Expert Group on Refugee, IDP and Statelessness Statistics

Defining the Comparator Population for Refugee and IDP Statistics: From the concept of the host community to statistical definitions

EGRISS
EXPERT GROUP ON REFUGEE, IDP AND STATELESSNESS STATISTICS

August 2025
Methodological Paper Series
Paper No. 4

The Expert Group on Refugee, IDP and Statelessness Statistics (EGRISS) is a UN Statistical Commission mandated, multi-stakeholder group. Since its establishment in 2016, it works with national statistical offices, international organisations and civil society to develop and support the implementation of international standards and guidance to improve official statistics on refugee, internally displaced and stateless persons.

With 61 country members and 40 institutional members, EGRISS welcomes collaboration with all relevant partners from the statistical community. EGRISS implements activities with and through its members, organised in Technical Subgroups (TSGs). This paper was developed by TSG 2 'Methodological Research and Guidance Development'.

The documents set is intended as the Expert Group's input to ongoing methodological debates. They do not constitute consensus or represent the official views of EGRISS, its members, TSG 2, or the institutions with which the authors are affiliated.

EGRISS would like to express particular appreciation to Hyunju Park (UNHCR) and Sigrid Weber (IE University) for leading the development of this methodological paper with the inputs and support from Anqi Pan (World Bank–UNHCR Joint Data Center on Forced Displacement), Braydon Madson (Duke University), and Charis Sijuwade (EGRISS Secretariat), in addition to Task Team members including Alvaro Sardiza (IMPACT Initiative), Ana Aguilera, Paula Andrea Rossiasco (World Bank), Felix Schmieding (World Bank–UNHCR Joint Data Center on Forced Displacement), Lorenza Rossi (UNHCR), Oscar Rico, Margharita Lundkvist-Houndoumadi, Mirjam Kuschnitzki (JIPS), Sebastian Steinmuller (UNRWA), Simone Holladay (IOM), Filip Mitrovic, Ladina Schlatter, and Natalia Baal (EGRISS Secretariat).

Contents

List of Acronyms

CRRF	Comprehensive Refugee Response Framework
DHS	Demographic and Health Survey
ECV	Encuesta National de Calidad de Vida [National Survey on Quality of Life]
EGRISS	Expert Group on Refugee, IDP and Statelessness Statistics
FDS	Forced Displacement Survey
GAIN Survey	Global Annual Inclusion Survey
GCR	Global Compact on Refugees
IASC	Inter-Agency Standing Committee
IDA	International Development Association
IDP	Internally Displaced People
IOM	International Organization for Migration
IRIS	International Recommendations on IDP Statistics
IROSS	International Recommendation on Statelessness Statistics
IRRS	International Recommendations on Refugee Statistics
KAP	Knowledge, Attitude, and Practice
SDG	Sustainable Development Goals
UNHCR	United Nations High Commissioner for Refugees
WASH	Water, Sanitation, and Hygiene
WFP	World Food Programme

Introduction

1. The Expert Group on Refugee, IDP and Statelessness Statistics (EGRISS) has developed international statistical recommendations to guide the production of official statistics on refugees, internally displaced persons (IDPs), and stateless populations, all of which have been endorsed by the United Nations Statistical Commission in 2018, 2020, and 2023 respectively. These recommendations aim to contribute to the improved availability of nationally produced statistics on all three groups with an emphasis on harmonisation or standardisation of key definitions. As part of these guidelines, the International Recommendations on Refugee Statistics (IRRS) and the International Recommendations on IDP Statistics (IRIS) highlight the need to compare statistics and indicators on the "protection and wellbeing" of refugees and on the "characteristics" of IDPs with those of other non-displaced groups for analyses concerning the integration of and durable solutions for forcibly displaced populations. This non-displaced comparator population, however, is loosely defined, sometimes referred to as the "general/national population" and other times a subset thereof – such as "benchmark populations within the host country", "a subset of the general population", the "host community", or "host populations".

2. Reference to such comparator populations, associated with the concept of hosting, in both sets of statistical recommendations reflects the policy discourse concerning displacement. The "host community" has increasingly become a focus of forced displacement policy and response across the humanitarian, development and peace nexus, in order to reduce the perceived economic and social burden or negative impact of hosting displaced populations in specific areas or countries, reduce inter-communal tensions and enable integration in a more holistic fashion. For refugees, this is evidenced in the New York Declaration for Refugees and Migrants and the associated Comprehensive Refugee Response Framework, or CRRF (United Nations, 2016), and the Global Compact on Refugees, or GCR (United Nations, 2018). For IDPs, this reference is also central in the Inter-Agency Standing Committee (IASC) Framework on Durable Solutions for Internally Displaced Persons, the United Nations Secretary-General's Action Agenda on Internal Displacement (United Nations, 2022) as well as in regional frameworks such as the African Union Convention for the Protection and Assistance of Internally Displaced Persons in Africa or Kampala Convention (African Union, 2009).

3. Despite the prominence of this concept, various definitions of the host community exist and there is no tangible consensus on a single definition. For example, the United Nations Refugee Agency (UNHCR), refers to the host community as "a community that hosts large populations of refugees or internally displaced persons (IDPs) whether in camps, integrated into households, or independently" (UNHCR n.d.-b). In operational practice at UNHCR, the host community is often defined as non-displaced recipients of assistance. As a different example, the International Organization for Migration (IOM) uses a definition of the host community as "non-displaced people living in the locations where displaced population reside or those sharing public infrastructure, services, or resources with the displaced people" (IOM, 2024). These varying definitions emphasise different aspects of the concept and underscore that the term host community lacks standardisation across contexts, organisations and applications.

4. Whilst it is challenging to unify how the policy community defines a host community given that this term is referenced for different purposes (e.g. assistance and targeting) across different types of displacement situations (e.g. settlements including camps or non-settlements), there remains a

need to clarify the concept for the purpose of standardised official statistics on displacement. Statistically defining the comparator population is necessary to improve the harmonisation of displacement statistics and enable better evidence-driven policies concerning (re-)integration and durable solutions (EGRISS, 2018; EGRISS, 2020), both of which require comparative analysis or benchmarking against a population who is not displaced.

5. Following IRIS, this paper embarks on an effort to forge a ***statistical definition of the comparator population*** in contexts of forced displacement, considering both the national or general population and other concepts of the host community. Although aware of the policy discourse, the paper does not aim to define the broader concept of a host community for other purposes. To focus on a statistical definition, the paper starts with a discussion of the relevance of the host community and distinguishes the wider concept from that of a statistical comparator population. It then systematically maps out how these comparison groups are defined in practice across existing data collections from humanitarians, governments, development actors, and academic researchers. The mapped definitions include:

    1. The general or national population,
    2. The population in affected administrative areas,
    3. Distance-based definitions,
    4. Criteria-based definitions, and
    5. Other definitions.

6. The advantages and disadvantages of these definitions are then discussed with a specific focus on their suitability to produce official statistics. Based on this, the paper finally presents empirical evidence by implementing different comparator populations across four spotlight cases.

7. This methodological contribution concludes that the commonly used comparator populations – the national population or subsets of it – often yield comparable results across empirical cases. **For official statistics, the paper recommends using the national population or a definition based on affected administrative areas.** Other definitions could be used for other purposes or when data is unavailable, but require additional methodological guidance to standardise the definitions (e.g. defining what distance to use in a distance-based approach). The paper maps out the needed methodological decisions required to ensure a standardised use of these alternative definitions.

Background

The host community in policy frameworks

8. This subsection briefly reviews the concept of the host community in policy frameworks to provide a clearer context. The term "host community" has increasingly become a focus of humanitarian and development action in forced displacement contexts. Depending on the framing, it is often used inter-changeably with similar concepts such as resident population in a certain area, non-displaced population, affected population, receiving communities (including in areas of return and other settlement locations), host families, host households, or other similar terminologies. Some of these terminologies already imply relatively narrow or clear definitions – for example, host families or households indicate the people living in the same residence as displaced persons – whilst others are vaguer. The general or national population [1] is also sometimes referred to as the host community; however, this paper considers the host community as a more narrowly defined subset of the general population who are affected by forced displacement. Overall, from a brief desk review, it is clear that there is no consensus on a concrete definition of the host community in the policy context despite in wide usage and growing importance.

9. Various policy frameworks guiding humanitarian and development action to address forced displacement reference the host community as a crucial population group of concern. It is the host communities that carry the largest burden and who are also best placed to support solutions, and therefore they must not be over-looked. For refugees, this population group is prominently featured in the New York Declaration for Refugees and Migrants, which set out the CRRF (United Nations, 2016), having led to the GCR (United Nations, 2018).

10. The CRRF calls for the support of refugees and the host community in various domains, including needs, capacities, infrastructure, assessments of infrastructure and institutions, targeted interventions, and monitoring systems (UNHCR, 2019). Based on the CRRF, the GCR aims to ensure that refugees and their host community receive the needed support and are not left behind in progress towards the Sustainable Development Goals (SDGs). Given its focus on predictable and equitable responsibility-sharing of refugee situations and on easing pressures on host countries, the GCR provides a general overview of different definitions of the host community (UNHCR, 2019). According to the GCR, the host community can mean the population in a district, the physical proximity to refugee populations, infrastructure and service catchment areas, areas that sustainably absorb refugees, households that cohabitate with refugees, the areas in which non-displaced individuals' livelihoods are impacted by displacement as well as a comparison with other people with specific needs (see Box 1).

