UNHCR
The UN Refugee Agency

Rapid Methods for Assessing Water, Sanitation and Hygiene (WASH) Services at Refugee Camps in Emergency Settings

Briefing Note

Kenya / Somali Refugees / IFO Camp Dadaab, The nothing but nets malaria survey in IFO. Respondents being interviewed. Nov 2010.
© UNHCR/Sarah Hoibak

COVER PHOTO:

Kenya. United Nations High Commissioner for Refugees Filippo Grandi visits the Dadaab refugee camps to assess the situation in the camps and meet with refugee representatives. And aerial view shows part of the Dadaab refugee camps in Kenya. Many of the refugees that talked to the High Commissioner said that they would like to return home to Somalia but that repatriation should be done in a dignified and humane manner. They also voiced their wish for a continued investment by the international community in the Dadaab refugee camps.
© UNHCR/Siegfried Modola

Graphic Design: Alessandro Mannocchi/Rome

Rapid Methods for Assessing Water, Sanitation and Hygiene (WASH) Services at Refugee Camps in Emergency Settings

Briefing Note

Prepared by:

Matthew E. Verbyla, Ph.D.
Ryan W. Schweitzer, Ph.D.

Prepared for:

Reporting Officer: Murray Burt, Senior WASH Officer
The United Nations High Commission on Refugees (UNHCR)
Section/Division: PHS/DPSM

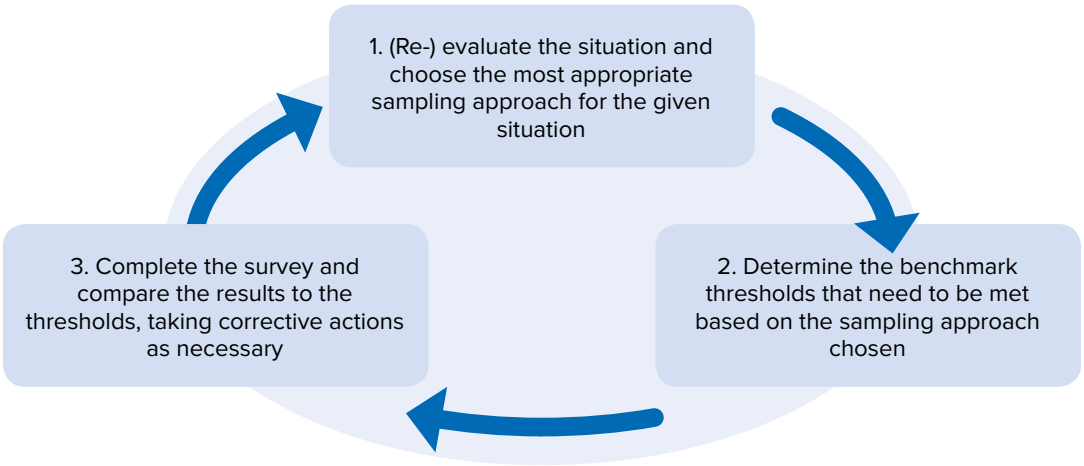
February 12, 2017

UNHCR
The UN Refugee Agency

Introduction

UNHCR and its partner organizations and institutions require reliable data that can be collected in a timely manner for planning needs during emergency situations. UNHCR utilizes household-based assessments of water, sanitation & hygiene (WASH) services for millions of refugees worldwide. UNHCR has invested in evaluating the most effective ways to collect data, including studies of different sampling approaches and the frequency of data collection. Based on the time and resources available during acute emergencies, it was determined that a sampling approach should be designed to maximize the potential for a 60 household (HH) sample to estimate the level of WASH services for the overall camp population in emergency contexts. This sample size is derived from the capacity of two trained interviewers working two days. Many camps/settlements have populations between 5,000 and 150,000 people, with a total of 1,000 to 30,000 households, although this varies widely for each situation. Camps are often subdivided into several different administrative "zones," and in some situations, it is necessary to estimate differences in the level of WASH services for each zone within a camp, in order to target corrective actions for zones with the lowest levels of service.

