



USAID GLOBAL HEALTH SUPPLY CHAIN PROGRAM

TECHNICAL ASSISTANCE, NATIONAL SUPPLY CHAIN ASSESSMENT TASK ORDER

Instructions for Using the Sampling Template

NSCA 2.0



DISCLAIMER: Development of the NSCA 2.0 toolkit was funded by the United States Agency for International Development (USAID). The authors' views expressed in this publication do not necessarily reflect the views of USAID or the United States Government.

INTRODUCTION

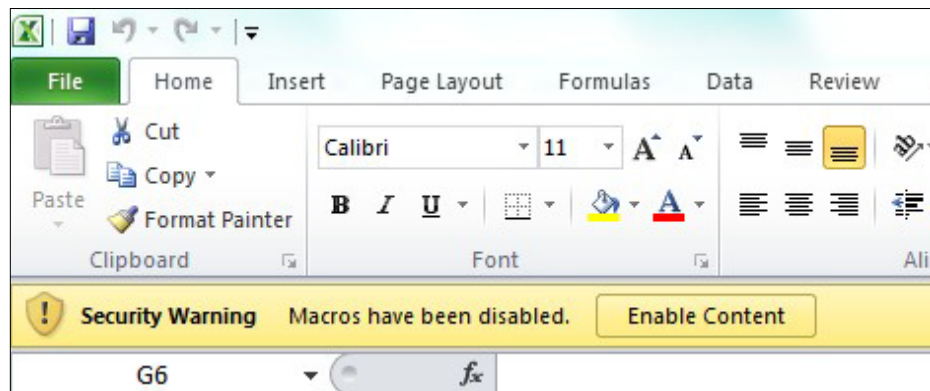
This document provides step-by-step instructions on how to use the sampling template provided as part of the NSCA 2.0 toolkit.

- The instruction guide is meant to be followed on a step-by-step basis. The instructions are meant to be completed in the order presented; completing steps out of order may result in formula errors or other errors that either fail to produce results or produce invalid results.
- The sampling template requires that the sample frame has been enumerated (see the Sampling Framework) – that is, that all of the entities that could be included in the sample for the assessment have been listed and checked. The template requires the use of Microsoft Excel 2010 (or later); the sample frame should be listed in an Excel workbook.
- Thick red arrows in this document indicate areas to click. Thin maroon arrows direct you to where you need to check data.
- This template can accommodate up to 5,000 rows of data, and 10 different types of entities (health centers, warehouses, etc.).
- The “automatic calculation” feature of Microsoft Excel has been deactivated because the calculations embedded in the workbook take several minutes to run. You may also want to adjust your settings on how often the workbook is ‘auto saved’ or backed up. If ‘auto save’ is activated, the workbook will attempt to calculate formulas; you can abort this process by hitting the “Esc” button.
- The sampling template was designed only for the convenience of users, and (provided skilled staff are available) other methods can be used to determine and draw the sample as appropriate.
- It is recommended that this guide and the accompanying Excel file “Sampling Template” be read and reviewed before starting the process of sampling; this will help the user understand the data that needs to be collected and the questions that need to be answered in order to complete the sample size calculation and the selection of the sample for the assessment.
- The sampling template is a tool for the provision of guidance; it is not meant to be used ‘blindly’. The assessment team conducting the NSCA is ultimately responsible for determining the sample size, drawing the sample, and determining the correctness and appropriateness of the resulting sample.
- It is advised that the tool not be uploaded into GoogleDocs and made into a Google sheet. This transition can handicap some of the functionality of the tool by disabling the macros.

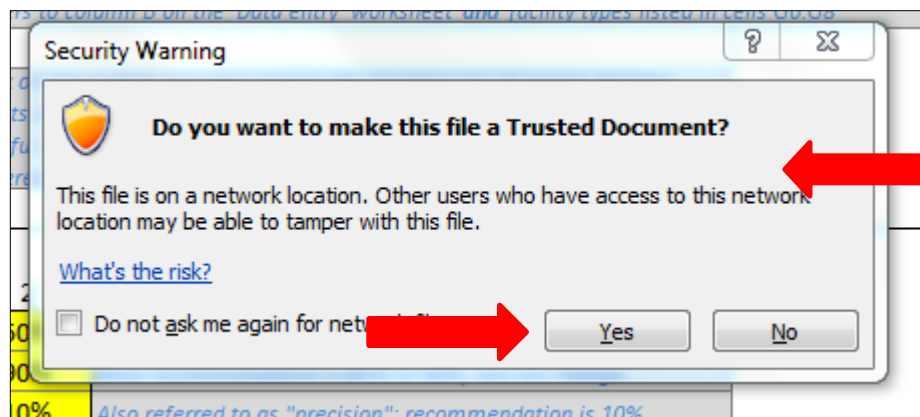
STEP ONE

STEP 1A: OPEN THE SAMPLING TEMPLATE WORKBOOK, ENABLE MACROS, AND SAVE THE SAMPLING TEMPLATE WORKBOOK

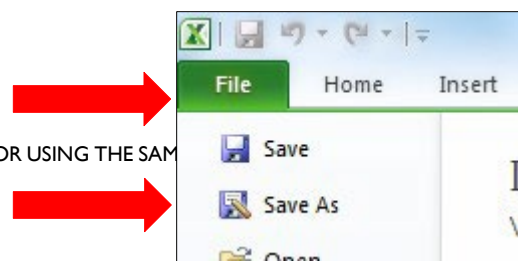
- Open the sampling template workbook in Microsoft Excel
- Enable the workbook's Macros by hitting the "Enable Content" button under the ribbon.



