Compilers' Manual on Forced Displacement Statistics

USE CASE INCLUDING REFUGEES IN A SAMPLE SURVEY OF THE NATIONAL POPULATION, OR RUNNING A STAND-ALONE SURVEY OF REFUGEES





This is Use Case B from the *Compiler's Manual on Forced Displacement Statistics*. The Use Case describes how refugees can be included in a sample survey of the national population and how to run a stand-alone survey of refugees.

The *Compiler's Manual* is aimed primarily at technical personnel in National Statistical Systems who want to include displaced populations – refugees and / or internally displaced persons (IDPs) – in official statistics. Each Use Case discusses a different scenario relevant to producing official statistics on refugees and IDPs, with a focus on the elements of statistical production cycles that are specific to refugee and IDP contexts. Spotlight examples of good practice in the production of refugee and IDP statistics are interwoven throughout the Use Case.

The Compiler's Manual and its individual Use Cases are intended to be a 'living document' which will be amended and extended as the body of expertise and knowledge develops worldwide.

Note: Paragraph numbering is per the complete version of the Compilers Manual.

The Expert Group on Refugee, IDP and Statelessness Statistics

The Expert Group on Refugee, IDP and Statelessness Statistics is a UN Statistical Commission mandated, multistakeholder group that works with National Statistical Offices, international organizations and civil society to develop and support implementation of international standards and guidance to improve official statistics on forcibly displaced and stateless persons.

The Compilers' Manual

The Compilers' Manual offers clear operational instructions on producing official statistics on refugees, asylum seekers, IDPs and related populations. It complements the content of the International Recommendations on Refugee Statistics and the International Recommendations on Internally Displaced Persons Statistics by providing hands on guidance.

Access the complete version of the Compilers' Manual'



TABLE OF CONTENTS

| Sample surveys – general background | 4 |
|---|----|
| Questionnaire design | 4 |
| Sampling considerations when surveying refugees | 13 |
| Fieldwork planning | 20 |
| Enumerator training | 22 |
| Fieldwork in practice | 23 |
| Data quality | 24 |
| Data protection / confidentiality / disclosure | 24 |
| Analysis and dissemination | 25 |

Sample surveys – general background

- 56. Sample surveys are an important data source for measuring the characteristics and living conditions of refugees and refugee related populations. When planning a representative survey of refugees, NSOs have the option of including refugees in a broader survey of the national population (e.g. LFS, MICS, DHS, integrated living conditions survey, etc.), or to design a specialized survey for refugee populations. Both approaches have their merit: a specially designed refugee survey has the potential to collect broader and deeper information, whereas inclusion in a national multi-topic survey allows refugees to be compared with other groups in the general population. The two options are not mutually exclusive, and countries may decide to implement a complementary mix of both options. (A detailed discussion of pros and cons of the two approaches, and of sample surveys more broadly, is provided in IRRS paras 156-159 and 176-179).
- **57.** The priorities when considering including refugees in a sample survey are:
 - In the design of the questionnaire, include recommended questions to identify refugees.
 - Identify which indicators of needs and integration are priority for data collection and analysis. If expanding an existing survey, establish whether any are already included in the data collection and add in question sets for any additional indicators.
 - Establish a sampling frame that includes refugees. Where necessary expand the existing sampling frame or find an alternative sampling frame that includes refugees, or the locations where refugees reside.
 - Undertake field visits to test sampling strategies, identify potential issues or concerns and to build relations between the parties involved in fieldwork implementation.
 - Ensure that enumerator training allows sufficient time to cover refugees, including the relevant definitions and concepts, practical survey considerations and data protection.
- **58.** Other than these special considerations during sample design and adjustments of the questionnaire and analysis, there are no further major technical implications from the statistical framework set out in IRRS for surveys of refugees. All the usual principles and best practices for running sample surveys for official statistics apply including on fieldwork security risk assessment and mitigation and are not discussed in this manual.
- **59.** Many of the issues discussed in this Use Case are similar to the considerations when including IDPs in a sample survey (Use Case C) and so are repeated in Use Case C to ensure both Use Cases serve as stand-alone resources.

Questionnaire design

- **60.** Designing an effective questionnaire is fundamental to the success of any survey. When considering including refugees in an existing or new survey the key factors are to ensure appropriate identification and classification questions are included and the question flow in the sections covering places of residence and migration status works accurately given the additional complexity. Typically, questions on refugee status should be introduced in the flow of questions on migration status.
- **61.** Wherever possible, when including refugees, survey development timescales should allow additional time for questionnaire development, to provide an opportunity to bring together relevant stakeholders and seek their input to the survey design. Investing time to engage

stakeholders (for example, relevant government agencies, key communities, organisations supporting refugees, UNHCR) early in the project can help ensure the success of the survey by bringing together relevant knowledge and expertise and appropriately accounting for any sensitivities. It also helps raise awareness of the data that will be collected and promotes its use when the survey is complete.

Refugee identification and classificatory questions

- **62.** The IRRS recommends questions for identifying refugees and refugee related populations in household surveys (IRRS paragraph 183) that are the same as those discussed in Use Case A for censuses:
 - a. Country of birth
 - b. Country of citizenship
 - c. Acquisition of citizenship
 - d. Year or period of arrival in the country
 - e. Reason for migration, with response categories:
 - » Employment (including military service)
 - » Education and training
 - » Marriage, family reunification or family formation
 - » Forced displacement (refugees, asylum seekers, temporary protection status, others)
 - » Other
- **63.** In addition to these core identification questions, IRRS recommends the following classificatory variables in collecting and compiling refugee statistics. Survey practitioners will need to assess which of these are relevant and can realistically be obtained through a survey interview in a given context:
 - a. Country of previous or last residence
 - b. Date of first displacement/leaving previous country of habitual residence
 - c. Parents' refugee statuses
 - **d.** If an unaccompanied child (under the age of 18 years and separated from both parents or legal guardian)
 - e. Legal residential/international protection status as applicable to the national context:
 - » Persons in need of international protection
 - prospective asylum seeker
 - asylum seeker
 - individual refugee status
 - prima facie refugee status
 - subsidiary or complementary protection status
 - temporary protection status
 - refugee resettled in a third country
 - refugee-like (state form of protection granted)

