AMF – PDM Sampling Methodologies

Document version: March 2020

Summary

This document details the way AMF carries out the identification of specific villages and households to be visited during monitoring activities. It covers the way villages and households are randomly selected and the requirements and potential flexibilities in the field when data is being gathered.

Review process: AMF frequently revisits these methodologies and operational practices to assess if work can be done more effectively i.e. more accurately and cost effectively.

Sections

1. Post-distribution monitoring (PDM)
    1) Village and household selection for Main ('1.5%') household visits
    2) Household selection for Subsequent ('5% of the 1.5%') household re-visits

1. Post-distribution monitoring (PDM)

1) Village and household selection for Main (‘1.5%’) household visits

For a given administrative area, typically a sub-district, within which typically 1.5% of the households will be visited, data are selected as follows:

1. Calculate the total number of households to be surveyed

1.5% of the total number of households in the sub-district.

2. Define with our partners ‘the number of households to be visited per day’ (N)

Rationale:
1. Cost-effectiveness. PDM activity is most cost-effective when all the work in a village is competed in one day. This avoids a second, costly visit including travel time.

The number of households that can be visited in a day varies from country to country, and from region to region within a country. Typically it is between 15 and 30 and is commonly 25 households.

3. Set a threshold size for a village as N + 30%

Rationale:
1. A quantity of pre-selected ‘spare’ households is required so that data collectors are not left to decide themselves which household to visit to achieve the target number of households if a selected household is not accessible.
2. From field experience, 30% is enough to cover the number of non-accessible households in most villages.

This typically gives a threshold of 33 households (assuming N=25).

Some households may not be accessible in a village due to absence, permission not granted or a household having moved to another village

4. Remove all villages with fewer than this threshold number of households

Rationale:
1. Cost-effectiveness, as above, to exclude villages where less than a day would be required in each of these villages to survey the selected households, increasing all of: the number of villages to be visited, travel time and therefore the overall time taken for the data collection and, importantly, the cost of achieving the 1.5% of sub-district households target.
2. Limited impact on results if these villages are excluded. On average, villages with fewer than the threshold number of households is around 3%.

5. Create a randomly ordered list of villages in the sub-district

Use registration data to create a list of all remaining villages in the sub-district and the number of households per village.

Order these villages based on a randomly assigned Globally Unique Identifier (GUID).

Note: We use registration data rather than distribution data to ensure that we reach households that were registered but may not have received nets.

6. Calculate the number of households to visit in each village

This is done in two stages:

1. Calculate 1.5% of the households in the village, rounded up to the nearest integer.

2. Allocate to each village 25, or a multiple of 25 (i.e. 'the number of households to be visited per day' N, as defined in 1, or a multiple thereof), representing the number of households to visit in that village if selected. The number of households is determined as follows:

Where 1.5% of # HHs is fewer than 25, 25 HHs are selected
Where 1.5% of # HHs is between 25 and 50, 50 HHs are selected
Where 1.5% of # HHs is between 50 and 75, 75 HHs are selected, etc

Note: Given the typical size of villages and a 1.5% sample size, it is uncommon that more than the minimum number of households (N, i.e. typically 25) is chosen. When N = 25, only 0.2% of all villages in our database are large enough that we would select 50 or more HHs.

7. Select a subset of villages to be visited, such that 1.5% of households in the area will be surveyed, with minimal overshoot

Select villages in order from the top of the list such that the sum of the households in these villages is less than the total number required, but that adding the households from the next village in the list would exceed the total. Then add the next village, including only the minimum number of households (as a multiple of 25), to bring the cumulative number of households to be selected just above, but as close as possible to, the target total to be visited (i.e. 1.5% of total number of households in the sub-district).

Rationale:
This approach is to guard against over-shooting the number of households required.

For example, assume 105 HHs were needed in an area and already had 4 villages x 25 HHs selected. If the fifth village in the randomly generated list was particularly large, such that the number of HHs to be selected (as calculated in Step 5) was 75, the addition of this village to the list would unnecessarily increase the cost of the PDM (i.e.

175 HHs rather than 105 HHs). As a result, our selection code would only select 25 HHs from the fifth village, to minimise the overshoot.

8. Randomly select main and spare households for each selected village

For each village selected in Step 6, randomly order the households (again using GUIDs).

Select the required number of main households (as calculated in Step 5). Then select the required number of spare households (as defined in Step 2, typically 30%), continuing in order down the randomly ordered list.

2) Household selection for Subsequent ('5% of 1.5%') household re-visits

To encourage accurate data collection, the 'main' data collectors that visit households for the 1.5% data collection, are informed during training that their work will be checked. They are told it will be via: 1) supervisor presence; and 2) a second set of data collectors visiting a random selection of 5% of the households the main data collectors have visited, with the second wave of data collectors not having any knowledge of the data collected during the main visits.

We consider that the majority of the impact of this approach in achieving the desired goal, encouraging accurate data collection, is achieved through letting the data collectors know that this checking will take place i.e. through the psychology of data collectors being aware their work will be checked. That the data is collected and data sets compared largely serves to deliver data to AMF to allow accuracy to be assessed.

1. Determine the proportion of households to be re-visited

This is typically selected as 5%.

Rationale:
This is a significant portion of the main ('1.5%') household visited and balances data accuracy and expense.

2. Calculate the total number of households to be re-visited

5% of the number of main ('1.5%') households to be visited in the sub-district.

3. Define the number of households to be visited per village

We assess a sensible number of HHs per village as 12.

Rationale:
Cost-effectiveness. PDM re-visit activity is cost-effective when we avoid visiting villages to survey a small number of households. Otherwise, a large number of villages would need to be visited, with associated travel time, leading to an unnecessarily expensive re-visit programme.

4. Calculate the total number of villages to be re-visited

Total number of re-visit households (from Step 2) divided by the households per village (step 3), rounded up to the nearest integer.

5. Randomly order the sub-districts and the randomly order the PDM villages within each of them

Using GUIDs, randomly order the sub-districts in the district. Randomly order the villages within each sub-district.

6. Select enough villages in each sub-district to exceed the total number of households required

Run through the list of sub-districts, selecting one village from each. If this exceeds the number of households required (from Step 2) i.e. 12 x the number of villages, stop. If not, select another village from every sub-district.

Randomly order all villages selected, using GUIDs. Calculate the number of villages selected that are excess to requirements (as calculated in Step 4). Select this many from the top of the list and delete.

Note: The number of villages is always comfortably larger than the number of sub-districts.

7. Randomly select re-visit households for each selected village

For each village selected in Step 6, randomly order the households (again using GUIDs).

Select the required number of re-visit households, (typically) 12.