UNHCR
The UN Refugee Agency

Forced Displacement Survey

METHODOLOGICAL PAPER 1

Questions to Identify Forcibly Displaced People in National Surveys

Findings from Three Case Studies, 2026

Acknowledgements

This methodological paper was produced by Lorenza Rossi (Survey Specialist, formerly UNHCR), Imane Chaara Thomas(Survey Specialist, formerly UNHCR), Nice (Data Analyst, formerly UNHCR), and Grayson Sakos (Associate Survey Specialist, UNHCR) with support and inputs from UNHCR's Global Survey Team.

The document benefited from inputs and comments from colleagues, including Imane Chaara and Andrej Kveder (UNHCR), Natalia Krynsky Baal and Filip Mitrovic (Secretariat for the Expert Group on Refugee, IDP and Statelessness Statistics (EGRISS)), and Felix Schmieding and Patrick Brock (World Bank-UNHCR Joint Data Center). The authors gratefully acknowledge the contributions of UNHCR country operations, the World Bank-UNHCR Joint Data Center on Forced Displacement, and government partners in South Sudan, Cameroon, and Pakistan where the pilot FDS surveys were realized.

Recommended citation: Rossi, L., Chaara, I., Nice, T. and Sakos, G. (2026). *Questions to identify forcibly displaced people in national surveys: Findings from three case studies.* FDS Methodological Paper No. 1. Copenhagen, Denmark: United Nations High Commissioner for Refugees.

Cover photo: Registration of refugees for government-issued biometric ID cards in Cameroon. (©UNHCR/Ruth Mbendong)

Contents

Introduction	**4**
Objective	**4**
Data	**6**
Reliability of self vs. proxy responses	**8**
Comparing legal status and response outcomes	**10**
Drivers of misidentification and bias	**14**
Discussion and conclusion	**17**
Appendices	**19**

Introduction

An accurate and standardized approach to identifying refugees, asylum-seekers, and internally displaced persons (IDPs) is critical for their inclusion in national statistical systems. This enables meaningful data disaggregation and targeted analysis, which are essential for developing effective policies that recognize and address the specific needs of forcibly displaced populations.

The Expert Group on Refugee, Internally Displaced Persons and Statelessness Statistics (EGRISS), UNHCR, and the World Bank-UNHCR Joint Data Center on Forced Displacement are leading efforts to develop internationally recognized guidelines and align data collection methodologies to advance the inclusion of forcibly displaced people in national statistical systems.

For statistical purposes, the International Recommendations on Refugee and IDP Statistics (IRRS and IRIS), endorsed by the United Nations Statistical Commission (UNSC) in 2018 and 2020 respectively, provide standardized criteria to identify forcibly displaced people in national statistics (EGRISS, 2018, 2020). Based on the criteria laid out in the International Recommendations, EGRISS published a methodological paper outlining the rationale for a sequence of identification questions in 2023, recommending measurement of indirect variables, such as citizenship, the reason for migration, and having crossed an international border (EGRISS, 2023). A standardized sequence of questions like the one outlined in the methodological paper is necessary to learn the characteristics and verify the criteria classifying an individual into a specific forcibly displaced population group. This is called the criteria-based approach to identification. Alternative approaches, like direct self-identification or declaration of forcibly displaced status, are not recommended or feasible, because individuals may misreport or be unaware of their status, leading to measurement error. This may be due to differences in interpretation or an unwillingness to disclose one's legal status (EGRISS, 2020, para. 107). Although the EGRISS methodological paper outlines a sequence of questions, there is little evidence on their application in practice. The formulation of the survey questions, the understanding of survey questions by respondents, their sensitivity to proxy or self-response, and the resulting inclusion and exclusion errors require further investigation.

UNHCR's new flagship survey programme, the Forced Displacement Survey (FDS), fills this gap by implementing the guidance from the International Recommendations and the EGRISS methodological paper to test the accuracy of its proposed set of questions and sequence for criteria-based identification of refugees. Based on field testing and findings in the three FDS pilot countries – South Sudan, Pakistan, and Cameroon – it is possible to evaluate the performance of the identification questions and criteria-based approach in detail across diverse displacement contexts.

Objective

This paper examines if the criteria-based approach and the set of identification questions aligned with international recommendations capture accurately the status of refugees. The following research questions guide our analysis:

- **Who is the most reliable respondent to the series of questions to identify refugees' legal status in a household survey with multiple respondents per household?** Does the classification outcome differ when asking a proxy respondent rather than a direct respondent?
- **Do registered refugees respond to the set of questions in a way that is consistent with their actual legal status/documentation?** If not, what are plausible explanations for their unexpected responses?
- **How does reliance on a criteria-based approach for identification affect population estimates of socio-economic outcomes**, compared to a document-based approach?

We choose to focus this analysis on registered refugees and asylum-seekers because, for registered refugees and asylum-seekers, the outcome of the identification questions and criteria-based approach

Box 1. Approaches to identification of Forcibly Displaced Persons (FDPs) in household surveys

Multiple approaches are possible to identify forcibly displaced populations in household surveys. The FDS identifies refugee status based on legal documentation, i.e. proof that the individual has successfully applied for and received refugee status. We call this the document-based approach to identification. At the same time, the survey asks respondents the sequence of identification questions discussed in this paper to understand their reported displacement history, whether they were forced to flee, and determine if their answers accurately classify respondents in accordance with their legal documentation. This is called the criteria-based approach to identification. A further approach focuses on respondents' fear of return to their country of origin but is not the focus of this paper. The table below outlines the three approaches.

	Legal documentation	Criteria-based approach	
		Forced to flee	Fear of return
Strengths	legally verified status; clear binary classification; simple verification with documentation	captures migration history; includes those without formal status	captures current protection needs; independent of legal processes
Limitations	legal status may not cover all protection needs and depend on political context	recall issues (e.g. for those who were forced to flee at young ages); difficulty capturing complex circumstances; challenges with abstract terminology	subjective interpretation; potential response bias depending on availability of information, culture etc.

can be validated against the true value as demonstrated by an individual's successful refugee registration, i.e. their possession of a valid refugee document. For IDPs, people in refugee-like situations, and others in need of international protection, this document-based approach is not usually possible as registries do not exist and/or there is no official legal document demonstrating their protection status against which outcomes of the criteria-based approach could be validated[1].

As such, although the identification questions within the FDS questionnaire and the sequence outlined in the 2023 EGRISS methodological paper identify numerous statistical categories, including refugee returnees, (prospective) asylum-seekers, and IDPs, this analysis does not cover all people forced to flee, or all statistical categories identified by the EGRISS recommendations (EGRISS, 2018, p.p. 31). Nevertheless, given there is overlap among the questions for identification of refugees and for other statistical categories, the lessons learned on the performance of the identification set in FDS for registered refugees is informative for the refinement of the questions for other categories (EGRISS, 2023, p.p. 23-26).

In the first section, we outline the data collected and the identification set of questions developed in the EGRISS methodological paper and tested through the Forced Displacement Survey, then we compare the outcomes of the identification questions for the proxy respondent (the household head or most knowledgeable member of the household) and the individuals themselves to assess the accuracy of their respective responses relative to the legal status determined by their documentation. Then, we assess the outcome of each question for the subgroup of respondents who own a valid refugee document to pinpoint sources of exclusion error. Lastly, we assess empirically the drivers of exclusion error and compare the estimates of key socio-economic indicators using the document- and criteria-based approaches to understand if errors in identification bias the estimation of survey-based indicators.