- » Persons with a refugee background
 - naturalised former refugee
 - · child born of refugee parents without own refugee status
 - reunified family member
 - others with refugee background
- » Persons returned from abroad after seeking international protection
 - repatriated asylum seeker
 - returning from other form of international protection
 - returning from other international protection
- **64.** EGRISS has published a methodological paper, *Towards a standardized approach to identify IDPs, refugees and related populations in household surveys*¹ that expands on the IRRS recommendation and suggests survey question flows for use in household surveys. The paper discusses the questions that need to be included in order to address each element of identification of refugees and related populations; the limitations and potential sources of inclusion and exclusion error; and the practical considerations around administering the questions within a survey. It also provides question flows for inclusion in MICS and DHS.

Capturing indicators of needs and integration, and priority SDG indicators

- **65.** Beyond basic socio-demographic variables (age, sex) and the identification and classificatory questions discussed in the previous section, IRRS also proposes a common set of indicators that can measure the satisfaction of refugees' needs and the progress towards their integration on a global basis. Most of these indicators can be obtained through surveys and should be considered for inclusion in a refugee survey as appropriate and feasible. It is also important to consider collecting GIS data where possible, to enable locational analysis.
- **66.** In addition to the indicators of needs and integration recommended by IRRS, the IAEG-SDGs in collaboration with EGRISS has highlighted 12 SDG indicators that are of particular relevance to refugees. For survey practitioners, this implies that these indicators should be captured in any survey including refugees (whether this is a national survey that includes refugees or a refugee-specific survey) if applicable and feasible within the scope of the planned survey. The 12 indicators overlap with IRRS's indicators of needs and integration. Measurement of these SDG indicators for refugees should be aligned wherever possible with the national statistical practice for capturing these indicators among the wider national population, to ensure comparability.
- **67.** Table B.1 outlines the IRRS indicators and the 12 SDG indicators (highlighted orange). EGRISS does not provide detailed definitions of the indicators but rather refers to the relevant international statistical standards for definitions and questionnaire specifications, such as the SDG metadata sheets². If including refugees in an existing survey, it is likely that some of these indicators are already included and the scope for adding additional indicators may be limited. If launching a new survey of refugees, Table B.1 will help inform which indicators to measure.

¹ EGRISS Methodological Paper Series (2023) <u>Towards a standardized approach to identify IDPs, refugees and related</u> populations in household surveys

² See IRRS paras 320-445 for a detailed discussion of the indicators, other than SDG indicators.

68. It will not be possible to cover all the topics listed in a single survey and the actual content of any survey should be determined by local priorities and users' needs. The identification of an appropriate comparator population is necessary, and this is discussed further in paragraph 78.

Table B.1: Priority indicators of satisfaction of immediate and ongoing needs and of integration (excluding classificatory variables). The 12 SDG indicators highlighted by IAEG-SDGs as particularly relevant to refugees are highlighted (red).

| Торіс | Indicators | Relevant references |
|-----------------------------------|--|---|
| Demographic and mi | gration | |
| Family and household situation | Marital status, Composition of the family, Information on absent family members. | |
| Civil | | |
| Legal access to labour market | | |
| Right to own property | SDG indicator 1.4.2: Proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure, by sex and by type of tenure. See following table. | SDG indicator metadata sheet |
| Eligibility to state benefits | SDG indicator 1.3.1 Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable | SDG indicator metadata sheet |
| Access to justice | SDG indicator 16.3.1 Proportion of victims of violence in the previous 12 months who reported their victimization to competent authorities or other officially recognized conflict resolution mechanisms | SDG indicator metadata sheet This can be a relatively small number compared to the overall sample of the surveyed individuals, and difficult to report if below a defined threshold of unweighted cases. |
| | 16.9.1 Proportion of children under 5 years of age whose births have been registered with a civil authority, by age | SDG indicator metadata sheet |

| Education | | |
|---|--|---|
| Educational attainment | SDG indicator 4.1.1: Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex. See following table. | SDG indicator metadata sheet Collecting data for this indicator requires detailed learning assessments at the three levels specified in the indicator, which in most countries are administered through participation in various national, regional and/ or international school-based assessments (as opposed to household-based surveys, which often will not sample sufficient numbers of children at a required age and education level). Inclusion of forcibly displaced children in these learning assessments is the preferred data source for this indicator, although if such an approach is likely to yield a small number of refugee children then an alternative approach should be considered. It will be important to ensure that reading proficiency tests are available in the language the child feels most comfortable reading. Countries wishing to include a learning assessment in a household survey covering forcibly displaced children should align the assessment method in the survey with the relevant school-based assessments that a country carries out. If not applicable, the MICS (Link, see "Foundational Learning: module") and EGRA&EGMA may also provide useful survey-based approaches relevant to the lower schooling levels. |
| Participation in education | SDG Indicator 4.1.2: Completion rate (primary education, lower secondary education, upper secondary education) | SDG indicator metadata sheet This indicator is often calculated by including education modules in household questionnaires. |
| | SDG Indicator 4.5.1: Parity indices (female/male, rural/urban, bottom/ top wealth quintile and others such as disability status, indigenous peoples and conflict-affected, as data become available) for all education indicators on this list that can be disaggregated | SDG indicator metadata sheet |
| Participation in preschool education and access to childcare | SDG Indicator 4.2.2: Participation rate in organized learning (one year before the official primary entry age), by sex | SDG indicator metadata sheet |
| (Host country) language proficiency | | |