- Make the sampling template a "Trusted Document"

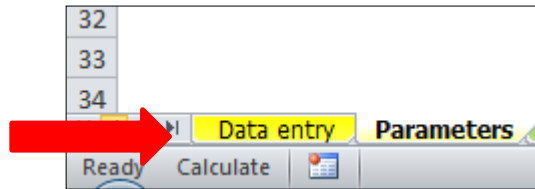


- Save the file under a new name. This will enable you to retain the original sampling template to return to if you need to re-start or re-do the sample selection. This may take a few minutes.



STEP 1B: COPY AND PASTE THE ENTITIES ELIGIBLE FOR INCLUSION IN THE ASSESSMENT ON THE “DATA ENTRY” WORKSHEET.

- a. Select the “Data Entry” worksheet at the lower left hand side of the workbook.



- b. Ensure that Cell A2 is selected on the “Data Entry” worksheet.

A screenshot of the 'Data Entry' worksheet. The table has two columns: 'Facility name' and 'Facility Type'. Row 1 contains the headers. Row 2 is selected, with a red arrow pointing to cell A2. The table structure is as follows:

	A	B
1	Facility name	Facility Type
2		
3		
4		
5		

- c. Open, if necessary, the workbook with the list of entities that could be included in the sample for the assessment. The enumeration list does not need to be in any type of order, nor do the columns need to be in the same order as presented in the Sampling Template – columns can be copied on one at a time. They should be inclusive of all entities (e.g., it should include district and provincial warehouses if these could be visited as part of the assessment).
- d. Paste the information from the enumeration list into the Sampling Template.

A screenshot of the 'Sampling Template' worksheet. The table has eight columns: 'Entity name', 'Entity Type', 'Intermediate Unit (e.g., Province)', 'Lowest Unit (e.g., District)', 'Intermediate level 1', 'Intermediate level 2', and 'Lower Strata'. Row 1 contains the headers. Row 2 is selected, with a red arrow pointing to cell A2. The table structure is as follows:

	A	B	C	D	E	F	G	H
1	Entity name	Entity Type	Intermediate Unit (e.g., Province)	Lowest Unit (e.g., District)	Intermediate level 1	Intermediate level 2		Lower Strata
2								
3								
4								
5								
6								
7								

- i. **Column A Entity Name:** This column should contain the unique name of every entity (E.g., Happy Health Center, Kabalee District Warehouse).
- ii. **Column B Entity Type:** In this column, list the type of facility (e.g., health facility, district hospital, referral hospital, district warehouse, provincial warehouse, central warehouse).

Important note: Entities that will be treated the same in the sampling should have the same “Entity Name”. Thus, if a country has two types of health centers (basic, advanced or similar), they should have the same name here if they are to be sample together. This is further explained in **Step 1d**.

- iii. **Column C Intermediate Unit (e.g., Province):** This lists the intermediate unit in which the entity is located. An intermediate unit is a unit between the central level and the “Lowest Distribution Point” or “Lowest Geopolitical Entity”, which is typically a province, state, or similar entity. If no such unit exists, this column can be left blank.
- iv. **Column D Lowest Unit (e.g., District):** In this column list the “Lowest Distribution Point” or “Lowest Geopolitical Entity” (which is typically a district, country, local government authority, or similar entity) associated with the entity listed in Column A.
- v. **Column E Intermediate level 1:** This column reflects the intermediate unit entered in column C, but the formulas in this worksheet use the information in column E rather than in column C. Every entity listed in Column A should have an entry in column E. This may mean that some new information is needed – for example, the central warehouse may need to have “Central” (if that is not the name of a province) entered here. This column can be made equal to column C, as shown below.

	A	B	C	D	E
	Entity name	Entity Type	Intermediate Unit (e.g., Province)	Lowest Unit (e.g., District)	Intermediate level 1
1					
2					=C2
3					
4					

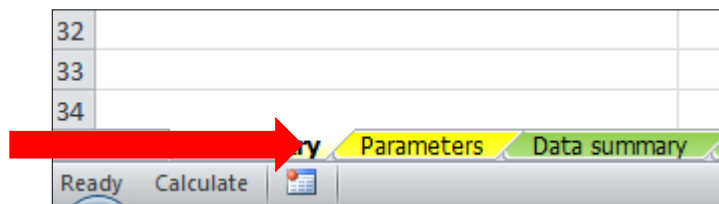
- vi. **Column F Intermediate level 2:** Similar to column E, this column reflected the lowest unit entered in column D, but needs to be filled in for all entities listed in Column A.
- vii. **Column H Lower Strata:** If applicable, enter strata for entities below the lowest unit level (e.g., health facilities). These strata may be types of health facilities (health centers, health posts) or other factors (e.g., urban and rural). Note that if different types of health facilities are to be considered for stratification, they should have the **same** name / type listed in column B. See also **Step 1d** for further explanation.

	A	B	C	D	E	F	G	H
					Intermedi ate level	Intermedi ate level		
1	Facility name	Facility Type	Province	District	1	2		Lower Stra
20	Health Center 8	Health Facility	Province 1	District 1	Province 1	District 1		Health Cente
21	Health Center 9	Health Facility	Province 1	District 1	Province 1	District 1		Health Cente
22	Health Post 23	Health Facility	Province 1	District 1	Province 1	District 1		Health Post
23	Health Post 24	Health Facility	Province 1	District 1	Province 1	District 1		Health Post

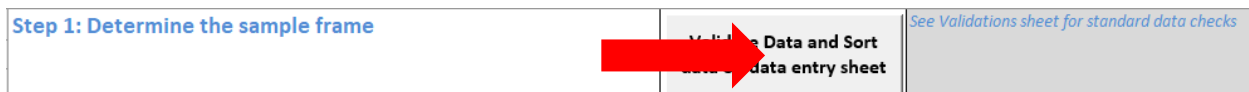
- viii. **Do NOT enter any data in Columns I, J, or K.** The columns contain formulas necessary for the sampling tool's calculations and altering these formulas will produce errors.

