UNHCR
The UN Refugee Agency

Forced Displacement Survey

SAMPLING NOTE

V 1

Forced Displacement Survey

ZAMBIA

2025

Table of content

Target populations / sampling universe 2

Disaggregation and statistical power 6

Sample Size 8

Sampling approaches and frames 8

Replacement Protocols 15

Sample Adjustment 16

Variance estimation and analysis 18

Target populations / sampling universe

The Forced Displacement Survey Programme, globally, aims to collect data on nationally representative samples of forcibly displaced and stateless people. The 2025 Forced Displacement Survey in Zambia collected nationally representative data on refugees and asylum seekers (RAS), host communities in proximity to refugee settlements, and former refugees in Zambia. The survey collects livelihood data and other data for calculation and reporting of SDG and other indicators as well as to inform programming and policy formulation targeting the needs of the aforementioned target populations.

Refugees and Asylum Seekers

Refugees and asylum seekers (RAS) are defined consistent with UNHCR and national status determination procedures. As such, refugees are those individuals who registered either with the Zambian host government as asylum seekers and underwent a status determination process and got a positive outcome of recognised refugee status, while asylum seekers are those individuals who are still in the process of status determination to become a recognised refugee. By including only registered RAS, FDS data will align to UNHCR official statistics. It should be noted, however, that the population of registered RAS may differ from the total population of potential refugees (i.e., all individuals who could qualify for refugee status should they register and complete status determination processes). In Zambia, concretely, we exclude a known population in Luapula province from the Democratic Republic of the Congo of so-called "self-settled refugees" who may qualify for international protection but whose inclusion is beyond the scope of the survey and would be inconsistent with FDS's focus on recognized, registered refugee populations.

RAS in Zambia are subject to a strict encampment policy which legally confines them to the settlements in which they are registered. There are three refugee settlements in Zambia, Mantapala, Mayukwayukwa, and Meheba, containing 62,680 RAS or 70 percent of the total refugee population. The remaining RAS are registered in urban areas, primarily Lusaka which houses 23,597 RAS or 26 percent of all registered RAS. 75% of RAS are from DRC, 16% from Burundi, and 6% from Somalia. Asylum seekers and Somali refugees are more likely to be in Lusaka. Table 1 shows the distribution of RAS by location and country of origin.

Table 1: Distribution of RAS by Settlement and COO

		Burundi		DR Congo		Rwanda		Somalia		Other		Total	
		HHs	Inds	HHs	Inds	HHs	Inds	HHs	Inds	HHs	Inds	HHs	Inds
Lusaka	AS	2,267	3,019	1,568	3,668	8	8	229	261	210	252	4,283	7,209
	Refugee	2,443	4,309	3,140	9,874	117	465	1,122	1,799	542	744	7,364	17,191
	Total	**4,710**	**7,328**	**4,708**	**13,542**	**125**	**473**	**1,351**	**2,060**	**752**	**996**	**11,646**	**24,399**
Mantapala	AS	0	0	323	727	0	0	0	0	1	1	324	728
	Refugee	0	0	2,231	8,461	0	0	0	0	0	0	2,231	8,461
	Total	**0**	**0**	**2,554**	**9,188**	**0**	**0**	**0**	**0**	**0**	**0**	**2,554**	**9,188**
Mayukway-ukwa	AS	18	30	36	74	0	0	0	1	0	0	54	105
	Refugee	1,594	2,501	5,931	16,880	10	57	4	7	4	8	7,543	19,453
	Total	**1,612**	**2,531**	**5,967**	**16,954**	**10**	**57**	**4**	**8**	**4**	**8**	**7,597**	**19,558**
Meheba	AS	37	97	291	987	1	7	83	88	19	20	431	1,199
	Refugee	2,282	4,351	6,791	26,848	126	438	1,470	1,764	96	149	10,765	33,550
	Total	**2,319**	**4,448**	**7,082**	**27,835**	**127**	**445**	**1,553**	**1,852**	**115**	**169**	**11,196**	**34,749**
Other	AS	135	223	157	375	1	1	131	147	20	24	444	770
	Refugee	106	191	219	658	2	13	866	1,099	125	151	1,318	2,112
	Total	**241**	**414**	**376**	**1,033**	**3**	**14**	**997**	**1,246**	**145**	**175**	**1,762**	**2,882**
Total	**Total**	**8,882**	**14,721**	**20,687**	**68,552**	**265**	**989**	**3,905**	**5,166**	**1,016**	**1,348**	**34,755**	**90,776**

Asylum seekers and refugees are both subject to mobility restrictions. For refugees, mobility rights are granted by the possession of either an Urban Refugee Pass or a gate pass. The Urban Refugee Pass is represented by a white refugee ID card with longer term validity. It gives the refugee permission to reside in an urban area in a specified province on the basis of unpublished criteria including possession of another type of immigration permit (e.g. employment, spouse), medical grounds, security considerations, higher education, or family unity. Gate passes are granted to refugees registered in settlements who have orange refugee IDs and has limited validity of one to ninety days. Asylum seekers registered in urban areas are permitted to be there by their Asylum Seeker Certificate designating the area as their residence while asylum seekers registered in settlements are confined to settlements unless they apply for and are granted a gate pass that allows them to leave the settlement temporarily. All RAS in Lusaka should possess either an urban refugee pass or urban asylum seeker certificate. These individuals will form the Urban sampling frame for the FDS. The remaining 1,762 RAS households registered outside settlement are excluded from the survey. As such, the urban RAS target population are the approximately 11,646 HHs[1] registered in Lusaka.