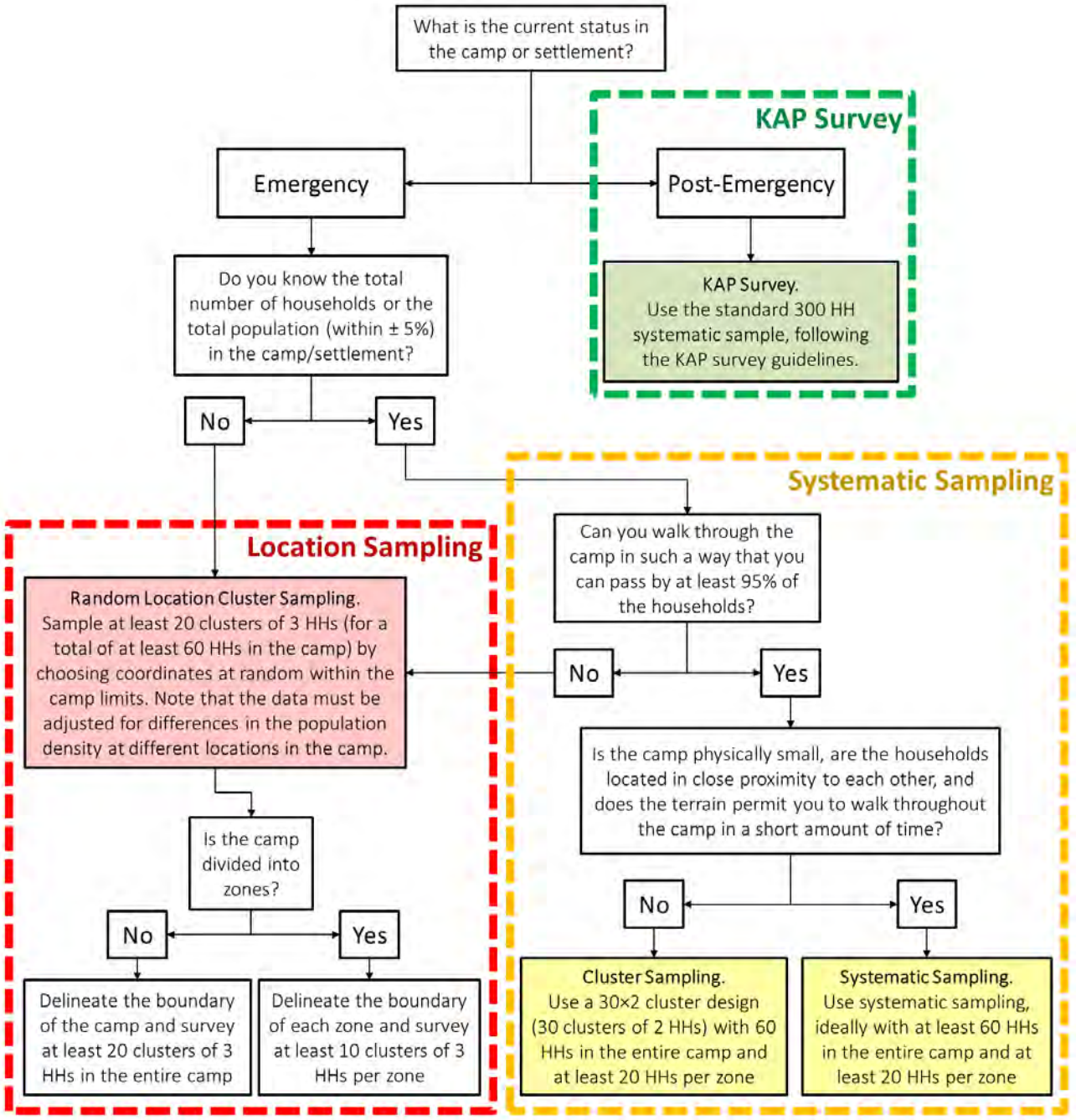
A research study commissioned in 2013 by UNHCR assessed four different options for obtaining 60 HH samples (random, systematic, and 30×2 or 20×3 cluster), concluding that the 60 HH sample size was adequate[1] for numerical indicators (e.g. water stored per person per day), and that there was a need to further investigate the precision of binary (yes/no) indicators. The authors noted considerable spatial variation in the level of services within camps, stating that households with better services "clustered towards the center of the camp, while the periphery seemed to have fewer services." This initial study had several limitations, particularly that the authors were unable to draw "definitive conclusions about the relative advantages of each [60 HH sampling] method" because they only performed five simulations for each sampling method. As a result, the present study was commissioned to make specific recommendations about which sampling approach is most appropriate during the initial phases of an emergency situation, based on the nature of the situation. The recommendations presented in this briefing note are supported by results from a total of 10,000 simulations per sampling approach, using data from a total of five different camps. More detailed information about the findings and methods used in the study can be found in the Working Paper. The recommendations presented in this briefing note are laid out as a three-step iterative process:



1 The term "adequate" was not defined by the authors of this report.

1. (Re-) evaluate the situation and choose the most appropriate sampling approach

To determine the most appropriate sampling approach for collecting information to assess WASH services in refugee camps or settlements, follow the decision tree below:



2. Determine the benchmark thresholds for the sampling approach chosen

After determining the most appropriate sampling approach using the decision tree, use the table on the following page to determine the benchmark thresholds that need to be met to comply with the UNHCR standards for each of the WASH indicators. These thresholds are based on the statistical precision of each sampling method. In general, if fewer HHs are surveyed, the benchmarks are lower and potentially more difficult to meet. Indeed, there is a tradeoff between the time saved by surveying fewer HHs and the greater likelihood that corrective action will be required due to the lower benchmark thresholds. Below are a few rules of thumb to follow when implementing the sampling approaches described in this briefing note.