STEP IC: GO TO THE “PARAMETERS” WORKSHEET AND USE THE MACRO BUTTON TO SORT THE DATA ON THE “DATA ENTRY” SHEET

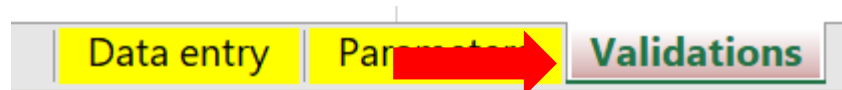
- a. Select the “Parameters” worksheet at the lower left of the workbook.



- b. Click on the macro button “Validate data and sort data on data entry sheet” at the top of the “Parameters” worksheet. This macro will automatically sort the data on the Data Entry Worksheet in the way that is needed for this workbook to operate correctly.



After pressing this button the “Validations” tab will appear showing a series of standard data checks that are performed automatically by the tool.



Validations:

1. Trim Cell Value – All data in the “Data Entry” tab will remove trailing and leading spaces
2. Count of Unique Values – The Validations tab will show the counts for unique values in each column. If there are more than **3** unique values in the

Entity Type and/or *Lower Strata*, then an error box will appear asking for the data to be corrected.

	1	16	260	16	260	0	3
Entity Type	Intermediate Unit (e.g., Provir)		Lowest Unit (e.g., District)	Intermediate level 1	Intermediate level 2	Lower Strata	
health facility	Ahafo	Asunafo North	Ahafo	Asunafo North		Hospital	
	Ashanti	Asunafo South	Ashanti	Asunafo South		Health Centre	

3. **Matching Columns** – The *Intermediate Unit* should match identically the *Intermediate Level 1*. Likewise, the *Lowest Unit* should match identically the *Intermediate Level 2*. If these columns do not have an identical match then an error message will appear asking for the data to be corrected.

Match	Match
Intermediate Check	Lowest Unit Check
0	0
0	0
0	0
0	0
0	0

- c. Note that cells on the “Parameters” worksheet are color coded:
 - i. Cells colored in **yellow** are meant to be altered or changed by the user.
 - ii. Cells colored in white are NOT meant to be changed by the user.
 - iii. Cells colored in **gray** with blue text provide instructions or explanations to help guide the user through the process of using this worksheet.
 - iv. The steps to follow are numbered along the right side of the worksheet.
- d. Also note that automatic calculation of formulas has been turned off in this workbook. It was turned off because the formulas in the workbook take several minutes to calculate on a typical laptop. This means that the results of what you type in will **not** automatically appear after you type them in.

STEP 1D: CHECK THE NUMBER OF INTERMEDIATE AND LOWER UNITS, ENTER THE TYPES OF ENTITIES BELOW THE LOWEST UNITS TO BE SAMPLED, AND RUN THE “CHECK FACILITY DATA” MACRO.

- a. Rows 4 and 5 of the “Parameters” worksheet displays the number of intermediate and lower units (e.g., provinces and districts) based on the information you entered on the “Data Entry” worksheet.

	A	B	C	D	E	F	G
1	Step 1: Determine the sample frame						Validate Data and Sort data on data entry sheet
2							
3							
4	Number of:	Intermediate levels (provinces or similar):					16
5		Lowest distribution points (districts or similar):					260

- i. **If the numbers are not what you expected:** Return to the “Data Entry” worksheet and check for errors. Common errors include multiple spellings of the same province/district name, or inconsistent spacing:

D3		District 1		There is an extra space after the “l” in this name; Excel will think that “District l” is different than “District I”	
	A	B	C	D	
1	Facility name	Facility Type	Province	District	
2	District Hospital 3	District Hospital	Province 1	District 1	
3	District Warehouse 4	District Warehouse	Province 1	District 1	

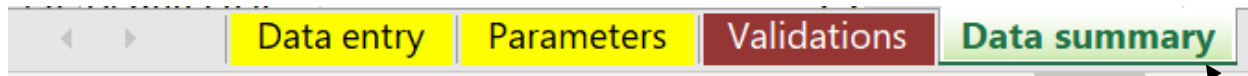
- ii. **Ensure errors are fixed:** After you correct errors on the “Data Entry” worksheet, you must re-run the “Sort data on data entry sheet” at the top of the “Parameters” worksheet.
- b. Enter the names of the entities below the lowest unit (typically health facilities) that you want to have sampled **at the full statistical power** (determined in Step 2) in Cells G6:G8. You may enter up to 3 types of entities.

	A	B	C	D	E	F	G
1	Step 1: Determine the sample frame						Sort data on data entry sheet
2							
3							
4	Number of:	Intermediate levels (provinces or similar):					10
5		Lowest distribution points (districts or similar):					38
6		Facilities to include in count:					1 Health Facility
7							2
8							3

- i. This means that, for example, if you determine the 60 entities need to be sampled, all of the entities named in this step will have 60 units selected for the final sample. If you enter “Health Center” and “Health Post” here, then a total of 120 entities (60 health centers and 60 health posts) will be selected.
 - ii. If you want in total 60 entities to be selected but want 20 to be health posts and 40 to be health centers, then you should select stratification in **Step 3** (and ensure that the names of the health centers and health posts are the same in Column B of the “Data Entry” worksheet).
 - iii. You should **NOT** enter here the names of entities where there is only a few, one, or less than one entity per lowest unit. For example, you would not enter district hospitals, provincial hospitals, district warehouses, provincial warehouses, national referral hospitals, etc. here.
- c. Click on the **“Check facility data”** macro

	A	B	C	D	E	F	G
1	Step 1: Determine the sample frame						Sort data on data entry sheet
2							
3							
4	Number of:	Intermediate levels (provinces or similar):					10
5		Lowest distribution points (districts or similar):					38
6		Facilities to include in count:					1 Health Facility
7							2
8							3
9		Count of facilities:					870
10		Average number of facilities per lowest distribution point:					22.9
11							Check facility data
12							
13							

- i. This will automatically take you to the “Data Summary” worksheet.