11. The concept of the host community has also gained momentum in the development community, with the Window for Host Communities and Refugees by the International Development Association (IDA), to create meaningful development opportunities and sustainable solutions for refugees and their hosts (IDA, n.d.). The World Bank also refers to the host community in its results framework, Scorecard, addressing the need to support the host community, and recommends using

[1] This paper used general population and national population interchangeably. It refers to the overall resident population of a country which is a well-known statistical concept.

subnational areas affected by displacement, that is, specific districts, counties, or other administrative units to ensure consistency across projects in the same country (World Bank, 2024).

Box 1. *Possible definitions of the host community in the GCR*

- Considering the entire population of districts as host community
- Physical proximity to refugee population
- Infrastructure and service catchment areas
- Absorption capacity and sustainability of refugee presence
- Cohabitating/co-existing with refugees
- Livelihood impact
- Persons with specific needs[2]

12. For IDPs, although the 1998 Guiding Principles on Internal Displacement are not explicit about the role of host communities (United Nations, 1998), their focus on non-discrimination, equal participation in public affairs and equal access to services (principle 29) implies the need for comparing the situation of IDPs and non-IDPs in the context of solutions. The later-developed IASC Framework on Durable Solutions for Internally Displaced Persons – arguably one of the most well-used IDP policy documents internationally – became more explicit in describing the need to consider the "affected population, including host and receiving communities" in all locations, (i.e. places of return, local integration, or settlement elsewhere) as a key principle for any meaningful durable solutions process as their "needs may be comparable and should not be neglected". The IASC states that "IDP-specific strategies should also take into account the needs of resident populations who share the burdens of displacement. These include host communities and host families that took in and supported displaced families as well as communities that receive IDPs who return, locally integrate, or settle elsewhere in the country" (Brookings Institute - University of Bern Project on Internal Displacement, 2010). Similarly, the Kampala Convention – a legally binding framework for African State Parties - contains a provision that obligates states to assess the needs and vulnerabilities of IDPs and those of their host community (African Union, 2009).

13. More recently, following efforts to re-invigorate global attention on internal displacement and re-energise international commitments to enabling durable solutions for IDPs, the role of host communities in this context has become even more prominent. The UN High Level Panel's Report on Internal Displacement includes a total of 51 references to host communities, although no definition is offered (United Nations, 2021). Also, the associated UN Secretary-General's Action Agenda on Internal Displacement further recognised the critical concept of a host community through the central guiding considerations of the overall agenda that calls for action that "recognize(s) the rights and agency of IDPs and host communities" and explains that "Action on internal displacement must be part of a whole-of-displacement approach that also considers the rights and needs of individuals who fled across international borders, individuals who returned after cross-border displacement and host communities" (United Nations, 2022).

[2] Persons with specific needs refers to a subset of a host community as one of the above possible definitions in Box 1.

14. As demonstrated through this brief overview, different policy frameworks and organisations increasingly incorporate the role of a host community in displacement contexts, however, a clear and common definition remains a gap. As statistical efforts strive to align with policy frameworks to enhance the impact of analysis produced, this gap constitutes a major challenge in efforts to develop a harmonised statistical definition for a comparator population to facilitate analysis of protection and wellbeing across different displacement situations.

The need for a statistical definition of comparator populations for refugees and IDPs

15. The need for a statistical definition of comparator populations is particularly critical to the production of official statistics on forcibly displaced populations, where comparability and accuracy is paramount. Given the plethora of definitions for the host community derived from the policy environment, it is critical to delink the wider concept of the host community from the more specific purpose of statistical application where the situation of displaced populations needs to be compared to other groups. This paper focuses on identifying and defining a comparator population for statistical purposes.

16. To guide the production of internationally comparable statistics on refugees and IDPs, EGRISS has developed two sets of internationally endorsed statistical recommendations, namely the IRRS and the IRIS, which were endorsed by the UN Statistical Commission in 2018 and 2020, respectively[3]. Both sets of recommendations highlight the value of comparing the vulnerability situation of refugees and IDPs with that of other population groups, however, both stop short of providing a fully-fledged definition. The IRIS makes clear that no definition of the benchmark population has yet been developed for statistical purposes (EGRISS, 2020), and the IRRS emphasises the importance of establishing a definition for a benchmark population against which to compare the situation of refugees in a given country (EGRISS, 2018). See Box 2 for the references to comparator populations from the IRRS and the IRIS.

17. The IRRS names this population for comparison as a "benchmark population within the host country" already implying the comparison population should be a subset of the general population (EGRISS, 2018). The IRIS describes the host community also as a subset of the general population, but given challenges surrounding this concept recommends using the general population for the purpose of official statistics (EGRISS, 2020). As an example, the IRIS recommends that IDPs are compared to the general population on core dimensions of vulnerability, such as shelter and housing conditions or economic self-reliance, to assess if a proportion of the IDP population has successfully overcome displacement-related vulnerabilities and to assess progress towards durable solutions (EGRISS, 2020). The logic of this assessment is that the comparator population sets the relative thresholds to determine if IDPs overcome their displacement-specific vulnerabilities for statistical purposes in line with the IASC Framework as described above. Thus, the comparator population becomes an integral part of how official measures of durable solutions and exits from the stock of IDPs should be conducted.

[3] The International Recommendations on Statelessness Statistics (IROSS) are the third set of recommendations developed by EGRISS, but the IROSS do not specify a comparator population and are hence outside of the scope of this methodological discussion.

Box 2. *Key references to comparator populations in the International Recommendation on IDP Statistics and the International Recommendations on Refugee Statistics*

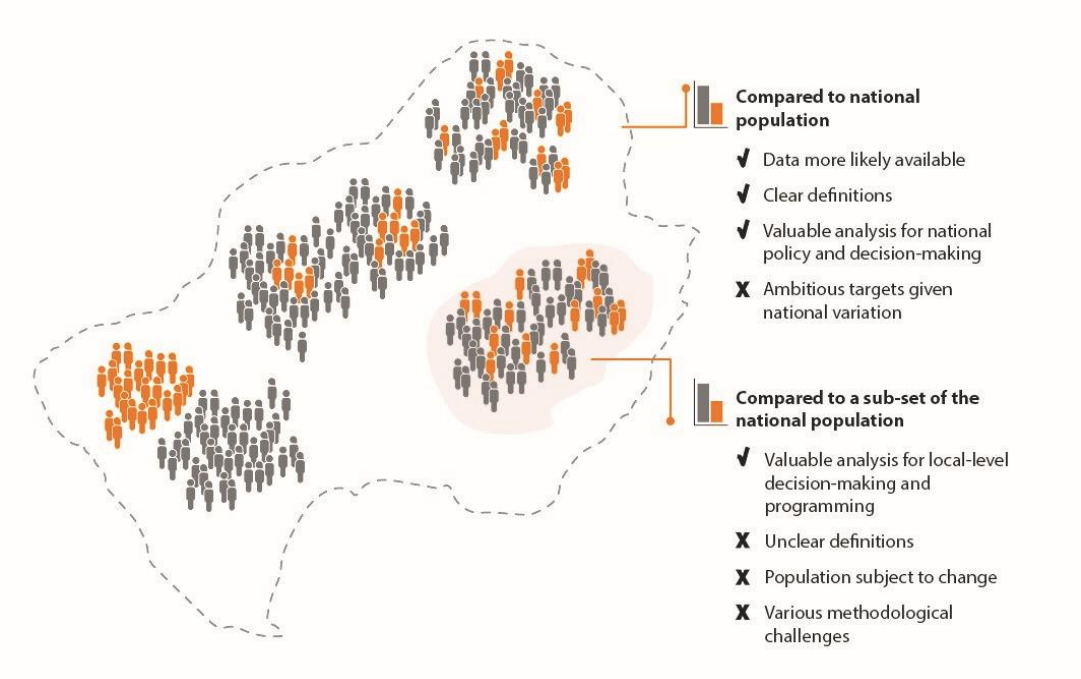
> *"Given the specificities of different displacement contexts, it can also be valuable to compare the vulnerability situation of IDPs with that of a different subset of the general population, commonly referred to as the 'host community' by governments and humanitarian and development organisations responding to internal displacement. This can provide valuable analysis for local-level decision-making and programme design; however, for the purposes of official statistics there are some challenges associated with this method. [...] The first challenge is that the host population has so far not been defined and different approaches are taken in practice. [...]. Common approaches include identifying non-displaced households that live in close proximity to IDPs or within the geographical area where IDPs reside; identifying non-displaced households within these geographical areas that have similar living conditions to those of IDPs; specifying a more literal meaning of hosting by identifying non-displaced households that physically host IDPs in their own homes; or identifying populations living in the surrounding area of IDP camps and settlements. [...]. Overall, for the purpose of official statistics, the general/national population is recommended as a comparison group when analysing the displacement-related vulnerabilities of IDPs."*
>
> *- International Recommendations on IDP Statistics, paragraph 142*

> *"In order to assess the basic needs, living conditions and integration of refugees, it is important to establish benchmark populations within the host country who can be compared with the refugee population. The characteristics of the total population can then be used to measure the conditions of the refugee population relative to other residents in the host country. This requires surveys and censuses that include both refugees and the host population."*
>
> *- International Recommendations on Refugee Statistics, paragraph 454*

18. When it comes to the production of official statistics on forcibly displaced populations with a comparator population, the application of different concepts and definitions can be problematic and necessitates standardisation to enable comparisons across countries and over time. For the purpose of official statistics, the IRIS hence recommends the general or national population as a comparison group when analysing the displacement-related vulnerabilities of IDPs. The key reason for this recommendation is that information about the general population is more likely to be available, making data collection and analysis simpler. More guidance is needed, however, on the choice of the comparator population, in particular if data from the national population is not available to use as a benchmark.