Never make decisions based on surveying fewer than 20 HHs (or for location sampling, no fewer than 30 HHs). Surveying fewer households can lead to misleading results, simply because of the random error associated with sampling. For example, if you use the systematic sampling approach and survey only 60 HHs in the camp, you can only evaluate WASH service levels in a maximum of three zones. There is no need to use weighting factors based on the size of the zone. However, ideally the population in each zone should be somewhat similar. In the schematic map shown to the right, the camp is administratively divided into seven blocks. Therefore, if you need to monitor and report WASH services in each block using the systematic or cluster sampling approach, then you should survey at least 20 HHs per block (for a total of 140 HHs in the entire camp). Alternatively, if you only have resources to sample 60 HHs in the entire camp, then the blocks can be organized into three zones (for example, Zone 1 including Blocks A, B, and D; Zone 2 including Blocks C and E; and Zone 3 including Blocks G and F). Then, a total of 20 HHs can be sampled in each zone. If location sampling is used, then you should survey a minimum of 30 HHs per zone. Thus, if only 60 HHs will be surveyed in the entire camp, then the camp should be divided into two zones (for example, Zone 1 including Blocks A, B, C, and E; Zone 2 including Blocks D, F, and G).

If the results are close to the thresholds, but not in compliance, survey more households. If fewer HHs are surveyed, the benchmarks are lower and potentially more difficult to meet. If you choose only 20 HHs per zone and do not meet the thresholds, increase the sample size per zone to 40 HHs for the next survey, and compare your results to the new benchmark thresholds for this larger sample size.

During non-emergency situations, use the standard Knowledge, Attitude and Practice (KAP) method. Standard KAP surveys provide much more precise assessments of WASH services relative to the approaches presented in this briefing note. Follow the UNHCR CDC standardized KAP Manual, which recommends the use of a systematic sample of 300 HHs. KAP survey methods are not covered any further in this briefing note.

WASH Service Indicator	UNHCR Standards	**Benchmark Thresholds for the Households Surveyed** (if the statements below are true, the camp/zone is likely to be in compliance with UNHCR standards)							
		Random Location Sampling Approach*		**Systematic Sampling Approach**			**Cluster Sampling Approach**		
		20×3	**10×3**				**30×2**	**20×2**	**10×2**
		Of 60 HHs surveyed…	Of 30 HHs surveyed…	Of 60 HHs surveyed…	Of 40 HHs surveyed…	Of 20 HHs surveyed…	Of 60 HHs surveyed…	Of 40 HHs surveyed…	Of 20 HHs surveyed…
Average liters of water collected per person daily	At least 15 L/p/d	Average of at least 19 L/p/d	Average of at least 20 L/p/d	Average of at least 18 L/p/d	Average of at least 19 L/p/d	Average of at least 20 L/p/d	Average of at least 19 L/p/d	Average of at least 20 L/p/d	Average of at least 21 L/p/d
Households with *less than* 10 L/p storage capacity	Less than 30%	No more than 8 HH	No more than 2 HH	No more than 10 HH	No more than 5 HH	No more than 1 HH	No more than 9 HH	No more than 2 HH	Not a single HH
Households *without* access to soap	Less than 30%	No more than 9 HH	No more than 3 HH	No more than 10 HH	No more than 5 HH	No more than 1 HH	No more than 10 HH	No more than 3 HH	No more than 1 HH
Households *without* proper handwashing knowledge	Less than 30%	No more than 9 HH	No more than 3 HH	No more than 10 HH	No more than 5 HH	No more than 1 HH	No more than 10 HH	No more than 3 HH	No more than 1 HH
Households that *do not* have a toilet	Less than 40%	No more than 15 HH	No more than 5 HH	No more than 16 HH	No more than 9 HH	No more than 3 HH	No more than 15 HH	No more than 5 HH	No more than 2 HH
Number of people sharing a toilet	No more than 50 p/toilet	Average of no more than 42 p/toilet	Average of no more than 39 p/toilet	Average of no more than 44 p/toilet	Average of no more than 43 p/ toilet	Average of no more than 40 p/ toilet	Average of no more than 43 p/ toilet	Average of no more than 40 p/ toilet	Average of no more than 38 p/toilet