| Economic | | |
|-------------------------------|---|--|
| Employment status | SDG indicator 8.5.2 Unemployment rate, by sex, age and persons with disabilities | SDG indicator metadata sheet. See ILO's model questionnaires (Link) for labour force surveys which include the required questions for capturing unemployment, as well as UNHCR's guidance to labour measurement in forced displacement contexts (Link) for a more specific discussion. |
| Informal sector employment | SDG indicator 8.3.1 Proportion of informal employment in total employment, by sex | SDG indicator metadata sheet. See ILO's model questionnaires (Link) for labour force surveys which include the required questions for capturing informality, as well as UNHCR's guidance to labour measurement in forced displacement contexts (Link) for a more specific discussion. |
| | | Informal employment refers to people that in their main or secondary job were in one of the following categories: |
| | | Own-account workers, employers, members of producers' cooperatives employed in their own informal sector enterprise |
| | | Own-account workers engaged in goods production for their own final use |
| | | Contributing family workers (no explicit written contract of employment, not subject to labour legislation or social security) |
| | | Employees holding informal jobs |

| Economic - continued | d | |
|---|--|---|
| Income and consumption | SDG indicator 1.2.1 Proportion of population living below the national poverty line, by sex and age | SDG indicator metadata sheet. Poverty measurement is a complex undertaking. No internationally standardized questionnaire module for capturing consumption expenditure exists (since food and non- food items consumed vary from country to country). When comparing consumption aggregates against a national poverty line, it is crucial to capture consumption in a way that aligns with the primary poverty survey instrument used in the country from which the poverty line was derived. This paper documents best practices around survey-based consumption and poverty measurement. |
| | | Compute a comprehensive estimate of total household consumption or income (with consumption being the preferred welfare indicator) and construct a correctly weighted aggregate of consumption or income per person. Compare this aggregate with the national poverty line. |
| | | To measure consumption, cover questions on food and non-food expenditure as well as food consumed from own production (either through recall questions using lists of consumption items or through diaries) |
| | | An alternative approach may be to use the Wealth Index constructed from data on asset ownership collected in DHS, or similar through MICS |
| Youth not in education/training/ employment | SDG indicator 8.6.1 Proportion of youth (aged 15-24 years) not in education, employment or training. | SDG indicator metadata sheet |
| Access to land for growing food | See "Right to own property" above. | |

| Social inclusion | | |
|--|---|---|
| Poverty | See "income and consumption" above. | |
| Material deprivation. This is similar to poverty, but while poverty is generally measured using information on income and consumption, material deprivation is measured by the lack of access to basic necessities. These can include food, water, sanitation, and housing, as well as some durable items. | Indicator 1.2.2: Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions | SDG indicator metadata sheet An alternative approach may be to use the Wealth Index constructed from data on asset ownership collected in DHS, or similar through MICS |
| Housing conditions | SDG indicator 11.1.1: Proportion of urban population living in slums, informal settlements or inadequate housing | SDG indicator metadata sheet |
| | SDG indicator 7.1.1 Proportion of the population with access to electricity | SDG indicator metadata sheet. A single survey question on primary source of lighting is sufficient to capture this indicator. See question HL2 in the broader set of energy- related questions recommended by WHO and the Word Bank (Link). Sources of electricity such as generators, candles, batteries, etc., are not considered due to their limited working capacities and their function as backup sources for lighting. |
| | SDG indicator 16.1.4 Proportion of the population that feel safe walking alone around the area they live | SDG indicator metadata sheet. Survey questionnaire specification provided directly in the "methodology" section of the metadata sheet. This is also available in the victimization module in MICS for both men and women 15- 49 years of age. |
| Overcrowding | See "housing conditions" above. | |

Health

| Self-reported health status | | |
|--------------------------------|---|---|
| Essential health | SDG indicator 3.8.1: Coverage of essential | SDG indicator metadata sheet |
| services | health services | The indicator is an index reported on a unitless scale of 0 to 100, which is computed as the geometric mean of 14 tracer indicators of health service coverage. |
| | | Household surveys collect about half of the indicators that go into this composite index. Some of the indictors listed are Level 2 indicators in IRRS, so they could be collected even if the composite indicator is not calculated |
| | | MICS collects Use of tobacco (3.a.1), ANC coverage (3.8.1) need for family planning (3.8.1) DTP coverage (3,8.1) PNV coverage (3.b.1) measles (3.b.1) and care seeking for pneumonia. |
| | | This means different indicators need to be collected with long questionnaires to estimate this composite one, and they include a variety of indicators from tobacco use, to vaccination for children up to 3 years of age and care seeking for pneumonia, which usually has very low denominator as only children that have had a disease episode within two weeks before the survey are taken into consideration. |
| | 6.1.1 Proportion of population using safely managed drinking water services | SDG indicator metadata sheet. For survey questionnaire specification, see table 1 and "core questions" W1-W5 of the JMP recommendations (Link) |
| | | The indicator also requires W6 from the JMP core questions, which includes E.coli testing for each household. This requires procurement of equipment and training ³ . |
| | 2.2.1 Prevalence of stunting among children under 5 years of age | SDG indicator metadata sheet. In terms of survey questionnaire design, only fields for age/DOB and height are required. However, it must be noted that height measurement is non-trivial. See guidance on anthropometric data collection from DHS (Link), MICS (Link) and the FANTA Project (Link). |

³ Basic information on the E.coli testing process is available through the <u>UNICEF website</u>.

