

# RMS Burkina Faso

## Sampling Methodology

### 1.2. General and Specific Objectives

#### 1.2.1. General Objective

The overall objective is to normatively integrate RMS surveys into the landscape of UNHCR surveys, in order to provide estimates on Results-Based Monitoring (RBM) indicators at the level of UNHCR target groups, including internally displaced persons (IDPs) and refugees and asylum seekers (RAS), in Burkina Faso.

#### 1.2.2. Specific Objectives

- Provide UNHCR with reliable databases on IDPs and RAS.
- Provide estimates of impact indicators for UNHCR's results monitoring.
- Ensure UNHCR's results monitoring within the framework of national multi-year strategies.
- Integrate RMS surveys into UNHCR's normative data collection system.

### 1.3. Expected Results

#### 1.3.1. Expected Results from CESAO

As the local organization responsible for collecting data on IDPs and RAS, CESAO is expected to achieve the following results:

- Data collection is completed.
- High-quality data on IDPs and RAS are available.

To achieve these results, CESAO will carry out the following tasks:

- Make every effort to obtain all necessary administrative and technical authorizations for conducting the survey.
- Translate the questionnaires from English to French by specialists in the field.
- Translate key or complex expressions from French to local languages in the survey areas, such as Fulfuldé, Arabic, Tamashek, Gulmacéma, and Moré.
- Recruit survey personnel (consultants, supervisors, enumerators).
- Conduct a pre-test of the questionnaire with supervisors using 10 to 20 respondents before training the enumerators.
- Train the enumerators.
- Carry out an information and awareness campaign among local authorities, IDPs, and RAS to gain their support for the survey.
- Manage security issues.
- Manage COVID-19-related matters.

### **1.3.2. Expected Results from TANGO**

As the contracting company with UNHCR, the expected results from TANGO International are:

- • Provide guidance to the local company (CESAO-AI) at each stage of the RMS.
- • Conduct train-the-trainer sessions on all survey components and the working protocol with RAS and IDPs.
- • Conduct remote data collection monitoring to ensure data quality.
- • Test and provide proposals for improving the data collection tools used in RMS implementation.
- • Test and provide proposals for improving the RMS methodology.

### **1.4. Sampling Plan**

The methodological approach used in this survey is structured around the following points:

- Sampling frame
- Choice of sampling method
- Calculation of sample size
- Stratification of the sample
- Sample selection

#### **1.4.1. Sampling Frame**

The sampling frame used for sampling is derived from two sources:

- The database on internally displaced persons (IDPs) from the National Emergency Relief and Rehabilitation Committee (CONASUR) within the Ministry responsible for humanitarian action.
- The database on refugees and asylum seekers (RAS) from the National Committee for Refugees (CONAREF) within the Ministry of Foreign Affairs.

These two databases provide lists of household heads of IDPs or RAS, distributed by region, province, municipality, and site (village, sector, neighborhood, etc.).

#### **1.4.2. Choice of Sampling Method**

The survey will be conducted among the population of households living in the target areas of UNHCR. This population is divided into two homogeneous sub-populations (strata):

- The sub-population of internally displaced persons (IDPs)
- The sub-population of refugees and asylum seekers (RAS)

The sampling strategy used in the pilot survey is two-stage stratified cluster random sampling (sectors/villages, neighborhoods, etc.). At the first stage, within each stratum, clusters are randomly

selected in each target municipality using probability proportional to size (PPS). At the second stage, a predetermined number of households will be selected from each cluster through systematic random sampling.

#### 1.4.3. Calculation of Sample Size

The calculation of the sample size for the pilot survey is based on the impact indicators:

- Indicator 2.2: Proportion of PoCs residing in physically safe and secure camps with access to basic facilities.
- Indicator 2.3: Proportion of PoCs with access to health services.
- Indicator 3.3: Proportion of PoCs feeling safe when walking alone in their neighborhood.

The formula for calculating the sample size to detect changes in proportion variables is as follows:

$$n = D [(Z_{\alpha} + Z_{\beta})^2 * (P_1 (1 - P_1) + P_2 (1 - P_2)) / (P_2 - P_1)^2]$$

Where:

n = Minimum sample size required per survey round or comparison group.

D = Cluster effect (design effect).

P1 = Estimated level of an indicator measured as a proportion at the time of the first survey (or for the control area).