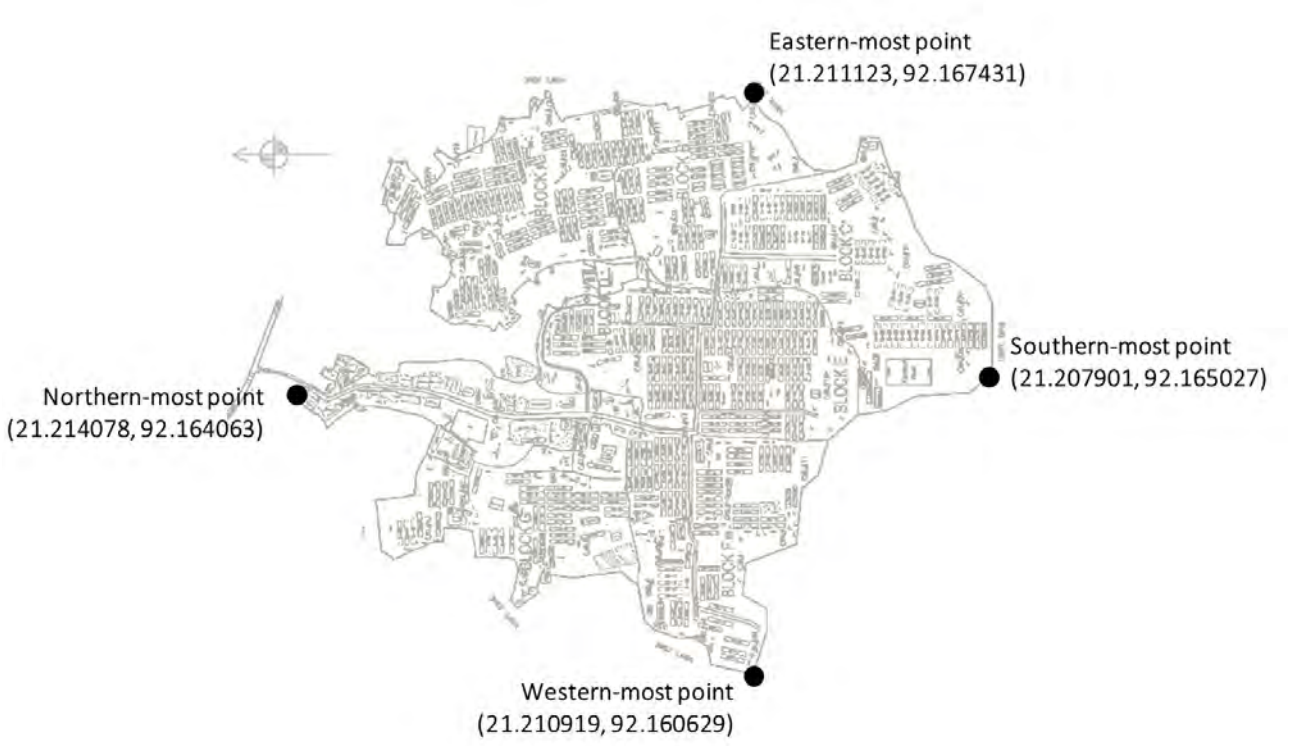
* The indicator values for location sampling must be adjusted using weighting factors based on the estimated population densities before comparing them with the benchmark thresholds presented in this table (see Section 3 for information about how to calculate the adjusted values)

3. Complete the survey and compare the results to the benchmark thresholds

Random Location Sampling Approach

Description. The location sampling method is based on the selection of random coordinates within the boundaries of the camp, and then finding the three households closest to each of those coordinates. The distance between each of the coordinates and the furthest of the three households is used to calculate a population density factor, which is used as a weighting factor to adjust the data values for a bias that results from differences in the population density throughout the camp.

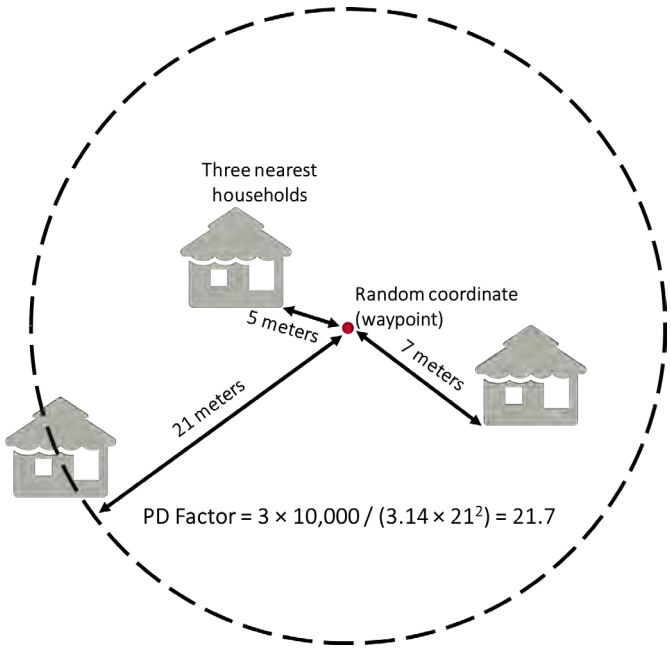
Procedure. Start by recording the camp's outer boundaries with a GPS. Walk around the perimeter of the camp and record the GPS coordinates (in decimal degrees), especially for the outermost points, as shown in the figure below. Determine the northern-most, eastern-most, southern-most, and western-most points. Coordinates are always presented as a pair of values in parentheses (first latitude, then longitude). The northern-most point has the greatest latitude value (e.g., 21.214078), and the southern-most point has the lowest latitude value (e.g. 21.207901). The eastern-most point has the greatest longitude value (e.g. 92.167431), and the western-most point has the lowest longitude value (e.g. 92.160629). Mark the coordinates of these four points. Use 5 decimal places if possible (this is equivalent to a resolution of ~1 meter). Record the minimum and maximum latitude values, and the minimum and maximum longitude values.