| 3.1.2 Proportion of births attended by skilled health personnel | SDG indicator metadata sheet The relevant question in DHS and MICS reads: 'Who assisted with delivery of (NAME)?' and answer choices include both health personnel and other persons. See figure 1 of <u>this article</u> for a direct comparison and discussion. MICS enquires about the last birth during the past 2 years, whereas DHS enquires about all births during the past 5 years. |
|---|--|
|---|--|

69. Examples of questionnaires from other refugee surveys are available in the UNHCR microdata library⁴.

Sampling considerations when surveying refugees

- **70.** In theory, sampling in forced displacement contexts is no different from sampling elsewhere, and the same general principles and methods apply⁵. However, in practice the selection from and application of these methods is often faced with specific challenges when it comes to refugees (IRRS paras 163-175).
- 71. Forcibly displaced persons may be hard to reach, either because they are difficult to identify, not willing to be identified, on the move or live in areas difficult to access. Moreover, displaced people often constitute a small proportion of the wider population and are not evenly distributed around a country or region making it difficult or impossible to rely on sampling frames which don't allow for the direct identification of forcibly displaced households, or at least of small geographical areas where they are predominantly located. Refugees can live either in dedicated camps or they can live in houses among the general population, either as part of another household or as an independent household.

Identifying a sampling frame

- 72. The first question that guides the decision on sampling from refugee populations is how to construct a frame from which to sample refugees in a targeted manner. In the simplest case, this can either be a list of the wider population which allows identifying forcibly displaced, or a list solely describing the refugee population. Such a list can often come from administrative records, either from relevant government agencies or operational data from international agencies such as UNHCR. In order to gain access to such sources it is likely that a data sharing agreement will be required, which can take some time to establish and should set out how data privacy will be protected for example, by basing sample selection on anonymised data, so that only the personal details of those invited to take part in the survey are shared.
- **73.** If one does not have an existing list of possible respondents, an area-based frame needs to be created. The following methods are commonly used:

⁴ UNHCR Microdata Library

⁵ Useful reading on survey sampling in general can be found in UNSD's "<u>Designing Household Survey Samples: Practical Guidelines</u>", "<u>Household Sample Surveys in Developing and Transition Countries</u>", and "<u>Sampling Rare and Elusive Population</u>".



QCASE STUDY: ADAPTIVE CLUSTER SAMPLING IN UGANDA AND ETHIOPIA

In 2022, Fafo Institute for Labour and Social Research undertook a survey to understand economic development and integration of refugee communities in urban settings in Uganda and Ethiopia, commissioned by the World Bank. The aim of the survey was to obtain detailed information that would help explore labour market outcomes for refugees and host communities living side by side and engaging in a shared labour market settings. The surveys were carried out with support from the NSI, who provided the initial list of PSUs.

Due to the difficulty of surveying refugees in urban settings where they tend to be in smaller proportions compared to the host population, the survey used an adaptive cluster sampling approach to obtain a sufficiently large, representative sample of refugees. The procedure begins by constructing the list of households in the initial randomly selected PSUs. The listing process asked only for the name of the head of household (for identification purposes), the number of household members and nationality. Nationality proved to be an effective proxy for refugees. A threshold of 10% of non-national households in a PSU was set for the listing exercise to be extended to all nearest neighbour PSUs. Theoretically the listing could continue to all neighbours so long as the threshold criteria is fulfilled. In the Fafo survey, a single neighbour listing was conducted for cost reasons but resulted in sufficient number of refugee households.

a. If the locations in which refugees reside are known, and the National Statistics Office maintains a list of Primary Sampling Units (PSUs) for the country, the two can be used to identify those PSUs where refugees reside. The first stage of sampling can then include either only those PSUs, or over-sampling from those PSUs. Relevant PSUs are selected randomly or systematically, and then a household listing is carried out or updated in the selected PSUs, capturing household's refugee status⁶.

⁶ Thematic report: Data disaggregation of SDG indicators by forced displacement, by Carlos Barahona-Zamora with the support of Emily Nevitt, Alex Thomson, and Ciara McHugh at Stats4SD, and Andrea Pellandra at UNHCR November 2020, <u>Data disaggregation of SDG indicators by forced displacement</u>.

Compilers' Manual on Forced Displacement Statistics



Q CASE STUDY: SAMPLING IN A CHALLENGING SECURITY SITUATION IN SOMALIA

In a survey in Somalia, it was not feasible to conduct a full listing of all households in an enumeration area, as this was too time-intensive and may have raised suspicion. Instead, a micro-listing approach was used, which required enumeration areas to be segmented into smaller enumeration blocks using satellite imagery. Enumeration blocks were small enough for enumerators to list and select households immediately before conducting the interview.

Source: https://openknowledge.worldbank.org/bitstream/handle/10986/31267/WPS8735.pdf

- b. A similar approach can be applied if the locations of refugees are unknown, by drawing a random selection of PSUs and undertaking a listing exercise, using a very short set of questions, for example with nationality as a proxy for refugee status. In those PSUs with higher concentrations of refugees above a pre-determined threshold such as 10% the household listing is extended to the neighbouring PSUs (or one nearest neighbour PSU). This process continues extending until PSUs no longer cross the threshold. In this way, PSUs with the highest concentration of refugees are identified and the listing can be used as the sampling frame. The effectiveness of this adaptive cluster sampling approach⁷ relies on the settlement distribution of refugee households in urban settings. When refugees are settled in clusters with enough concentration, adaptive cluster sampling is a useful method to efficiently and effectively sample refugee populations.
- c. If a list of PSUs of reasonable quality does not exist, alternative ways to identify smaller geographical areas to sample from exist. Satellite, aerial or drone photos of areas in which refugees are known to reside can be used to divide geographical areas into sampling units to be sampled and the households existing there listed to be sampled in a second stage.
- d. If refugees' locations are (partially) unknown, Mobile Phone Tracing may be an option. Digital trace data generated by refugees' mobile phones can help identify the areas in which refugees are located and thus the geographical areas to list/sample from. This is not the same as identifying refugees' phones and then sampling them for a phone survey (see section (e) for comparison). Approaches such as these naturally require careful consideration around data privacy. Further discussion of non-traditional data sources is included in Use Case F.

