



ENHANCING UNHCR SOCIOECONOMIC ASSESSMENTS

RESOLVING THE “CASE VERSUS HOUSEHOLD” ISSUE WHEN CONDUCTING A SOCIOECONOMIC ASSESSMENT (SEA)



SEA RESOURCE PACK

This note is part of a resource pack for planning and implementing SEAs.

Other planning aids available include:

- Guide to starting a SEA
- Standardized questionnaire modules
- Template to create a terms of reference for hiring a survey firm

REGISTER YOUR SEA

It takes just 5 minutes and will allow UNHCR's Division of Resilience and Solutions - Research, Analysis and Knowledge Management Unit to offer you **customized support**. Register here:
<https://enketo.unhcr.org/x/#T90Bwg70>

GET IN TOUCH

livelihoods@unhcr.org

Contents

Background	1
The difference between a <i>case</i> and a <i>household</i>	2
Why is this a problem?	2
How can this be resolved?	3
Option 1: Don't use proGres as a sampling frame	3
Option 2: With proGres as a sampling frame: sample cases but interview households, and mitigate the difference	4
Option 3: With proGres as a sampling frame: Individual-level sampling	5
Which option is best for my situation?	7

This socioeconomic assessment planning resource is produced by the Division of Resilience and Solutions – Analytics, Research and Knowledge Management Unit, UNHCR. Questions about planning a socioeconomic assessment or regarding the use of this document and related resources should be written to livelihoods@unhcr.org.

Resolving the “case versus household” issue in Socioeconomic Assessments

Background

In the effort to promote socioeconomic inclusion of persons of concern, UNHCR operations routinely conduct quantitative surveys, often under the label *Socioeconomic Assessments (SEAs)*.¹ These surveys provide the basis for selecting and designing interventions, targeting, measuring impact, and advocacy. Their thematic scope can differ substantially; nevertheless, common topics include demographic characteristics of household members, dwelling characteristics, asset ownership, access to services, food consumption score/groups, coping strategies, debt/finance, assistance received, social participation and perceptions, priority needs, and consumption expenditure (with a view to computing monetary poverty). Currently, around 10-15 SEAs are carried out each year by UNHCR operations globally.

Despite the importance of SEAs and the data they deliver, an internal review of recent SEAs found that most of them were at the weak end of the quality spectrum; and that their quality, comparability and cost-efficiency is severely impeded by a lack of standardized methodological approaches and harmonized guidance material. To help address this shortcoming, this note – as part of a wider resource pack for SEAs – discusses a technical challenge commonly faced by SEAs, namely the difference between a *registration case* and a *household* as usually referred to in survey methodology. It also provides information to help operations select and implement the sampling approach that is most appropriate for their situation.

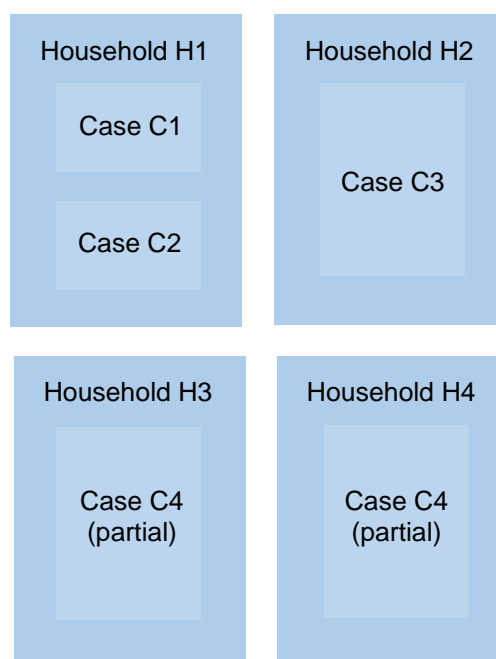
While this resource package is aimed towards socioeconomic assessments, the lessons and approaches apply to a broader range of household surveys, for which the case versus household issues in relevant. This technical note is designed to be shared with either in-house UNHCR technical leads familiar with statistics; or with the relevant survey firm’s technical lead on sampling for the SEA; or with partner agencies who are conducting surveys on forcibly displaced populations. The aim of this note is to provide detailed information on designing a sampling frame for a SEA.