Despite mobility restrictions and the limited validity and availability of gate passes, a portion of refugees in settlements still exercise mobility. In fact, it is understood that there are a substantial number of refugees and asylum seekers are outside of the settlements either temporarily or for longer periods. As such, Table 1's estimates from proGres almost certainly overestimate the population of settlements. Estimates from UNHCR operations based on the 2022 validation are that 15-30 percent of RAS in Meheba and Mayukwayukwa are not regularly in the settlements. In Mantapala, on the other hand, the estimated proportion is significantly lower. To comply with Zambian policy and for protection considerations, RAS are only able to register/validate their registration in the place of their legal residence, so no exact figures are available.

The FDS will aim to be representative of all RAS registered in settlement regardless of their routine physical presence. To do so, the survey will collect data on household member mobility and be powered to compute statistics for RAS households registered in settlements with and without mobile members separately. The instrument will be tailored to collect information on mobile members. Mobile and non-mobile RAS households will be sampled from a area frame of buildings in Meheba and Mayukwayukwa settlements. In so doing, the FDS will exclude RAS households fully outside the settlement illegally. These are households made up of individuals that live outside the settlement but do not maintain connection to any household physically in the settlement.

In expectation and anecdotally, this is quite rare as mobility passes are granted on individual basis and RAS registered in settlements cannot access services outside settlements, including schooling for children, healthcare, and protection services. Therefore, it is believed that most individuals outside settlements are members of split households. Consequently, we expect that the most likely to reside outside settlements would be single member households without family in settlement. Table 2 shows the share of single member registration groups by settlement (i.e., individuals who completed registration alone).

As expected, the highest share of single member groups are outside settlements, in part because urban refugee passes are granted often on individual basis (i.e., for education). Within settlements, the share ranges from 10-20 percent. This likely overestimates the true share of single member households as members of the same households may arrive at different times and complete registration separately. Nevertheless, if 30 percent of single member households are outside the settlement as estimated by the operation, under coverage would be about 5 percent.

[1] The number of households here is equal to the number of ProGres registration groups. The definition of a registration group is however not the same as the FDS household definition. As such, the sampling plan when sampling ProGres groups survey the household around the case regardless of whether it includes or excludes some case members, correcting selection probability as needed.

Table 2: Prevalence of Single Member Registration Groups by settlement

	Single RAS Registration Groups	Total RAS	Percent
Lusaka	7,148	23,597	30.29%
Mantapala	883	9,226	9.57%
Mayukwayukwa	4,000	19,511	20.50%
Meheba	5,471	33,943	16.12%
Other	1,415	2,902	48.76%

In any case, the FDS will thus sample RAS registered in settlements from three geographic strata, (1) Meheba, (2) Mantapala, and (3) Mayukwayukwa. To facilitate disaggregation of households by the presence of mobile members, in Meheba and Mayukwayukwa where mobility is most common, a sufficient number of households with mobile members will be interviewed to allow for disaggregation at the national level.

Former Refugees

The FDS will include former refugees from Angola and Rwanda who have lost their refugee protections, however, remain in Zambia and are at risk of statelessness. Angolan refugees arrived in Zambia beginning in the 1970s while Rwandans arrived beginning in the 1990s. Due to the resolution of the conflicts leading to their displacement, they have lost refugee protection. Despite the end of their refugee protection, they have been allowed to stay and pursue permanent residence. Those permitted by Zambian authorities to remain have been issued a Registration Certificate by the Zambian Commission for Refugees. The registration certificate allows them to continue to access services like other refugees, however, is not sufficient to regularize their stay in Zambia. To do so, former refugees must pursue Alien registration and temporary or permanent residence. This requires overcoming both financial (application fees) and practical (having a national passport) hurdles.

Former refugees residing in settlements are subject to the same mobility restrictions as refugees, unless they have obtained documentation conferring additional rights (i.e. permanent residence). Former refugees remain registered in ProGres with data updated regularly by CoR. Table 3 shows the distribution of former refugees by settlement and their share of the total settlement population (i.e., RAS plus former refugees). There are no former refugees in Mantapala while there are small numbers in Lusaka and other areas out of camp. The FDS will be representative of all former refugees in Mayukwayukwa and Meheba. The 264 former refugees not in camps are excluded from the sample frame. Further, the FDS excludes former refugees not registered (i.e., not in possession of registration certificates or otherwise registered with CoR).