⁷ Thompson, Steven K. "Adaptive Cluster Sampling." Journal of the American Statistical Association 85, no. 412 (1990): 1050–59. Adaptive Cluster Sampling.



QCASE STUDY: SURVEY OF NEWLY ARRIVED VENEZUELANS IN PERU

National authorities in Peru, supported by the UN inter-agency coordination platform, initiated ENPOVE - a household survey on the living conditions and experiences of newly arrived Venezuelan migrants and refugees in 2018. To overcome the challenge of not having a dedicated sampling frame for the survey, the Peruvian National Statistical office identified areas with high concentrations of Venezuelans based on the findings of the national population and housing census conducted in late 2017. This helped narrow down the geographical scope of the survey to the national capital and five provinces that were home to an estimated 85% of the Venezuelan population. Those areas were then oversampled in a conventional two-stage design.

In 2019, with a view to expanding the sampling options of the ENPOVE in light of a high and spacially dispersed influx of Venezuelans since the 2017 Census, a Mobile Phone Tracing exercise was carried out to establish alternative ways of identifying areas with high concentration of Venezuelans. Based on 15 million anonymized call detail records obtained from a major telecommunication company, mobile phones were tagged as likely to belong to a Venezuelan, if they were (a) registered under the name of a Venezuelan national, or (b) were used at least 30 times over the past month to make/receive a voice call or text message to/from Venezuela; or accessed a website of interest to Venezuelans. Using georeferencing, these phones were then associated with census PSUs – providing an additional indication of Venezuelans present in these PSUs and thereby complementing the census data from 2017.

Source: <u>https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1666/</u> <u>libro.pdf and https://openknowledge.worldbank.org/bitstream/handle/10986/34175/Big-Data-</u> <u>for-Sampling-Design-The-Venezuelan-Migration-Crisis-in-Ecuador.pdf</u>

e. Moving away from area-based sampling approaches, it may be possible to undertake a phone survey, assembling a list frame of phone numbers of the refugee population, by Mobile Phone Tracing or via Random Digit Dialling⁸, or to use phone numbers collected through previous surveys. If launching a face-to-face refugee survey, it may also be worth collecting phone numbers and consent for future phone-based panel surveys at the same time.

⁸ The usual risks of sample undercoverage resulting from phone ownership and network connectivity rates among the target population need to be taken into account.

AFGHANISTAN

QCASE STUDY: CREATING A SAMPLE BASED ON CALLING RANDOM MOBILE PHONE NUMBERS IN AFGHANISTAN

Over two million Afghan refugees had returned to their homes between 2014 and 2017 as a result of the gradually stabilizing situation in their country. Limited information was available on the conditions of those returnees and the hazardous security situation in many parts of the country made it difficult to send enumerators to the field to collect data for a household survey. This prompted the World Bank to explore the feasibility of conducting remote interviews through phone surveys given the rising trend in mobile phone ownership rates within the Afghan population. The sampling frame was constructed through a Random Digit Dialling (RDD) process to ensure that all mobile phone owning households had an equal probability of inclusion in the sample. Once a household was contacted, a short automated Interactive Voice Response (IVR) program would screen the respondent for basic demographic information and displacement status to confirm their eligibility for the survey. Some 3,254 returnees were identified through the process who were then contacted by a live enumerator to complete the main survey questionnaire. Validation of the sample design and weight calibration was done by comparing sample composition and characteristics to data from a census of Afghan refugees in Pakistan and records maintained by humanitarian agencies involved in facilitating the return of Afghans from Pakistan.

Source: <u>http://documents1.worldbank.org/curated/en/298881533562809348/pdf/Afghanistan-</u> World-Bank-Phone-Survey-of-Afghan-Returnees-Methodology-and-Representativeness.pdf

74. Sometimes a combination of different approaches is required. For example, if expanding a national survey to include refugees living in camps and outside camps, it is possible that PSUs will be available but that they exclude the camps. In this case, it may be necessary to use registration lists of refugees in the camps as a complementary sampling frame.

Drawing a sample

- **75.** Constructing a sampling frame of refugees allows a survey to apply probability-based sampling methods. This ensures that every unit in the population has a known, non-zero chance of being selected into the sample, and its probability of selection can be accurately determined. This makes it possible to produce unbiased estimates of population totals, by weighting sampled units according to their probability of selection. It also allows for estimates of uncertainty, like confidence intervals. Probability-based sampling is considered best practice when sampling for surveys and ought to be the first choice whenever it is possible to construct a sampling frame, i.e. a list of units to draw a sample from.
- 76. Including refugees in a general household survey is likely to require a boost to the sample ratio for refugees to ensure that a sufficient number are reached through the survey because refugees tend to be a small proportion of the overall population. The size of a boost will depend on the analysis requirements for example the disaggregations and level of precision that are expected.

