



Forced  
Displacement  
Survey

**SAMPLING  
NOTE**

**V1**



Forced Displacement Survey  
**SOUTH SUDAN**

2023

# Contents

<b>Target populations / sampling universe</b>	<b>3</b>
Refugees . . . . .	3
Asylum seekers . . . . .	5
Host community . . . . .	5
Refugee returnees . . . . .	5
Internally Displaced Persons . . . . .	6
Sample representativity . . . . .	6
<b>Disaggregations and statistical power</b>	<b>7</b>
<b>Sampling approaches and frames</b>	<b>9</b>
Refugees . . . . .	9
Refugees in camps and settlements in RAA and Upper Nile . . . . .	9
Refugees in settlements and camp in Western and Central Equatoria and Jonglei . . . . .	10
Hosts . . . . .	10
Returnees . . . . .	10
<b>Sample size</b>	<b>12</b>
<b>Replacement protocol</b>	<b>14</b>
<b>Sample adjustment</b>	<b>15</b>
Survey setup . . . . .	15
Process of weight estimation . . . . .	16
Use of weights in analysis . . . . .	17
Variance estimation and analysis . . . . .	18

# Target populations / sampling universe

The Forced Displacement Survey Programme, globally speaking, aims to collect data on a nationally representative sample of the refugee and asylum seeker population as well as a sample of the host population, i.e. nationals who live in close proximity to the refugee population and whose lives are affected by their presence. Refugees, asylum seekers and host communities thereby constitute the primary populations of interest to the survey programme.

In the case of South Sudan, many South Sudanese are themselves refugees in neighbouring countries. Since the signature of the Comprehensive Peace Agreement (CPA) in October 2018, many of them decided to return to South Sudan spontaneously, and currently these refugee returnees outnumber the refugee population. For this reason, they have been identified as an important emerging population of concern in South Sudan about which little data is available, and they have been included as an additional target population in the FDS. However, as the exact numbers, characteristics and locations of return are not known, it was decided not to aim for full representativeness at the national level, and to opt for an experimental sample focusing on areas of high returns.

Below is a brief description of the two main target populations and one experimental for the survey in South Sudan: refugees, host community and returnees.

## Refugees

The definition of the refugee population is based on UNHCR's determination of their status. Hence all of the individuals registered by either UNHCR or host governments that have undergone a successful process of status determination are considered eligible for the purpose of the Forced Displacement survey. This definition aligns with UNHCR official statistics statistical based on vital statistics emanating from registration such as the Global Trends Report and Refugee Data Finder. This definition will also provide a reliable base for comparison of the Forced Displacement data across contexts.

It needs to be noted that in reality the UNHCR status determination activities may not capture all of the refugees in a given host country due to various reasons. The Forced Displacement survey will continue exploring capturing these groups as experimental samples in order to provide a better understanding of the total potential refugee population in the host country and with this drive narrative of possible under-coverage of the refugee population in UNHCR registration systems as well as provide evidence on how to improve registration through operational means.

FDS also collects household level information for the refugees. It is therefore important to also define what constitutes a refugee household for the purpose of data collection as well as analysis. To-date there is no fully agreed upon definition of the refugee household. The definition of the refugee household for the FDS is one where the head of household or his/her spouse is a refugee.

The main sampling frame for the refugee population is thus the UNHCR registration data system proGres.

The main sample for the refugees in South Sudan is spread across the following strata geographically:

- Rweng Administrative Area,
- Upper Nile state and
- Western and Central Equatoria and Jonglei.

These populations are divided between formal camps and sites or settlements. There are formal camps in the Rweng Administrative Area and Upper Nile state as well as Gorom camp in Central Equatoria, while the other formal settings are all considered settlements. Settlements are formally managed and supported by UNHCR and are typically associated with a host community village, adjacent or attached to the village, with free movement of people and goods between the village and settlement. The settlements are not always clearly distinct from the villages. Refugees living in settlements are very mobile often spill into nearby communities.