1 The FDS sampling frame consists of registered refugees and asylum-seekers whose records are maintained by UNHCR and/or host country governments and who own documents testifying their legal status i.e. being a recognized refugee or an asylum-seeker.

Data

In 2021, UNHCR launched the Forced Displacement Survey (FDS) as a new flagship survey initiative, aimed at filling the socio-economic data gap on forcibly displaced populations in low- and lower-middle income countries. The FDS pilot rounds were conducted in South Sudan, Pakistan, and Cameroon and provide nationally representative insights into the living conditions of refugees. The refugee sampling frames come from two primary data sources, depending on the context:

- registered refugees and asylum-seekers whose records are maintained by UNHCR and/or host government and who own recognized refugee documents; and
- Google Building Footprint data within refugee camps or sites[2].

Four respondents are interviewed in each household[3]: the head (or the most knowledgeable respondent) who answers on behalf of the household and its members (in the roster), a random adult member aged over 15, a woman aged 15 to 49 who gave birth in the past two years, and the primary caregiver of a child younger than five. Table 1 shows sample sizes of registered refugees examined in this paper by country of asylum: 1) individuals aged over 15 from the roster and 2) among the randomly selected adults.

Table 1. Sampled individuals aged 15 and over with a valid refugee document, by country and questionnaire

Country	Adults in household roster	Randomly selected adult†
Cameroon	13,914	5,007
Pakistan	23,200	5,060
South Sudan	11,523	3,072

† In Pakistan, the set of identification questions were administered in the questionnaire for the randomly selected woman, rather than for the randomly selected adult in South Sudan and Cameroon. See Pakistan FDS High-Level Report for further information.

Identification sequence

Following the criteria identified in the International Recommendations, the sequence of questions proposed in the methodological paper (EGRISS, 2023) and implemented in FDS cover citizenship, being forced to flee (for a valid reason requiring international protection), migration history, and application/pursuit of international protection. This sequence constitutes the criteria-based approach and aims to identify individuals who are outside their country of citizenship for reasons of feared persecution, conflict, generalized violence, or other circumstances that have seriously disturbed public order and who, as a result, require international protection (UN, 1994, paras. 8, 10-11, 31-3). There is a separate question, independent of the criteria-based approach, that asks for each household member if the individual has a valid refugee documentation which constitutes the document-based approach.

The identification sequence first appears in the survey's household roster, where the head of household responds about themself and about all members of the household over 15 as a proxy. One direct respondent per household then receives the same set of questions and answers about him/herself. In South Sudan and Cameroon, the direct respondent was the randomly selected adult, while in Pakistan it was the randomly selected woman. If the randomly selected adult/woman is the head of household who already responded to the household roster, the questions are only asked in the roster. The repetition of the sequence both in the roster and to the randomly selected adult or woman allows us to test whether there exists a difference in proxy and self-response to the sequence.

2 Even if sampled from sites or camps using building footprints, the eligibility criteria for participation are the same for refugees, i.e. ownership of a recognized refugee document. Those sampled from the footprint frame who are not registered refugees are ineligible.

3 For the purposes of the FDS, a refugee household is defined as a household in which the head or one of his/her spouses is a registered refugee based on documentation.

Table 2 provides an overview of the questions in the identification module by country in which the surveys were implemented (UNHCR, 2024, 2025)[4,5]. In South Sudan, the sequence exactly replicates the proposal of the 2023 EGRISS methodological paper (EGRISS, 2023). In subsequent surveys, we adjusted the questions to improve performance and test alternative specifications based on issues flagged during analysis of the identification modules in previous surveys and field testing. To improve clarity, we simplified the wording of three questions (1, 2, 3, and 4). We also introduced two recovery questions to offer alternative phrasing in case the respondents misunderstood the initial wording (2b and 3b).

Table 2. Overview of the data elements and related questions in the identification module by FDS country

Criteria	Q	South Sudan	Pakistan	Cameroon	Response
Citizenship	1	Of which country [are you/is NAME] a citizen?	What is (your/NAME's) nationality?		1. (Host Country) →NOT AS or REF† 2. Other, →Q2
Forced displacement	2	While living abroad, (have you/has NAME) ever had to flee home?	While living outside (host country), (have you/has NAME) ever had to flee (your/their) home? Home is (your/their) house and/or land.		1. Yes →Q3 2. No →Q2b
	2b		RECOVERY: While living outside (host country), (have you/has NAME) ever had to abandon (your/their) home? Home is (your/their) house and/or land.		1. Yes →Q3 2. No →NOT AS or REF†
	3	What is the main reason (you/NAME) had to flee home?	What is the main reason (you/NAME) fled or abandoned your home?		1. Valid reason →Q4 2. Invalid →Q3b
	3b			RECOVERY: You said (you were/NAME was) never forced to flee or abandon (your/their) home; however, (you/they) now live in a foreign country. What is the reason that made (you/NAME) leave (country of origin) most recently? OR What is the most recent reason that made (you/NAME) leave (country of origin)?	1. Valid reason →Q4 2. Invalid →NOT AS or REF†
Migration history	4	After (you/NAME) were forced to flee home, did (you/NAME) have to cross an international border?	After (you/NAME) forced to leave or abandon (your/their) home, did (you/NAME) move to another country?		1. Yes →Q5 2. No →NOT AS or REF†

† AS - asylum-seekers; REF - refugees

4 https://www.unhcr.org/media/questionnaire-forced-displacement-survey-south-sudan-2023

5 https://www.unhcr.org/media/questionnaire-forced-displacement-survey-pakistan-2024

Table 2. Overview of the data elements continued…

Criteria	Q	South Sudan	Pakistan	Cameroon	Response
Legal status	5	While in (host country), did (you/NAME) apply for international protection?	While in (host country), did (you/NAME) apply for international protection? For example, did (you/NAME) go to an office like (host country local examples) to register?		1. Yes →Q5 2. No →Q5b
	5b	Did (you/NAME) have the intention to apply for international protection?	While in (host country), did (you/NAME) plan to apply for international protection? For example, did (you/NAME) go to an office like (host country local examples) to register?	While in (host country), did (you/NAME) plan to apply for international protection? For example, did (you/NAME) go to an office like (host country local examples) to register?	1. Yes →potential AS 2. No →NOT AS or REF†
	6	What is the outcome of (your/NAME's) application?			1. Approved →REF 2. Not approved →NOT AS or REF† 3. Awaiting resp. →AS†
Legal documentation‡		What is the main document that allows you to stay in this country?			1. Valid doc →REF or AS† 2. Invalid doc →Other