Choose coordinates for random locations by choosing random numbers between the minimum and maximum latitude and longitude values. *Hint: if you sort the points first by latitude, then by longitude, it will make it easier to walk to them in order.* Random numbers can be generated in Excel with the following equation: =RAND()*([maximum value] – [minimum value]) + [minimum value]

The number of random coordinates needed is 1/3rd of the desired sample size. For example, if you need 60 HHs, then choose a total of 20 pairs of random coordinates. If you need 30 HHs, choose 10 pairs of coordinates.

Enter each pair of coordinates, one at a time, as waypoints into the GPS unit. Then, navigate throughout the camp until you arrive at the specified point. When you arrive at Point 1, find and survey the nearest three HHs, and measure the distance between Point 1 and each of the three HHs. These will be called the triangulation distances. After completing the surveys, enter the second pair of coordinates as a waypoint, and navigate to Point 2. Survey the three nearest households and record the distances between Point 2 and each HH. Continue with the remaining waypoints, until you have completed surveys at all 60 HHs. At the end, when the data are analyzed, the maximum triangulation distances (the distances between each of the waypoints and the furthest of the three nearest HHs) will be used to generate a weighting factor (called the population density or PD factor). This PD factor is a rough estimate of the population density based on the three households closest to each waypoint. It is calculated as the number of HHs in the cluster (3) multiplied by 10,000 and divided by 3.14 times the furthest distance, squared (see example below). It is used to adjust for a bias caused by differences in population density throughout the camp. The raw data are multiplied by the corresponding PD factor for each point to get the adjusted data. The final adjusted average is then calculated as the average of the adjusted data divided by the average PD factor. For binary data responses (for example, is the HH meeting the standard; yes/no), just input 100% for yes and 0% for no, and the adjusted average will be the adjusted percentage.



Example. Suppose you are assessing WASH services in a camp with no defined zones. Therefore you choose to survey 60 HHs in the entire camp. You determine the minimum and maximum latitude and longitude values, and select coordinates at random between the minimum and maximum latitude and longitude values. As you select these points, nine of the coordinates chosen end up being located outside of the camp boundaries, therefore you replace them with new coordinates until you have 20 coordinates within the camp boundaries (see figure below). You walk to each of the locations and measure the distances between each of the random coordinates and the three nearest households (i.e. the triangulation distances; note that if the location is right on top of the household, you record a distance of zero). You record the data for each household. Then, you return back to the office and calculate the population density (PD) factors as: 3*10,000/(3.14*[maximum trianulation distance]^2). You multiply the raw

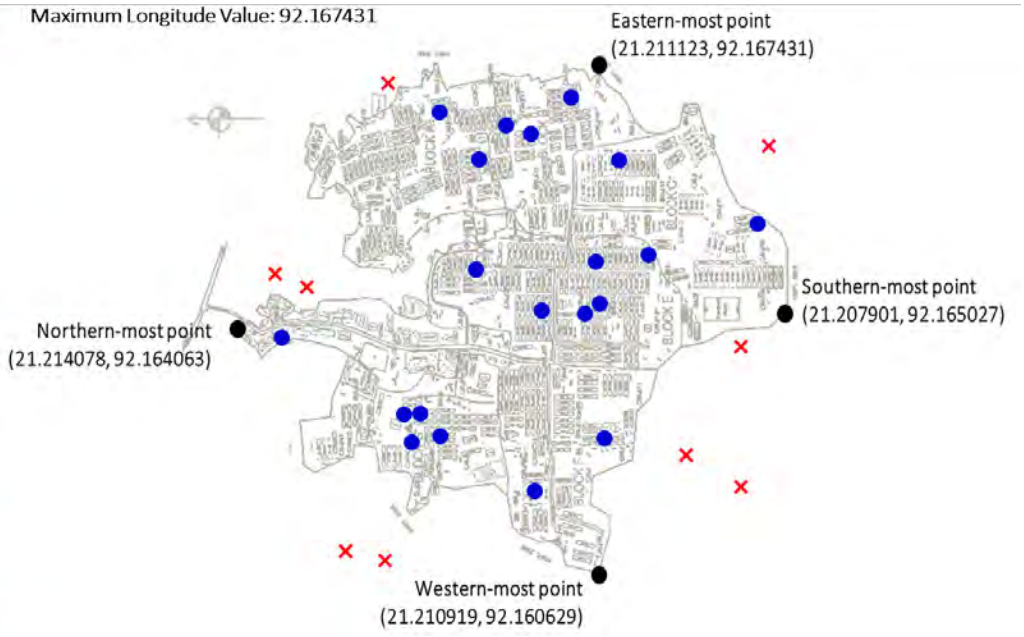
data by the corresponding PD factor to get the adjusted data, then take the average of the adjusted data and divide it by the average of the PD factors. This gives you the adjusted average, which you compare to the appropriate threshold value.