UNHCR estimates that South Sudan hosts around 335,000 refugees as of March 2023. Sudanese refugees are mainly located in the north (Rweng administrative area and Upper Nile state) with some small pockets in the south. Congolese and Central African refugees are primarily located in the south (Central and Western Equatoria). Ethiopian refugees are primarily located in Jonglei (Pochalla) and Central Equatoria (Gorom). 80% reside in camps, 18% in settlements, and 2% in urban areas - especially in the capital city Juba.

This table summarizes the refugees by region, site, and nationality. The survey (sampling universe) is

**Table 1.** Refugee population in South Sudan

Region	County	Type	Site/Department	Primary Nationality	Count	Percent
Rweng Administrative Area	Pariang	Settlement	Yida	Sudan	33,847	10.2%
Rweng Administrative Area	Pariang	Camp	Ajuong Thok	Sudan	46,167	13.9%
Rweng Administrative Area	Pariang	Camp	Pamir	Sudan	45,507	13.7%
Sub-total (Rweng AA)						
Upper Nile	Maban	Camp	Doro	Sudan	73,071	21.9%
Upper Nile	Maban	Camp	Batil	Sudan	53,583	16.1%
Upper Nile	Maban	Camp	Kaya	Sudan	27,327	8.2%
Upper Nile	Maban	Camp	Gendrassa	Sudan	20,472	6.1%
Upper Nile	Fashoda	Settlement	Kodok	Sudan	327	0.1%
Sub-total (Upper Nile)						
Jonglei	Pochalla	Settlement	Pochalla	Ethiopia	2,214	0.7%
Western Equatoria	Yambio	Settlement	Makpandu	DRC	7,940	2.4%
Western Equatoria	Ezo	Settlement	Ezo Centre	DRC	2,276	0.7%
Western Equatoria	Ezo	Settlement	Tambura	DRC	631	0.2%
Western Equatoria	Tambura	Settlement	Source Yubu	DRC	961	0.3%
Western Equatoria	Tambura	Settlement	Andari	DRC	297	0.1%
Western Equatoria	Tambura	Settlement	Naandi	DRC	272	0.1%
Western Equatoria	Maridi	Urban	Maridi	DRC	118	0.0%
Central Equatoria	Yei	Settlement	Lasu	DRC	8,638	2.6%
Central Equatoria	Yei	Urban	Yei	DRC/Sudan	1,775	0.5%
Central Equatoria	Juba	Urban	Juba	Sudan	4,440	1.3%
Central Equatoria	Rejaf	Camp	Gorom	Ethiopia	2,186	0.7%
Central Equatoria	Morobo	Settlement	Morobo (Panyume)	DRC	925	0.3%
Sub-total (W & C Equatoria, Jonglei))						
Total						

representative of the entire registered refugee population in the country.

## Asylum seekers

An asylum seeker is a person who is seeking international protection. In some countries, it is used as a legal term referring to a person who has applied for refugee status or a complementary international protection status and has not yet received a final decision on their claim. It can also refer to a person who has not yet submitted an application but may intend to do so, or may be in need of international protection. Not every asylum-seeker will ultimately be recognized as a refugee, but every refugee is initially an asylum seeker. However, an asylum-seeker may not be sent back to their country of origin until their asylum claim has been examined in a fair procedure, and is entitled to certain minimum standards of treatment pending determination of their status.”<sup>1</sup>

UNHCR estimates that South Sudan hosts a total of 4,240 asylum seekers, mainly located in Juba (4,059). They are primarily from Burundi, Eritrea, and Ethiopia. The universe definition does not consider prospective asylum seekers, or those who have the intention to apply but have not done it yet.

The asylum seeker population is excluded from the final sampling design due to operational and situational constraints.

## Host community

Definition of host population is very vague without commonly standardised and agreed-upon formulation. FDS globally and in South Sudan particularly aims to contribute to the discourse on the operational definition of the host communities that can be used both in the survey as well as operational context. FDS aims in capturing the living conditions of South Sudan nationals both compared with refugees, and as a consequence of living in proximity to the refugees, especially in proximity to camps and settlements.