† AS - asylum-seekers; REF - refugees

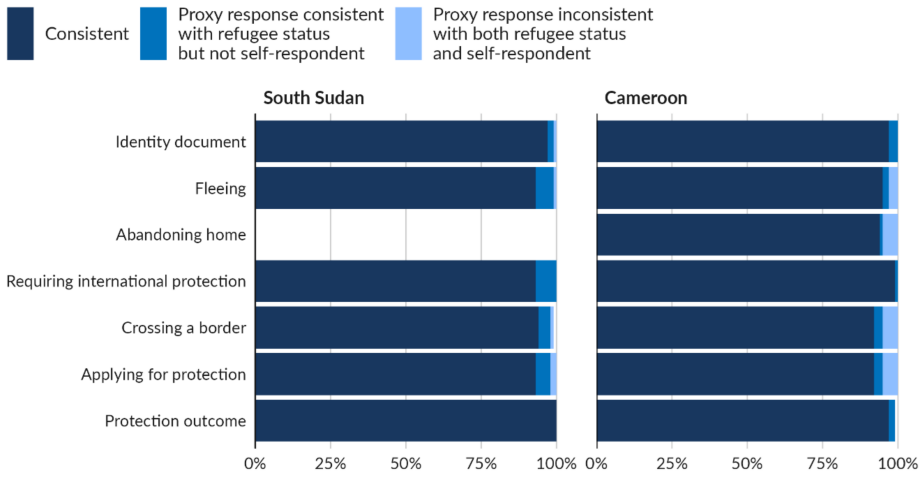
‡ not part of criteria-based approach

Reliability of self vs. proxy responses

Refugee status and all status determinations or identification for other categories of FDPs are made and granted to an individual. Therefore, it is not possible in a household survey to make a single status determination for all household members. The identification sequence, therefore, must be asked for each individual in the household. This begs the question who can or should respond to the identification questions. Must they be asked to each member directly or can a knowledgeable member, i.e. the household head respond on behalf of other members? The choice of the respondent or respondents for the identification questions in a household survey with multiple respondents can, however, impact the reliability and accuracy of responses. It is a priori unclear if the household head or each individual member themself may be more or equally knowledgeable about the individual's displacement history. That said, placing the identification sequence in the household roster and asking the household head or a single knowledgeable respondent has the advantage that we can efficiently gather information on all adults in the household, rather than just the one randomly selected adult respondent or having to ask each member directly. Moreover, by collecting data on all household members, it is possible to identify mixed households composed of both nationals and refugees. This enables more in-depth and nuanced analysis of the relationship between socio-economic outcomes and household composition.

To test the reliability of proxy responses, we look at the difference between responses from the household head in the household roster and the responses of the randomly selected individual respondent or woman. Overall, our analysis of proxy versus self-reporting reveals high consistency rates with the responses of the household head matching those of the randomly selected respondent 95 per cent of the time, although important discrepancies affecting identification do remain. Note that, as a large share of the sample for the FDS in Pakistan was born in Pakistan (see next section), there are few

Figure 1. Proportion of consistent responses by proxy and self-respondents



cases in which respondent who is not the head of household makes it through the entire identification sequence, so we focus the analysis of proxy response reliability on data from Cameroon and South Sudan.

In South Sudan, responses matched in 95 per cent of cases on average. There are two types of discrepancy:

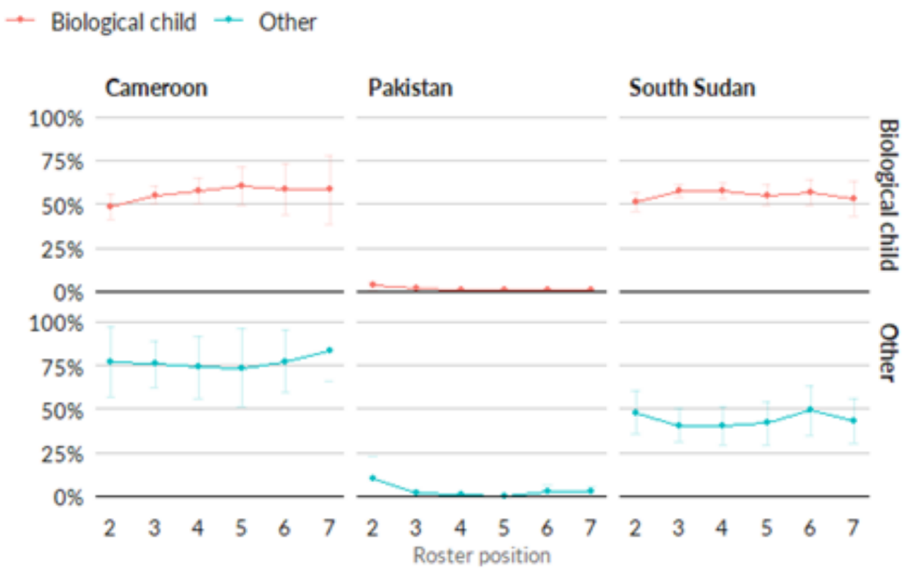
- responses from the proxy that are consistent with the legal status as a refugee, but responses from self-reporting that are inconsistent with the legal status (on average 4 per cent of cases across the identification sequence) and
- responses from the proxy that are inconsistent with the legal status as a refugee, but responses from self-reporting that are consistent with the legal status (on average 1 per cent of cases across the identification sequence)

Cameroon showed similar results, with 95 per cent of responses matching on average, though with a slightly different error distribution (2 per cent consistent proxy but inconsistent self-response and 3 per cent inconsistent proxy but consistent self-response). However, specific questions showed higher rates of inconsistency in both countries – notably those concerning border crossings and applications for international protection, where matching decreased to 92 per cent. This is likely explained by difficulty understanding the terminology of these questions, as discussed in the next section.

Although individual questions show relatively low inconsistency rates, the overall inconsistency rate in final outcomes, i.e. classification as a refugee or non-refugee, is higher at 16 per cent. This is because the final outcome aggregates multiple responses, so even small inconsistencies in individual items compound lead to a greater overall discrepancy. In 10 per cent of cases, the self-respondent's information does not match the legal status while the proxy respondent's information does. Conversely, in 6 per cent of cases, the self-respondent's information matches the legal status, while the proxy respondent's information does not. This suggests that the proxy respondent may in fact be providing more accurate responses, as they are more likely to be in line with those expected based on the refugee document.

Overall, the difference between proxy response and self-response indicates high reliability. However, we may expect that the reliability could vary across different types of household members, i.e. responses are more reliable for those more closely related to the head, or position in the household roster, i.e. due to respondent fatigue induced by answering for many members. Figure 2 shows that likelihood of completing the identification sequence does not vary neither across positions in the household roster nor by relation to household head. Note that, because the majority of refugees in Pakistan were born in Pakistan, the identification sequence is rarely completed regardless of roster position as most respondents state in Q2 and Q2b they were never forced to flee.

Figure 2. Share of refugees with a valid document completing the identification questions, by member position and relation to household head



Comparing legal status and response outcomes

In the following section, we analyse each question of the identification set in detail to identify potential gaps in the identification process across the three FDS pilot countries. We highlight key discrepancies between known refugee status – identified independently of the identification cascade through each individual's ownership of refugee documentation – and responses to the identification questions. See Appendices for an overview of the exclusion error resulting from the identification questions in each survey.

As shown in the previous section, differences between the proxy and direct respondent are small. We therefore only present the proxy responses from the head of household about all household members in this section.