Coordinates selected randomly between the minimum and maximum latitude and longitude points, and example calculation of the adjusted average based on the PD factors at each point

	Random Coordinates	
	Latitude	Longitude
1	21.21120	92.16312
2	21.21388	92.16330
3	21.21192	92.16234
X	~~21.21153~~	~~92.16655~~
X	~~21.20944~~	~~92.16187~~
X	~~21.20812~~	~~92.16559~~
4	21.21371	92.16491
5	21.21132	92.16294
6	21.20967	92.16491
7	21.20994	92.16536
X	~~21.21079~~	~~92.16119~~
8	21.21124	92.16365
9	21.21354	92.16409
10	21.20899	92.16374
11	21.21343	92.16379
X	~~21.20904~~	~~92.16119~~
12	21.21231	92.16099
13	21.21089	92.16687
14	21.21184	92.16298
X	~~21.21328~~	~~92.16084~~
X	~~21.20824~~	~~92.16607~~
15	21.21174	92.16447
16	21.20995	92.16516
X	~~21.21408~~	~~92.16555~~
17	21.20959	92.16526
18	21.21312	92.16443
X	~~21.21169~~	~~92.16694~~
19	21.21311	92.16269
20	21.20964	92.16288

Minimum Latitude Value: 21.207901
Maximum Latitude Value: 21.214078

Minimum Longitude Value: 92.160629
Maximum Longitude Value: 92.167431



Random Location Point	Triangulation Distances (m)			Population Density (PD) Factor (HH / hectare)
	HH 1	HH 2	HH 3	
1	37	23	42	5
2	6	7	0	195
3	13	20	26	14
4	14	13	0	49
5	9	5	0	118
6	9	7	0	118
7	7	6	0	195
8	13	19	18	26
9	10	9	5	95
10	9	13	14	49
11	9	6	13	57
12	6	5	0	265
13	7	6	0	195
14	6	9	14	49
15	6	13	13	57
16	7	6	5	195
17	5	4	0	382
18	9	8	14	49
19	9	10	14	49
20	6	3	0	265
				Average PD Factor
				121