P2 = Expected level of the indicator either at a future date or for the project area, so that the quantity (P2 - P1) represents the magnitude of change to be detected.

Z<sub>α</sub> = The corresponding Z-score for the desired level of confidence to conclude that an observed change of size (P2 - P1) would not have occurred by chance (α - statistical significance level).

Z<sub>β</sub> = The corresponding Z-score for the desired level of confidence to accurately detect a change of size (P2 - P1) if such a change occurred (β - statistical power). Z<sub>α</sub> and Z<sub>β</sub> have "standard" values based on the desired reliability. They are provided in Figure 3-2. Note that higher percentages provide more certainty in measuring accurate results.

The following parameter values were applied in the calculations:

- Cluster effect D is 2.
- Statistical significance level = 95%, so Z<sub>α</sub> = 1.645.
- Statistical power β = 80%, so Z<sub>β</sub> = 0.840.
- Expected change (P2 - P1) is 10 percentage points between baseline and possible endline.
- Non-response factor of 9 percentage points to account for the estimated household non-response rate.

Since there are no current estimates for the baseline level of UNHCR's impact indicators, the default value of P1 was set at 50%. Given the expected change (P2 - P1) of 10 percentage points, the corresponding value of P2 is 60%.

By applying the above formula and incorporating the values of the different parameters, the minimum sample size  $n$  is calculated as 605. Considering the specific characteristics of UNHCR's target groups, a non-response rate of 9% (informed consent) is expected. Therefore, the sample size increases to 666 households for each of the two strata considered as independent homogeneous entities.

Furthermore, since data collection will be done through direct interviews, either in-person (face-to-face) or by phone, it is possible that households may be untraceable or inaccessible for various reasons. To account for this, reserve households are planned at a rate of 36% of the parent sample size, totalling 899 households, rounded up to 900 households to be drawn in each of the two strata, as follows:

- Parent Sample = 660 households
- Reserve Sample = 240 households

In each cluster, 30 households will be selected, resulting in 30 clusters per stratum and 60 clusters in total, corresponding to the 1800 households of the two strata. Table 1 below provides the distribution of clusters and households per stratum (IDP and RAS sub-populations).

Type d'échantillon	Number of clusters			Number of Households par grappe	Total Number of Households		
	<i>PDI</i>	<i>RAS</i>	<i>TOTAL</i>		<i>PDI</i>	<i>RAS</i>	<i>TOTAL</i>
Parent Sample	22	22	44	30	660	660	1320
Reserve Sample	8	8	16	30	240	240	480
<b>Total</b>	<b>30</b>	<b>30</b>	<b>60</b>	<b>30</b>	<b>900</b>	<b>900</b>	<b>1800</b>

Table 1: Number of Clusters and households in the pilot survey sample

#### 1.4.4. Sample Selection

##### 1.4.4.1. Selection of Primary Units (Clusters)

In each of the two strata, sample selection will be carried out in two stages: 1) selection of clusters, and 2) selection of households. Clusters are hosting sites that can be villages, sectors, neighborhoods, etc., which are the smallest geographical units available in the sampling frames of each sub-population.

At the first stage, clusters are randomly selected within each sampling frame of the respective populations using probability proportional to size (PPS). To account for potential inaccessibility of certain clusters, a "randomly generated reserve sample" will be drawn using a two-phase approach. This involves selecting the required 22 clusters for each population (RAS and IDP), totalling 44 clusters for the overall sample. Additionally, 8 reserve clusters (36% of the required number) will be selected for each stratum (RAS and IDP), totalling 16 reserve clusters, in case any clusters need to be replaced, resulting in a total of 30 selected clusters per stratum (22 + 8).

The sampling frame will include all clusters (villages/sectors, neighborhoods, etc.) in the HCR intervention areas, with a population greater than or equal to 5% of the total population of interest.

#### **1.4.4.1. Selection of Primary Units (Households)**

At the second stage, 30 households will be selected through simple random sampling, specifically systematic sampling, from the complete nominal list of household heads in the sampling frame of CONASUR for IDPs and CONAREF for RAS. Thus, the sampling of the parent sample households will be done within each cluster of the parent sample clusters, and the sampling of reserve sample households will be done within each cluster of the reserve sample clusters. In total, 900 households will be selected in each stratum, including 660 households from the parent sample and 240 households from the reserve sample.