Table 3: Distribution of Former Refugees by Settlement and Country of Origin

		# of HHs	Share of HHs in Settlement	# of Inds	Share of Inds in Settlement
Lusaka	*Angola*	10	0.08%	15	0.06%
	Rwanda	232	1.95%	685	2.73%
	Total	242	2.04%	700	2.79%
Mayukwayukwa	*Angola*	1,925	20.14%	7,189	26.74%
	Rwanda	37	0.39%	137	0.51%
	Total	1,962	20.53%	7,326	27.25%
Meheba	*Angola*	1,625	11.82%	6,621	14.85%
	Rwanda	929	6.76%	3,221	7.22%
	Total	2,554	18.57%	9,842	22.07%
Other	*Angola*	6	0.34%	24	0.81%
	Rwanda	17	0.95%	45	1.52%
	Total	23	1.29%	69	2.34%

As shown in table 3, most Rwandan former refugees are in Meheba where they account for 7 percent of the settlement's population while Angolans are more likely to reside in Mayukwayukwa (26 percent of the population), however there is also a substantial Angolan population in Meheba (15 percent of the population). It should be noted that, like RAS, former refugees may also exercise mobility and may also be outside settlements. Given the length of their stay in Zambia, however, many former refugees have established more permanent dwellings in settlements. Still, it is expected that former refugee households may also have mobile members. Households with mobile members will be included in the former refugee sample, although without explicit strata allowing for disaggregation. Instead, the former refugee sample will be stratified by country of origin and powered to detect country of origin-specific effects nationally.

Host Communities

There is no standard definition of refugee host communities or populations. Operationally, in the strictest definition, the host community is the national/legal resident non-refugees/persons of concern of the host country accessing UNHCR services. However, more commonly and more broadly, it is defined as any host country nationals/legal resident non-refugees/persons of concern affected by or frequently in contact with refugee populations. We adopt the latter definition for the purposes of the FDS in Zambia focusing on individuals living in proximity to the three refugee settlements, Meheba, Mantapala, and Mayukwayukwa.

By consequence, we exclude communities not in proximity to settlements that may still host refugees, including Lusaka. This is for two main reasons. First, communities (and areas within cities) hosting refugees outside settlements are not easily identifiable as refugees are widely dispersed. And second, the number of refugees relative to the size host community (i.e., the population of Lusaka) away from settlements and in cities is much smaller meaning their impact on the host community is comparatively minimal.

For purposes of the FDS, the host community sampling frame will be an area frame defined by distance to settlement boundaries. In Meheba and Mayukwayukwa, all households residing within a 30 km radius will be included in the frame. In Mantapala, all houesholds within a 15 km radius will be included in the frame. Table 4 shows the distribution of buildings by distance using the Overture opensource building database.[2]

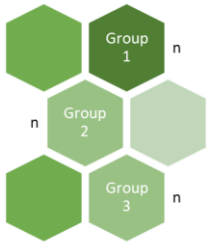
Table 4: Distribution of Buildings within 30 Kilometers of settlement by distance

		Mantapala	Mayukwayukwa	Meheba
0<=Distance<=5	*n*	1,416	5,153	7,904
	%	12%	15%	17%
5<=Distance<=10	*n*	1,786	3,680	11,659
	%	15%	11%	26%
10<=Distance<=15	*n*	8,483	4,611	5,623
	%	73%	13%	12%
15<=Distance<=20	*n*		5,084	8,474
	%		15%	19%
20<=Distance<=25	*n*		7,164	5,217
	%		21%	12%
25<=Distance<=30	*n*		8,558	6,295
	%		25%	14%
Total		11,685	34,250	45,172

[2] Overture Maps Foundation, overturemaps.org.

Disaggregation and statistical power

As described above, the FDS in Zambia will be statistically representative of three main target populations: refugees and asylum seekers (RAS), refugee host communities, and former refugees. For RAS and former refugees, FDS will be nationally representative while for host communities FDS will be representative of communities in proximity to Zambia's three refugee settlements. A survey's representativeness is determined by the extent to which the selected sample is similar to (or differs from) the target population. Random selection of respondents from a complete sampling frame is central to representativeness. To achieve representativeness, FDS will develop list or area frames for each target population using spatial data combined and UNHCR registration databases, then randomly select respondent households. As detailed below, a strict replacement protocol to address non-response/ineligibility as well as ex-post sample adjustment will also be employed to ensure representativeness.



sample size = 3 * group sample size

In addition to representativeness, the quality and utility of a survey is dictated by its precision and power. The acceptable level of precision and power is dictated by tolerance for error in the survey's measurement, i.e., the level of confidence in reported statistics. Error tolerance in turn inversely dictates the required sample size. Both low and high confidence surveys can fail to be representative while representative studies can have low or high confidence. As such, the two main questions that dictate sample design are: "of whom or what does the survey need to representative?" and "what does it need to be powered to measure?".