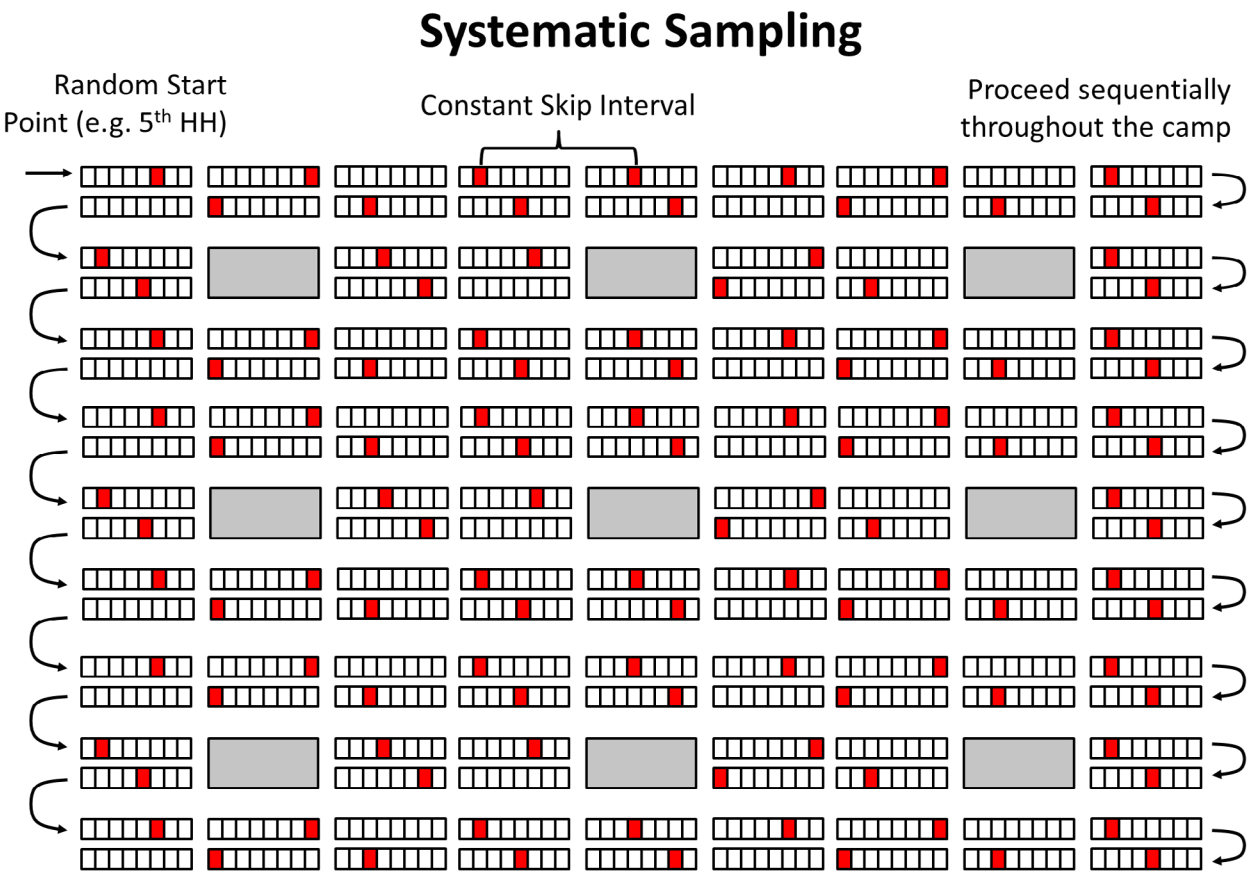
Raw Data (people per HH)			Adjusted Data (RawData * PD Factor)		
HH 1	HH 2	HH 3	HH 1	HH 2	HH 3
1	1	1	5	5	5
5	6	8	974	1169	1559
2	2	3	28	28	42
3	2	5	146	97	244
6	4	4	707	472	472
2	3	4	236	354	472
4	5	7	780	974	1364
1	2	1	26	53	26
2	1	2	191	95	191
1	2	2	49	97	97
3	3	6	170	170	339
7	8	9	1857	2122	2387
4	4	2	780	780	390
2	2	1	97	97	49
1	2	3	57	113	170
7	5	10	1364	974	1949
12	7	7	4584	2674	2674
2	2	2	97	97	97
2	3	4	97	146	195
6	7	8	1592	1857	2122
Raw Data Average (average of raw data)			**Adjusted Data Average** (average of adjusted data)		
3.9 people/HH			684		
			People per HH Adjusted Avg. = 5.6 p/HH (adjusted data avg. / avg. PD factor)		

Raw Data (female head of HH)			Adjusted Data (RawData * PD Factor)		
HH 1	HH 2	HH 3	HH 1	HH 2	HH 3
			0	0	0
100		100	19488	0	19488
			0	0	0
	100		0	4872	0
100	100	100	11789	11789	11789
100		100	11789	0	11789
100		100	19488	0	19488
			0	0	0
		100	0	0	9549
			0	0	0
	100		0	5650	0
		100	0	0	26526
100	100	100	19488	19488	19488
			0	0	0
			0	0	0
100		100	19488	0	19488
100		100	38197	0	38197
		100	0	0	4872
			0	0	0
100	100		26526	26526	0
Raw Data Average (average of raw data)			**Adjusted Data Average** (average of adjusted data)		
23 (of 60) --> 38%			6921		
			% Female Heads of HH Adjusted Value = 57% (adjusted data avg. / avg. PD factor)		

Systematic Sampling Approach

Description. The systematic sampling approach is based on the selection of households to survey along a fixed skip interval, until the desired number of households are surveyed, and the households selected are spaced evenly throughout the entire camp.