¹ Socioeconomic Assessments (SEAs) were originally introduced in 2015 within the framework of the *Operational Guidelines on the Minimum Criteria for Livelihoods Programming*. While the *Minimum Criteria* are no longer required, the principles that they introduced, including socioeconomic assessments, have been mainstreamed into UNHCR Livelihoods programming and remain widely relevant across operations, including beyond the remit of the Livelihoods sector.

The difference between a *case* and a *household*

UNHCR regularly conducts quantitative surveys of its population-of-concern. Its registration database, proGres, often serves as the sampling frame from which a systematic or random sample is drawn – which is, in principle, a statistically robust approach.

However, the units in the proGres list are *cases*, which are different from a *household* as defined in standard survey methodology (which is built around the concepts of shared dwelling and housekeeping arrangements, often simplified to “people who live together and eat from the same pot”; family ties are irrelevant in this definition).² A “case” is a grouping of people considered together for a specific purpose, usually in relation to a decision or action, such as in status determinations or resettlement. Most of the time, the family, household and cases that an individual belongs to will all be composed of the same persons, however, this may not always be the case, especially as people move and adapt to their new environment.³



It is frequently observed that several *cases*, in particular those comprised of single individuals, join up to effectively form a *household* (living and eating together).⁴ See illustration of households H1 and H2 to the right – household H1 contains two *cases* and creates challenges for survey sampling.

Moreover, one or several members of a *case* may leave the other members to form a new or join another *household*. See illustration of households H3 and H4 to the right – both *households* contain parts of the same original *case* and this also create challenges for survey sampling.

The extent of this problem will differ from survey to survey, but we get an indicative insight from surveys commissioned by UNHCR in Kakuma Camp, Kenya, in 2017 and in Uganda in 2019. Due to concerns about the proGres list frame being out of date, this survey applied an area-frame approach instead, resulting in the sampling units being “true” *households* from a methodological point of view. The survey then checked the number of *cases* in each household. It found that 28% of households contained more than one *case*.⁵ In Uganda, 35 percent of households were found not to correspond to their composition in proGres in a 2019 household vulnerability survey.

² For more on the definition of a UNHCR case, see the [UNHCR Integrated Framework for Household Survey](#) and the UNHCR-World Bank paper on [Improving Targeting and Welfare of the Syrian Refugees \(Annex 1\)](#).

³ See UNHCR Registration Handbook (2003), “[Families, households and cases](#)”.

⁴ Further complicating matters, several households may occupy one dwelling. For a discussion of disentangling data at the dwelling level, see [Creating household data: a conundrum of feasibility](#), VAF January 2015.

⁵ The Kenya study found that following the “case” definition underestimated the average household size and overestimated the proportion of households with one member as well as child-headed households.

Why is this a problem?

Ignoring this discrepancy between the sampling unit and the analytical unit and attempting to simply carry out the interview at the *case* level, can create two types of problems, each with severe impacts on data quality. Doing so also limits the ability to compare the results to standard surveys -- such as national surveys used to monitor host communities – which add critical context to the situation of refugees and are often comparable across time and location.

- **Improper measurement for household-level data.** Many modules that are included in standard survey questionnaires rely on the unit under observation being the *household*. A good example is consumption analysis, where data is collected on the food and non-food items consumed by the people that live and eat together. If the survey tries to capture this only for the members of a *case* (which may live and eat together with people from other cases), it becomes close to impossible to disentangle who ate which proportion of the shared meals.
- **Improper measurement for individual-level data.** If case focal persons are asked to report on individual-level information for all other members of their *case*, this may be referring to persons with whom they no longer interact on a regular basis because they are not necessarily members of their functional *household*. Such information would undoubtedly be of low quality.

On the other hand, ignoring the discrepancy between the sampling unit and the analytical unit, and simply carrying out the interview at the *household* level without proper adjustments, can create a different type of problem:

- **Double-counting and biased inference from sample to population.** If the sampling unit in the frame is the *case*, but the questionnaire is administered to an entire *household*, then *households* containing two *cases* have a double probability of selection. Firstly, this can result in fieldwork inefficiencies and redundant interviews. Secondly, if left unadjusted, the *household* weights calculated to allow inference from the sample to the population would be inflated and imprecise.

How can this be resolved?