FDS aims to be nationally representative but must facilitate sufficient disaggregation across population features of interest to inform operations, policymaking, and other reporting needs with sufficient statistical confidence. Concretely, this means that for each target population disaggregations (or analytical domains) must be defined. These are the sub-groups within the target population for which analytical results will be able to derived in a statistically robust way (i.e. with confidence that the results obtained from the data reflect the reality in the population). As such, the total sample size of the survey is a function of the number of disaggregations and the desired level of precision not the size of the total target population or the size of the subgroup. Therefore, although disaggregation across many domains is desirable, priority analytical must be selected. The size of these analytical groups or disaggregations are referred to in this text as MAES or Minimum Acceptable Effective Size.

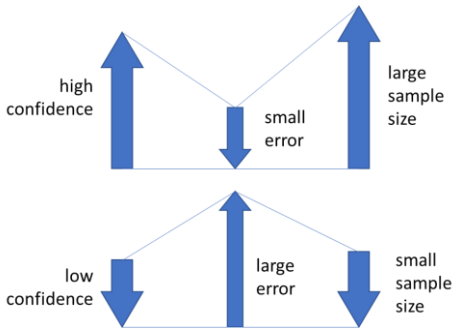


Table 5 summarizes the main analytical groups selected for each target population and MAES.

Table 5: Analytical Groups and Sample Sizes (MAES)

#	Population Groups	Sample Size
In Settlement Refugees and Asylum Seekers (Settlement-specific statistics)		
1	**Meheba**	**600**
	Wholly in settlement HHs	350
	HHs with mobile members	250
2	**Mayukwayukwa**	**600**
	Wholly in settlement HHs	350
	HHs with mobile members	250
3	**Mantapala**	**500**
Out of Settlement Refugees and Asylum Seekers		
4	**Lusaka**	**500**
	Refugees	350
	Asylum-seekers[3]	150
Host Communities (Settlement-specific statistics)		
5	**Meheba**	**400**
6	**Mayukwayukwa**	**400**
7	**Mantapala**	**400**
Former refugees		
8	**Rwanda**	**150**
	Meheba	144
	Mayukwayukwa	6
9	**Angola**	**500**
	Meheba	230
	Mayukwayukwa	270
Total Sample Size		4,050

In total, there are 9 analytical groups identified for the FDS in Zambia. For refugee and asylum seekers, the FDS will facilitate disaggregation to report settlement-specific results. It will also include an urban analytical group from Lusaka and support disaggregation between refugee and asylum seeker households with and without mobile members jointly in Meheba and Mayukwayukwa settlements. The survey is also powered to measure settlement-specific statistics for each of the three host communities around Meheba, Mantapala, and Mayukwayukwa settlements. Former refugees will be reported together with country-of-origin specific statistics for Angolans. There will be no disaggregation by settlement for former refugees given the relatively small size of the population and the sample size it would require.

[3] As an in-field decision, asylum seekers were removed from the survey design due to contact rate of less than 2 percent. As such, only refugees were surveyed in Lusaka and the target revised to be 500 refugee interviews.

Sample Size

As described, the overall approach to sample size calculation and composition is based on FDS's analytical requirements and the identified analytical domains and needed disaggregations. The required sample size or MAES is therefore estimated at the level of each disaggregation and depends on the proposed sampling strategy.

The proposed sample size per analytical domain is calculated using the following formula:

$$n = \frac{z^2 \times p(1-p)}{\propto^2} \times deff$$

where:
$n =$ net sample size
$z =$ z-score
$p =$ proportion of children with a given trait
$\propto =$ margin of error
$deff =$ design effect

For analytical domains using clustered sampling, 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 - and the similarity of households within each cluster or the intra-cluster correlation (ICC).

$$deff = 1 + (m-1) \times ICC$$

In the FDS in Zambia, the proposed sampling method for the analytical groups is simple systematic stratified random sampling. Therefore, the minimum sample size for each group with a 5 percent confidence level is 385 (assuming population proportion of 50%), however, because the true population proportion is not known for most FDS indicators and because FDS collects on many indicators with ex-ante unknown variances the MAES for each group is increased to at least 500 observations for all domains that will be reported separately. Table 4 summarizes the required sample sizes for each explicit stratum. Note that for HHs with mobile members and former refugee analytical groups the sample sizes within each settlement are less than the required 385 households as the decisions was made to not disaggregate the results for these populations by settlement. Similarly, the sample size for Rwandan former refugees is proportional to their share of the population and no country-of-origin specific statistics are foreseen.