Compilers' Manual on Forced Displacement Statistics

UGANDA

A boost was applied during the 2018 Uganda DHS survey to ensure that sufficient data on refugees would be collected. A boosted sample for host communities was also needed, to ensure the sample was still representative at the regional level (the regular DHS sample is not representative at the district level but only at the regional level). This was done in a way that was consistent with the selection protocols of the primary sample while enabling the production of robust estimates at the level of the host community. The envisaged sample size for refugees and the host community booster sample for the Northern stratum is shown below:

| | Northern stratum | | |
|------------------|--|---|------------------------------------|
| | # of sample households in regular DHS | # of sample households added via boost | Total # of households in domain |
| Refugees | 0 | 1,200 | 1,200 |
| Host communities | 690 | 510 | 1,200 |

- 77. If sampling from highly concentrated population such as refugees living in camps, it is generally not necessary to consider clustering the sample. Clustering is a technique that one uses to bring down sample dispersion and make travel during fieldwork manageable, but it comes at the cost of losing precision and driving up design effects. When sampling from a highly concentrated population there is no real cost saving from clustering but there is still a loss of precision, so it is preferable to obtain a registration list for the camp (if it exists and is sufficiently up to date) and use it as a complementary sampling frame from which to draw a systematic and unclustered selection of dwellings.
- **78.** If carrying out a dedicated survey of refugees, it is also important to identify an appropriate "host community" to include in the sample as a comparator population. There is no standard definition of host community, so at present this will need to be locally determined. Examples include the wider population that lives within a certain radius of a refugee camp, or households within the same administrative area.
- **79.** It should also be noted that responses to questions on attitudes, intentions to move, and decision-making power in the household may differ substantially by gender. When feasible and appropriate, using a random selection protocol of female and male respondents within a sampled household, or interviewing more than one member of each household, are ways to ensure more reliable data. The decision on eligible household members to be included in the survey will also need to reflect which variables are to be collected, since the eligible respondents vary across different SDG indicators, for example.

Non-probability sampling methods

80. When constructing a sampling frame is not feasible, reverting to non-probability-based sampling is sometimes considered. Commonly used non-probability sampling techniques are respondent driven sampling (RDS) or snowball sampling. If refugees are difficult to find, either because they are few in numbers, are hiding, are homeless, on the move or live in an area difficult to access, such methods may be considered as an option of last resort. They should however be used with caution, as they do not allow for statistically representative, unbiased inference from the sample to the population. Application of non-probability sampling methods in official statistics production is uncommon.

Limitations of registration systems as sampling frames

- **81.** If a potential sampling frame exists from registration records, a frequent challenge is that many registration systems group individuals into "cases" or "registration groups", which are different from a household as defined in standard survey methodology (which is built around the concepts of shared dwelling and housekeeping arrangements). A detailed discussion of the issue as well as possible solutions is provided in the UNHCR guidance note titled "Resolving the 'case versus household' issue when conducting a Socioeconomic Assessment"⁹.
- 82. In a dedicated survey of refugees or IDPs that includes those both inside and outside camps, consideration needs to be given to the relative accuracy of sampling frame sources. It is possible that registration systems for those living in camps will be more accurate and complete than for those living outside camps, and respondents will be easier to reach inside camps than outside introducing a risk of non-response bias. Conversely, another frequent challenge with sampling frames derived from refugee registration systems is that address/contact information can be outdated.

Box B.1

Guidance material on sampling in forced displacement contexts

Two dedicated resources on sampling in forced displacement contexts have recently been published and may be considered for further details and examples:

- Stephanie Eckman and Kristen Himelein (2022): <u>Innovative Sample Designs for Studies of</u> Refugees and Internally Displaced Persons
- Joint IDP Profiling Service (JIPS, 2020): <u>Sampling Guide for Displacement Situations and</u> <u>Practical Examples</u>

⁹ UNHCR proGres Database: Handling differences between cases and households

Fieldwork planning

- **83.** Before any survey fieldwork commences, it is advisable to conduct one or more field visits to fully understand and plan for the locations and practicalities involved in reaching refugees. This can be important in overcoming the additional challenges that surveying refugees can entail. Ideally one or more of the most challenging locations / scenarios should be visited, with the following objectives:
 - Observe the situation on the ground;
 - Test whether the sampled households can in practice be located and accessed and make contingency plans if not, for example through engaging local community leaders, local authorities or landlords for access;

ETHIOPIA

CASE STUDY: FIELD VISITS FOR THE SOCIO-ECONOMIC SURVEY OF REFUGEES IN ETHIOPIA

In 2022 Ethiopia launched a national refugee survey—the Socio-Economic Survey of Refugees in Ethiopia (SESRE) which was funded by the Joint Data Center—using the same methodology and content as its national household poverty survey, the Household Welfare Statistics Survey (HoWStat). The objective of SESRE was to sample refugees in most refugee camps across Ethiopia and implement it, for the first time, through the Ethiopian Statistics Service (ESS). In order to prepare for the survey, a team from ESS undertook field visits to 3 camps with the following objectives:

- a. Better understand the situation in the camps with respect to how camps are set up;
- Better understand the challenges around identifying sampled households in the camps;
- c. Discuss the objectives of SESRE with camp management and refugee community leaders;
- d. Engage refugee community leaders to start

information campaigns on SESRE in the camps; and

e. Link ESS field teams with the field office teams of the Refugees and Returnees Service who is responsible for managing refugee camps in Ethiopia.

UNHCR and RRS supplied the sampling frame based on the most up-to-date registration lists, but thanks to the field visits, it quickly became apparent that the registration lists were outdated and did not contain sufficient information to locate respondents. Based on information from the field visits, a new approach to locate respondents was defined which relied on the knowledge and contacts of refugee community leaders inside the camps to locate respondents rather than location information from registration lists. Moreover, relationships between different government entities (ESS and RRS) at the survey locations could be established early to ensure full collaboration across institutions. Refugee community leaders received information and an understanding of SESRE before implementation was started which was helpful in receiving information which improved survey implementation. The knowledge from field visits was further built into the training of enumerators and ensured a more efficient and effective way to successfully roll out the survey.