- ii. On the “Data Summary” worksheet, there are three columns that need to be assessed for accuracy:

1. **Column A and B** list the lowest geopolitical or delivery units (e.g., districts) and the number of entities listed on the “Parameters” worksheet Cells G6:G8 listed

	A	B
1	Lowest Unit (e.g., District)	
2	District 1	21
3	District 10	22
4	District 11	22
5	District 12	22
6	District 13	21
7	District 14	22

This is the number of health facilities in “District 1”, as listed on the “Data Entry” worksheet.

2. Ensure that the names of all districts are correct, and that there are no ‘duplicated’ district names. Per above, ensure that there are not extra spaces at the end of names. Any errors detected must be corrected on the “Data Entry” worksheet, and the process of sorting and checking the data must be re-started.
3. **Columns D and E** list the different entities (as listed in Column B of the “Data Entry” worksheet) included in the sample frame and the **total** number of each entity included.

D	E
Entity Type	
District Hospital	4
District Hospital	39
District Pharmacy	13
District Pharmacy	17
Health Center	514
Referral Hospital	4

This is the total number of district hospitals included in the entire sample frame (e.g., country). Note, however, that “District Hospital” appears two times, indicating that there is an error (likely one has an extra space at the end of

4. Again, ensure that the names of all entities are corrects, and that there are no 'duplicated' entity names. Any errors detected must be corrected on the "Data Entry" worksheet, and the process of sorting and checking the data must be re-started.
5. **Columns G, H, and I** list the names of the intermediate (e.g., province) units included in the data entry sheet, the number of entities listed on the "Parameters" worksheet Cells G6:G8 listed within each intermediate unit, and the number of lowest distribution points (e.g., districts) within each intermediate unit (e.g., the number of districts in a provinces).

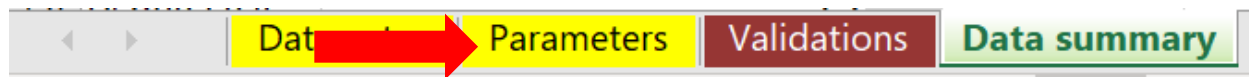
G	H	I
Intermediate	Number of counted entities	Number of Lowest distribution points
Province 1	87	5
Province 10	87	4
Province 2	87	4
Province 3	87	4
Province 4	87	4
Province 5	87	4
Province 6	87	4

6. Again, ensure that the names of all entities are corrects, and that there are no 'duplicated' entity names. Any errors detected must be corrected on the "Data Entry" worksheet, and the process of sorting and checking the data must be re-started.

STEP TWO

STEP 2A: ENTER THE RATIO, LEVEL OF CONFIDENCE, AND MARGIN OF ERROR TO USE TO DETERMINE THE SAMPLE SIZE.

- a. Return to the “Parameters” worksheet by clicking on the tab in the lower left of the screen.



- b. Enter the **ratio** in row 17 of the “Parameters” worksheet.

	A	B	C	D	E	F	G	H
15	Step 2: Determine the sample size							
16	Sampling parameters						Scenario 1	2
17						Ratio:	50%	50%
18						Level of Confidence (alpha):	85%	90%
19						Margin of error (±amount):	10%	10%

- You can enter two ratios, but it is recommended that you leave this ratio at 50% unless you have a strong belief that this is not relevant for the country you are assessing (e.g., past NSCA results from that country are available to inform this number). Using 50% will maximize the sample size as compared to other ratios. Note that **ONLY** the data entered in Column G will be used for determining the overall sample size; Column H is provided to allow you to compare the effect of different parameters on the sample size.
- The ratio should be between 0% and 100%. Entering a negative number or a number greater than 100% will result in calculation errors.

- c. Enter the **level of confidence** in row 18 of the “Parameters” worksheet.

	A	B	C	D	E	F	G	H
15	Step 2: Determine the sample size							
16	Sampling parameters						Scenario 1	2
17						Ratio:	50%	50%
18						Level of Confidence (alpha):	85%	90%
19						Margin of error (±amount):	10%	10%

- i. The NSCA 2.0 recommends a minimum level of confidence of 85%; higher levels of confidence can be used if necessary.
- ii. The level of confidence should be between 0% and 100%. Entering a negative number or a number greater than 100% will result in calculation errors.

d. Enter the **Margin of Error** in row 19 of the “Parameters” worksheet.

	A	B	C	D	E	F	G	H
15	Step 2: Determine the sample size							
16	Sampling parameters						Scenario 1	2
17						Ratio:	50%	50%
18						Level of Confidence (alpha):	85%	90%
19						Margin of error (±amount)	10%	10%

- i. The NSCA 2.0 recommends a minimum margin of error of 10%; other levels of confidence can be used if necessary.
- ii. The margin of error should be between 1% and 50%. Entering a negative number or a number greater than 50% will result in calculation errors.

16	Sampling parameters						Scenario 1	2
17						Ratio:	50%	50%
18						Level of Confidence (alpha):	85%	90%
19						Margin of error (±amount):	10%	10%
20								
21						Estimated # of lowest distribution points for sample:	15	18
22						Estimated # of entities per lowest distribution point:	3	4
23								

- e. Click on the **“Update Calculations” macro button** located beginning in Cell B30 in order to adjust the results to reflect the numbers

Row 21 reflects the number of lowest distribution points / lowest units (e.g., districts) that will be included in the sample.