19. In specific circumstances, for example local decision-making or producing official statistics at the subnational level, comparisons with a subset of the national population (e.g. the non-displaced population in a given area) may also be considered. For example, if a humanitarian actor is planning a local initiative to improve access to schools for forcibly displaced populations in a specific geographical area, using information on the access to schools for host community members in this locality may be a better operational benchmark than the access to schools of the entire population. This approach is often considered to be the most relevant and feasible approach for operational purposes (EGRISS, 2020). See Figure 1 for an illustration of these options.

Figure 1. *Advantages and disadvantages for identifying comparative population groups*



Source: *International Recommendations on IDP Statistics (IRIS)*

Mapping of empirical definitions

20. Operational data collection has increasingly included both displaced and comparator populations in situational analyses, comparative reports and planning for assistance to accommodate the rising policy demand to compare the situation of displaced populations with others. Examples include the Refugee and Host Communities Household Surveys in Uganda (World Bank, 2018), the Survey of Syrian Refugees and Host Communities in Jordan (World Bank, 2019), the Forced Displacement Survey in South Sudan (UNHCR, 2024), and the Durable Solutions Analysis Survey in North Darfur (UNHCR and JIPS, 2021a) and in South Darfur (UNHCR and JIPS, 2021b). The joint investigation by UNHCR and the World Bank of the poverty impact of COVID-19 in Jordan, Lebanon, and the Kurdistan Region of Iraq demonstrates a use case enabled by comparable data between forcibly displaced people and comparator populations across different countries and areas (Joint Data Center, World Bank, and UNHCR, 2020). Similar initiatives have emerged that collect nationally representative data on both the comparator populations but also relevant forcibly displaced populations. For example, the Jordan Demographic and Health Survey (DHS) 2017-18 (Department of Statistics and ICF, 2019), the Colombia Encuesta Nacional de Calidad de Vida (Departamento Administrativo Nacional de Estadística, 2019), and the Uganda DHS 2019 (Uganda Bureau of Statistics, 2023) are nationally representative household surveys including samples of forcibly displaced populations in each country.

21. This section systematically describes the definition of host community used in datasets included in UNHCR's Microdata Library (as of September 2024), a collection of academic studies by Verme and Schuettler (2021) and a smaller collection of case studies shared by members of EGRISS' TSG2 during the development of this paper. The Microdata Library of UNHCR is an online data repository containing microdata on forcibly displaced populations and other relevant groups, which the

authors considered as a key repository to help identify the definitions of the host community used in quantitative data processes in practice. As of 6 September 2024, UNHCR's Microdata Library contained 845 datasets. The number of datasets was then narrowed down to 144 by searching for the keyword of "host". Out of 144 datasets, 88 do not have data of comparator populations while containing the keyword of "host" in their descriptions, and they are classified as false positive in Table 1.

22. Verme and Schuettler's (2021) review of 59 empirical papers that studied the impact of forced displacement on host communities in the economics literature, covering various countries and definitions of the host community in academic research, also stood out as a relevant resource for this mapping. Lastly, contributors of the paper identified ten additional datasets to review (see Table 1). Importantly, due to the sources used, nationally representative surveys and census-like data collections that also include displaced persons by National Statistical Offices are not well captured here although some examples are included. These data collections often use the full national population as comparator populations for IDPs and refugees. As more nationally representative datasets including forcibly displaced populations become available, further examples could be included in this mapping effort to update findings accordingly[4].

23. The definition of host community used in applied work found in the above sources – essentially a collection of both operational and statistical research – were then classified into five categories to provide an overview of the types of definitions commonly used (see Box 3).

Box 3. *Category of definitions of the host community*

1. National population;
2. Population in affected administrative areas;
3. Physical proximity;
4. Cohabitation or criteria-based definition; and
5. Other definitions

24. The most commonly used definition is the entire population in affected administrative areas at the subnational level. We found 70 out of 213 studies using this definition. For example, Refugee and Host Communities Household Survey 2018 in Uganda defines the host population as the native population in districts where refugee settlements are situated (World Bank, 2018). Ceritoglu et al. (2017) compare heavily affected administrative areas by refugee influx ("treatment" areas) and less affected areas ("control" areas). Some studies define the host community in some subnational areas which are not administrative areas such as commuting zones in the United States (Mayda, Parsons, Peri, & Wagner, 2017), and enumeration areas in Niger (Coulibaly, Hoogeveen, Hopper, & Savadogo, 2024). Physical proximity is also common with 25 examples identified. For instance, the Joint Assessment Mission 2020-21 in Mozambique sampled the host community within a 6km radius from the Maratane settlement in Nampula, Mozambique (UNHCR and WFP, 2022). Betts, et

[4] The **EGRISS GAIN Survey** monitors statistical projects related to forcibly displaced people and statelessness within national data systems or using the IRRS, IRIS, and IROSS, covering from 2020 to 2024 at the time of writing. GAIN Survey results could help inform future efforts to update the current mapping given the increasing number of national surveys and census being reported.

al. (2023) also use population living nearby refugees as the host community in Uganda, Kenya, and Ethiopia. Other definitions are those different from the first four definitions in Box 3 such as infrastructure and service catchment areas. See Table 1 for the summary of mapping results.

25. The full mapping results are presented in Appendices (see Appendix 1, 2, and 3). The applicability of these definitions and their usages are discussed in the following section.

Table 1. *Definitions of host communities of datasets from the UNHCR Microdata Library, Verme and Schuettler (2021), and other sources*

Classification of the definitions	UNHCR Microdata Library	Verme and Schuettler (2021)	Additional sources	Total
1. National population	3	5	1	**9**
2. Population in affected administrative areas	21	48	1	**70**
3. Physical proximity	18	4	3	**25**
4. Cohabitation or criteria-based	3	-	-	**3**
5. Other definitions	7	-	-	**7**
National and affected administrative areas (1 and 2 combined)	-	2	3	**5**
Physical proximity and criteria-based definitions (3 and 4 combined)	-	-	2	**2**
Unclear	4	-	-	**4**
False positive	88	-	-	**88**
Total	**144**	**59**	**10**	**213**

Evaluation of a statistical definition of the comparator populations

26. Which definitions of the comparator population, from the various approaches used in practice and research, are suitable for the purpose of generating official statistics on forced displacement? This section uses three main criteria to evaluate which definition may be recommended: the definition should allow for standardisation across contexts; should provide a good representation of the actual host community in statistical terms; and should not be volatile to changes over time.