In accordance with operational priorities, and to accommodate budget constraints, the survey is representative of host communities living in close proximity to refugees in Rweng Administrative Area and Upper Nile state only, where the majority of refugees live. The focus of the host population is further limited to counties hosting to camps and settlements.

## Refugee returnees

As of March 2023, around 528,093 refugees have returned to SSD. The returns are spontaneous. It is estimated that about 58% returned from Uganda, 25% from Sudan, 11% from Ethiopia, 4% from DRC, and the rest from other countries (including Kenya and CAR).

A large part of these returnees arrived relatively recently, with around 270,000 individuals returning over the course of 2021.

Most of the verified returnees who came from Uganda went to Rweng administrative area and Jonglei state, and most of those who came from Sudan went to Rweng administrative area and Upper Nile state.

---

<sup>1</sup>UNHCR. UNHCR master glossary of terms.<https://www.unhcr.org/glossary>. Accessed on January 15<sup>th</sup>, 2024

**Table 2.** Number and distribution of returnees across the states

State	N of returnees	%
Central Equatoria	218,275	41%
Eastern Equatoria	111,256	21%
Western Equatoria	11,292	2%
Rweng Administrative Area	82,483	16%
Upper Nile	75,760	14%
Jonglei	29,027	6%
TOTAL	528,093	100%

## Internally Displaced Persons

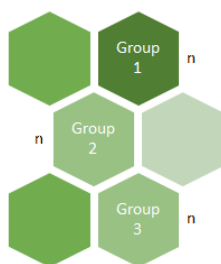
Finally, a note on Internally Displaced Persons (IDPs). Although this population is numerically and programmatically very important in South Sudan, it did not constitute a target population for the FDS survey in 2023. However, the questionnaire includes one module on identification of forced displaced populations (FDPs) designed based on the IRIS and IRRS recommendations drafted by EGRISS. The set of questions will allow the identification of IDPs, therefore their presence in host community or refugee household may be identified. However, this population group does not constitute a sample per se.

The UNHCR Results Monitoring Survey (RMS) series, implemented in South Sudan in 2022, covers the IDP population.

## Sample representativity

The notion of sample representativity is very often associated directly with the sample size and the notion of statistical power or robustness, reliability of statistical estimates – i.e. how confident are we that the number derived from the data actually represents the reality in the population. Sample representativity is actually not the result of increased sample size but is rather a reflection of the structure of the sample and how similar (or different) is the sample from the actual population in terms of key socio-economic and geographical strata (groups).

# Disaggregations and statistical power

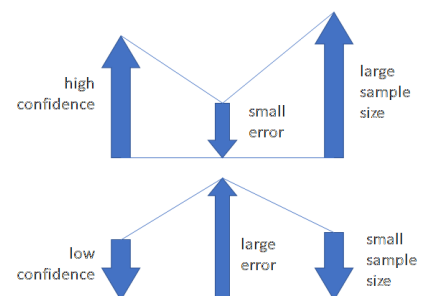


sample size = 3 \* group sample size

The preceding section described the sampling universe of the survey, i.e. the target populations of which the survey will be statistically representative. A further essential sampling parameter to define is disaggregations (or analytical domains) and with it the statistical power with which results will be presented for each disaggregation. This defines the sub-groups within a given target population for which analytical results can be derived in a statistically robust way (i.e. with confidence that the results obtained from the data reflect the reality in the population). For example, typical disaggregations of the findings on a country's refugee population as a whole could be by province or by in-camp versus out-of-camp. See Annex A for an illustrative example. While greater granularity is of course always desirable, it must be noted that additional

disaggregations will inevitably drive-up sample sizes considerably – and hence need to be selected cautiously.

Sample size requirement is driven by the required statistical precision to deliver the analytical results in a statistically robust manner –i.e. with high confidence. The higher the confidence in results is required, the larger sample size needs to be. Sample size calculations are discussed in more detail below. In order to have high confidence in the results produced at a certain disaggregation level, the sample size of that disaggregation needs to meet the required minimum. The size of these analytical groups or disaggregations are referred to in this text as MAES or Minimum Acceptable Effective Size.



