:

Error in UNFPA Listing

The UNFPA Hergaisa listing operation had problems in two Hergaisa locations. These had to be re-listed at the last moment by the UNFPA survey. The SHS 2013 survey dropped the old frame for these two areas and waited for the updated results to be provided and reselected. The two locations that presented problems were:

- Axmed Dhagax had ten (10) clusters reselected
- Gacan Libax had three (3) clusters reselected

Cluster not Found

In four (4) cases the teams were unable to find the assigned cluster in the field. The settlement had moved or did not exist. This could have been a problem with the polling station sampling frame.

Security

The most prevalent problem in the rural area were in the Sool, Sannag and Sahil zones, and these were identified as problems with security. The problem was persistent throughout the survey and at times the replacement cluster selected also had to be replaced. In order to avoid this problem, the Field Coordinator reviewed the rural polling stations and indicated which ones had security problems. In the

end these were dropped from the frame in order to assure that the selection process would not become overly burdensome. In the end, 68% of the polling stations in these zones were considered to be areas prone to security problems. These were frequently associated with different clans and sub-clans, and the security problems would have likely been in the nature of bodily threat or abduction. The workable frame in these regions was narrowed to 82 polling stations. These adjustments have been accounted for in the weights.

Response Rates

The table below illustrates the response rates that were attained for the survey. A total of 1,843 households were visited of which 1726 were interviewed. In two cases, the interviews were initiated and sufficient information was collected for the consumption module. Two households therefore cut-off the interview however as consumption information was considered complete, they were retained for the survey.

Table 4: Response rates

	Households	Percent
Completed interview	1726	93.7
Respondent not available	47	2.6
Nobody home	34	1.8
Interview interrupted	1	0.1
Refused to participate	28	1.5
Dwelling vacant	2	0.1
Household not found	5	0.3
Total	1843	100.0

*Kimetrica, SHS 2013

As can be seen from the table, 117 households were replaced with an effective response rate of 93.7%.

Weighting

In order for the sample estimates from the SHS to be representative of the population, it is necessary to multiply the data by sampling weights. The basic weight for each sample household would be equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage). A household weight will be attached to each sample household record in the data file.

Based on the stratified multi-stage sample design, the overall probability of selection for the SHS sample households can be expressed as follows:

$$P_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{1}{S_{hi}} \times \frac{m_{hi}}{M'_{hi}},$$

where:

i =PSU

h =stratum

p_{hi} = probability of selection for the sample households in the i -th sample PSU. In the case of the polling station, this would refer to the population of registered voters. (

n_h = number of sample PSUs selected (and enumerated) in stratum h

M_{hi} = number of registered voters in the case of the rural stratum or households for Hergaisa and the other urban stratum) for the i -th PSU in stratum h

M_h = total number of registered voters or households) in the sampling frame for stratum h

S_{hi} = number of segments delineated in the i -th PSU in stratum h

m_{hi} = number of sample households selected in the i -th PSU (cluster) in stratum h

M'_{hi} = total number of households in the i -th PSU in stratum h

The three components of this probability of selection correspond to the individual sampling stages. The second term is applicable for polling stations that were subdivided into two or more segments, where one segment was selected at random to be listed. For the remaining sample PSUs, S_{hi} is equal to 1, so the second term in this probability would be equal to 1.

The basic sampling weight is calculated as the inverse of this probability of selection. Based on the previous expression for the probability, the weight can be simplified as follows:

$$W_{hi} = \frac{M_h \times S_{hi} \times M'_{hi}}{n_h \times M_{hi} \times m_{hi}},$$

where:

W_{hi} = basic weight for the sample households in the i-th sample cluster in stratum h

The variability in the weights within each stratum depends on the correlation between the number of registered voters (or households) in the frame and the number of registered voters? and households listed in each cluster.