Criterion 1: Fleeing or abandoning home (forced displacement)

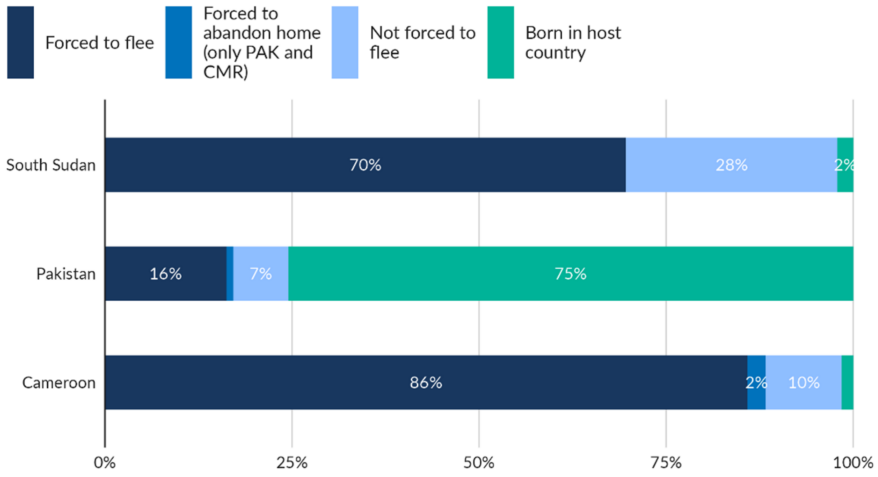
South Sudan	While living abroad (have you/ has NAME) ever had to flee home?
Pakistan and Cameroon	While living outside (host country) (have you/has NAME) ever had to flee home? Home is (your/their) house and/or land.
	RECOVERY: While living outside host country have you/has NAME ever had to abandon your home? Home is your/their house and/or land.

After the first question of the identification sequence on nationality, we ask if the respondent was forced to flee. In South Sudan, 30 per cent of adults holding refugee cards report that they did not have to flee, an unexpected response given their documented refugee status. We examined possible reasons why that might be. For example, this negative response could be explained by the fact that the South Sudanese state was founded in 2011, potentially creating confusion among respondents who were born in what is now South Sudan. Although there does not appear to be a systematic difference among those who respond as expected that they were forced to flee and those that did not based on year of arrival, there is a correlation with the location of residence among refugees in the North of South Sudan: Among those located in Pariang county, two-thirds (65 per cent), responded that they were not forced

to flee. In Maban, the other county with a large number of refugees, the share is just 19 per cent.

The high share of unexpected responses may reflect that the expression "forced to flee" is too complex or may not fully capture the multifaceted drivers of decisions to leave home. Conflict and livelihood pressures can reinforce each other, meaning that people may decide to move even if they are not directly affected by a single conflict event (Crawley, 2018; Erdal, 2018). This can make respondents less likely to describe their displacement as "forced", even if induced by a valid reason for statistical classification, and therefore fail to capture many people's actual experiences.

Figure 3. Proportion of individuals who were forced to flee/abandon their home among those with a valid refugee document



In Pakistan and Cameroon, learning from the experience in South Sudan, a recovery question was introduced with the phrasing "abandoning home" to recover respondents who did not self-identify as having been forced to flee.

In Pakistan, 84 per cent of the Afghans with refugee cards responded that they never had to flee. The recovery question only captured an additional 1 per cent that responded positively or that they had to abandon their homes. Most recognized refugees who report that they did not experience displacement were born in Pakistan (90 per cent). This highlights that in countries with protracted, intergenerational displacement, the sequence of identification questions requires a condition to identify those born in the country of displacement/asylum, who likely have "derivative refugee status" (EGRISS, 2018). Learning from this experience, we modified the questionnaire for future FDS rounds, with the questions on being forced to flee or abandoning home not asked to those born in the country of asylum.

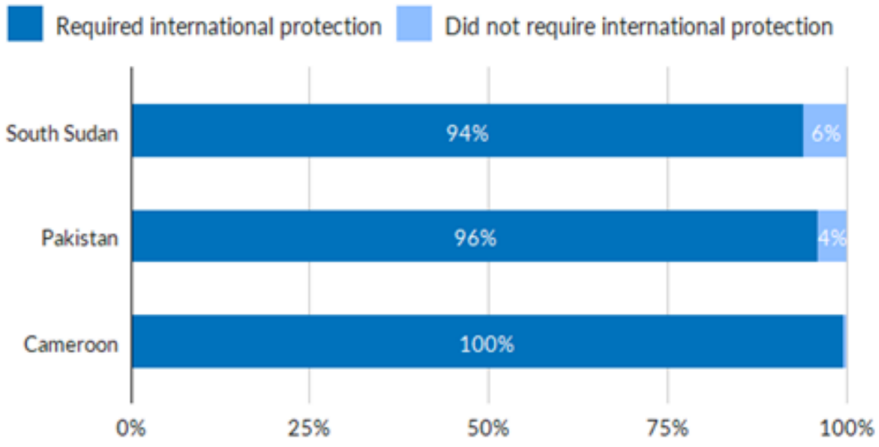
In Cameroon, 14 per cent of refugees with valid documentation responded that they never had to flee. The recovery question on "abandoning home" led to the inclusion of an additional 2 per cent of refugee card holders who would otherwise have dropped out of the identification sequence. Most individuals who responded that they never had to flee or abandon their homes were born abroad. It seems that neither the terminology "fleeing" nor "abandoning" fully captures how the respondents perceive the move from their country of origin to the country of displacement. Therefore, a further recovery question was added (see next section), which is more effective at capturing multifaceted displacement experiences.

Criterion 2: Reason for fleeing†

South Sudan	What is the main reason (you/NAME) had to flee home?
Pakistan and Cameroon	What is the main reason (you/NAME) fled or abandoned (your/their) home?

† The phrasing of the question was changed after the survey in South Sudan to reflect the added recovery question about having to abandon one's home.

Figure 4. Proportion of individuals with a reason for fleeing requiring international protection among adults with refugee documents who were forced to flee



According to the 1951 Refugee Convention, valid reasons to be recognized a refugee are "*race, religion, nationality, membership of a particular social group or political opinion*" (UNHCR, 2010). These translate into response categories 1) security reasons, 2) armed conflict and generalized violence, and 3) fear of persecution in the FDS questionnaire, as recommended in the IRRS (EGRISS, 2018). All other reasons are considered invalid to obtain refugee status and international protection.

Over 9 in 10 recognized adult refugees from South Sudan who fled their homes state that they did so due to security threats, violence, or conflict – situations that qualify them for international refugee protection. However, about 6 per cent of individuals left for "family reasons", meaning that they would not strictly fulfil the criteria for classification as refugees. This trend correlates with family separation patterns. Among individuals from households that experienced separation during displacement, around 10 per cent report that they fled to reunite with family members. In contrast, only 1 per cent of people from households that remained together during displacement cited family reunification as their reason for fleeing. In Pakistan, 4 per cent of adults remaining in the cascade reported leaving for personal reasons, while in Cameroon this share is less than 1 per cent.

This suggests that to accurately identify all individuals eligible for refugee status, the sequence of questions should either continue despite the response that appears to be a non-valid reason for asylum or include a follow-up question that allows respondents to indicate if they moved to reunite with family members who are refugees.

Recovery question: Reason for moving abroad (forced displacement)

In Cameroon, a further recovery question was introduced at this point in the identification sequence to identify individuals who may not have responded as expected to the previous questions about being forced to flee or to abandon home.

> You said (you were/NAME was) never forced to flee or had to abandon (your/their) home; however, (you/they) now live in a foreign country. What is the reason that made (you/NAME) leave Cameroon most recently? OR What is the most recent reason that made (you/NAME) leave Cameroon?