Administrative units in South Sudan:

- State and administrative areas
- County
- Payam
- Boma

This Forced Displacement survey is designed and powered to produce representative estimates of the following sub-groups:

- Refugees

- By location

- Ruweng Administrative Area
- Upper Nile
- Rural Jonglei, Western Equatoria and Central Equatoria combined (excluding urban Juba)

- Hosts

- By County

- Pariang county in Ruweng Administrative Area
  - Mabane county in Upper Nile state

- By proximity

- Those living close-by
  - Those living further away (but within the same county, as per sampling universe specified above)

- Returnees

- Magwi County in Eastern Equatoria: No disaggregation – one MAES only

Where more than one first-level disaggregation is listed per target population, these constitute parallel and not cumulative disaggregations.



# Sampling approaches and frames

## Refugees

There are four key locations for sampling refugees:

- refugee camps and settlement in RAA and Upper Nile,
- refugee settlements and camp in Western & Central Equatoria and Jonglei
- rural at-large living in townships
- and urban areas.

Given the differences among these three locations, the sampling frames and approaches used to select and reach each sub-population vary.

## Refugees in camps and settlements in RAA and Upper Nile

For the camps and settlements in RAA and Upper Nile, single stage systematic random stratified sampling across all locations is used and using Google Building footprint data as the sampling frame.

Google Building Footprints is a database of geocoded building data, that is acquired by processing high resolution aerial photography. For the purpose of the FDS and this particular stratum, any buildings allocated within the estimated boundaries of the quarters of townships, where refugees live, has been considered to belong to a refugee household. These buildings therefore constitute a sampling frame for this stratum. The samples was drawn with stratified systematic random sampling method. The main stratification criterium were the 100 by 100 m grid on which the camps are placed, ensuring an even spread across the area each camp occupies.

For Rweng administrative area the initial sampling strategy assumed proportional allocation of the sampled units across the two camps (Ajuong Thok and Pamir) and the Yida settlement. As Yida settlement is no longer directly supported by UNHCR, the sample size within this stratum was boosted to allow the minimum required sample size for the specified precision be allocated to the two supported camps. In Upper Nile the initial allocation of the sampled units was proportional across the four camps. Due to the operational requirements to better understand intricacies at the camp level, the sample size allocated to bigger camps was increased to allow for more analytical power at some of the individual camps. In terms of sampling methodology, the separate camps were treated as explicit strata within the broader Upper Nile refugee stratum.

## Refugees in settlements and camp in Western and Central Equatoria and Jonglei

For refugees living in Jonglei, Western Equatoria and Central Equatoria, a one stage simple systematic random sampling from a nominative list based on ProGres registration system is used.

Sampling frame for the stratum of refugees in in Jonglei, Western Equatoria and Central Equatoria was an extract of all refugees (individuals) living in the designated geographic area. The records represent registered individuals who are organised into registration cases as identified by the combination of individual and case IDs. Registration cases approximately resemble a family but are not entirely equivalent to the definition of household as an economic unit used in the household surveys. Selection of the primary sampling units is determined on the basis of the list of all the individuals. Determination of the PSU as the sampling unit is based on its geographic and administrative classification. Admin levels 0 to 2 are standardised and use nationally and internationally compatible classification. Any details at admin level 3 or below are recorded in an open text field stored under admin 6 field within the registration database. Admin 6 field would hold the lowest possible details down to the street address.

## Hosts

The sampling of the representatives of the host communities is organised into two distinct strata:

- ◆ Host communities in close proximity to camps and settlement in Rweng Administrative Area
- ◆ Host communities in close proximity to camps in Upper Nile state

Sampling of the host population is based on the Google Building Footprints database. The sampling frame includes all the building objects from the database within the borders of counties (Pariang and Mabane) in which the camps and settlement are located and outside of the boundaries of these settlements or camps. Sample of hosts is selected with systematic random sampling method using probabilities proportional to proximity of the respective buildings to the closest settlement/camp boundary. Proximity (distance) estimation was modified (with power transformation) in order to achieve approximate allocation of 50% of the sample within 10 km distance and 50% from over 10 km distance to the outer administrative boundary of the county.