- i. **Row 22** reflects the number of entities (e.g., health facilities) that will be included in the sample for each of the lowest units included in the sample. In the example above, 15 lowest units (e.g., districts) will be included in the sample, and 3 health entities from each lowest unit (for a total of $3 * 15 = 45$ health entities.)
 - ii. **NOTE** that the formula employed here assumes that the clustering of entities within lowest units will result in an increase in the total number of entities needed by a factor of 1.6. This number is based on piloting experience with the NSCA 2.0. If more or less than the number of entities per lowest unit is desired, it is recommended that this be changed in **Step 2c** and the numbers entered here **NOT** be changed.
- f. Update the parameters as needed by repeating steps **a** through **e** above.

STEP 2B: DETERMINE THE SEQUENCE OF SAMPLING.

- a. The sampling tool allows two methods for selecting the lowest units to be included in the sample: (i) The lowest units are drawn directly, regardless of which intermediate level they are located in, and (ii) the selection first of intermediate levels (e.g., provinces) and then the selection of lowest units (e.g., districts) from those intermediate levels included.
 - i. Typically, the first option is preferable for smaller countries where travel to different provinces is not overly costly or onerous, while the second option would be preferable in larger countries, where time and travel costs may be prohibitive in allowing the assessment teams to travel throughout the country. The sampling may be done both ways to determine whether or not the first option is feasible.
 - ii. Selection of the second option may (and likely will) increase the number of lower units (e.g., districts) and health entities (e.g., health facilities) that need to be visited.
 - iii. To select the first option, in which the lowest units (e.g., districts) are selected without considering which intermediate level (e.g., province) they are located in, leave Cell G24 as “No”.

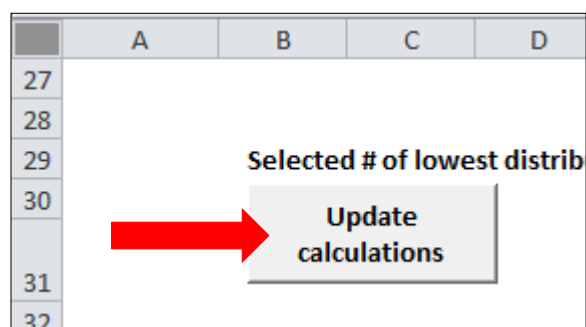
	A	B	C	D	E	F	G	
21		Estimated # of lowest distribution points for sample:					15	
22		Estimated # of entities per lowest distribution point:					3	
23								
24		Select intermediate levels and then lowest distribution points?					No	

Cell G29: Using this approach, **Cell G29** should be the same as **Cell G21**.

- iv. To select the second option, in which intermediate units (e.g., provinces) are first selected, and then the lowest units (e.g., districts) are selected from the included intermediate levels (e.g., province), enter “Yes” in Cell G24.

	A	B	C	D	E	F	G	
21		Estimated # of lowest distribution points for sample:					15	
22		Estimated # of entities per lowest distribution point:					3	
23								
24		Select intermediate levels and then lowest distribution points?					Yes	

- v. Click on the “Update Calculations” macro button located beginning in Cell B30.



- vi. Review and update the results

23			
24	Select intermediate levels and then lowest distribution points?	Yes	
25	List expected design effect:	1.1	
26	List number of lowest distribution points per intermediate level:	2	
27	Number of LDPs with design effect:	17	
28	Number of intermediate levels to include:	9	
29	Selected # of lowest distribution points for sample:	18	

Cell G25: In this cell, enter the design effect due to clustering of lowest units (e.g., districts) in intermediate units (e.g., provinces) that is anticipated due to the use of this sampling approach. The number reflects how much the sample size of lowest units must increase in order to account for the clustering. The NSCA 2.0 pilots **did not** use this approach when drawing samples and data to anticipate the magnitude of this design effect are not available. Generally, if lowest units (e.g., districts) operate fairly independently from intermediate units (e.g., provinces) then a smaller number is appropriate.

This may be modified in later versions of this tool as more data is amassed.

The number entered should not be less than 1.

Cell G26: The formula in this cell calculates the number of lowest units (e.g., districts) to be sampled from each intermediate unit (e.g., provinces). The number is based on the minimum number of lowest units in any of the intermediate units (as found in the “Data entry” worksheet). Thus, if the intermediate unit with the least number of lowest units is selected, this will ensure that the sample size is maintained.

Cell G27: This is an intermediate cell used for calculation; it can be ignored.

Cell G28: Based on the data entered in Cells G25 and G26, this cell indicates the number of intermediate units (e.g., provinces) that will be included in the sample.

Cell G29: Based on the data entered in Cells G25, G26, and calculated in G28, this cell gives the final number of lowest units (e.g., districts) that will be included in the sample.

- vii. If you change any of the numbers in cells G25:G29, you need to **Click on the “Update Calculations” macro button** located beginning in Cell B30 for the changes to be reflected in the worksheet.

STEP 2C: DETERMINE THE FINAL SAMPLE SIZE.

- a. You now need to determine how many of the different types of entities you will include in the final sample. Cells F32:F41 should list the different, unique types of entities to be included in the assessment. If not, please **click on the “check Facility data” macro button** and / or **click on the “Update Calculations” macro button**.

	Update calculations				
31		Type of facility:	How sampled:	Facilities per LDP	-OR- Total number of facilities
32		District Hospital	Per LDP	1	
33		District Warehouse	Per LDP	1	
34		Health Facility	Per LDP	3	
35		National Hospital	Total number		2
36		Provincial Hospital	Per Intermediate Level	1	
37		Provincial Warehouse	Per Intermediate Level	1	
38		✓			
39		✓			
40		✓			
41		✓			

- b. There are three methods for determining the number of each type of entity to visit: (i) per Lowest Distribution Point (LDP), (ii) per Intermediate Level, and (iii) Total number. For each entity listed in column F, one method needs to be selected. This can be done by clicking on the cell, then clicking on the gray box with a downward chevron located in the center, and then selecting the method from the resulting drop-down list.