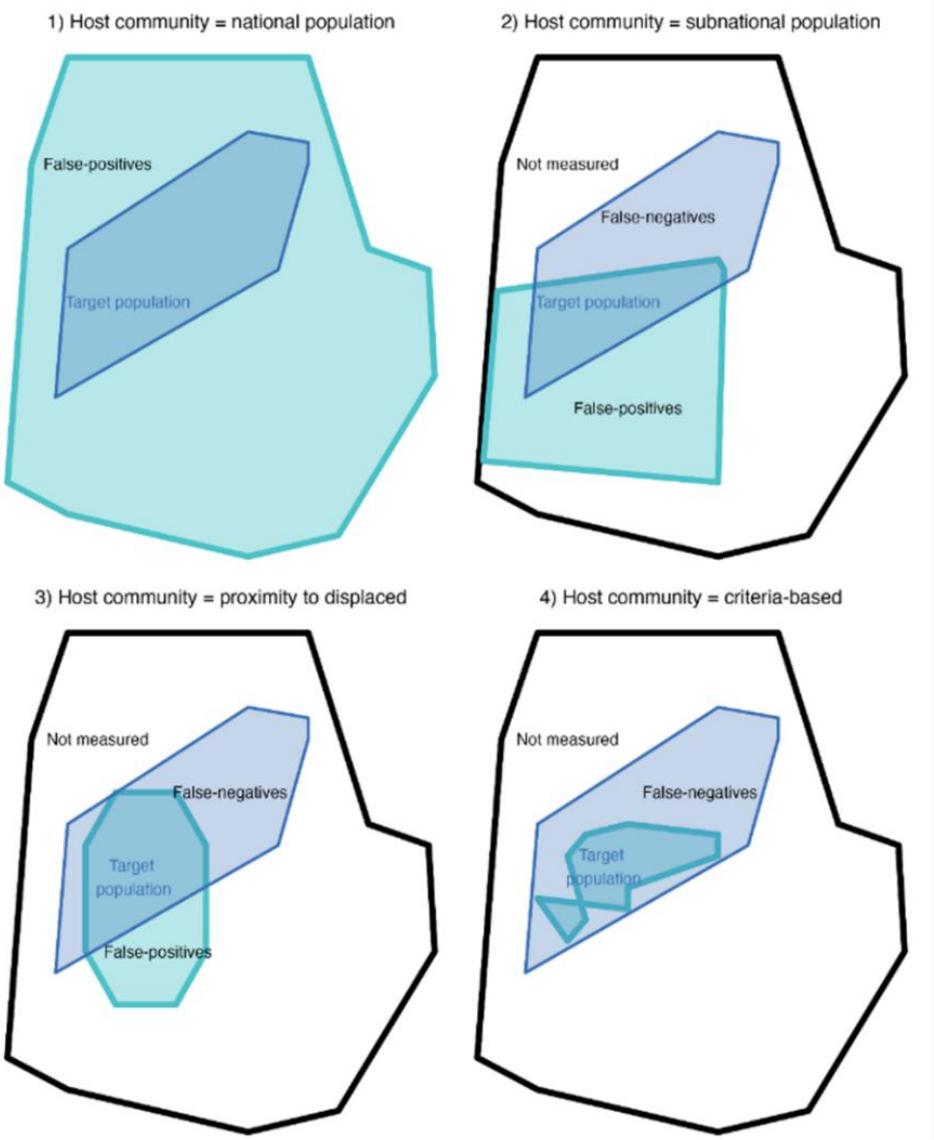
Evaluation criteria for comparator populations in official statistics

27. Official statistics must be objective, reliable, and comparable across time and context, but they must also be realistically implementable across various data environments. This section outlines three criteria used to assess the suitability of a comparator population definition for the purpose of producing official statistics on forced displacement:

- **Standardisation potential:** A central aim of national statistics on forced displacement is to have comparative numbers across different country contexts. In other words, a comparison of forcibly displaced people to a comparator population in one context to another context should reveal differences between the contexts but not differences in the measurement practice itself. To achieve comparability, the comparator population should be general. Choosing a definition that can only be implemented in some but not in other contexts does not allow for standardisation. Also, a statistical measurement approach for the comparator population must be simple so contexts with limited resources can still conduct benchmarking exercises.

- **Benchmark bias:** Statistical representations and measurements translate larger concepts, in this case the concept of the host community, into statistical practices so they can be implemented through quantitative data systems. A good statistical definition of the comparator population is an unbiased translation of the host community into a measurement strategy. Bias in the context of identifying a comparator population refers to the risk of reporting a better or worse performance of the comparator population on key statistical indicators in comparison to the theoretical population of the host community. The more measurement bias, the more distorted the comparison to forcibly displaced population becomes. A major challenge in identifying a statistical definition of the comparator population that has a low benchmark bias is the continued conceptual contestation of what constitutes the host community, which also makes the translation into a statistical concept contested. The different definitions used in practice vary in their level of inclusiveness. While some definitions are general and potentially include a larger proportion of the local national population as the host community, other are more exclusive and identify a smaller and specific set as the host community within a national population. This has implications for the size of the host community, as well as for the potential bias introduced when using the identified comparator population as a benchmark in vulnerability assessments. This paper assumes a theoretical concept of the host community as those who are not displaced but affected by forced displacement to describe the benchmark bias, but this concept is controversial, and other concepts could be used. Compared to a theoretical concept of the host community, also referred as “target population” in this paper, inclusive statistical definitions for the comparator population **include “false positives” referring to the households or areas that are theoretically not part of the host community** but are counted as host community in the created statistic. This sets an ambitious benchmark because “false positive” host community members may fare significantly better than forcibly displaced populations. In contrast, very exclusive definitions that only include host community members that fulfil certain criteria risk excluding households that are **part of the theoretical concept of a host community as “false negatives”.** This may risk making assessments of the vulnerability of forcibly displaced populations only in comparison to particularly vulnerable hosts, introducing a low benchmark and potentially biasing the results. There is hence a critical trade-off between the inclusion of false positive households and the exclusion of false negative households in the statistical definition of the comparator population. Any chosen definition should minimise this benchmark bias (see Figure 2).

Figure 2. *Visualisation of the benchmark bias in different definitions of the comparator population*



Notes: *Imagine a country as outlined in black. The blue area describes the theoretical host community that we aim to operationalize in statistics. The turquoise area describes households that are defined as hosts per definition. This visualisation does not present 5) Other definitions.*

- **Volatility risks:** Related to the potential bias introduced by the inclusiveness of each definition, we also need to consider the extent to which an employed definition is volatile to changes. In an ideal scenario, the comparator population – that is the size and the individual households making up the benchmark population – remains relatively constant across data collections. Definitions that are very flexible and exclusive may imply that only certain households and different households over time constitute the host community. General and inclusive definitions of the comparator population are more stable over time. For a statistical definition, we may want to choose a definition in which the benchmark population is not continuously changing but in which the composition of the comparators is stable to facilitate comparison. In repeated comparison and benchmark exercises, any recorded progress of forcibly displaced populations in comparison to the comparator population of hosts should be due to true progress of the forcibly displaced population and not due to a change in what constitutes the benchmark. We hence discuss to what extent volatility is a risk in commonly used definitions.

28. Good statistical definitions have high standardisation potential, low benchmark bias, and low volatility risks. The following subsections evaluate the five comparator population definitions identified through our mapping exercise against these three criteria including other definitions.

1. National population

29. For official statistics, IRIS highlights that the usage of a representative sample of the national population or a census is beneficial as a benchmark to assess displacement-related vulnerabilities (EGRISS, 2020). **Several factors support this approach.** First, in many settings, data on the general population is typically accessible thanks to ongoing monitoring efforts linked to the Sustainable Development Goals or country-specific information systems. Moreover, the proposed definition offers a clear and transparent distinction between host and non-host populations by designating the entire national population as the comparator group. This also means that new data collection efforts can rely on existing sampling frames to gather benchmark information. **Given both the accessibility of data and the clarity of the definition, designating the national population as the comparator group presents strong potential for standardisation.** This approach is comparatively more stable than alternative comparator populations and can serve as a consistent reference point to guide national planning over time. Although the areas in a country affected by displacement may change over time, using the entire population as a benchmark means that the same population will continuously be included for comparison. As such, the benchmark population remains the same and only their performance on key vulnerability and well-being criteria changes over time. For example, Angrist and Kugler (2003) use the national population as the comparator population to estimate the effect of labour migration on European countries, while Carrington and de Lima (1996) and Mäkelä (2017) also use the national averages in the context of migrants from Angola and Mozambique in Portugal, respectively.

30. Despite the simplicity and robustness, the national population is not often used as a key comparator population across operational and research contexts as Table 1 suggests. One reason that many studies and operations apply other definitions may lie in the potentially high benchmark bias introduced by this definition.

31. Panel 1 in Figure 2 shows a fictive country in which the population in blue is a non-displaced population that is truly impacted by forced displacement. This is the theoretical host community for the purpose of visualising this benchmark bias. In defining the whole national population as the comparator (see turquoise shape), we introduce many false positives: A substantial proportion of the population has no exposure to displaced individuals, is neither negatively or positively impacted by the presence of forcibly displaced persons in their communities, but is nonetheless used to assess the progress towards solutions for forcibly displaced people.

32. This approach may set ambitious targets for evaluating the vulnerability of forcibly displaced populations relative to the comparator populations from a theoretical perspective. For instance, benchmark values may be shaped by individuals living in more stable regions or closer to the capital, which could inflate the national standard. Nonetheless, the national benchmark is often used as a reference point for overcoming vulnerabilities, in line with IRIS recommendations. In the empirical analysis below, we examine to what extent national benchmarks do indeed sets a high standard. We find no significant difference, at least given the available data and the contexts analysed.

2. Population in affected administrative areas

33. A more commonly adopted approach in our mapping reverts to a sub-national definition of the comparator population, whereby all individuals residing in a specific administrative area or within certain geographic boundary are designated the comparator population. For example, the 2018 Refugee and Host Communities Household Survey by the World Bank in Uganda defines the "host population" as the native population in districts in which refugee settlements were situated. Morales (2018) identifies the comparator population level as all individuals in displacement-affected municipalities in Colombia.

34. There are considerable challenges when trying to standardise this definition across contexts. First, it is unclear which geographical unit may be best suited as a benchmark population since geographical and administrative units vary distinctively across countries; a simple recommendation to use a certain administrative unit may not be feasible across contexts. More detailed methodological guidance is needed to identify the right level of administrative units proportional to the overall country size. Standardisation may be even more challenging when other types of geographical areas are used such as commuting zones or enumeration areas rather than administrative units. Another aspect requiring more guidance regarding standardisation is the distinction between displacement-affected and non-affected administrative units or catchment areas. In contexts where forcibly displaced populations reside in camps with known locations, identifying relevant administrative units is relatively straightforward. However, in other contexts which are characterised by informal settlements, integration into host communities beyond camps and continuously changing housing patterns – such delineation becomes far more complex as these conditions obscure clear geographic boundaries and complicate efforts to consistently define host units for statistical or operational purposes. Nonetheless, with proper guidance on how to identify the location of significant displaced populations and the right level of administrative aggregation, the feasibility of this approach is generally high. A disadvantage may be that no subnational sampling frame exists in certain contexts, but with the increasing availability of census data and subnational reporting, many such issues can be solved.