## Returnees

Sampling of returnees in South Sudan is geographically limited to those returnees residing in Magwi county. The sample is thus NOT nationally representative of all the returnees living in South Sudan. The sample, however, is representative of the returnees who settled in Magwi county in the state of Eastern Equatoria. Magwi county was selected based on operational priorities as one of the identified “Pockets of hope”. Sampling approach for the returnees is essentially a two-stage sampling approach where the first stage sample of clusters is based on the Google building footprint database and the second stage is based on a generated frame based on household listing carried out in these clusters.

A database with records of returning refugees does not exist. UNCHR collects data on this population group through key informants and maintains updated information on their total numbers by payam of return, which informed the stage one selection of clusters designated to be listed. Selection was systematic random and stratified by payam with allocation proportional to the payam’s known returnee population.

Therefore, a listing exercise was done to create a sampling frame for this population group, using the building footprint database (Google Open Building (v2)).

500 building location seed samples were drawn in admin 3 areas (payams) within Magwi County that host returnee refugees. Within admin 3 areas (payams), seed building location samples were drawn randomly such that circular buffers with radii of distance to the 8<sup>th</sup> nearest building location to seed building locations are non-overlapping, and so that the number of seed building location samples per area is proportional to the number of returnees reported in the UNHCR data.

The buildings identified with the above-described method were visited and a short questionnaire was administered to verify the presence of returnees refugees, who constitute the frame used in the sampling stage.

This Listing survey constitutes the sampling frame for returnees' refugees.

In the second stage of selection only one eligible household was selected for interview from each cluster.

## Sample size

Overall approach to sample size and composition is based on the analytical requirements and subsequently identified essential analytical domains and disaggregations as described earlier in this document. This leads to optimal sample allocation which is based on balanced allocation of sample units per identified domain which results in an identified explicit stratum. The required sample size is therefore estimated at the level of each disaggregation. For this purpose, the Minimum Acceptable Effective Size (MAES) is used. MAES identifies the minimum number of households required for a desired level of precision in each disaggregation.

The proposed sample size per MAES were be calculated using the following formula:

$$n = \frac{z^2 \times p(1 - p)}{\alpha^2} \times def f$$

where

$n$  = net sample size

$z$  = z-score

$p$  = proportion of population with a given trait

$\alpha$  = margin of error

$def f$  = design effect

The value of the design effect is a direct function of the proposed size of the survey cluster ( $m$ ) – i.e. the number of households sampled in each PSU in the case of multistage sampling method.

$$def f = 1 + (m - 1) \times ICC$$

Effective sample size per MAES depends on the sampling method proposed for a particular stratum. Proposed sampling method for most of the explicit strata for the FDS in South Sudan used single stage sampling and thus would require smaller size per MAES of  $n = 500$ .

The table below summarizes a suggested allocation as expressed in MAES.

**Table 3.** Sample allocations per MAES (explicit stratum)

	MAES	Clustered	Frame & Approach	listing	n
Refugees					
Rweng Administrative Area	1				600
Ajuong Thok		No	Building footprints	No	250
Pamir		No	Building footprints	No	250
Yida		No	Building footprints	No	100
Upper Nile	2				1,000
Doro		No	Building footprints	No	370
Batil		No	Building footprints	No	370
Kaya		No	Building footprints	No	204
Gendrassa		No	Building footprints	No	56
West & Central Equatoria and Jonglei	1				500
all areas		No	proGres	No	500
Hosts					1,000
Rweng Administrative Area	1	No	Building footprints	No	500
Upper Nile	1	No	Building footprints	No	500
Total	6				3,000
Returnee experimental sample					
Magwi county	1	Yes	Building footprints & Listing	Yes	500
Total experimental sample	1				500

# Replacement protocol

In order to address any nonresponse or ineligibility of sampled households a replacement protocol has been put in place. Replacement has been chosen as the approach to address the problems of nonresponse, particularly non-contact due to poorer quality of contact information in the sampling frames. Oversampling is often preferred as a better and methodologically more robust method of addressing nonresponse. However, due to largely unknown rates of noncontact in the target populations and to avoid too low statistical power due to higher than expected nonresponse, as well as to avoid excessive costs of data collection due too high oversampling rates, the replacement protocol was chosen over the oversampling method as the most appropriate operational solution.