Type of facility:	How sampled:	Facilities per LDP
District Hospital		
District Hospital		
District Pharmacy		

- Per LDP:** This method will draw the number of entities listed in column H from each of the lowest units selected. Thus, this should be used for entities that have one or more entities per lowest unit (e.g., district warehouse, district hospital, or health facilities).
- Per Intermediate Level:** Similar to “Per LDP”, this option will select the number of entities listed in column H from each of the intermediate units selected. Thus, this should be used for entities that have one or more entities per intermediate unit (e.g., provincial warehouse or provincial hospital). ***This option can only be used if there are data in column E of the “Data entry” worksheet.***

- iii. **Total Number:** This indicates that the number of entities selected is listed in column K, and typically will be used for entities that do not exist in every intermediate level unit (e.g., central warehouse, national referral hospitals, Ministry of Health).
- c. Once you have entered how the number of entities will be determined, **click on the “Update Calculations” macro button.**
- d. Review the numbers and enter new numbers as appropriate in Columns H and K.

	Update calculations	Type of facility:	How sampled:	Facilities per LDP	-OR-	Total number of facilities
31						
32		District Hospital	Per LDP	1		
33		District Warehouse	Per LDP	1		
34		Health Facility	Per LDP	3		
35		National Hospital	Total number			2
36		Provincial Hospital	Per Intermediate Level	1		
37		Provincial Warehouse	Per Intermediate Level	1		
38						
39						
40						
41						


- i. The formula in **Column H** automatically enters the number from Cell G22 in those entities that have been listed in Cells G6:G8. These are the entities for which you will have the same sampling characteristics (level of confidence, margin of error) as specified above. It is not recommended that you change these numbers.
- ii. If an entity listed as “Per LDP” is not listed in Cells G6:G8 and for those entities listed as “Per Intermediate Level”, the default value is 1; this can be changed if necessary.
- iii. For entities listed as “Total Number” there is no default value; users are required to enter the sample size directly into the worksheet.
- iv. **Ensure** that a number is listed for all entities before proceeding.

e.


STEP THREE

DETERMINE THE METHOD FOR SAMPLING

- a. There are two decisions that need to be made for this step: (i) whether or not to stratify entities below the lowest unit, and (ii) whether to use sampling with probability of selection proportional to the size of the population (PPS) or simple random sampling with selecting the sample. It is recommended to use PPS sampling; this is typically done in many survey sample selection processes and ensures that every member of the population served by the 'lowest distribution point' has an equal chance of being included in the sample.
- b. **Whether or not to stratify entities below the lowest unit:** In cell G45, select "No" if you do not want to include stratification in the sample, and select "Yes" if you do.

43	Step 3: Determine the method for sampling	
44	Are there Strata included for:	
45	Facilities below the lowest delivery point?	No 
46		

- i. To select "Yes" for this option, you must have entered in column H on the "Data entry" worksheet values for entities that have the same name in Column B of the Data entry" worksheet.
 - ii. You can enter up to 3 different values (strata) in column H. More than 3 will cause the template to incorrectly stratify.
 - iii. If you have met these requirements, and have selected "Yes" in cell G45, then the template will draw a sample that, as closely as possible, will reflect the proportion of each strata as present in the overall population. Thus, if the strata are "Urban" and "Rural" and "Rural" is 30% of all entities, the template will work to draw a sample that has 30% of selected facilities from the "Rural" strata.
- c. **Select the sample PPS or using simple random sampling: The NSCA 2.0 recommendation is to sample PPS, and this is the default setting in the template (Cell G47 = "Yes").** However, you may decide that simple random sampling is more appropriate for your assessment. Simple random sampling does not consider the 'population' of sites within a geographic / political area but may be applicable in very small countries or for assessments that have a very limited scope. In this case, you should change Cell G47 to "No". The intermediate and lowest units (e.g., provinces and districts) as appropriate will be selected at random, without consideration of how many entities are located in each unit.

Select data with the probability of selection proportional to the size of the population (PPS): Yes 

- d. Click on the “Verify / enter population and strata” macro button located in cell G53.

53		Verify / enter population and strata
54		
55		
56		

- i. Clicking on this button will automatically move you to the “Population” worksheet. You will need to review or update this worksheet to confirm the data are correct.
- ii. **Columns A and B** present the lowest unit (e.g., district) in column A, and the number of ‘entities’ (those listed in in Cells G6:G8 of the “Parameters” worksheet) in each lowest unit in column B. If you are using simple random sampling (Cell G47 on the “Parameters” worksheet is equal to “No”) the population of the lowest units will not matter.

	A	B
1	Lowests Distribution Point	Number of counted entities
2	District 1	21
3	District 10	22
4	District 11	22
5	District 12	22
6	District 13	21



You can change the population in Column B. The population presented is only one possible population that could be used to select facilities. If you are using PPS sampling, you may consider a population other than the number of entities (e.g., number of patients served). You will need to enter these data in column B for each of the lowest units listed.

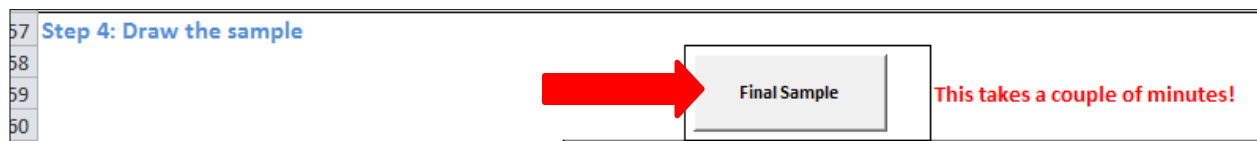
- iii. **Columns G and H** present the strata you are using; if cell G45 on the “Parameters” worksheet is set to “No”, this will not be relevant. Further, if there are no data in column H on the “Data entry” worksheet, these columns will be empty. If you are using stratified sampling, the numbers presented should be verified as accurate.