Sampling approaches and frames

The FDS in Zambia will employ both area and list frames to sample its four target populations. The main source for the development of area frames will be the Overture Building Footprints database. The Overture Building Footprints is a database of geocoded building data that is acquired by processing high resolution aerial photography aggregating across numerous data sources, including OpenStreetMap, Esri Community Maps, Microsoft's Global ML Building Footprints, Google Open Buildings, USGS 3D Elevation Program, and other. For areas with observed under coverage, the database is augmented through manual identification of buildings from aerial photography. The list frame will be sourced from registration databases. From each frame, households will be randomly sampled. This section outlines the frames and sampling strategies for each target population and analytical group.

Refugees and Asylum Seekers

The Refugee and Asylum Seeker population is divided among three settlements and Lusaka. Within the population registered to settlements, the population is further divided between

households with members exercising mobility outside the settlement and those wholly in the settlement. For households wholly or partially in settlement, the primary frame will be an area frame of building footprints within settlement boundaries. In Lusaka, the frame will be a list of cases from the ProGres registration database.

For the purposes of FDS, a refugee household is defined as a household where the household head or a spouse of the household head is a refugee or asylum seeker. For cases of mixed households, the refugee or asylum seeker status dominates. This means that households in which the head or spouse is a refugee or asylum seeker, but their partner is a former refugee, host community member, or returnee form part of the refugee and asylum seeker sampling frame. Sampled cases not meeting the above definition are replaced. The population of refugee households is further subdivided between households with and without mobile members. Mobile members are individuals from the household that reside habitually outside the settlement with or without a valid mobility pass while continuing to share common resources with the household (see box 1).

RAS in Settlements

As explained, despite registration in settlements, a number of refugees reside informally outside settlements. This results in split households where some members remain in settlement while others are outside. As such, within the population of households registered in Meheba and Mayukwayukwa, FDS will distinguish households with mobile members from those without mobile members. Then, at national level, the survey will support disaggregated statistics and indicators for these two types of households. Additionally, to learn about the livelihood conditions of mobile members, the FDS will include an individual interview of a randomly selected mobile member in households identified as having one or more mobile members.

Box 1: Identification of RAS Households with and without Mobile Members

Households with and without mobile members will be identified through two main mechanisms:

1. HH Roster of Head of HH questionnaire
2. Random listing of building footprints

In the HH roster, members will be listed with prompting to include members present and not present. For members, identified as away a migration roster will be created. The same migration roster will be used for the listing survey. In migration roster, respondents will be asked to identify:

1. All members who have resided outside the settlement to work or look for work for at least 1 month in the last 12 and have returned.
2. All members currently residing outside settlement to work or look for work

Households meeting with at least one member meeting either condition are classified as HHs with mobile members.

The sample of RAS in settlements will therefore be representative of all refugees registered in settlements, including refugees residing outside settlement that maintain connection to households in settlements. The only potential source of exclusion error is the existence of households that are fully outside the settlement and do not maintain a connection to a household in the settlement. As explained, we expect this number to be relatively small. Further, it is beyond the resources of FDS to develop a list frame and track refugees that maintain no connection to settlements.

Box 2: Individual Survey of mobile members
In each household where a mobile member is identified, the FDS will interview one randomly selected mobile member. If the member is present (i.e., they were working outside the settlement and have returned), the survey will take place at the same time as the household interview. If the mobile member is away and will not return in the period of data collection, they will be traced and interviewed by phone.

The survey of mobile member include a subset of modules typically addressed to the random member, specifically those on employment, job search, financial inclusion, acceptance, plus modules on access to housing and food and nutrition to understand living conditions outside the settlement. The survey will allows for disaggregated reporting of indicators for the mobile population.

In total, a sample of 1,900 RAS households will be surveyed. In Meheba and Mayukwayukwa, the 700 households will be selected composed of 300 households with mobile members and 400 households without mobile members. In Mantapala, 500 households will be surveyed where no stratification by existence of mobile member is foreseen. The main sampling frame for selection of RAS households in settlements will be an area frame of buildings within the settlements. The sample is selected though systematic random sample from the whole frame of buildings stratified by settlement.

Within the building frame, the only information available is the location of the building. Information on the household inhabiting and unit eligibility can only be obtained at time of interview and the selection probabilities adjusted accordingly ex-post. Owing to the limited information in the frame, stratification below the settlement level at the stage of sample selection is not possible. In Mantapala, this is not a problem as no stratification is foreseen, and it is expected that all households in the settlement are RAS households and eligible as there are no former refugees in Mantapala. In Meheba and Mayukwayukwa on the other hand, this is a not the case because we aim to stratify between RAS HHs with and without mobile members and the building frame contains households inhabited by both RAS HHs and former refugee households. This means that different sampling strategies are required in Meheba and Mayukwayukwa than in Mantapala.