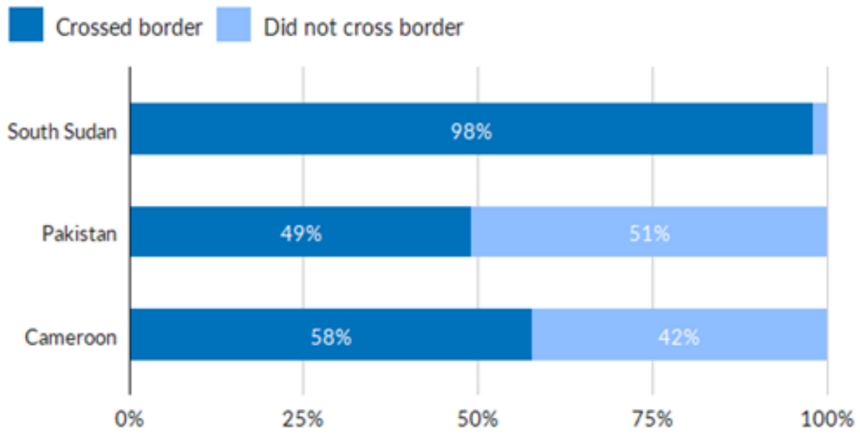
If the respondent provides a reason that would grant international protection, this response supersedes the negative response to whether an individual fled or abandoned their home, and the sequence of identification questions continues. This recovery question captures over two-thirds (69 per cent) of individuals about whom the proxy respondent initially said that they never had to flee or abandon their homes.

Criterion 3: Crossing an international border (migration history)

South Sudan	After you were forced to flee home, did (you/NAME) have to cross an international border?
Pakistan and Cameroon	After (you/NAME) were forced to leave or abandon (you/their) home, did you move to another country?

In South Sudan, almost all individuals with a valid reason for fleeing report having crossed an international border. Nonetheless, the phrasing was changed after the first pilot survey because cognitive interviews showed that refugees may not know they have crossed a border, for example because there is continuous cross-border mobility or there is no clearly demarcated immigration post (Vancluysen, 2021). It is also possible that respondents do not understand the abstract concept of an "international border", which may not be relevant to their daily lives.

Figure 5. Proportion of individuals who crossed an international border among adults with refugee documents with a reason for fleeing requiring protection



Therefore, in Pakistan and Cameroon, the term "cross an international border" was removed and simplified to "move to another country". Despite this rephrasing, there remains a high share of unexpected responses in both surveys. In Cameroon, 42 per cent of refugee card holders who are still in the identification sequence are classified as not having crossed a border. In Cameroon, the share of unexpected responses is particularly high among refugees in sites closest to the border with the Central African Republic (58 per cent) and in rural areas in the Far North, close to the Nigerian border (52 per cent). In these locations, people may still be able to speak the same languages as in their regions of origin and pendular migration may be more common. The share of recognized refugees who state that they never had to cross a border is considerably lower among refugees in the urban stratum of Yaoundé and Douala (11 per cent), far from the countries' frontiers.

In Pakistan, 51 per cent of individuals who have a refugee card and remain in the identification sequence state that they never had to move abroad. The unexpected response is particularly frequent among individuals living in refugee villages (59 per cent compared to 34 per cent of those living outside refugee villages).

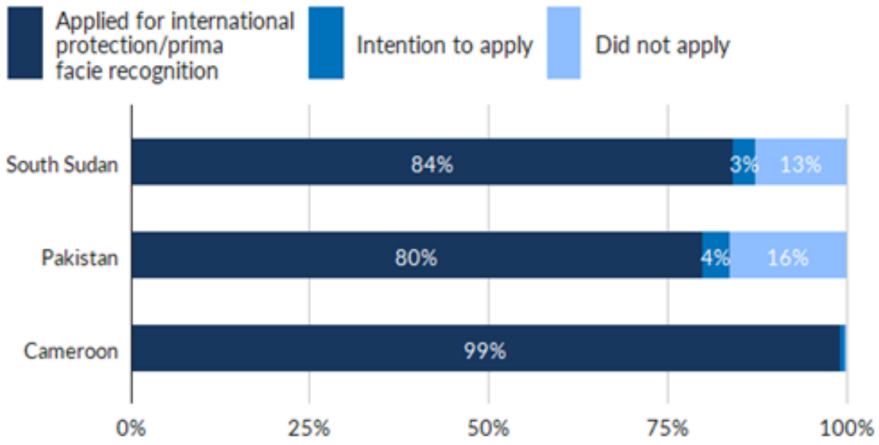
Due to the high rate of inconsistent responses in the pilot surveys, this question is removed from subsequent FDS surveys for refugee identification. In any case, the question is redundant in the sequence of questions for refugees, as an individual born outside the country in which they live must have crossed a border. However, the finding remains informative for adaptation of the identification questions to identify internally displaced people and refugee returnees whose status determination would require this or a similar question[6].

6 For refugee returnees, we must know if they crossed an international border to enable questions on application for asylum in another country. For IDPs, we must similarly know to distinguish them from returnees and because individuals forced to flee that cross an international border but do not stay for longer than 12 months, return to their country of origin, and did not apply for international protection remain IDPs.

Criterion 4: Applying or intending to apply for international protection (legal status)

South Sudan	While in (host country), did (you/NAME) apply for international protection?
Pakistan and Cameroon	While in (host country), did (you/NAME) apply for international protection? For example, did (you/NAME) go to an office like (host country local examples) to register?

Figure 6. Proportion of individuals who applied for international protection among adults with refugee documents who crossed an international border



The share of recognized refugees reporting that they applied for international protection varies by country. This may reflect different understandings of registration, depending on the country of origin and the asylum process in the host country. In South Sudan, where refugees from Sudan received prima facie recognition, 87 per cent of individuals with a valid refugee card who crossed an international border state that they applied for protection, received prima facie recognition or intend to apply. The high share of inconsistent responses is likely again due to the misunderstanding of the question phrasing. Cognitive interviews showed that many respondents are not familiar with the terminology "international protection".

Therefore, we modified the wording in the Cameroon and Pakistan surveys, providing contextual examples of registration locations. This may have helped respondents better understand the question. In Cameroon, almost all individuals (over 99 per cent) with a valid refugee card who crossed a border applied for international protection or intend to apply.

In Pakistan the share is lower, at 80 per cent. However, this likely reflects the country context, rather than an issue with the wording. Many refugees were not able to officially register with the Government of Pakistan, outside a short window in 2006. In fact, all refugees arriving after 2006 have either received a different document from a refugee card, the Afghan Citizen Card (ACC), or been considered undocumented. In this context, it makes sense for a respondent to say that they neither applied nor had the intention to apply for protection, knowing the government did not provide this opportunity after 2006. The share is similar for men and women, although household heads are slightly more likely to have applied for protection than other household members (82 per cent vs. 78 per cent).

Drivers of misidentification and bias

The previous analysis shows that the criteria-based approach in FDS fails to identify accurately refugees in a number of circumstances, and individuals are erroneously excluded at several points in the sequence on different criteria. It is also apparent that this exclusion is non-random with certain groups, for example children of refugees, households which were separated and subsequently reunited, and those in fluid border regions seemingly systematically excluded. As a result, relying on the identification

questions as implemented in the FDS pilot countries would misrepresent the true population share of important population subgroups.