Option 1: Don't use proGres as a sampling frame

One way to resolve the *case-versus-household* conundrum is to avoid using the proGres list frame altogether. There appears to be an exaggerated focus in many past UNHCR surveys on this frame despite the fact that it has known quality concerns in some countries (for example, due to the time elapsed since the last verification, or changes in the location of households) and the fact that the population under analysis actually lends itself well to alternative sampling approaches.

In particular, where surveys attempt to provide representative estimates of the population of a refugee camp, which is a geographically clearly delimited area,⁶ it would appear perfectly acceptable and statistically robust to follow an area-frame approach, e.g. selecting dwellings from a map, from a recent shelter listing, or from a systematic walk path through the camp under a pre-defined sampling interval. Given that multiple *households* can occupy the same dwelling, a two-stage selection process would be required (first select dwelling, then randomly select a *household* from within dwelling), which is a common and uncontroversial procedure in standard survey practice.⁷

An added advantage of this sampling approach is that it has the potential to capture non-registered individuals residing within a camp, which would be excluded by default from the survey universe in a proGres-based sampling approach.

Option 2: Sample cases but interview households, and mitigate the difference

Even in situations where the sample must be drawn from the proGres frame (e.g. because the survey is an add-on to a verification exercise and absolutely must follow the *case-based* data collection setup), the problem can still be resolved. One option is to draw a sample of *cases* from the proGres list but administer the survey to the *household* found “around” the focal point of the sampled *case*. This option may be considered in particular where the survey is administered at the location where a *household* resides, i.e. where the survey teams have a direct overview of the actual composition of a *household*.

A few arrangements are in order when using this approach, to mitigate the incongruity between the sampling unit and the analytical unit under this approach:

1. The survey questionnaire absolutely must be administered to the *household* (as defined in standard survey methodology – be it congruent with or larger or smaller than the sampled case) found around the focal point of the sampled *case*. Survey enumerators need to be well trained and clearly aware of the appropriate *household* definition as well as the difference between the sampling unit and the analytical unit to avoid confusion, and in order to be well prepared for the discussion likely to be required at the beginning of many interviews to identify and delineate the actual household to be interviewed.
2. In order to account for the double-selection probability where a *household* contains the focal points of multiple *cases*, the introductory sections of the questionnaire should include a brief module that documents the number of proGres case focal points present in the *household* under observation. This may ideally record the *case* and individual ID numbers of all *household* members for later checking with the proGres database, but this is not absolutely required for later sampling weights adjustment (the number of *case* focal points present in the interviewed *household* is the most important thing to record here).⁸

⁶ Most past SEAs did not exclude or oversample specific sub-populations in camps, both of which would require prior stratification of the sampling frame along pre-known characteristics (i.e. requiring a list frame with additional background information on the respondents, like proGres).

⁷ The [Demographic and Health Survey: Sampling and Household Listing Manual](#) provides a detailed discussion of practices for listing and sampling household members, see section beginning on page 32.

⁸ Recent survey experience has shown that in some settings it can be difficult on the ground to actually figure out how many cases make up a household. Households may not know, may not have the right documentation, or have reasons to obscure the fact that their household consists of multiple cases.

3. A good supervision system could ensure that a *household* containing two or more *case* focal points, and where both focal points were sampled for the survey, is not interviewed multiple times.
4. In cases where original members of the sampled *case* are no longer part of the *household* found around its focal point, but still reside within the survey's target area, the field team should trace their whereabouts. As long as these individuals have joined the *households* of other *case* focal points, no further action is required. However, where these individuals have split off to form new *households*, they should be traced and interviewed.^{9,10}
5. At the analysis stage, sampling weights need to be adjusted for *households* containing more than one *case*, to reflect their true "share" in the overall *household* population.¹¹

Option 3: Individual-level sampling

In situations where the sample must be drawn from the proGres frame but the survey is not administered at the location where a *household* resides (e.g. because the survey is administered centrally at a registration centre in parallel to a verification exercise), the above option would be difficult to implement because survey teams have no direct overview of the actual composition of a *household*, and no clear picture of which individuals form a functional *household* around the sampled *case* focal point. Moreover, since verification is often carried out in a pre-defined order (e.g. by *case* size) different members of a *household* may appear at the central location on different days, further complicating matters.