Table 6: Number of buildings vs. number of ProGres cases by settlement

	# of Buildings	# of ProGres Registration Groups
Mantapala	*3,373*	*2,558*
Mayukwayukwa	*9,049*	*9,611*
Meheba	*16,619*	*13,869*
Total	29,041	26,038

Sample selection in Mantapala

As shown in table 1, the entire population of Mantapala settlement is refugees and asylum seekers from the Democratic Republic of Congo. Compared to Meheba and Mayukwayukwa, it receives more new arrivals and is Zambia's newest refugee settlement. Mobility out of the camp among refugees is also considerably lower due to the settlement's remoteness and the recent arrival of much of the population. Residents of the settlement are also much less likely to have a gate pass with less than 1 percent having an active pass and only five percent ever having a pass. For these reasons, the sample is not stratified explicitly among HHs with and without mobile members. And, since the population within the settlement is homogeneous in terms of eligibility, the sampling strategy is systematic random sampling of buildings with ineligibility and non-response addressed according to the replacement protocol outlined below. In total, 500 RAS households will be interviewed. To ensure sufficient geographic coverage of the settlement, the settlement is placed a 500 m by 500 m grid where each cell in the grid will be one implicit stratum. In terms of coverage of the Overture building footprints database, review of satellite imagery confirms accuracy and table 6 shows that the number of buildings exceeds the number of ProGres groups, i.e., households, as we would expect.

Sample selection in Meheba and Mayukwayukwa

In Mayukuwayukwa and Meheba, there are four sub-populations of interest in each settlement by which the FDS sample is stratified:

1. Refugee and Asylum Seeker Households without Mobile Members
2. Refugee and Asylum Seeker Household with Mobile Members
3. Angolan Former Refugee Households
4. Rwandan Former Refugee Households.[4]

It is not possible ex-ante to distinguish among these households from the building footprints frame. Even if we were to use ProGres as the frame, it would still not be possible to distinguish between RAS HHs with and without mobile members and due to mobility ProGres would not identify households in settlement at time of data collection. Therefore, the preferred option is to draw one systematic random sample of buildings from the building footprints frame to cover all four strata in Meheba and Mayukwayukwa. Then, to classify households, households in selected buildings will be screened in random sequence and will be interviewed if their stratum has not been completed. Interviews will continue until the desired sample size in each stratum is reached. Once a stratum is complete, all households in the selected sample identified subsequently will be screened out.

Because all households in the frame are eligible owing to the fact they live in the settlement, the random selection and random ordered screening maintains equal probability of selection and allows us to estimate the population size of each stratum. To ensure the probabilities of selection are maintained, a strict protocol will be maintained to ensure that the random ordered screening protocol is maintained and the screening will continue until all strata are complete. Further, because the order of the screening is random, the results can be used to estimate the proportion of households falling into each stratum in each settlement which is unknown ex-ante for the RAS HH with and without mobile member strata and required for sample adjustment and reweighting. The sampling procedure is as follows:

1. Systematic Random Selection of Buildings in each settlement with probability,
$$p_i = \frac{n_s}{N_s}$$
where $n_s$is the total number of sample units drawn and N_s is the total number of buildings in the frame and is approximately equal to
$$N_s \cong N_{RAS_{mobile}} + N_{RAS_{nonmobile}} + N_{formerrefugee}$$
2. Randomization of selected units.
3. Screen each unit following prescribed order. Continue screening until target sample size for all strata reached.
4. If eligible and the stratum is not complete, the unit is interviewed.
5. Estimate population size of each stratum based on results of listing to facilitate ex-post sample adjustment,
$$N_{RAS_{mobile}} = N_S * p_{RAS_{mobile}}$$
$$N_{RAS_{nonmobile}} = N_S * p_{RAS_{nonmobile}}$$
$$N_{formerrefugee} = N_S * p_{formerrefugee}$$
Where p is the estimated proportion based on the prevalence of each group in the screening.

Overall, this procedure allows us to efficiently identify and classify households into each stratum while avoiding oversampling which is inefficient from a cost perspective and quota sampling which is non-probabilistic.

[4] Note that the nationality of a former refugee household is defined by the nationality of the head of household. If the head of household is not a former refugee, then it is defined by the nationality of the spouse. If both the head and spouse are former refugees and their nationalities differ, then the nationality of the head defines the nationality of the household.

The main sample size to cover all 4 strata in each settlement will be 2,000 buildings in Mayukwayukwa and 2,500 buildings in Meheba. Should the main sample be exhausted before all strata are complete a top-up sample(s) will be drawn from building frame, with size $N_s - n_s$, for which the same procedure will be followed with screening proceeding in random order upon completion of the main sample. The screening questions will be a short questionnaire that identifies the presence of mobile members in the households and the status of the household head and his/her spouse(s) to determine whether they are a RAS or former refugee according to the FDS definition above. Like in Mantapala, the overture building frame as been verified against recent satellite imagery and the number of buildings identified in Mayukwayukwa and Meheba exceeds the number of registration groups in ProGres as expected.