Table 3 presents a summary of key socio-demographic statistics among refugees with a valid document for the three countries, disaggregating by those who complete and those who do not complete the identification cascade. It shows a number of significant differences. In all three countries, refugees who do not complete the criteria-based questions are more likely to live in households that were separated following their displacement and are younger (in Pakistan, considerably younger). In South Sudan and Cameroon, refugees who do not respond to the criteria-based questions as expected also live in smaller households.

Table 3. Overview of the data elements and related questions in the identification module by FDS country

	South Sudan			Cameroon			Pakistan		
Variable	Incomplete cascade	Complete cascade	t-/z-stat	Incomplete cascade	Complete cascade	t-/z-stat	Incomplete cascade	Complete cascade	t-/z-stat
Year of arrival (median)	2011	2011	1.1	2014	2014	3.3**	1985	1983	5.0***
Age (median)	25	28	-5.1***	36	36	2.5*	28	54	-49.5***
Female	53.3%	53.9%	-0.4	56.2%	56.0%	0.1	48.3%	47.8%	0.3
Separated house-holds	38.1%	30.1%	6.7***	20.7%	20.4%	0.3	9.4%	5.9%	2.4*
Citizenship (top 3 countries)	SDN 82.4% COD 8.9% ETH 5.4%	SDN 82.4% COD 8.9% ETH 5.4%		CAF 72.0% NGA 25.9%	CAF 72.0% NGA 25.9%		AFG 99.9%	AFG 99.9%	
Household size (mean)	8.1	8.6	-5.4***	6.1	6.6	-5.6***	11.8	11.7	0.5
n	404	746		2,395	3,291		2,137	955	

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Countries: SDN-Sudan; CAF-Central African Republic; AFG-Afghanistan; COD-Democratic Republic of Congo

NGA-Nigeria; ETH-Ethiopia

To assess both the drivers and the implications of this misidentification bias in criteria-based refugee identification, we first model the likelihood of cascade completion using demographic and displacement history variables to determine whether exclusion is systematically correlated with specific household or individual characteristics. Second, we conduct difference-in-means tests comparing key socio-economic outcome indicators collected in the FDS between samples identified through the criteria-based cascade versus the sample identified via documentation to understand how misidentification biases the surveys' population estimates.

Appendix 2 presents a series of logistic regression models predicting the likelihood of completing the identification questions as expected using a number of demographic and contextual variables including sex, household head status, country of birth, years since arrival, separation status, and refugee location (the sample stratum). The results are presented as odds ratios (OR) rather than log-odds coefficients for ease of interpretation.[7]

7 An OR of 1 indicates no effect, values greater than 1 indicate increased odds (e.g., 2 means double the odds), and values less than 1 indicate decreased odds (e.g., 0.5 means half the odds).

The models confirm that the bias is systematic and non-random. However, they also show that the drivers of misidentification are not consistent across the three countries where FDS was implemented. The geographic location (sampling stratum) is consistently the strongest predictor across all three models. In Pakistan, respondents in refugee villages and urban area were less likely to be correctly identified. In Cameroon, on the other hand, there is a positive correlation with identification and residing in peri-urban and urban areas. In South Sudan, which did not include an urban sample of refugees, refugees in Maban are much more likely to be identified correctly as refugees compared to those in Pariang. Individual demographics (sex, household head status) are generally not significant predictors of correct identification nor is the relationship to the household head. As expected, being born in the host country has strong negative correlation in Pakistan and Cameroon confirming the finding that the identification questions do not perform well for those with derivative refugee status. In Cameroon, longer stays in country are correlated with higher likelihood of completing the cascade as expected while in Pakistan and South Sudan duration of stay is mostly insignificant. Given the complexity in some of the terminology used for refugee identification, we may expect that more educated household heads perform better. This appears to be the case only in Cameroon. In South Sudan, there is no significant relationship between education of the household head and identification, although there is low variability in educational achievement overall. In Pakistan, the inverse is true, however this is likely due to younger cohorts being better educated but also significantly more likely to be born in Pakistan.

Figure 7. Difference in means between document-based and criteria-based approaches in estimates of socio-economic indicators

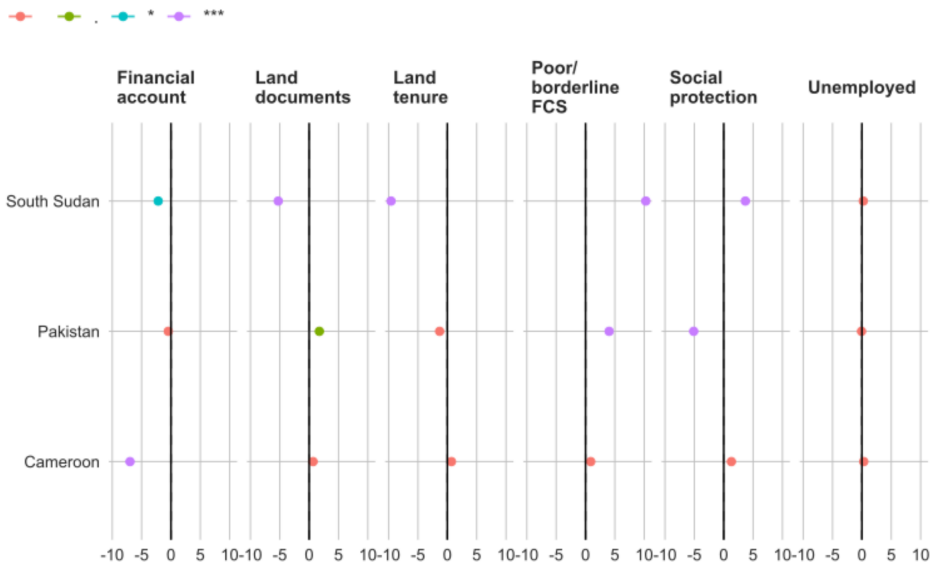


Figure 7 shows the results of difference-in-means tests for six socio-economic indicators measured through the FDS showing the difference between estimates using the document- and criteria-based approaches. A positive difference indicates that the estimate using the document-based approach exceeds the criteria-based approach (underestimation) while a negative value indicates that the estimate of the document-based approach is less than the criteria-based approach (overestimation). The findings are consistent with non-random, systematic bias however indicate that the direction of the bias varies across contexts. In Cameroon, the difference between the two approaches is only significant for estimates of the prevalence of refugees with access to a financial account which is overestimated under the criteria-based approach. In Pakistan, there are statistically significant differences in estimates for access to social protection, food security, and access to land. Importantly, the criteria-based approach underestimates food insecurity and overestimates access to social protection. In South Sudan, there is the largest bias in the estimates likely owing to the high correlation between misidentification and location. The criteria-based approach overestimates refugee's access to financial accounts and secure land tenure while underestimating food insecurity and access to social protection.

Discussion and conclusion

Development of a standardized criteria-based approach to identification of refugees as well as other forcibly displaced persons is necessary for statistical inclusion. However, full inclusion requires this approach to be accurate and reliable. The experience of the Forced Displacement Survey (FDS) shows that further refinement is still required to ensure accuracy.