In such a setting, it may be advisable to draw the survey sample and administer the survey not at the case/household level, but purely at the individual level instead. In consequence, data would be collected as follows:

- **Individual level questions:** the sampled individual is asked only about himself/herself, no proxy responses are collected for other members of his/her *household*, i.e. no roster-type module is applied. The resulting individual-level dataset is fully representative of the overall population of registered individuals, without any methodological caveats.
- **Household level questions:** the sampled individual is asked questions about the characteristics of his/her functional *household* (e.g. household size, dwelling characteristics, water and sanitation access, etc.). The resulting dataset (of individuals providing information about their respective *household*) may be considered representative of the overall population of households, with the following caveats/adjustments:
 - *No more than one sampled individual per household:* One would have to assume that each sampled individual represents exactly one *household*. In other words, the sample of individuals cannot contain multiple members of the same *household*, otherwise this household would be "overcounted" in the aggregate results. Given

⁹ Otherwise, households with no case focal point would have a zero-selection probability. Households comprising split-off members of a sampled case would be included in the sample by default, without further randomization. Their sampling weight would be equal to the sampling weight of the case from which they split off.

¹⁰ A problem would occur where various non-focal members of different cases have split off from their original cases to form an independent household together, without any case focal being present. Such households would also have a multiple selection probability, but this could not be easily accounted for. However, it appears unlikely that such a scenario would occur very frequently, hence it appears acceptable to ignore it.

¹¹ For example, a household containing two case focal points should have its sampling weight divided by two, a household containing three case focal points should have its sampling weight divided by three.

that no ex-ante information on *household* affiliation is available from the sampling frame, this is impossible to rule out in advance, but can be mitigated through the selection protocol, as follows:

- The proGres-based sampling frame of individuals should be ordered by cases and geographic location (where available), then a systematic selection is made (i.e. every n-th person from the list is selected). While cases are obviously not congruent with households, there is a strong correlation between the two concepts, hence this should reduce the likelihood of selecting multiple persons from the same household.
 - This individual-level sampling approach should only be applied where the sampling fraction is less than 13% of the overall population (since very few households have more than 8 members; hence, the likelihood of selecting multiple members of the same household further reduces when selecting every 8-th individual from a list ordered by cases).
 - To further control for the selection of multiple members of a household, sampled individuals may be handed out simple and recognizable cards (e.g. blue card with a star on it) that they are requested to pass on to their other household members after the interview completes. The other household members should be advised to carry these cards with them at all times when visiting the central location where they could potentially be sampled for interview as well. At the start of each interview, the sample respondent would be asked whether they carry a blue card with a star, and if so, the interview would be cancelled (since that person would represent a household for which another member was already interviewed).¹²
- *Children:* If the individual-level questions of the survey are only relevant for adults, children may be excluded from the sampling frame before making the selection. However, if the individual-level questions are also relevant to children, then they should be included in the sampling frame. In consequence, one might draw a baby or young child that is unable to answer the survey questions by themselves. However, this would be relatively easy to resolve given that such children would always be accompanied to the registration centre by a caregiver. In such cases, while the interview questions would refer to only the sampled child, the answers could be provided by the accompanying caregiver as a proxy respondent. This is unproblematic from a methodological point of view.
 - *Informed respondents:* Under this approach, one may sample an individual that is not well informed about all the functionings of his/her wider household. While almost all members of a household will be able to answer household-level questions about the household's dwelling, water, sanitation, etc., only select members of the household will be able to provide accurate responses on issues like consumption expenditure. As a result, household-level questions that are not easily answerable by all/most household members should be avoided under an individual-level sampling approach.

¹² To avoid the diffusion of cards among the target population, it will be important to clearly explain to the sample respondent that these cards are not connected to any kind of assistance or benefits. An even more rigid control would be to write the names of the other household members – as stated by an interviewed individual – on the cards when they are given out, to ensure that they are not accidentally or deliberately passed on to other people in the erroneous expectation that they are tied to some sort of benefits.

- *Sampling weights:* At the individual level, each sampled respondent will have the same selection probability from the frame, and thereby the same sampling weight. However, when analysing the household-level data, larger households will have a higher probability of selection into the sample. This can be corrected by asking each sampled individual the size of their respective household, and then dividing the sampling weights at the household level by that number.

Which option is best for my situation?

The following diagram presents scenarios related to camp and non-camp settings, the contexts within which SEA surveys will be conducted, and the probable best option for the respective scenarios.

Diagrammatic representation of scenarios related to camp and non-camp settings