- Build relations between the different agencies and communities involved in implementing the survey trust and rapport can be essential to success of the survey;
- Identify any language barriers and how best to address them (questionnaire translations, availability of interpreters).
- Test the questionnaire, both in terms of content and flow. Do the questions lead to correct identification and classification of refugees; are there any issues with translations; are there any sections of the questionnaire that are seen as more sensitive and potentially might result in non-response.
- **84.** It is advisable to allow additional budget for overcoming language barriers, including translation of questionnaires; hiring enumerators with relevant languages; and increased availability of interpreters. There may also be additional costs for extended duration of fieldwork compared with a standard survey, reflecting the challenges of identifying refugees, and potentially greater use of supervisors to train, monitor and support enumerators.
- **85.** Surveys may involve a communication plan at a national or local level, to ensure that residents are aware the survey is taking place and primed to take part at the correct time. It is important to ensure that displaced people are considered as part of the communication plan, if good coverage is to be achieved. There are no fixed prescriptions that will work everywhere, but experience from various countries should be observed and evaluated. The main messages to convey include:
 - That the survey will include displaced people and refugees and that their responses are important.
 - How the data will be used (positive messages) and clarity that it is not linked to allocating assistance.
 - That confidentiality and data security are assured
 - Any special arrangements that have been made to include displaced people (such as availability of interpreters).
 - What opportunities there will be for respondents to hear about the survey results following data collection. Depending on the scale of the survey, it may be appropriate to offer some dissemination aimed specifically at the target population.

The communication plan may also consider raising awareness of any existing referral mechanisms available to the community, for feedback or protection concerns, according to the principles of "do no harm".

- **86.** To reach displaced people, communication at local and community level is often the most effective. Examples of communication strategies include:
 - Leaflets or visits to relevant institutions (refugee camps, etc) in the lead up to Census, with interpreters available if language is likely to be a barrier.
 - Identification of and engagement with community leaders before the survey, to establish the most effective way of reaching displaced populations and to gain their buy-in and support.
 - Acknowledge the potentially sensitive nature of the question topics for displaced populations and explain the purpose of collecting the data and principles of data protection.
- **87.** If it is possible to include refugees within the survey workforce, for example as enumerators, there can be many benefits, including greater success reaching and effectively communicating with the refugee community.

Enumerator training

- **88.** Displaced people are often harder to reach and can need support in order to take part in a survey and this requires specific content in the enumerator training course. Allowing sufficient time to address these issues in the enumerator training is key to delivering a successful survey. It may take between one and two additional days to cover the appropriate material.
- **89.** Topics to cover in enumerator training include:
 - The definitions and concepts used in statistics about displacement, which may be unfamiliar to many experienced field staff. For example, refugee concepts such as habitual residence, usual residence, migration concepts, refugee categories, meaning of displacement, meaning of refugee and IDP, and reasons for displacement and migration.
 - Accurate recording of all members of the household, including temporary members who may be displaced. This is important as displaced people may fail the interviewers standard understanding of 'usual members of the household', leading to them being inaccurately treated as visitors and excluded from household rosters.
 - The sensitivities around interviewing refugees and the data protection implications. When conducting surveys, the displaced population has special needs or conditions of interview, and the 'no-harm' principles and UN ethical standards on data collection and data protection must be applied. Enumerators must be aware of the existing referral mechanisms for community feedback or protection concerns, should any issues present themselves during the course of the data collection.
 - Building trust and rapport, and skills in interviewing vulnerable people, such as children and disabled people. People who have been displaced have suffered trauma and any interview may ask people to relive painful experiences.
 - Strategies for managing personal well-being for the enumerators, if they are likely to be exposed to traumatic testimonies during the data collection which can become emotionally exhausting.





CASE STUDY: TRAINING ENUMERATORS FOR THE SOCIO-ECONOMIC SURVEY OF REFUGEES IN ETHIOPIA

The 2022 Socio-Economic Survey of Refugees in Ethiopia (SESRE) was implemented, for the first time, by the Ethiopia Statistics Service (ESS). Though ESS has much experience carrying out national household surveys, it was the first time, they surveyed refugee populations. To ensure sensitivities and special protection needs are taken into consideration, the two week training of enumerators included one day specifically focused on the intricacies of collecting data from refugee populations. This content was taught by representatives from the government's Refugees and Returnees Service (RRS) and UNHCR, and aimed at providing:

- Definitions and concepts that are specific to refugees;
- b. Overview of the situation of refugees in the country as well as how camps are organized and managed;
- c. Practicalities of enumerating refugees in large camps;
- d. Advice on locating respondents;
- e. Advice on effectively communicating the purpose of the survey to avoid creating expectations by respondents;
- f. Information on special protection needs of respondents.

The feedback received from enumerators and ESS staff showed that the time invested in these specific aspects of the training was instrumental in providing enumerators with the right context and sensitivities to successfully implement the survey.

Fieldwork in practice

- **90.** On the whole, provided fieldwork planning and enumerator training has taken full account of the need to cover displaced people, there are few issues specific to refugees in terms of the actual fieldwork. However, it is worth noting that effective feedback from enumerators during any survey can be crucial in identifying any emerging issues in reaching refugees, which might be addressed by continued or improved communication and community engagement.
- **91.** If budget allows, it is advisable to undertake more monitoring of fieldwork than for a standard survey, reflecting the challenges associated with identifying refugees and potential issues overcoming sensitivities around data collection. Increased availability of supervisors can help to monitor the quality and consistency of data collection and play a facilitating role, communicating the purpose of the survey with more authority, to gain trust and avoid non-response.

