

Socio-Economic Survey of Refugees in Ethiopia (SESRE)

Survey Design and Methodology

SESRE is a separate but integrated survey alongside the Ethiopian Household Welfare Statistics Survey (HoWStat),¹ the national household survey to measure poverty and other socio-economic outcomes. Like most national poverty surveys, HoWStat excludes displaced populations—Internally Displaced People (IDPs) or refugees—including in Ethiopia. To have up-to-date information on the socio-economic outcomes and poverty levels of refugees and to allow comparison to Ethiopian host communities, the SESRE applied the same questionnaire and data collection methods as the HoWStat, with some modifications. Training of the enumerator team and implementation arrangements of the survey followed the same standards and procedures as the HoWStat. SESRE data was not collected alongside HoWStat due to security concerns at the time of data collection for HoWStat, especially in the refugee areas.

The SESRE aimed to solve two problems: (i) gaps in data on the socioeconomic dimensions of refugees, and (ii) gaps in analytical studies presenting the socioeconomic outcomes of refugees and hosts. Lack of up-to-date evidence is a significant obstacle to designing effective policies and support for refugees and host communities. To this end, the availability of the data helps to analyze refugee hosting areas' social dynamics and longer-term socioeconomic viability by focusing on the: (i) social impact of refugees on host communities, (ii) socioeconomic interaction, (iii) social inclusion, and (iv) social relations among refugees and between refugees and host communities. The data provides valuable information to development partners and governments to inform policies to facilitate refugees' integration and improve their lives, along with refugee hosting communities.

The SESRE covers all current major refugee camps: Eritreans, South Sudanese, and Somalis, as well as the out-of-camp refugees in Addis Ababa. In addition, the survey covers the respective host communities around the camps, including the host communities of Addis Ababa. Due to the conflict in the Tigray region of Ethiopia between 2020 and 2022, Eritrean refugees living in camps in Tigray could not be included in this survey. To avoid exclusion of Eritrean refugees in Ethiopia, we included Eritrean refugees living in camps in the Afar region and the newly established refugee hosting zone *Alemwach*. Eritrean refugees who were in the Tigray region prior to the conflict are included in this survey in two ways: we sampled (i) refugees from *Alemwach*, where most of the refugees previously located in Tigray moved after conflict broke out and (ii) from Addis Ababa, namely those refugees who arrived in Addis Ababa after November 2020. Data collection took place between November 2022 and January 2023.

Sample population

The SESRE covers three types of groups, all of which require a distinct sampling procedure:² (i) refugees in camps; (ii) refugees out-of-camps; and (iii) host communities. This section discusses the sampling frames of each group.

¹ Formerly the Household Consumption and Expenditure Survey and Welfare Monitoring Survey.

² Prior to the sampling process, the survey team conducted a pre-sampling assessment by visiting the camps to verify on-the-ground conditions.

(a) Refugees in Camps

The sampling frame for refugee camps is based on UNHCR's proGRES database. The refugee camps were grouped into three domains based on the concentration of refugees from the three major origin countries: South Sudan, Somalia, and Eritrea.³ The first sampling stage divided each camp into enumeration areas (EAs). Based on the proGRES database, we created pseudo EAs by taking 150-200 households in a row from the list; that is, 50-200 HHs grouped as EA1 and the next 150-200 households grouped as EA2, and so on. EAs and households from each sampled EAs were selected.

(b) Refugees in Addis Ababa

We used a slightly different approach for refugees in Addis Ababa because of the difficulty of obtaining a reliable, complete list of locations for refugees living there. The refugee sampling frame in Addis was based on UNHCR's proGRES registration data, sorted by location. The UNHCR list has information about how many refugee households live in each *Woreda* in Addis Ababa, their contact details, location, and other information. We developed pseudo-EAs from the list by location (sub-city and *Woreda*); some EAs covered more than one *Woreda*, and multiple EAs were in a single *Woreda*. We selected a sample of EAs and households from each EAs in collaboration with UNHCR. Finding refugees in Addis Ababa was challenging, as they change their location frequently. To minimize the burden of searching for selected refugees, representatives of the selected households were contacted before the survey to ask them to come to a UNHCR center to collect preliminary information, including their current residential address. Since many Eritrean refugees in Addis Ababa had fled from the conflict in Tigray, out-of-camp refugees in Addis Ababa were stratified into two domains: refugees who arrived before the start of the conflict in November 2020, and those who arrived after November 2020.

(c) Host Communities

Populations around the refugee camps under each domain were meticulously identified in consultation with UNHCR and RRS. We used the ESS EA maps to assess the settlement of communities around camps, ensuring a precise fit to the definition of a "host community". The assessment highlighted that using the list of EAs obtained from the new cartographic frame⁴ meets the definition of host community. In the SESRE, host community members are defined as those who live adjacent to a refugee camp but within a radius of 5km. We use the updated Ethiopian Statistics Service 2018 cartographic database of enumeration areas (EAs) to define them. An EA is a defined area where 100-150 households live in rural areas, while in urban areas, it is an area where 150-200 households live. The first stage of sampling for the host community involved using simple random sampling to select EAs—the primary sampling unit—from the list of EAs that are adjacent but within a radius of 5km.