It is also important to adjust the basic weights for the sample households to take into account the non-response in each sample cluster. Since the weights will be calculated at the level of the sample cluster, it is advantageous to adjust the weights at this level. The final weight (W'_{hi}) for the sample households in the i-th sample cluster in stratum h can be expressed as follows:

$$W'_{hi} = W_{hi} \times \frac{m_{hi}}{m'_{hi}},$$

where:

m'_{hi} = number of sample households with completed interviews in the i-th sample cluster in stratum h

The sampling probabilities at each stage of selection and the final weights were calculated in an Excel spreadsheet with information from the sampling frame for each sample cluster. Following the SHS fieldwork it was necessary to enter in this spreadsheet the total number of households listed in each sample cluster.

Would be useful here to add a table with non-response rates at the district level.

PSU and cluster is used interchangeably in the document and this makes the reading confusing. Might be good to homogenize this in the final document.

Annex A:

Listing Operation

The listing operation is a very important component of assuring that the quality of results is accurate. We want to assure that the households we interview are properly delineated and identified within its boundaries. The boundary is important as these are parts of the whole that represent a nation.

Listing consist in physically counting and marking each household contained in a geographic boundary. These are used as the basis of selecting your 9 households randomly.

Please make sure you are able to confirm your location with the aerial map that is provided. You will use the GPS points to orient your arrival. The listing operation has various steps:

- Village Orientation
- Initial Count
- Segmenting
- Listing
- Selection
- Interview

Orientation

As stated before, orientation is a critical part of the arrival into a polling station. Locate yourself on the aerial map and make sure that that location can be verified. Meet with the village official and begin to discuss the activities. You will need to likely enlist the assistance of two guides for the work. These guides are local persons provided at the recommendation of the village official.

Be aware of your surroundings as you drive up and scan for landmarks and the general layout of the area.

Before arriving you should be expected. Make sure you call the village officials before hand and confirm the location (polling station). You should not arrive someplace that does not exist! The current GPS coordinates may not be accurate. They are used to help guide you. You are expected to contact the village officials 3-days before arriving.

No good pilot accepts the reading on the fuel gauge without physically checking the level of fuel in the tank. Gauges can give a faulty reading. You do not want to be flying and discover you have no fuel because you trusted a gauge. The same is with the GPS reading. GPS reading serve to localize and orient but assuring that the officials are aware of your arrival is another matter altogether.

Initial Count

Aerial maps are provided for you to locate your position in general terms. These may also be inaccurate. Once you have confirmed the location you can validate the position and structures using the aerial map. But you will proceed to draw a new map based on the actual reality as indicated by the guide.

! YOUR FIRST TASK IS TO MAP THE BOUNDARIES OF THE POLLING STATION AS THEY REALLY ARE.

Once you have met the village official you will then proceed to the initial count. You may need to recruit assistance from local guides.

You will draw a sketch maps for each EA. This is the large scale map and is based on the aerial map.

The enumerators will become your cartographic assistants. Choose the best one to mark the boundaries (you will likely accompany him with the guide and driver).

1. In the urban areas you will walk the boundaries (in rural you will drive) and mark the points boundaries and take GPS points. If the polling station is large, you may have to drive the boundaries. Have the village official accompany you to define the boundaries.
2. Take note of landmarks as You will sketch primary boundary markers on the map. (See symbols provided) .
3. The two cartographers that remain behind will map internal features of the polling station. This will help you understand the lay out and segment the polling station.
4. Make sure you use the standard symbols to mark land marks and also clearly note the clusters of villages with their estimated number (see GPS exercise)

Segmenting

1. With the village official you will estimate the area (delineating on the map) where 100-120 households can be listed. Identify key boundaries and mark these on the larger scale sketch match. You will then prepare to assign this segment to be listed. A smaller map will be drawn. This will include all structures of the segment.
2. Use the segmentation form to segment the number of households and draw their boundaries on the large scale map.
3. Select a segment using the random number table.