Replacement samples were drawn together with the main sample and selection into a replacement sample was randomised. Issuing of replacement has been further randomised at the level of the explicit sampling stratum definition and was strictly controlled – i.e. issuing of replacement households was handled by the central survey coordination team from the implementing agency. Use of households from the replacement pool was strictly documented. Using of the replacement methodology enabled the FDS field teams to ensure that the target sample sizes are achieved while managing costs of data collection.

The replacement sample is included together with the main sample and encoded into the Kobo Questionnaire form. Any household belonging to the replacement sample are protected with a validation code. If an interviewer selects a replacement sample he is prompted for a validation code, which is matched to the encoded value. Only if the interviewer is provided with the validation code, the questionnaire can advance to the interview. This setup supports flexibility as both the main and replacement samples are readily available at interview time, while maintaining strictly controlled issuing of the replacements – issuing of validation codes is under the control of the survey coordination team.

# Sample adjustment

As part of the data processing tasks, the sample was adjusted using weighting procedures. All the weights used in FDS data are analytical weights – i.e. total sum of weights is equal to the sample size. The weights used for analysis are composite weights comprising of basic sampling weights as well as structural adjustment weights. Sampling weights correct for unequal probabilities of selection across different strata, while structural adjustment weights adjust to basic population structures such as geographic distribution. Structural weights are also used to adjust the sample of balanced size sample strata to population proportions in order to derive national estimates.

FDS data analysis does not use population weights – i.e. weights that sum up to the population totals and thus enable the analyst to estimate true population numbers for indicators and across population groups. FDS is not meant to be the source of population data and as such does not provide weight for such estimations.

## Survey setup

FDS survey features up to 4 distinct interviews and aims to represent their respective populations:

- Refugee households,
- Adults aged 15 and over,
- Children under the age of 5 and
- Women who gave birth within 2 years prior to the interview.

The microdata features weights to correctly represent these populations.

The sampling design identified 5 distinctive sampling strata:

- Refugees in Rweng Administrative Area
- Refugees in Upper Nile state
- Refugees in Central Equatoria, West Equatoria and Jonglei
- Hosts in Pariang county of Rweng Administrative Area
- Hosts in Mabane county of Upper Nile state

The main sampling principle assumed equal sample size for each identified stratum. Due to late changes in the survey design oversampling of refugees in Pariang and Mabane county was introduced in order to

achieve more analytical power at the levels below the identified strata.

## Process of weight estimation

As mentioned, final analytical weights in FDS are composite weights composed of base weights and structural adjustment weights. The weight estimation closely follows the sampling methods used in the selection process as well as adjustments of protocols, if any, as implemented in the field. In the strata where single stage systematic random selection is used there are four basic steps that are followed:

1. In the first step the base probabilities of selection are estimated separately for each explicit stratum (sample allocation stratum - 3). Selection probability is calculated as follows:

$$p_i = \frac{n_s}{N_s}$$

Where  $n_s$  is the total number of sample units drawn into both the main and replacement samples and  $N_s$  is the total number of units in the frame. The base weight is the inverse of the selection probability

$$w_i = \frac{1}{p_i}$$

2. In the second step the initial base weights are rescaled to the realised sample size of respondents ( $r_s$ ).

$$w_i^* = \frac{w_i}{\sum w_i} \times r_s$$

3. In the third step structural adjustments are made. As the full information matrices on geographical distributions of the sampled population exist, poststratification is used to adjust the weights to the correct population proportions.

Due to in-field adjustments to issuing of replacements, the weight estimation process assumes that the sampling has been carried out at sub-stratum level. The weight estimation reflects that. The sub-strata units remain represented proportionally.