In field change of approach

The above approach was challenged by higher-than-expected mobility of households outside and within the settlements, especially in Mayukwayukwa. In fact, contrary to expectation the target sample sizes for former refugees were reached prior to the targets for refugees despite refugees making up a larger share of the population according to ProGres registration data. As such, in both settlements, screening out occurred only for former refugees. Additionally, the planned approach to identify households with and without mobile members and follow up with mobile members proved infeasible as a relatively low share of households were identified having mobile members. Maintain the approach would have required lengthening the field window and screening out many refugee households without mobile members. The reason for the lower-than-expected share appears to be an unwillingness to identify mobile members, not considering those longer-term away as household members, and whole households travelling outside the settlement together. Additionally, in Meheba, given the relative proximity to economic opportunities members may not have been away long enough for respondents to understand their absence as time away from the household (which was how the questions were phrased).

Given these challenges, an in-field decision was made to remove the disaggregation of households with and without mobile members with the settlements. They were replaced with a sample of households registered in Mayukwayukwa but located in Lusaka during the data collection window. Sadly, as most households from Meheba travel to Solwezi, Ndola, Kitwe, and other locations, a similar approach could not be adopted as the survey had not field presence in those locations. However, it is understood that duration away from the settlement is shorter in Meheba and given the survey team's permanent presence in the settlement for approximately 3 months, many initially sampled absent households were captured upon their return to their settlement.

The sample of households registered in Mayukwayukwa but residing in Lusaka was selected from the ProGres registration database. The sample was selected by systematic random sampling with a sample size of 400. The same protocol for contacting and locating households in Lusaka was employed, however there was no community outreach because residence was less permanent with weaker community networks and each case's presence in the city was no known ex-ante. There were also protection-related concerns which made phone outreach the safest possible option. Sampled households contacted that were located in the settlement, already interviewed, or otherwise not in Lusaka were replaced. The approach achieved a reasonable number of interviews and compared to the mobile member approach has the advantage of having the full interview for mobile households.

Registered Urban RAS in Lusaka

Legal urban residency of refugees is predicated on possession of an urban residence permit (white refugee ID). It permits a refugee to live in an urban area in a specified province on the basis of unpublished criteria including possession of another type of immigration permit (e.g. employment, spouse), medical grounds, security considerations, or family unity. Asylum seekers cannot obtain urban residency cards and obtain their rights to reside in an urban area based on the location of their asylum, application. The Lusaka sample of RAS will be representative of all RAS legal authorized to reside there. RAS informally residing in Lusaka

are excluded from this stratum but are included among the mobile members sampled from selected households in Meheba and Mayukwayukwa.

The sampling frame for Lusaka will be a list frame of ProGres refugee and asylum seeker cases. As in other surveys sampled from ProGres, the household interviewed will be the household around the case focal point regardless of whether all members have urban residence passes, refugee, or asylum seeker status. From the ProGres list frame, a simple systematic stratified random sample will be drawn of 500 cases with explicit strata for refugees and asylum seekers to account for the higher prevalence of asylum seekers in Lusaka compared to settlements. Three hundred and fifty refugee households and 150 asylum seeker households will be interviewed. In cases of non-response and ineligibility in the main sample, cases will be replaced from a randomly selected replacement sample.

In field change of approach

Locating households from the ProGres registration database in Lusaka proved challenging. The search protocol was as follows. First, if available, the phone number for the case focal point was called, followed by all other phone numbers available for other case members. At least, three attempts were made for each phone number on different days and at different times a day. Second, if the case was not reached by phone or no phone numbers were available in ProGres, outreach to refugee community leaders and through reception centers was conducted. A protocol by which for each case, based on the country-of-origin and neighborhood of residence (if available), relevant community members were contacted and sites where refugees receive services (i.e., reception centers) and transact or work (i.e., local markets) were visited. Only upon exhaustion of relevant community outreach were cases deemed unreachable.

Contact rates were low for both refugees and asylum seekers with most cases being reached by phone and few cases without phone numbers reached through community outreach. For asylum seekers, the contact rate was especially low as only 10 percent of cases in ProGres had a phone number recorded for any member. Out of 1000 sampled cases, only 15 interviews were completed. As such, the in-field decision was made to remove asylum seekers in Lusaka from the survey design as achieving the target sample size let alone representativeness would be infeasible. Without the stratum for Lusaka asylum seekers, the sample size for Lusaka refugees was revised to 500.

Former Refugees

Former refugees will be sampled following the procedure outlined above for refugees and asylum seekers in the two settlements that host former refugees, Meheba and Mayukwayukwa. First, a main sample of building footprints will be selected through systematic random sampling without distinction between RAS and former refugees. Then, each selected building will be screened in random sequence and interviews completed with former refugees until the sample sizes listed in table 5 are reached. Should the target sample sizes for former refugees be reached prior to completion of all strata in the settlement, former refugee households identified subsequently will be screened out. The sample sizes for Rwandan and Angolan former refugees in each settlement are proportional to the distribution of each population between the two settlements for a total of 500 surveys with Angolan former refugees and 150 with Rwandan former refugees.