G	H
Entity-level strata	Number
Health Center	720
Health Post	150

STEP FOUR

DRAW YOUR SAMPLE

- a. Once the population has been verified and you are satisfied with all of the input data on the “Parameters” worksheet, you can select the final sample by clicking on the “Final Sample” macro button located in cell G58 in the “Parameters” worksheet.



- i. The calculations take a few minutes. You cannot use Microsoft Excel during this time, but usually other programs will be operational.
 - ii. Clicking on the macro button will automatically take you to the “Final Sample (Long)” worksheet.
- b. The “**Final Sample (Long)**” worksheet lists the names of the entities included in the sample (Column A), the sample weights associated with each entity (Column B), the type of entity (Column C), the intermediate unit (Column D, if appropriate), the lowest unit (Column E), and the strata for entities below the lowest unit (Column G, if appropriate).

	A	B	C	D	E	G
1	Entity name	Sampling Weight	Entity Type	Intermediate unit	Lowest unit	Lower Strata
2	National Hospital 1		2. National Hospital	Province 1	District 1	0
3	Provincial Hospital 1	1.666666667	Provincial Hospital	Province 1	District 1	0
4	Provincial Warehouse 2	1.666666667	Provincial Warehouse	Province 1	District 1	0
5	District Hospital 77	8.734467088	District Hospital	Province 1	District 3	0
6	District Warehouse 53	1.767279147	District Warehouse	Province 1	District 3	0
7	Health Center 57	33.15874087	Health Facility	Province 1	District 3	Health Center
8	Health Center 86	33.15874087	Health Facility	Province 1	District 3	Health Center
9	District Hospital 903	2.384648157	District Hospital	Province 10	District 37	0

- i. Column H is a place holder to help sort the data; it is not meant for use.
 - ii. The entities should be sorted according to intermediate unit, then lowest unit, the entity type, then the entity name
 - iii. Sample weights are calculated as the inverse of the probability of an entity to be selected. These can be used in the analysis to derive estimates for the overall population (refer to instructions on data analysis for further information on how to use these data).
- c. The “**Final Sample**” worksheet contains the same data as the “Final Sample (Long)” worksheet, but organized in a different manner:

- i. Columns A through D list the entities selected that are below the lowest unit, along with their sample weights, lowest unit, and type of entity.

	A	B	C	D
1	Entities below the lowest distribution point	Weight	Lowest Unit	Type of facility
2	Health Center 843	20.61	District 34	Health Facility
3	Health Post 774	16.44	District 31	Health Facility
4				
5	Health Post 974	21.68	District 39	Health Facility
6	Health Center 958	20.61	District 39	Health Facility
7	Health Center 839	20.61	District 34	Health Facility

- ii. Columns F through I list the lowest units' entities included in the sample, along with their sample weights, lowest unit, and type of entity.

F	G	H	I
Lowest distribution point level entitites	Weight	Lowest Unit	Type of entity
District Hospital 203	2.412377	District 9	District Hospital
District Warehouse 228	2.329595	District 10	District Warehouse
District Hospital 827	4.605446	District 34	District Hospital
District Hospital 227	4.605446	District 10	District Hospital
District Hospital 552	4.658382	District 23	District Hospital
District Warehouse 353	1.767279	District 15	District Warehouse
District Hospital 727	4.658382	District 30	District Hospital

- iii. Columns K through N list the entities selected that are above the lowest and intermediate units, along with their sample weights, the lowest unit where they are located, and type of entity.


K	L	M	N
Higher level entitites	Weight	Lowest Unit	Type of entity
National Hospital 3	2	District 25	National Hospital
National Hospital 1	2	District 1	National Hospital

- iv. Columns P through S list the entities selected that at the intermediate unit level (if appropriate), along with their sample weights, the lowest unit where they are located, and type of entity.

P	Q	R	S	T
Intermediate level entities	Weight	Lowest Unit	Type of entity	
Provincial Hospital 801	1.666667	District 33	Provincial Hospital	
Provincial Warehouse 702	1.666667	District 29	Provincial Warehouse	
Provincial Hospital 101	1.666667	District 5	Provincial Hospital	
Provincial Hospital 601	1.666667	District 25	Provincial Hospital	
Provincial Hospital 301	1.666667	District 13	Provincial Hospital	
Provincial Warehouse 302	1.666667	District 13	Provincial Warehouse	

Be sure to save your results!

Lastly, make a copy of the final sample so that the site list can be shared with in-country stakeholders for planning purposes without sharing the overall sampling tool, which is a large file. This can be done by copying the entire “**Final Sample (Long)**” worksheet and pasting the values into a new workbook. To copy the entire sheet, select the gray box in the upper left-hand corner of the worksheet:



	A
1	Entities below the lowest distributi
2	Health Center 843
3	Health Post 774

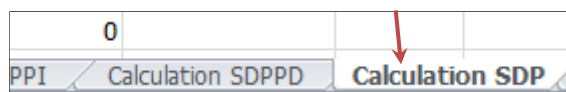
Then, right click the mouse and select “Copy”. Open a new work book, select cell A1, right click the mouse, and select “Paste Special” and then select “Values”.

OTHER CONSIDERATIONS

REPLACEMENT FACILITIES

In some cases, entities selected for inclusion in the assessment may need to be ‘replaced’ at the last moment (due, e.g., to travel considerations or additional information indicating that they are unsuitable for the assessment). In order to find replacement facilities, the original calculation sheets will need to be accessed. The facilities, e.g., on the SDP sheets are arranged in a random order which reflects the order in which they should be selected.