## Google footprint samples of hosts based on proximity

The samples of hosts in proximity to the camps in Rweng administrative area and Upper Nile state, are approximately self-weighting according to the selection gravity coefficient. Selection gravity coefficient is power adjusted (6<sup>th</sup> root) simple Euclidean distance from the building to the closest camp boundary. As the selection was carried out over two explicit strata (0 – under 10 km & 10km to county boundary), there may be a need to structurally adjust the sample according to unrestricted (not split over two strata) gravity coefficient. Self-weighted sample represents the population that lives in proximity of the refugees (selection gravity coefficient based on proximity favours those living closer in terms of selection).

$$\begin{aligned} dist_{ijk} &= |b_{ij} - P_{ik}| \\ dist_{jk}^* &= \sqrt[6]{dist_{jk}} \\ prox_{ijk}^* &= round(1,000 \times \frac{\max_{1 \leq j_k \leq J_k} (dist_{ijk}^* - dist_{ijk}^*)}{\max_{1 \leq j_k \leq J_k} dist_{ijk}^*}) \\ p &= \frac{1}{prox_{ijk}^*} \end{aligned}$$



## Number of buildings adjustment for building footprint samples

Weights for building footprint samples of hosts and refugees in camps are further adjusted for the number of buildings the household owns (lives in). Adjustment coefficient is one (1) over the number of buildings owned. The building ownership is self-reported as part of the interview process. Households who own multiple building have higher probability of selection, which needs to be adjusted as part of the unequal probability adjustment weight estimation.

Additional weights are estimated for the analysis of individual datasets (random adult, random child under 5 and random woman who gave birth in the last 2 years). Individual weights are composite weights of household selection weights with within household selection weights.

## Nonresponse adjustment

Due to adverse conditions on the ground in West & Central Equatoria and Jonglei and challenges in locating the refugee populations, the response rates vary greatly across the sampling strata. As the great majority of the population is located in the camps in northern counties, where the response rates were very high, the nonresponse adjustment component of the composite analytical weights was not deemed necessary.

Based on the available information on the final dispositions of the sampled units and structure of the target population based on registration data (proGres), nonresponse adjustment models if needed can use the socio-demographic characteristics of respondents and nonrespondents available in drive to predict response. This includes not only demographic features of the case, like household size, family composition, age of the head, education, etc, but also a selection of economic characteristics such as employment.

## Structural adjustment

For structural adjustment, the source of the target population distribution is the registration database proGres. It needs to be noted that the population distribution based on the registration data is not deemed to be the most reliable, particularly based on the observed differential survey outcomes across different strata. This is predominantly due to high mobility of the population, particularly in West & Central Equatoria and Jonglei. Sampled individuals were often not found in locations where they were registered. During the survey data collection, it could not be established whether the nonrespondents have merely moved internally or have left the country entirely. Despite these concerns proGres database remains the most reliable existing source of population vital statistics for refugees and asylum seekers. As no other source of population data is available, proGres remains the source for the estimation of the nonresponse calibration weights.

## Use of weights in analysis

Any analysis using FDS data should use the supplied or equivalently estimated weights in order to derive unbiased statistics.

The use of weights depends on the target population of a given indicator as well as on the purpose of analysis. The following analytical purposes are considered:

- estimation of national or sub-national indicators or models – structurally adjusted national weights (proportion of sampling strata is adjusted according to national population structure)
- comparative analysis of sampling strata – structurally adjusted strata weights (size of the strata is not adjusted to population proportions and left at the size sampled to maximise statistical power)

of analysis).

The following target populations are defined:

- Population of refugee households – household weight
- Population of all household members in the households – household weight
- Population of refugees aged 15 and above – individual weight
- Population of children under the age of 5 years – child weight
- Population of women who gave birth in the last two years – woman weight

## Variance estimation and analysis

Variance estimation will be facilitated by the use of survey design specification commands imbedded in the Statistical software like Stata or R. In Stata `svyset` command will be used to specify the survey design, associated weights and strata. For analysis `svy:` prefix will be used to estimated complex variances. Similar functions exist in R.