35. While a national definition outperforms a subnational definition in terms of its standardisation potential, a subnational definition reduces the potential benchmark bias. Panel 2 in Figure 2 shows that capturing a specific administrative area (visualised in turquoise) reduces false positives. Sampling only certain administrative units as a comparator population successfully excludes areas that are not directly affected by displacement. A downside is that certain areas that may only see smaller numbers of displaced populations or have been misclassified as non-affected in an assessment of hosting areas may no longer be sampled, introducing false negatives: households that are in fact affected by displacement are excluded from the comparator population. With sufficient information on the location of forcibly displaced populations, the number of such false negatives can be reduced. With regards to the actual bias introduced in the benchmarking, we can expect some bias from the excluded and included populations that are not part of the "true" theoretical comparator population, but this bias may be lower than in the case of a nation-wide definition.

36. In comparison to the definition of the comparator population as the national population, the subnational approach displays higher volatility risks. In dynamic displacement scenarios, the areas in which IDPs or refugees reside may change significantly over time. When information on displacement locations is not updated in information management systems, the wrong

administrative units may be sampled as host communities, introducing a greater benchmark bias. On the other hand, when information on displacement locations is updated, the benchmark population shifts over time, meaning that a household may be part of the comparator population in one round of data collection but not in a subsequent one. Hence, not only the comparator population's performance on key indicators changes but also who contributes data to these indicators changes over time, potentially leading to some inconsistencies over time. For example, forcibly displaced populations may improve their access to education by moving to new areas but subsequently perform worse in comparison to this new comparator population. Such shifts in the comparator population require continuously updated sampling frames. It is important to note, however, that this problem may be overstated depending on the geographical granularity. While the exact location of forcibly displaced populations may change over time, the affected administrative units may be more stable.

3. Population in physical proximity

37. A third type of commonly applied definition is the use of a distance-based approach in which administrative units are not the main identifier of the comparator population but the population living within a certain radius around displacement-affected locations, especially in forced displacement contexts in which a large number of refugees and/or IDPs are hosted in camps or settlements. For example, Loschman et al. (2019) randomly select communities within a 10km and 20km distance from the Gihembe, Kigeme and Kiziba camps in Rwanda to determine the comparator population.

38. Conceptually, this approach is very similar to the approach focusing on subnational administrative units but has its own advantages and disadvantages. The core theoretical advantage of distance-based approaches is that a camp located at an administrative border likely affects the population in the administrative unit that the camp falls within, as well as the population located across the administrative border. In theory, this could reduce the potential benchmark bias because the number of false positive and false negative host community members declines (see panel 3 in Figure 2). However, this is fundamentally based on the assumption that an updated list of major displacement locations exists which allows for fine-grained distinctions between host communities and non-hosts based on geography and distance. Geographical barriers and infrastructure such as rivers and roads should also be considered in this approach since they affect the interaction between forcibly displaced people and the comparator population, thus affecting the benchmark bias.

39. Similar to definitions of the comparator population based on administrative units, there is a need to provide further guidance on how to best identify the physical proximity that qualifies towns and villages as part of the comparator population. Without standardised guidance – for example, recommendations for a specific ratio between overall country size, average travel speed, and distance to the main settlements, proximity-based definitions are not easily standardised across different contexts. It remains unclear to what extent existing sampling frames and population information can be used to select such distance-based definition of comparator populations to sample for surveys. While this may be easy to implement in certain contexts, in others, no accurate and up to date geo-located list of settlements and households may exist, requiring listing and census exercises.

40. In summary, and despite the conceptual similarities with administrative definitions of the comparator population, distance-based approaches reduce the potential benchmark bias but are slightly harder to implement and standardize across contexts. In addition, they are more difficult to apply, particularly in dynamic displacement scenarios as it is crucial that the core areas in which refugees and IDPs reside are known for this definition. Even then, and as is the case with the administrative definition of the comparator population, the volatility risks are high; changes in displacement locations mean that individuals included as hosts at one point in time may not host refugees or IDPs at a subsequent point in time anymore.

4. Cohabitation or criterion-based population

41. There are other approaches in which the comparator population is defined based on specific criteria, in particular co-habitation. The South Sudan Forced Displacement Survey (FDS) 2023 is such an example. The comparator population can encompass all households that have hosted or are currently hosting forcibly displaced populations within their households or dwellings. Such a definition of the comparator population requires identification in household surveys and can hence be costly and time-consuming as a larger population must be surveyed to identify the target population. The conceptual advantage of this approach is that hosting displaced individuals as an inclusion criterion eliminates false positives. There may still be cases in which not all hosts are correctly identified and are thus missed (false negatives, see panel 4 of Figure 2).

42. Overall, cohabitation or criterion-based definitions of the comparator population are theoretically appealing because they reduce benchmark bias, and the criteria can be applied across contexts, but they are operationally hard to implement and suffer from low feasibility. Finally, they may also be volatile to changes over time as some households may stop hosting while others may start hosting displaced individuals, moving the benchmark and who constitutes the comparator population. Another limitation is this definition's inability to be used for household-level indicators as displaced individuals and non-displaced individuals cohabitate and constitute one household. Due to these operational difficulties, application of such a criterion-based definition is rather rare or only implemented in addition to geography-based definitions.

5. Other definitions

43. The GCR highlights a number of additional definitions as listed in Box 1, such as using infrastructure service or catchment areas, the population having livelihood impacted by forced displacement, focusing on comparisons with persons with specific needs, or focusing on the absorption capacity of areas that host refugees. In practice, neither of these additional definitions are frequently used. They share a common problem which is that they are often highly specific to a policy area (e.g. catchment areas of hospitals, comparisons to people with mental health needs, etc.) and are hence less suitable for statistical analysis that captures multiple policy areas, which would be needed for the composite index on overcoming of displacement-specific vulnerabilities. In addition, these definitions are volatile to changes and hard to standardise across country contexts, thus making them ill-suited for official statistics.

Summary of the theoretical advantages and disadvantages of comparator population definitions

44. A summary of the advantages and disadvantages of commonly used definitions for the comparator population can be found in Table 2. This assessment considers the use of definitions for the production of official statistics and may not apply to other more operational goals.

Table 2. Summary of advantages and disadvantages of the identified definitions

	Definition	Advantages	Disadvantages
1.	**National population**	**High standardisation potential:** Clear definition, easy to implement, data likely available, and valuable for national planning. **Low volatility:** Households in the comparator population stock remain stable. National population has a clear advantage as a comparator population when it comes to setting targets and thresholds to analyse the displacement-related vulnerabilities, especially for IDPs. This population group is also widely used for the national-level comparison such as SDG indicators. In case of including forcibly displaced population into national censuses or household surveys, cost implication is relatively low.	**High benchmark bias:** Introduction of many false positives (not all households are affected by forced displacement). This may cause overly ambitious targets. **Feasibility:** Potentially expensive if the whole country has to be surveyed and sampled to identify benchmark values for operational purposes.
2.	**Population in affected administrative areas**	**Medium standardisation potential:** Clear definition, mostly easy to implement, valuable for subnational planning, but requires guidance on which administrative levels to select across contexts. **Relatively low volatility:** Most households remain part of the comparator population stock over time. Similarly to the national population, in case of including forcibly displaced population into national censuses or household surveys, cost implication is relatively low.	**Relatively high benchmark bias:** Many individuals in administrative units are false positives and individuals in excluded admin units are false negatives. **Medium standardisation potential:** Unclear which level of administrative area to use/can make comparisons more uncertain for situations with less standardised admin units at the local level. It may cause ambitious targets. **Feasibility:** Uncertain needs for subnational sampling frames in dynamic forced displacement situations. Affected administrative areas could change over time. In case of expanding forced displacement situations, it may cause ambitious targets for operational purposes
3.	**Physical proximity**	**Medium standardisation potential:** Easy to implement once agreement on distance threshold and "significant population threshold" is reached.	**High volatility:** Sensitive to definition of physical proximity, hard to standardise maximal distance from e.g. camp to settlement given that travel distance and size of countries vary profoundly.