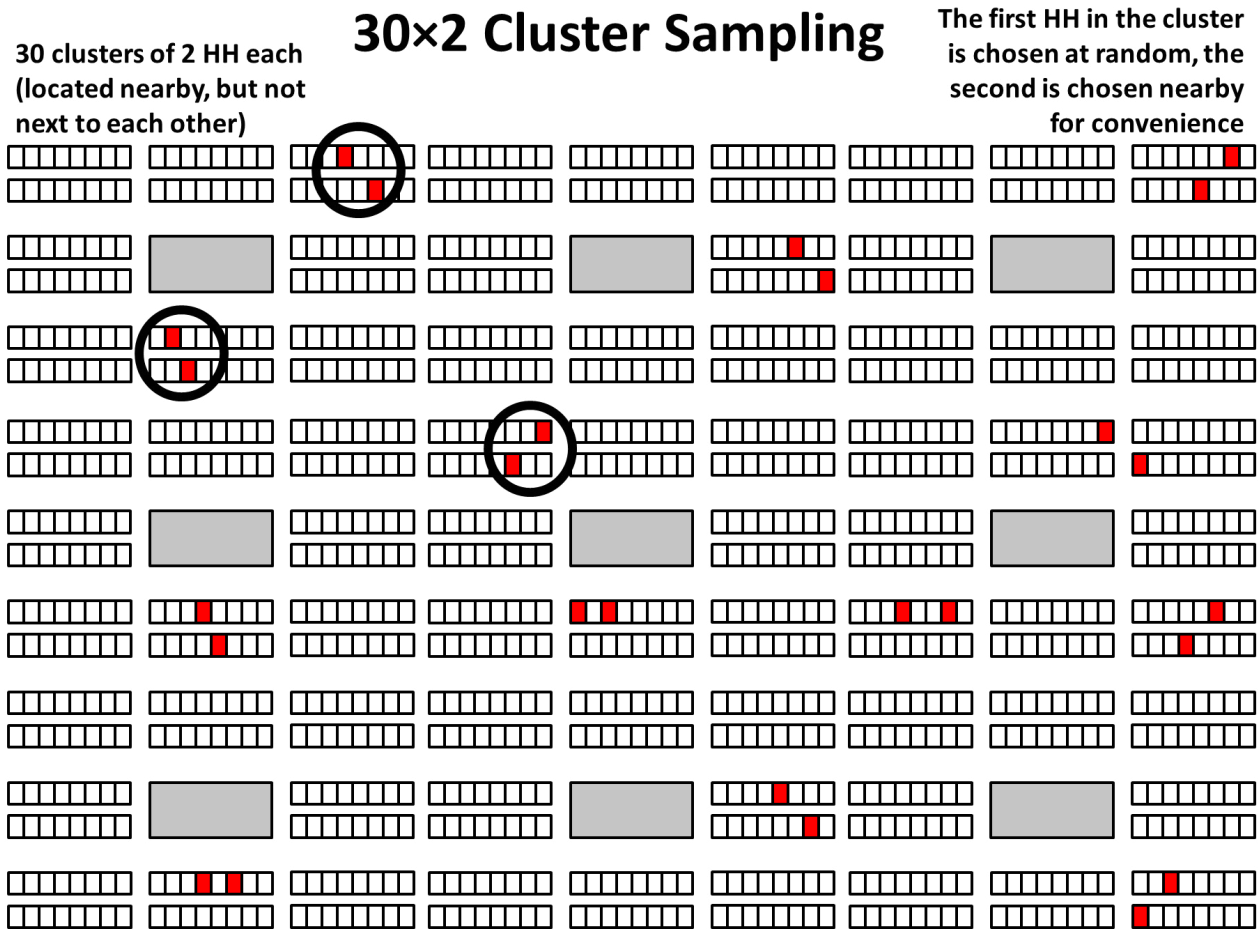
Procedure. Divide the total number of households by the sample size to get the skip interval. For example, in a zone with 500 HH, for a 60 HH sample, the skip interval is 500/60 = 8 HH. Choose a random number between 1 and 8 (inclusive) as the start point. For example, if the random number is 3, survey the 3rd HH, and then move throughout the entire camp, stopping at every 8th HH to collect data until you have visited a total of 60 HH (e.g. 3rd, 11th, 19th HH etc).



Cluster Sampling Approach

Description. The systematic cluster sampling approach is based on the selection of pairs of households chosen randomly from the camp. A list of the number of households is generated, and half of the households are chosen randomly from that list. The other half of the households are chosen due to their proximity to the randomly selected households. This reduces the amount of time associated with walking to each of the households, as two interviewers can survey pairs of nearby households simultaneously, saving time to complete the assessment.

Procedure. Choose 30 random numbers between 1 and the total number of households in the area to be surveyed. For example, if there are 500 HH, then choose 30 random numbers between 1 and 500 (inclusive). Send a group of two people through the camp to each of those thirty households. One person surveys one while the other person chooses another household located nearby in the same block as the first household (but not adjacent). Each pair of nearby houses is technically a "cluster of two", but the benchmark thresholds for the WASH service indicators are still assessed on all 60 HH. Note that on the figure, the cluster is shown as two adjacent houses, but in the field, the houses should be close to each other, but not necessarily adjacent.



UNHCR
The UN Refugee Agency

Rapid Methods for Assessing Water, Sanitation and Hygiene (WASH) Services at Refugee Camps in Emergency Settings

Briefing Note