Listing

Note the following terms:

A **cluster** is the smallest geographical statistical unit which groups a number of adjacent households in a geographical area. A cluster typically corresponds to a census enumeration area (EA) or to a segment of an EA.

An **enumeration area** (EA) is a geographical statistical unit that contains a certain number of households. In our case, the EA is the polling station. An EA is usually a group of small villages, or a village, or a part of a large village in the rural areas, and a city block in urban areas, with its location and boundaries well defined and recorded on census maps.

A **segment** is a sub-grouping of a polling station based on the threshold of 100 occupied households.

A **structure** is a free-standing building that can have one or more rooms, for residential or commercial use. Residential structures can have one or more dwelling units (for example: single house, apartment building). In the case where one household inhabits several small structures, as in the rural area, all the structures together, whether they are fenced in or not, constitute a structure.

A **dwelling** unit is a room or a group of rooms normally intended as a residence for one or more households (for example: a single house, an apartment, a group of rooms in a house).

A **household** consists of a person or a group of related or unrelated persons, who live together in the same dwelling unit, who acknowledge one adult male or female as the head of the household, who share the same housekeeping arrangements, and are considered as one unit. In some cases one may find a group of people living together in the same house, but each person has separate eating arrangements; they should be counted as separate one-person households. Collective living arrangements such as army camps, boarding schools, or prisons will not be considered as households. Examples of households are:

- A man with his wife or his wives with or without children
- A man with his wife or his wives, his children and his parents
- A man with his wife or his wives, his married children living together for some social or economical reasons (the group recognize one person as household head)
- An unmarried man or woman with his or her children and provides himself or herself the essentials and foods for their life
- A widowed or divorced man or woman with or without children

The **head of household** is the person who is acknowledged as such by members of the household and who is usually responsible for the upkeep and maintenance of the household.

A **cluster map** is a sketch map (with potentially varying levels of detail) produced in the household listing operation, with location or marks of all structures found in the listing operation which helps the interviewer to locate the selected households. A cluster map also contains the cluster identification information, location information, access information, principal physical features and land marks such as mountains, rivers, roads and electric poles.

1. Once the supervisor has assigned a segment to be listed, the team will move to a selected “most extreme point”. It should have a clear bound (whether north or another point).
2. With the enumerators the supervisor will assign a “BLOCK” letter to an enumerator and discuss the routing he/she should take to list. They will use the map to define boundaries. This rough sketch will serve as guide for the team when they begin the main work.

3. Each enumerator will canvas his/her block by selecting a landmark and moving clockwise. They should be accompanied with a local guide. Canvassing is the process where you divide up the EA and verify the number of households within an EA.
 - The enumerators will knock on the door and gather the necessary information as required on their listing sheet. They will mark the home using chalk or another marker so that it is known that the house has been visited (and not duplicate efforts). They will mark the house with their assigned letter (A,B,C) and the sequential number. Each enumerator should mark at least 40 households (i.e. notice the difference between a structure and a household).
 - The enumerator will fill out the sheet and mark the household as follows: if two households are present, they will be separated by a comma. If the housing structure is empty mark it with an X.

A-1 or B-2,3 or X

4. The enumerator will use a Household Listing Form (see packet). Begin by entering the identification codes of the cluster. The first two columns are reserved for the supervisor use only, leave them blank. Note: **LEAVE A BLANK ROW BETWEEN LISTED UNITS. MAKE SURE THE HEADERS ARE PROPERLY FILLED IN.**
5. **There are three forms to fill out: C1-C2-C3.**
6. **C1 provides the information of the polling station. Fill out the header as provided.**

HOUSEHOLD LISTING FORM: FORM C-1

GPS UNIT		L2. GPS mark of village administrator	WP	Easting	Northing	
CLUSTER						
ID 1: REGION CODE		L3. Name of Village Elder or Official				
ID 2: DISTRICT		L4. Contact phone				