			Medium benchmark bias: Hard to apply to non-camp dwelling populations or populations with less clear settlement patterns; distance to camps for example may leave out benchmark host communities that have informally or spontaneously absorbed forcibly displaced populations but are far from camps.
4.	**Cohabitation or Criteria-based**	**Conceptually highly specific:** captures theoretical concept of the comparator population to forcibly displaced population.	**Low feasibility:** Requires potentially expensive pre-screening/household listing on who is hosting; difficult for populations living in camps and who may not clearly share accommodation or land with hosts. **Low benchmark bias:** Ideally few false positives but potentially sets benchmark for vulnerability performance rather low as conflict-affected communities are also more likely to host; potentially a lot of false negatives: individuals who are not hosting displaced individuals themselves but whose communities are heavily impacted by displacement are not counted as part of the comparator population. **High volatility**: Sensitive to criterion. Certain households may exit and enter the comparator population because they do or do not fulfil criteria at the time of data collection.
5.	**Other definitions**	Useful for specific targeted intervention designs or analyses Conceptually appealing emphasis on resource sharing problem	**Low standardisation potential:** Ill-suited for general population statistics as identification criteria keep changing; Definition varies potentially by policy area (e.g. users of the same hospital to assess access to doctors or access to schools for educational purposes). **High volatility:** Potentially prone to continuous changes in the benchmark depending on local development (e.g. new schools) or destruction (e.g. schools destroyed). **Feasibility:** Hardly used in practice.

Empirical evidence on statistical definitions

45. What are the consequences of using different comparator populations in practice? In this section, four selected datasets are analysed with the aim to understand if some definitions of the comparator population are more suitable in practice, yield particularly high benchmarks for forcibly displaced populations or result in low benchmarks. The datasets are selected based on the coverage of different world regions, contexts, applicable definitions, and available indicators (see Table 3). All datasets allow a comparison of the application of at least two different comparator population definitions. The analysis is focused on indicators that are commonly used in humanitarian and development assessments regarding the wellbeing and protection of forcibly displaced populations (see Table 4).

Table 3. *Selected datasets for statistical definitions*

Dataset	Population type of forced displacement	Context of forced displacement shelters*	Available definitions
Jordan Demographic and Health Survey 2017-18	Refugees and asylum-seekers	Settlements	1) National population
			2) Population in administrative areas
Colombia Encuesta Nacional de Calidad de Vida 2019	IDPs	Non-settlements	1) National population
			2) Population in administrative areas
South Sudan Forced Displacement Survey 2023	Refugees and asylum-seekers	Mix of settlements and non-settlements	3) Physical proximity
			4) Cohabitation or criteria-based
Dem. Rep. of the Congo Socio-economic survey in Grand Kasai 2022-2023	IDPs	Non-settlements	3) Physical proximity
			4) Cohabitation or criteria-based

Table 4. *Selected indicators of the selected datasets*

Dataset	Indicators
Jordan Demographic and Health Survey 2017-18	Proportion of births attended by skilled health personnel (SDG 3.1.2)
	Years of education
Colombia Encuesta Nacional de Calidad de Vida 2019	Proportion of population that feel safe walking alone around the area they live (SDG 16.1.4)
	Household income per capita
South Sudan Forced Displacement Survey 2023	Proportion of population using improved drinking water sources[5]
	Household food insecurity experience scale
Dem. Rep. of the Congo Socio-economic survey in Grand Kasai 2022-2023	Proportion of population using improved drinking water sources[6]
	Household food insecurity experience scale

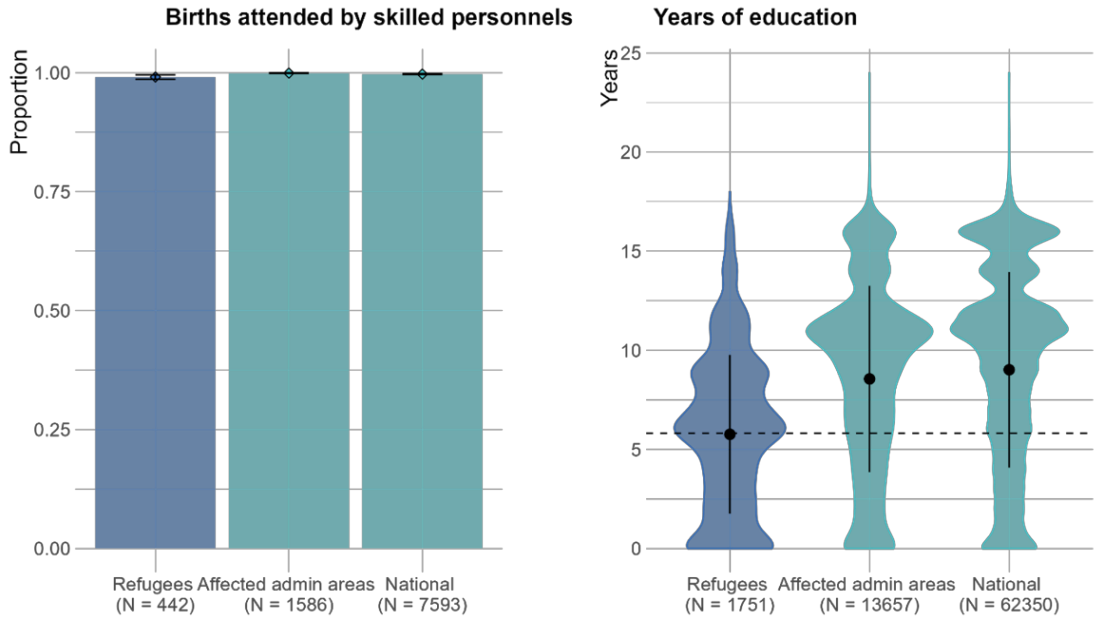
[5] Without water testing, the indicator does not meet the standard of SDG 6.1.1, Proportion of population using improved drinking water sources.

[6] Same as above, the survey did conduct water testing, therefore the indicator is not SDG 6.1.1.

Spotlight I: Jordan Demographic and Health Survey 2017-18

46. Jordan's Department of Statistics conducted the seventh DHS in 2017-2018. This comprehensive household survey covered the national population, providing an opportunity to analyse data on how refugees in camp settings compare to the general population and those living in administrative areas known to host refugees. Since identification questions were not included in this edition of the DHS, identifying the refugee population relied on the assumption that they are of Syrian nationality and reside in the governorates of Zarqa and Mafraq, which include major Syrian refugee camps. This assumption is based on the understanding that the majority of refugees in Jordan are of Syrian nationality and that the sampling approach of Jordan DHS 2017-18 considered each Syrian camp in the governorates of Zarqa and Mafraq as independent sampling strata (Department of Statistics and ICF, 2019).

Figure 3. Proportion of population with a birth attended by a skilled health personnel and single years of education in Jordan DHS 2017-18



Source: Jordan DHS 2017-18

47. We focus on two definitions of the comparator population: We compared the refugee population to the "host population" in administrative areas, focusing on non-refugee populations living in Zarqa and Mafraq (definition 2), as well as the national population (definition 1). Our analysis considered two key criteria: first, we evaluated the proportion of the population that had attained a high level of education, specifically those who completed secondary education or higher; second, we assessed the proportion of births attended by skilled healthcare personnel.

48. The right panel in Figure 3 compares the years of education completed across refugees, populations in displacement-affected administrative areas and the national population. In this dataset, completing five years of education is equivalent to completing primary school education. The figure shows that refugees completed fewer years of education in comparison to the national population and populations in affected administrative areas. In this case, we see that the definitions used for the comparator population yield slightly different but comparable results.

49. The left panel in Figure 3 shows that, across the three population groups — refugees, those in affected administrative areas, and nationals — a large proportion of births in Jordan between 2017 and 2018 were attended by skilled health personnel. In this analysis, skilled health personnel include doctors, nurses, and midwives. The data indicate that a slightly lower proportion of refugee births were attended by skilled personnel. Jordan's similarly high rate of births attended by skilled personnel across all three groups can be attributed to several factors, including the Jordanian government's investment in the healthcare system, which ensures that medical facilities are well-equipped and services accessible, as well as health policies aimed at improving maternal and child health.

50. This analysis thus suggests that, in this particular context, using the definition of the comparator population as the national population and the population in affected administrative areas yields similar results. However, differences observed in the proportion of the population that attained higher levels of education indicate that the specific variable being analysed can impact outcomes for each population group. This underscores the importance of adopting a context-specific approach when defining comparator populations.