Host Communities

Host communities for each of the three settlements will be included in FDS. No urban host community or mobile refugee host community will be included. The host community will be defined on the basis of proximity to the settlement. All non-refugee, non-asylum seeker, and non-former refugee households within thirty kilometers of the Meheba and Mayukwayukwa and 15 kilometers of Mantapala are eligible members of the target population, including both

national Zambians and other legal resident households. To account for the inverse relationship between distance and interaction with the settlement, the sample will be selected with probability proportional to proximity through simple systematic random sample of building footprints within the settlement buffer.

The area frame is sourced from the Overture Building Footprints database of all buildings using a Euclidean distance buffer around the settlement. The frame's coverage was verified through visual inspection of recent satellite imagery. The use of the footprint frame has the advantage of being a complete listing and allows us to avoid defining smaller geographic clusters and enumerating those clusters which would require larger overall sample sizes and more resources for listing. In total, 1500 host community objects will be selected equally divided among the settlements with probability proportional to proximity. Ex-post adjustments will be made to account for unequal probability of selection for households associated to two or more objects in the frame and objects associated to two or more households.

Replacement Protocols

In order to address any nonresponse or ineligibility of sampled households a strict replacement protocol has been put in place for the host community and Mantapala settlement samples. As explained top-up sampling will be used for samples in Meheba and Mayukwayukwa due to the inability to classify households in the footprint frame a-priori.

Although replacements as the main protocol to address potential unit nonresponse are often discouraged by producers of official and academic as they can result in biased samples if used inappropriately, they are most applicable for FDS as we must ensure sufficient statistical power for each analytical group and resource constraints mean that sample sizes cannot exceed by much the initial budgeted sizes. Further, oversampling would require a-priori estimation of expected response rate which is unknown for FDS target populations.

Replacement samples will be drawn together with the main sample and selection into a replacement sample was randomized. Issuing of replacement was further randomized at the level of the explicit sampling stratum definition and was to be strictly controlled – i.e. issuing of replacement households would have been handled by the central survey coordination team from the implementing agency. Use of households from the replacement pool needs to be strictly documented. Using of the replacement methodology will enable 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 is 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 weights for such estimations.

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:

In the first step the base probabilities of selection are estimated separately for each explicit stratum. Selection probability is the calculated as follows:

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

Where ns is the total number of sample units drawn into both the main and replacement samples and Ns is the total number of units in the frame.

The base weight is the inverse of the selection probability

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

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

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

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. This estimation strategy applies to both systematic random samples from ProGres and systematic samples of buildings within settlements.

For the Lusaka and Mayukwayukwa out of camp refugee samples selected from ProGres, there is an additional step of non-response adjustment made prior to structural adjustments. Due to relatively high response rates in camps and among hosts and the paucity of data

available in the frame, non-response adjustment is limited to interviews conducted in Lusaka. The non-response adjustments was made using a probit regression. The model was specified as follows:

$$Pr(R_i = 1|X) = \Phi(X_i'\beta)$$

Where R_i is a binary variable indicating the response outcome of the each household i, X_i' is a vector of explanatory variables including the gender of the case lead, the case size, age of the case lead, marital status of case lead, the dependency ratio of the case, the arrival year of case lead, and country of origin fixed effects, β is an equal sized vector of coefficients, and Φ is the cumulative distribution function of the standard normal distribution. The resulted predicted response probabilities are used to adjust the base selection probabilities.

Building footprint samples of hosts based on proximity

The samples of host in proximity to the settlements are approximately self-weighting according to the selection gravity coefficient. Selection gravity coefficient is power adjusted (6th root) simple Euclidean distance from the building to the closest settlement boundary. Calculated as follows.

$$dist_{ijk} = \left|b_{ij} - P_{ik}\right|$$

$$dist^*_{ijk} = \sqrt[6]{dist_{ijk}}$$

$$prox_{ijk} = round\left(1{,}000 \cdot \left(\frac{\left(\max_{1\le j_k\le J_k} dist^*_{ijk} - dist^*_{ijk}\right)}{\max_{1\le j_k\le J_k} dist^*_{ijk}}\right)\right)$$

$$p = \frac{1}{prox_{ijk}}$$

The self-weighted sample is approximately representative of the population that lives in proximity of the refugees as the selection gravity coefficient favors those living closer to the settlement boundary, meaning that they have higher weights in estimates on the host population. This is consistent with our theoretical conceptualization of host/affected communities.

Weights for building footprint samples of hosts and refugees in settlements are further adjusted for the number of buildings the household occupies. The adjustment coefficient is one 1 over the number of buildings owned. The number of buildings occupied by the household is reported during the interview, and the adjustment corrects for the fact that household occupying more buildings have higher likelihood of selection intro the sample.

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.

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 maximize 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 and other comparable statistical software.