7. **C2 is the assigned enumerator listing form.**

ID 1. Cluster		Number of pages		
ID 3. Enumerator Code		Current page		
ID 4. Circle block code	A B C			
		Total number of households (sum all pages)		

Serial ID (To be filled by Supervisor)	HH ID (to be filled by enumerator)	G.1	G.2	G.3	G.4	G.5
		NAME OF HH HEAD	GPS EASTING	GPS NORTHING	How many household members	How many registered voters

Complete the rest of the form as follows:

Column (1) [Serial ID of the household]: To be completed by the supervisor and is the sequential number of each valid household.

Column (2) [HHID Number of the housing structure] For each structure, record the serial number (and make sure it is the same as the chalked number).

Column (3) [G1: Name of the Head of Household]: Write the name of the head of the household. There can only be one head per household. If no one is home, ask neighbors for the name of the head of the household. If a name cannot be determined, leave this column blank. Note that it is not the name of the landlord or owner of the structure that is needed, but the name of the head of the household that lives there. Remember this is a longitudinal survey. This means that the household will be visited again in 2 years and in 5-years' time.

Column (4 and 5) [G2 and G3: GOS readings]: Write the UTM Easting and northing for each household visited. Mark the entry into the compound (if a compound); threshold of the door if a standalone home; the entry to the building if a high rise or multi home complex.

Column (6) [G4: Number of household members: Ask about the household members as per the definition provided. This should be all household members.

Column (7) [G4: Number of registered voters]: This is a very important question and is used to make sure that the household provides information relevant to weighting of the survey. The weighting will confirm the number of voters in the list with the reported number in the household.

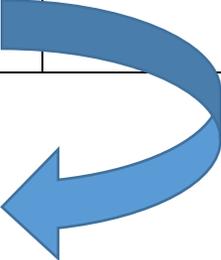
If the structure is an apartment building, assign one serial number to the entire structure but complete Columns (3) through (7) for each apartment in the building individually. Each apartment unit should be recorded as a structure having multiple households but identified with the prefix A and then a sequential number for the household residing in a specific apartment.

Selecting Households

The supervisor will collect all the listing sheets from the different blocks. He will make sure that the minimum number of households are available per the selected segment (120). If not, he will have to verify the reason and check against any existing lists. This may be indicative of an error in the listing. A blank row is provided to help the supervisor order any errors. Every eligible household should be given a sequential number in column 1 and all non-eligible structures should have an X in column (1). Place the total number of households listed in G0. Transfer this number to form C3 field G2.

Form C2:

ID 1. Cluster		Number of pages		
ID 3. Enumerator Code		Current page		
ID 4. Circle block code	A B C			
		G0. Total number of households (sum all pages)		



--	--	--	--	--

G2. Total number of households in segment	132	
G3. DIVIDE G2/9. If a decimal select the whole number (10.8 is 10)	Whole 14	Decimal 66
G4. Random start From table	12	

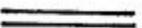
Now, using the random start from the random number table insert into G4. Select the listing number that has this random number (in this case "12" and then select every 14th household after that. So in this case, the households to be selected are provided in the table below:

HHID	Listing
1	12
2	26
3	40
4	54
5	68
6	82
7	96
8	110
9	124

Once selected, you will assign to the enumerator with preference given to the block assigned to the original enumerator.

If possible, the enumeration should start as soon as convenient after the household selection.

Appendix 2.2 Symbols for mapping and listing

Orientation to the North	
Boundaries of the cluster	
Paved road	
Unpaved (dirt) road	
Footpath	
River, creek, etc.	
Bridge	
Lake, pond, etc.	
Mountains, hills	
Water point (wells, fountain, etc.)	
Market	
School	
Administrative structure	
Church, temple	
Mosque	
Cemetery	
Residential structure	
Non-residential structure	
Vacant structure	
Hospital, clinic, etc.	
Electric pole	
Tree or bush	