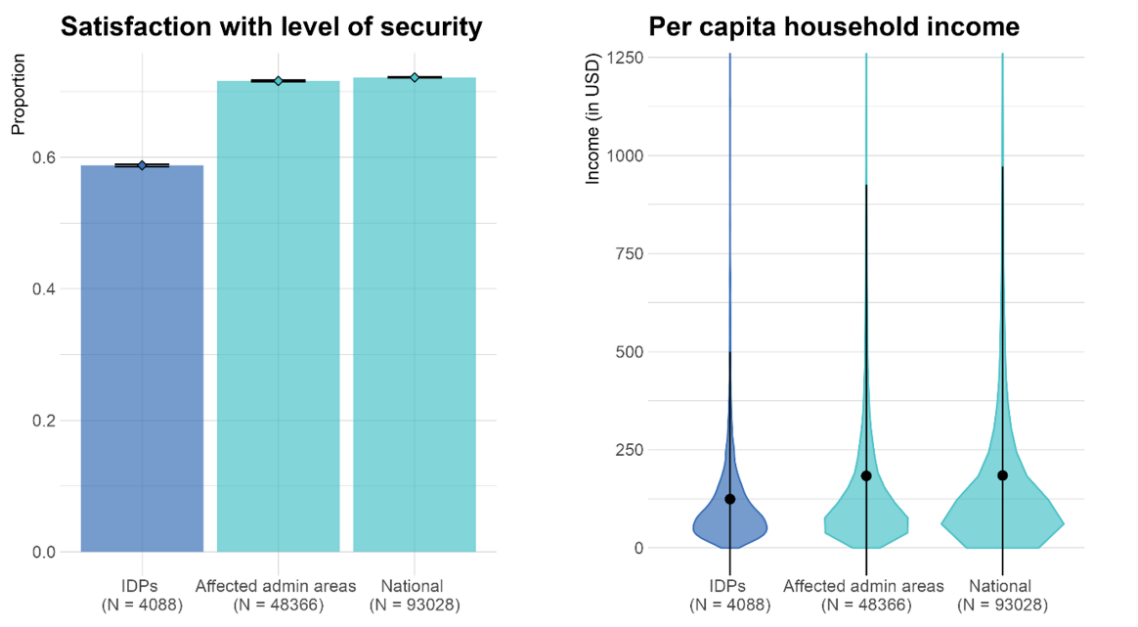
Spotlight II: Colombia Encuesta Nacional de Calidad de Vida 2019

51. The *Encuesta National de Calidad de Vida* 2019 (ECV) implemented by the National Statistical Office of Colombia provides an opportunity to compare two common definitions for the comparator populations for IDPs in a non-camp based setting (Departamento Administrativo Nacional de Estadística, 2019). More specifically, we compare the performance of IDP households to the comparator population applying two different definitions: First, we use the national population in the data collection as the benchmark (definition 1), second, we focus on hosts defined as all non-displaced individuals who live in administrative areas with above mean numbers of IDPs (definition 2). We focus on the 16 departments with most IDPs according to the ECV. We show comparisons between IDPs and these two comparator population definitions for two indicators: the proportion of households that are satisfied with their current level of security and the per capita household income (in USD).

52. The left panel in Figure 4 compares the proportion of IDP households in Colombia that are satisfied with security levels with the national population and the population in displacement-affected administrative areas. In this case, both definitions of the comparator population yield very similar results.

53. Figure 4 also shows the distribution of the per capita household income of IDPs, of the national average, and of the population in administrative areas affected by forced displacement in the panel on the right. The figure indicates that the national population and the subnational population living in areas affected by forced displacement provide very similar benchmarks in this context. One should note that this might partly be due to the nature of the displacement situation in Colombia, with IDPs living freely across the country in non-camp settings and with a lot of urban displacement, making IDPs comparable to the comparator population in many ways. Using the national population or specific administrative units does not affect the benchmark for IDPs strongly. This indicates that both definitions are easily implementable, and result in relatively stable and standardised benchmarks in this comparison.

Figure 4. Proportion of population satisfied with current levels of security and proportion of per capita household income in ECV 2019



Source: DANE - Encuesta National de Calidad de Vida 2019

Spotlight III: South Sudan Forced Displacement Survey 2023

54. The South Sudan FDS 2023 targets a representative sample of registered refugees, as well as a sample of the population living near refugee settlements in the northern region of South Sudan. The dataset comprises approximately 3,000 households. In this paper, we focus solely on the comparator population and the refugees in the north, where populations are concentrated in the Ruweng Administrative area and Upper Nile region (UNHCR, 2024)[7].

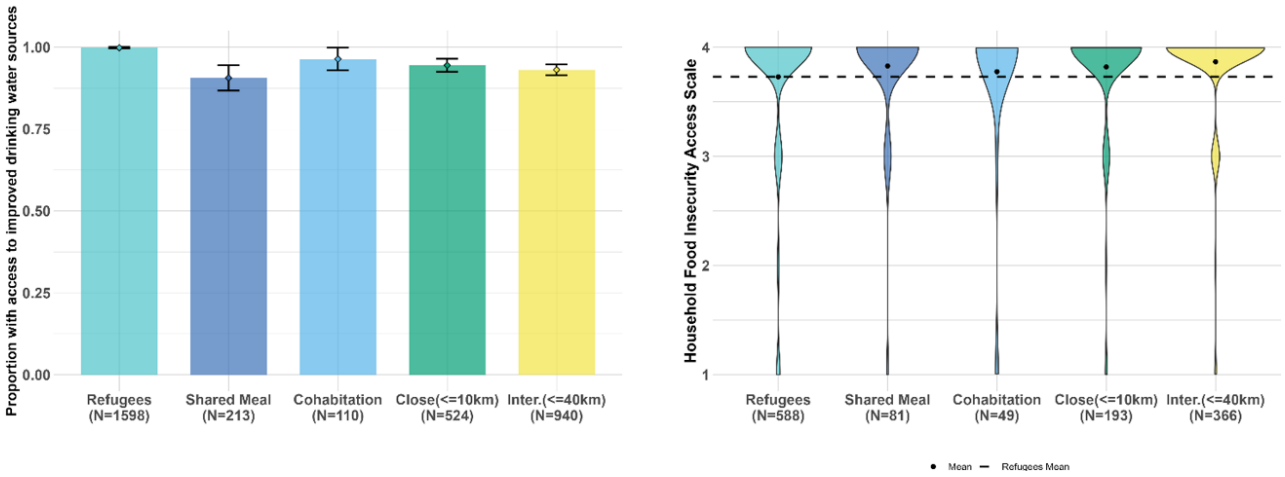
55. Two definitions of comparator population are applied. First, we focus on two criterion-based definitions: meal sharing and cohabitation (definition 4). Second, we focus on the geographical proximity without considering geographical barriers to refugee camps as a distance-based definition of the comparator population (definition 3). Proximity is further divided into two groups: those living within 10 km of the nearest refugee camp (close) and those within 40 km (intermediate). Meal sharing is determined by grouping comparator households that shared meals with refugees, while distance indicates the household's proximity to the nearest refugee camp. Since the FDS lacks a direct variable identifying mixed households, a combination of indicators is used to identify them:

- **Citizenship Question:** If a household is classified as "Refugees" but some members identify as "Nationals of the host country", this suggests a mixed household of cohabitation.
- **Crossing International Borders:** If members of a comparator population household have crossed international borders after being forced to flee, they should be classified as returned refugees, also indicating a mixed household of cohabitation.

The two indicators used for this comparison are accessibility to improved water resources and the Household Food Insecurity Access Scale.

[7] Although the sample includes refugees in the south, the host community in the south was not included in the sampling.

Figure 5. Proportion of refugees and different definitions of host communities with access to improved water sources and distribution Household Food Insecurity Access Scale in the Upper Nile and Ruweng, South Sudan



Source: South Sudan Forced Displacement Survey 2023

56. Figure 5 illustrates that refugees in South Sudan report the highest access to improved water resources, suggesting nearly universal access to improved water within this group. The group cohabiting with refugees shows the highest water accessibility, also the closest to refugees, followed by those residing within a 10 km proximity to the refugee camps. Including the 10-40 km distance group results in a slight decline in water accessibility. This may indicate that access to safe water increases as the distance to refugee camps decreases, which may be due to shared water resources near the camps. The comparator group defined as those sharing meal with refugees has the lowest access to water.

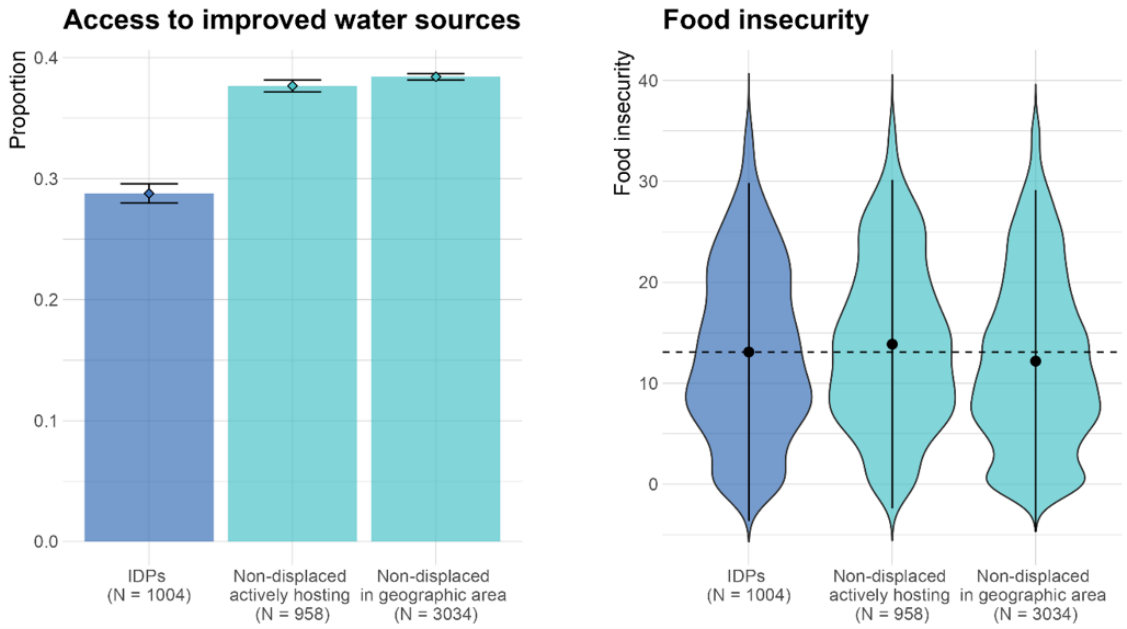
57. Regarding food security, we adopted the Household Food Insecurity Access Scale, where 1 represents food security, 2 mild food insecurity, 3 moderate food insecurity, and 4 severe hunger (Coates, Swindale, & Bilinsky, 2007). All groups in South Sudan experience severe level of hunger, but the comparator population showing the most closely aligned values with the refugees is the one defined by the cohabitation criterion. Those living within 10 km of refugees and those sharing meals with them report comparatively better food security, while the comparator population living in 10–40 km radius from the nearest refugee camps reports the highest levels of food insecurity. These results suggest that refugees have a slightly better food security, and as a result of that, cohabitating with or living closer to them may offer benefits and lead to a minor improvement in food security and water accessibility for non-displacement affected populations.