The comparison between proxy and self-reporting in the identification process shows fairly high consistency across the identification sequence, although for specific questions, for example about border crossings and applications for international protection, there are more discrepancies in responses. Although the causes of these differences require further examination, the FDS shows that reliance on proxy response is not the main source of misidentification bias in criteria-based identification and proxy response is sufficiently reliable and consistent with individual responses for identification of refugees.

Despite the high consistency of proxy and self-response, analysis shows that many registered refugees holding valid documentation of their legal status do not respond to the criteria-based questions as expected, contradicting their legal status. The differences are largely not a result of features of the respondent within the household, instead they result largely from misunderstanding of the questions and appear systematic across population subgroups, correlated with household features. Findings suggest misunderstanding or different interpretations of the terminology "forced to flee" that often leads to misidentification. Rephrasing the question as "abandoning home" in a recovery question had a positive but small effect. A further recovery question, added to the survey in Cameroon, showed that asking specifically about reasons for no longer living in the country of origin can capture a large share of individuals who would otherwise drop out of the identification cascade. As the question asking whether or not a person was forced to flee is used not only for identification of refugees but also other forcibly displaced people, further investigation is required to minimize exclusion error.

The case of South Sudan highlights the risk of relying on single-category reasons for fleeing and particularly the exclusion of "family reasons", which can mask legitimate claims such as family reunification with relatives who were forced to flee and have already obtained refugee status. It also shows the need for sensitivity to the multi-faceted nature of migration decisions even in contexts of forced displacement.

When looking at the "citizenship" and "migration history" criteria, the case of Pakistan highlights the challenges of identifying individuals born in the host country, who may not have crossed a border but still require international protection. The current set of questions focuses on adults, and there is a need to develop criteria-based questions that can accurately identify children of refugees. This includes considering the legal status of their parents and their own potentially separate displacement experiences.

The pilot FDS surveys also highlight possible difficulties in understanding cross-border displacement in contexts where administrative borders may not feature prominently in everyday lives and respondents may frequently travel back and forth. This shows that either a more nuanced, context-specific phrasing is needed with concrete geographical references, or that this question should be skipped altogether for refugees. The latter is the approach taken in subsequent FDS survey rounds. This question nevertheless requires further investigation and revision as it will remain necessary for identification of IDPs and refugee returnees among other categories of FDPs who may have similar difficulties in understanding. More testing is still required for other categories of FDPs, but we may expect that for identification questions common across categories (like reason for displacement and forced to flee) results would be similar.

Turning to the implications of misidentification bias for measurement, the results indicate that the direction of misidentification bias under the criteria-based approach is indeterminate a priori but in all three FDS surveys significant difference arise between the document- and criteria-based approaches for one or more socio-economic indicators. This confirms that the criteria-based approach as formulated in FDS suffers from systematic bias and that the measurement error is non-random. It also indicates that heterogeneity in survey context dictates the performance of the questions, obviating the

need for additional adaptation and contextualization of the questions across contexts. This means that standardized approaches to identification need to be sensitive to the displacement context in their formulation and require sufficient flexibility in adaptation.

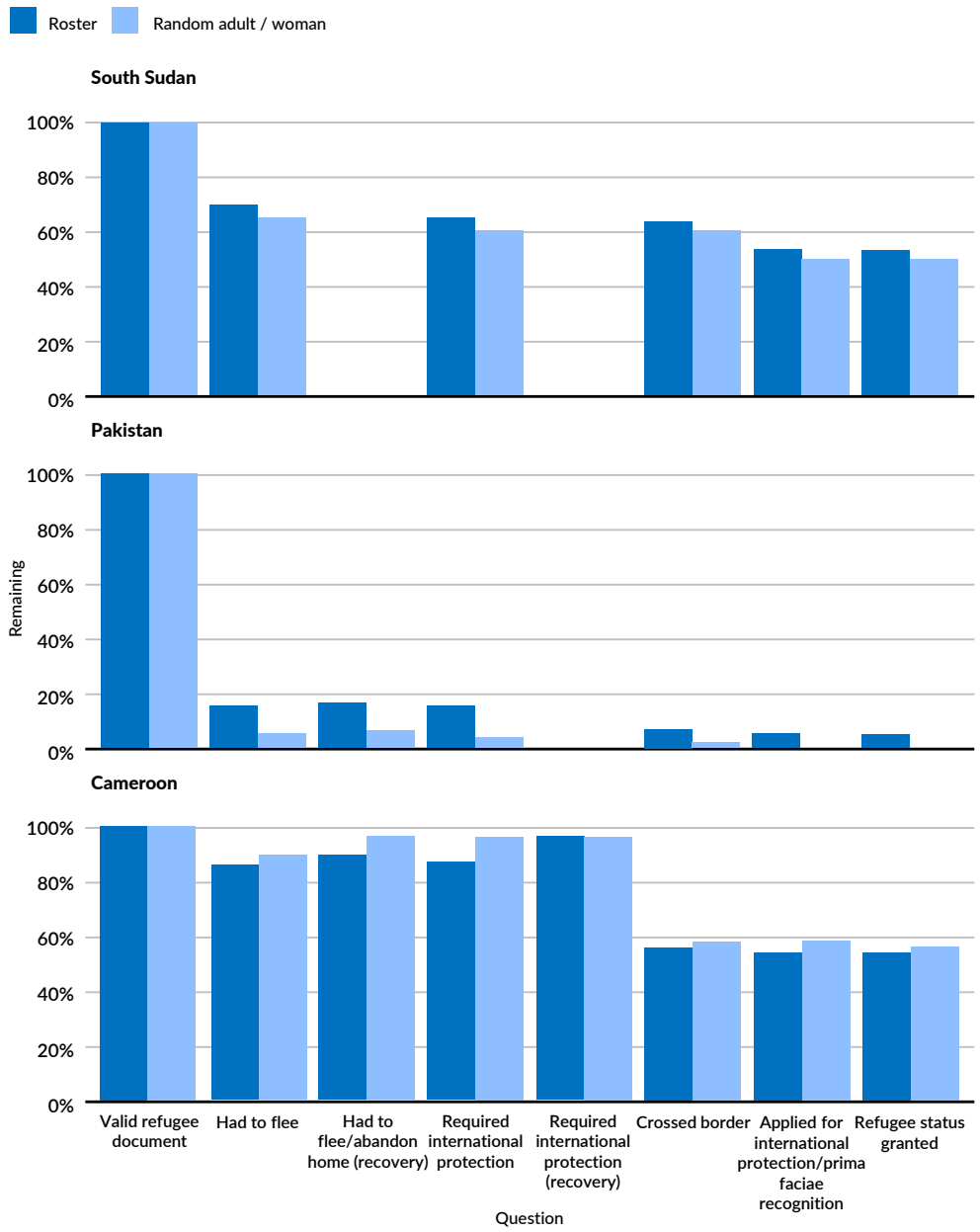
Finally, it is important to note that the recommendations of the EGRISS 2023 methodological paper (EGRISS, 2023) and the identification sequence in the FDS were designed to capture most, but not all, of the statistical categories identified in the IRIS and IRRS. Some categories may still be excluded in future refinements, particularly those relating to smaller population groups or more nuanced aspects of migration and family history. These would require a longer and more detailed set of questions, which may be better suited for thematic surveys. Further refinements will also investigate improved question wording and sequencing to minimize exclusion, as well as strategies to reduce the length of the identification sequence for inclusion in other national household surveys where FDP inclusion is necessary but addition of the full sequence is infeasible.