Data quality

- **92.** There are few data quality issues that are specific to surveys including refugees, over and above other surveys in challenging locations, and so data quality assurance should follow standard processes. The only additional data quality risk is around the accuracy of survey responses if respondents consider that their responses will influence their refugee status or outcomes in some way. This should be mitigated through accurate communication and good enumerator training.
- **93.** Often it is possible to triangulate survey data with data from other sources, such as administrative data from other government agencies or operational data produced during the regular activities of humanitarian organisations, to make sure that high level trends match between the various available datasets without any major deviation. For example, if the variation between two stocks is not significant, but administrative data identifies a high number of smaller, shorter term movements (e.g. in the case of preventative evacuations leading to quick returns, or mass movements leading to cross border returns) then this should help to verify or correct official statistics on flows or alert the need for better or more frequent data collection. Similarly, through comparing with cross border data, if movements lead to significant cross-border flight or returns, the datasets on both sides of the border should correspond with each other to ensure validity. Further discussion of administrative data, operational data and non-traditional data sources is included in Use Cases D, E and F respectively.

Data protection / confidentiality / disclosure

94. Due consideration should be given to privacy in accordance with national and international laws. The UN Fundamental Principles of Official Statistics states that "individual data collected by statistical agencies for statistical compilation ... are to be strictly confidential and used exclusively for statistical purposes"¹⁰. Disclosure should also include considerations of interoperability: how datasets on different population types talk to each other, their ability to share information on those populations, or even on specific individuals, their characteristics and their needs.

¹⁰ United Nations Principles and Recommendations for Population and Housing Censuses Revision 3.

Analysis and dissemination

- **95.** The process of analysing and presenting the survey data collected on refugees is no different from that for any other data and should be planned to meet the needs of users. This section briefly describes the outputs that are commonly produced to meet users' needs, with references for further information if required.
- 96. Data collected on identification and classification of refugees is used to produce analysis of stocks and flows for the different categories of refugees and refugee related populations, presented as counts and proportions cross-tabulated by sex and age group. Details are in IRRS paragraphs 114 125.
- **97.** Analysis of stocks is recommended for all three categories of refugee and refugee related populations.
 - a. Counts of each type of refugee and refugee related populations, by sex and age, or age group. If age group is used, then it is important to distinguish children from adults aged 18 years and over, and for some purposes it may also be helpful to identify those under 16 years of age.
 - **b.** Proportions of each type of refugee and proportions who are female. IRRS elaborates on these, drawing out a recommended set of indicators see Box B.2.
- **98.** Flows are inherently more difficult to collect through surveys and in practice the recommended indicators are likely to need to come from administrative data. Where it is possible to collect data, flows are recommended only for those in need of international protection and those who have returned to their country of habitual residence after having sought international protection abroad, not for those with a refugee background.
 - a. Counts of inflows during a specified reference period, for each sub-category of refugees, for those in need of international protection or who have returned to their country of habitual residence after having sought international protection abroad, by sex and age.
 - b. Proportions IRRS elaborates on these, drawing out a recommended set of indicators – see Box B.2.



Box B.2

Refugee stock and flow indicators (from IRRS, 2018)

- i. Stock of refugees in need of international protection (IRRS, 2018 paragraph 115):
 - a. Percentage of the total population who are in a country for international protection.
 - b. Proportion females among persons in a country for international protection.
 - c. Percentage of persons present in the country for international protection who are asylum seekers.
 - d. Percentage of persons present in the country for international protection who have determined status.
 - e. Total number of unaccompanied and separated children under 18 in a country for international protection, by sex and age.
 - f. Proportion of persons in a country for international protection who have remained for up to 5 years.
 - g. Proportion of persons in a country for international protection who have remained for up to 10 years.
 - h. Proportion of persons in a country for international protection who have remained for over 10 years.
- ii. Stocks of persons with a refugee background (IRRS, 2018 paragraph 120):
 - a. Percentage of the total population who have a refugee background.
- iii. Stocks of persons returned after having sought international protection abroad (IRRS, 2018 paragraph 123):
 - a. Percentage of persons returned from having sought international protection abroad among all returned citizens.
 - b. Proportion females among persons returned from having sought international protection abroad.
- iv. Flows of refugees in need of international protection (IRRS, 2018 paragraph 119):
 - a. Total number of asylum seekers who received a decision during a period of time, by sex and age.
 - b. Percentage of asylum applications decisions that are positive (or negative) during a period of time, by sex and age.
 - c. Percentage of determined refugee status granted during a period of time, by type (Prima facie, individual, derivative, complementary/subsidiary, temporary protection), by sex and age.
 - d. Percentage of rejected asylum seekers who left the country during a period of time, by sex and age.
 - e. Recognition rate during a period of time [the added value of the use of longitudinal information in more adequate calculation of the recognition rate should be acknowledged].
 - f. Percentage of persons entered for international protection during a period of time who were resettled elsewhere.



Q CASE STUDY: STATISTICS NORWAY'S GEOSPATIAL ANALYSIS

Persons with refugee background as a percent of municipal population, 1 January 2015.



The best examples of data dissemination are often through online resources, which enable dynamic access to the data and can tell the data story in an engaging and interactive way. This Case Study captures one image from an online resource, to illustrate and inspire. *The full details are available in the following source material: Refugees in Norway* <u>https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/refugees-in-norway</u>

- 99. Geospatial analysis is also recommended, to visualise the stocks and flows.
- 100. There are no specific recommendations relevant to refugees on how other data on socioeconomic indicators for measuring refugees' needs and integration (including SDGs) should be analysed and presented. Typical tabulations could include disaggregation by category of refugee, sex and age group. More complex analysis and commentary might address the integration of displaced people into host communities, the progress made towards overcoming key displacement-related vulnerabilities, and other key policy issues.
- **101.** In planning how the survey analysis will be disseminated, if part of a wider survey it is preferable for refugee analysis to be disseminated in the main survey report as a standard disaggregation but it may also be impactful to produce a separate report focussing just on refugees. There is no right answer here, it will depend how users' needs can best be met in the specific national context. It is also important to consider how survey findings can be shared with the refugee communities themselves. This feedback loop is important: as well as meeting the needs of a key user community it demonstrates what the survey is for and promotes the importance of the survey, encouraging future engagement and participation.