Following EA selection, a fresh list of households was prepared at the beginning of this survey, which was used as a frame to choose sampled households from each sample EA. In Addis Ababa, a separate host domain was developed as refugees spatially concentrate in a few sub-cities and *Woredas*. We applied the

³ Refugees from Sudan were not included in SESRE as, at the time of sampling, there were less than 50,000 Sudanese refugees in Ethiopia and inclusion was not deemed cost effective.

⁴ The cartographic map (frame) was prepared in 2018 for the upcoming Population and Housing Census

ESS EA maps around the area where refugees in Addis Ababa are located. We selected EAs in the first stage and then conducted a complete listing.

Sampling design

The sample for this survey was 3,456 households from eight domains, with data was collected from 3,452 households (Table C.3).⁵ There are three domains for the three largest in-camp refugee groups—Eritreans, Somalis, and South Sudanese—three for host communities of these major refugee groups, and one for refugees and one for host communities in Addis Ababa. In all categories, a stratified, two-stage cluster sample design technique was used to select EAs and 12 households per EA, whereby the EAs were considered a Primary Sampling Unit and the households as the Secondary Sampling Unit. The SESRE is designed to estimate demographic, socioeconomic, welfare, and refugee-specific indicators of the eight domains.

Table C.1: The distribution of sampled and surveyed households by domains

Domain	EA		HH	
	Sampled	Covered	Sampled	Covered
Eritrean refugee domain	36	36	432	432
Somalian refugee domain	36	36	432	432
South Sudanese refugee domain	36	36	432	432
Eritrean host domain	36	36	432	430
Somali host domain	36	36	432	431
South Sudanese host domain	36	36	432	432
Addis Ababa refugee domain	36	36	432	431
Addis Ababa host domain	36	36	432	432
Total	288	288	3,456	3,452

Note: We have three segmented EAs in Somali host domain.

Sample size estimation

(a) First Stage Sampling

In the first stage sampling, each domain is considered an explicit sampling domain. We used the list of all EAs as a sampling frame and their estimated population as a Measure of Size (MoS). A sample is selected with Probability Proportional to Size (PPS). The sample size is evaluated regarding the expected precision of the key indicator for the SESRE, the national household survey to measure poverty as the percentage is 0.235 (2016 Household Consumption and Expenditure). In the calculation, values for the measuring poverty rate (P) and design factor (deft) 1.5, the expected Relative Standard Error (RSE) of 4.63%, and finally, an adjusted Response Rate of 99% at a 95% Confidence level used to represent the expected precision is acceptable at the domain level. To select a representative sample from this population, first, the initial sample size was determined by using the following scientific formula:

⁵ See Annex C for sample size estimation.

$$n = def t^2 \frac{(1/p - 1)}{\alpha^2}$$

where the *deft* is the design factor defined as the ratio between the square root of standard error using the given sample design and the standard error resulting from a simple random sample used. Based on the above scenario, total sample size $n = 3,456$ Households, and $EAs = 288$.

An equal allocation method was used to ensure that the survey precision was comparable across domains, where 36 EAs were selected from each domain. Based on a fixed sample take of 12 households per cluster,

Equal Allocation formula

$$n_j = \frac{n}{j} \quad j = 1,2,3, \dots, 8 \text{ (number of domains)}$$

Where:

n = total number of sample households and n_j = Number of sample households allocated to stratum j

(b) Second Stage Sampling

In the second stage sampling, we selected 12 households per selected EA. The probability of p_{hij} of selecting a household in segment hij of EA hi of domain h is given by

$$p_{hij} = \frac{k_h n_{hi} s_{hi} m_{hij}}{N_h s_{hi} M_{hij}}, \text{ where}$$

- k_h is the number of EAs in the domain's sample,
- n_{hi} is the estimated population of the EA,
- N_h is the estimated population of the domain,
- s_{hi} is the number of segments listed in the EA (normatively always 2, or 1 if the EA is not segmented),
- m_{hij} is the number of households visited in the EA (normatively always 12), and
- M_{hij} is the total number of households listed in the EA.

There are 36 EAs per domain. With 8 survey domains, and 12 households per EA, a total of 3,456 households in the sample.

(c) Replacement of Households

We implemented a two-layer replacement strategy: First, in each EA, 12 additional households were sampled to serve as replacement households. We sampled the replacement households in the sample allocation and sample size determination stage. As part of the data collection protocol, each household needs to be visited at least three times before replacing a household from the list of replacements. The list of replacement households was only provided to the enumerators upon demonstrating that three visits were attempted. In the case of in-camp refugees, if the enumerators, together with the focal person from RRS and UNHCR, could not identify the selected household within the camp, they were provided with a list of replacements. Second, in case of missing to identify even the replaced sampled households, the enumerators were requested to go back to the original sampled household and skip ten households using a counterclockwise rule to find a new replacement household.