Additionally, it should be noted that, the findings here speak mainly to the implications of exclusion errors on measurement. More research is needed on the impact of inclusion errors. Also, more work is necessary to define and measure “inclusion errors”. As the FDS pilots surveyed the registered refugee population, an inclusion error would be classification of non-refugees or unregistered refugees as registered refugees. There is little evidence of non-refugees completing the sequence, however unregistered refugees should complete the sequence and their identification and inclusion in surveys may be important and desirable, especially in context, like Pakistan, where a large number of refugees are unregistered. Moreover, although the evaluation of the identification set for refugees presented here contains lessons for other population groups, more research is needed on the questions’ performance in other contexts, especially for slow onset and non-conflict induced displacement.

Appendices

Appendix 1: Proportion of individuals remaining in the identification cascade after each question

Figure 8. Proportion of individuals remaining in the identification cascade after each question



Appendix 2: Likelihood of completing Identification Cascade

Table 4. Likelihood of completing Identification Cascade

		South Sudan		Pakistan		Cameroon	
		Odds Ratio	SE	Odds Ratio	SE	Odds Ratio	SE
Male Member		1.01	(0.18)	1.07	(0.17)	1.12	(0.07)
HoH is Male		0.81	(0.19)	1.06	(0.17)	0.89+	(0.06)
HH Size		0.95*	(0.02)	1.01	(0.01)	1.03**	(0.01)
Less than 5		1.89	(0.58)	12.67***	(0.35)	2.74**	(0.31)
Years since arrival	5 to 9	5.41***	(0.43)	16.18***	(0.36)	4.39***	(0.29)
	10 to14	6.78***	(0.27)	15.09***	(0.44)	4.67***	(0.28)
	15 to 19	4.03***	(0.29)	4.21*	(0.59)	4.28***	(0.29)
	20 to 25	0.70	(0.40)	16.27***	(0.34)	3.87***	(0.33)
	25+	8.23+	(1.15)	22.39***	(0.23)	1.29	(0.55)
Separated HH		1.74**	(0.19)	0.57*	(0.25)	0.50***	(0.08)
South Sudan	Pariang	0.01***	(0.42)				
	Maban	6.21***	(0.25)				
Pakistan	Out of RV, Balochistan			2.13**	(0.28)		
	Ref. Village, KPK			1.46	(0.28)		
	Out of RV, KPK			2.17**	(0.26)		
	Ref. Village, Punjab			0.03***	(0.63)		
	Out of RV, Punjab			0.50*	(0.27)		
	Islamabad and Karachi			0.48*	(0.29)		
Cameroon	Rural, Far North					0.77*	(0.11)
	Sites North, East, Adamawa					0.51***	(0.10)
	Townships North, East, Adamawa					2.15***	(0.10)
	Yaoundé, Douala					5.98***	(0.13)
	Rural North, East, Adamawa					1.11	(0.10)
Relationship to HoH	Wife/Husband/Partner	0.99	(0.30)	0.97	(0.21)	1.14	(0.08)
	Biological child	0.86	(0.22)	0.80	(0.24)	0.96	(0.09)
	Mother/Father	3.08	(0.74)	1.27	(0.21)	1.48+	(0.23)
	Sister/Brother	1.17	(0.36)	0.67	(0.37)	1.19	(0.18)
	Grandchild	2.61	(0.95)	0.00	(300.10)	1.18	(0.42)
	Other	0.62+	(0.29)	0.56+	(0.33)	1.79**	(0.19)
Education of HoH	Lower Secondary	1.19	(0.42)	0.37**	(0.37)	1.94***	(0.19)
	Upper Secondary	1.53	(0.58)	0.28***	(0.35)	2.09***	(0.19)
	None	0.78	(0.35)	0.63+	(0.25)	1.29+	(0.13)
(Intercept)		0.81	(0.48)	0.04***	(0.46)	0.17***	(0.33)

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

References

Crawley, H.; Skleparis, D. (2018). Refugees, migrants, neither, both: categorical fetishism and the politics of bounding in Europe's 'migration crisis'. *Journal of Ethnic and Migration Studies*, 44(1):48–64.

EGRISS (2018). *International recommendations on refugee statistics*. European Union and United Nations, Luxembourg.

EGRISS (2020). *International recommendations on internally displaced persons statistics*. European Union and United Nations, Luxembourg.

EGRISS (2023). Towards a standardized approach to identify IDPs, refugees and related populations in household surveys. *Methodological Paper Series*, (1).

Erdal, M. B.; Oeppen, C. (2018). Forced to leave? the discursive and analytical significance of describing migration as forced and voluntary. *Journal of Ethnic and Migration Studies*, 44(6):981–998.

UN (1994). Note on international protection (submitted by the high commissioner) executive committee of the high commissioner's programme forty-fifth session un doc. a/ac.96 /830 7 september 1994. *International Journal of Refugee Law*, 6(4):679–713.

UNHCR (2010). *The 1951 Refugee Convention and 1967 Protocol relating to the status of refugees*. UNHCR.

UNHCR (2024). *Forced Displacement Survey South Sudan 2023 - Questionnaire*. United Nations High Commissioner for Refugees.

UNHCR (2025). *Forced Displacement Survey Pakistan 2024 - Questionnaire*. United Nations High Commissioner for Refugees.

Vancluysen, S. (2021). Deconstructing borders: Mobility strategies of South Sudanese refugees in northern Uganda. *Global Networks*, 22(1):20–35.

Published April 2026. Some rights reserved.

This work is made available under the Creative Commons Attribution 4.0 IGO license (CC BY 4.0 IGO) https://creativecommons.org/ licenses/by/4.0/.

In any use of this work, there shall be no suggestion that UNHCR endorses any specific organization, products, or services.

If you create a translation of this work, please add the following disclaimer: This translation was not created by the United Nations High Commissioner for Refugees (UNHCR). UNHCR is not responsible for the content or accuracy of this translation.

If the work is adapted, please add the following disclaimer along with the citation: This is an adaptation of an original work by the United Nations High Commissioner for Refugees (UNHCR). Views and opinions expressed in the adaptation are the sole responsibility of the author or authors of the adaptation and are not endorsed by the UNHCR.

Please cite the work as follows: Rossi, L., Chaara, I., Nice, T. and Sakos, G. (2026). *Questions to identify forcibly displaced people in national surveys: Findings from three case studies.* FDS Methodological Paper No. 1. Copenhagen, Denmark: United Nations High Commissioner for Refugees.

The use of the UNHCR logo is not permitted.

Nothing herein shall constitute or be construed or considered to be a limitation upon or waiver, express or implied, of any of the privileges and immunities of the United Nations including its subsidiary organs, or of the United Nations High Commissioner for Refugees (as a subsidiary organ of the United Nations).

This document along with further information on Forced Displacement Survey is available on UNHCR's statistics website: https://www.unhcr.org/fds.

Forced Displacement Survey

©2026 United Nations High Commissioner for Refugees

Some rights reserved (see previous page).

Statistics, Data Science, and Survey Section
UNHCR Global Data Service
UN City, Marmorvej 51
2100 Copenhagen, Denmark
stats@unhcr.org

Further information on Forced Displacement Survey is available on UNHCR's statistics website:
https://www.unhcr.org/fds