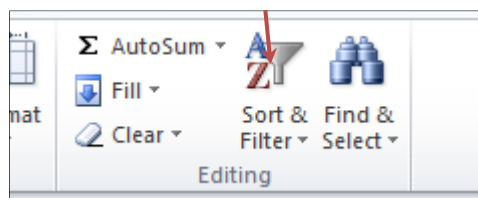
Thus, if you need to replace one facility in one lowest unit (e.g., district), first select the calculation sheet for SDPs:



Next, filter the data for the particular lowest unit that needs a replacement facility. To do this, select cell D1 on the Calculation SDP worksheet:

	A	B	C	D	E
1	Entity name	Entity Type	Intermediate Unit (e.g., Province)	Lowest Unit (e.g., District)	Interm lev

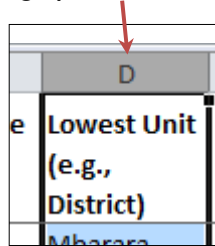
Then select the “Sort and Filter” button on the home ribbon in Excel:



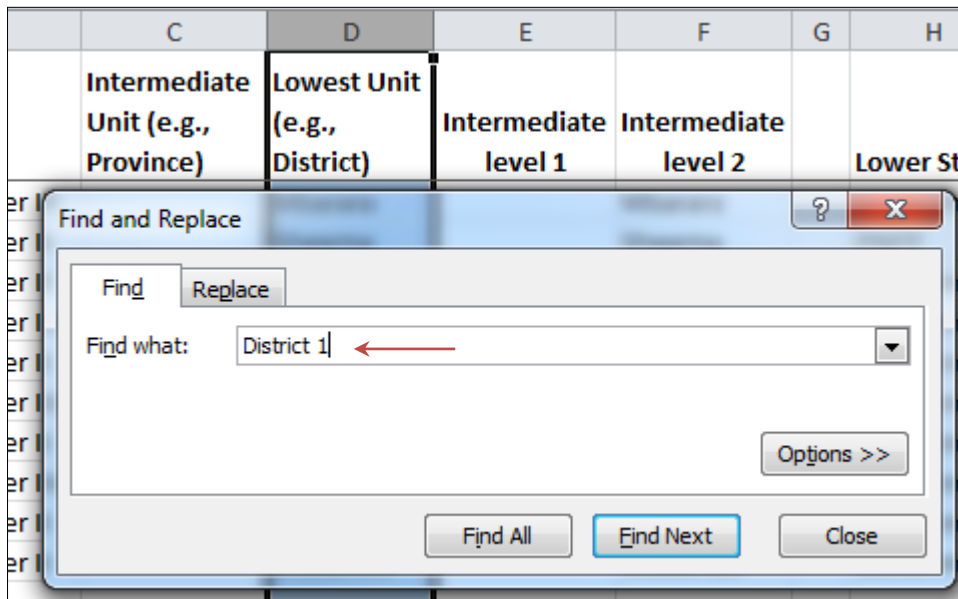
Select the “Filter” option from the drop down list. This should result in chevrons appearing in Row 1. First, you need to select district that have been selected by filtering column L to include only values with a 1:

	Count of Col	for calcul	ion	District Selected?	Firs
nt	1	0.508076	0.000635	1	Hea

Select the entire Column D by clicking on the gray cell at the top of Column D:



Next use the “Find” function to locate entities located in the lowest unit. Simultaneously press “Ctrl” and “F” keys on the keyboard:



Type in the name of the lowest unit where you need to find a replacement facility and click “Find Next”.

Check the data in Columns P and Q to find the replacement facility. Column P lists the type of entity (e.g., Health center), while Column Q lists the selection order. You are looking for a number in Column Q that is one greater than the number of entities selected per lowest unit. For example, if you have selected 3 health centers per lowest unit, you are then looking for an entity in the lowest unit that lists “Health Center” in Column P and has the number “4” in Column Q, and lists the correct lowest unit in Column D:

P	Q	R
Total Select1	Total Select1num	Selected?1
Health Center	4	0

Note that Column R should be zero. You can consecutively check the entities in the lowest unit but clicking the “Find Next” button in the “Find and Replace” dialogue box.

- a. **Note** that this method works only for individual entities. If entire districts or similar entities need to be replaced, then the entire sample should be re-done.

SAMPLE WEIGHT FORMULAS

The sample weights calculated by this template are standardized to reflect the entire population in the sample frame using the following formulae:

Probability of selecting an intermediate or lowest unit:

$$P_d = \frac{y * n_d}{N}$$

Where

P_d = Probability of selecting area d,

y = The number of intermediate or lowest units included in the sample (or for lowest units, in the intermediate level),

n_d = Number of health facilities / total population in area d, and

N = Total number of health facilities/total population in the country that have been included in the sample frame.

Probability of selecting an SDP:

$$P_{dif} = \frac{x_{id}}{n_{id}} * P_d$$

Where

P_{dif} = Probability of selecting entity f of type i in area d,

X_{id} = the number of entities of type i selected for inclusion in the sample in area d , and

n_{id} = the number of entities of type i in area d .

Initial sample weights

For intermediate or lowest units:

$$w_d = \frac{1}{P_d}$$

For SDPs:

$$w_{dif} = \frac{1}{P_{dif}}$$

Scaled sample weights:

To standardize sample weights for SDPs:

$$\hat{w}_{dif} = \frac{w_{dif}}{\left[\sum_f^{ns_i} (w_{dif}) / ns_i \right]} * \frac{ns_i}{N}$$

Where:

ns_i = the number of entities of type i included in the sample.