Spotlight IV: Socio-Economic Survey in Grand Kasai, DRC, 2022-2023

58. In the Grand Kasai area of the Democratic Republic of Congo, a household survey including non-displaced resident population, returned refugees, IDPs, and returned IDPs provides the possibility to analyse how IDPs in a non-camp setting in sub-Saharan Africa compare to non-displaced households in geographical proximity to them (UNHCR, 2023). Firstly, we compare IDPs to the non-displaced population that live in the same locality defined as 1x1km grid cells where IDPs reside in (definition 3). Secondly, we can compare IDPs to a criteria-based definition of the comparator population by focusing on non-displaced households in the area that are actively hosting displaced

persons within their dwelling or household (definition 4). Households were asked in the survey whether they host a displaced person in their dwelling. We do this along two criteria: First, we assess the levels of food insecurity (consumption-based coping strategy index) amongst IDPs, non-displaced households in the geographic area, and those actively hosting IDPs. Second, we compare the proportion of households with access to improved water sources.

Figure 6. Proportion of IDPs and comparator population members with access to improved water sources and distribution of food insecurity amongst IDPs and comparator population households in the Kasai, DRC



Source: Socio-Economic Survey in Grand Kasai, DRC, 2022-2023

59. Figure 6 shows that a significantly lower proportion of IDPs in the Kasai region have access to improved water sources compared to non-displaced households in the geographical proximity and non-displaced households actively hosting IDPs in their dwellings. Importantly, we find that active hosts – applying a criterion-based definition – are more similar to IDPs and are hence a lower threshold than the non-displaced in geographic proximity who do not host IDPs. Nevertheless, the differences are minor across comparator definitions.

60. In terms of food insecurity, we do not find that the two possible comparator population definitions yield significantly different results. This may suggest that a simpler, more easily implementable definition – such as a geography-based definition over a criterion-based definition – is more favourable as it does not, require asking hosts about their generosity towards displaced populations and does not risk volatility over time.

Summary of empirical results

61. From the selected datasets, between the first two definitions, the national population (definition 1) and the population in affected administrative areas (definition 2), Jordan and Colombia cases show that the different benchmark populations are significantly different on the selected indicators, pointing to little differences in benchmark bias and volatility risks across the two definitions. More evidence is required to conclude, however, that the national population could be chosen as the comparator population with its higher potential for standardisation than population in affected administrative areas. Physical proximity (definition 3) and cohabitation or criteria-based definition (definition 4) proved more difficult to standardise across the contexts from this analysis; the approaches in the South Sudan FDS and in the DRC survey only allow for an implementation of physical proximity and cohabitation in considerably different ways. While statistically insignificant, the differences in point estimates between these two definitions also show that the benchmark biases could be different for these alternative comparator populations (i.e. physical proximity vs. cohabitation or criteria-based definition).

Summary, conclusions and recommendations

62. Building on the growing significance of the host community in forced displacement policy and response mechanisms, as well as the IRRS and the IRIS, this paper embarked on an effort to establish a statistical definition of the comparator population in contexts of forced displacement, considering both the national or general population and the concept of a host community. First, the paper reviewed existing definitions of the host community or comparator population used in existing household- and individual-level datasets available in the UNHCR Microdata Library or used for academic research. It grouped the definitions into five broad categories: 1) the general or national population, 2) the population in affected administrative areas, 3) populations in physical proximity, 4) populations in cohabitation or criteria-based definitions, and 5) other definitions. Although none of the five provides a definition that fits all contexts, each present an opportunity to improve standardisation of the comparator population in official statistics and therefore should be considered.

63. Three main criteria were then used to evaluate which definition can be recommended for use in official statistics: standardisation potential, benchmark bias, and volatility risks. An empirical analysis was then conducted with datasets collected in Jordan, Colombia, South Sudan, and the DRC in order to assess the impact of the choice between the four main definitions: 1) national population, 2) population in affected administrative areas; 3) physical proximity, and 4) cohabitation or criteria-based definition on the comparison with displaced populations. Specifically, the empirical analysis compares the performance of displaced populations to differently defined comparator populations in the same context (e.g. comparing IDPs once to the national population and once to the population cohabitating with IDPs).

64. Interestingly, the empirical analysis shows that the definitions often yield comparable results to each other and across contexts. The results of the analysis show that the national population (definition 1) has the highest standardisation potential, and there is no significant difference between using the population in affected administrative areas (definition 2) compared to the national population as a comparator population. Benchmark bias and volatility risks of these two

definitions are low according to this empirical analysis. Physical proximity (definition 3) and cohabitation or criteria-based definition (definition 4) are far less easy to standardise compared to the first two definitions, proving their low standardisation potential. Although the results are limited in assessing the benchmark bias and volatility risks of all definitions, there is some indication that physical proximity sets a lower benchmark bias, implying potentially higher benchmark biases for the national population and those in affected administrative areas. Hence, to move forward with defining the comparator population for statistical purposes, more empirical analysis is needed.

65. Given that the analysis overall finds that differences across definitions in the benchmark they set are minor, it is critical to also discuss feasibility. From the practical perspective of feasibility, all definitions impose separate considerations for data collection and analysis. For instance, a definition based on the national population requires a sampling frame that is representative on the national level while a definition based on populations in affected administrative areas requires a representative sampling frame on the sub-national level. Beyond sampling, all definitions, with the exception of the national population, require further methodological guidance should they be selected for wider application as the statistical benchmark (see Box 4). Selecting and applying a specific definition requires a careful review of the methodology, especially for definitions with a high volatility risk. For example, using administrative units affected by displacement requires a careful identification of the units to use and which ones to classify as affected. To use a population in physical proximity to displaced populations similarly requires identifying the appropriate geographical distance and boundaries to use. There is currently no empirical evidence on how these methodological choices affect comparisons between comparator populations and forcibly displaced people. However, all additional methodological needs decrease the feasibility of using the definitions easily and consistently across official statistics in different country contexts.

Box 4. *Additional methodological guidance needed*

Population in affected administrative areas:

- Choice of level of administrative area to use (e.g. admin level 1 or 2)
- Criteria for determining whether an administrative area is affected or not (threshold for forcibly displaced people in camp and non-camp settings in the area)

Physical proximity:

- Choice of geographical distance to use (e.g. 50 or 100km radius)
- Criteria for determining whether an area is affected or not (threshold for forcibly displaced people in camp and non-camp settings in the proximity)

Cohabitation or criteria-based definition:

- Choice of criterion (e.g. meal sharing or co-habitation)
- Flexibility of criterion across contexts (e.g. context-specific or standardised criteria)
- Guidance on identification question and sampling frames

Other definitions:

- Specific guidance needed

66. **For official statistics, the first two definitions – national population and those in administrative areas affected by displacement – are the most suitable given their relatively high standardisation potential.** To be able to confidently assess the benchmark bias and volatility risks of more inclusive definitions that include more of the national population in the benchmark population, it is crucial to conduct data collections that allow systematically for a comparison to other benchmark populations. Including relevant dimensions that can help to make comparisons to comparator populations based on physical proximity, cohabitation, and other criteria in data collections could move this methodological debate forward. For example, censuses and national-level household surveys that cover granular administrative areas, geolocations, distances to defined settlements, and exposure variables allow a systematic comparison on how different definitions of the comparator populations change vulnerability assessments.

67. For physical proximity and cohabitation or criteria-based definitions, the standardisation issue – the need to define which distance is affected and which criteria to use – should be addressed to make these definitions suitable for comparable statistics. It is also recommended to include clear identification questions in data collection forms used (EGRISS, 2023)[8].

68. For operational data, all definitions identified in principle could be used depending on the main aim of the data collection. However, cohabitation or criteria-based definitions are more suitable to apply for analysis than data collection considering that this group would be close to the host community theoretically while collecting data on this group is less feasible in representative sampling approaches. To enable the comparison between comparator populations and forcibly displaced populations, questionnaires must be designed properly, and data quality control should be ensured during data collection.

69. This paper aimed to recommend a statistical definition of comparator populations in refugee and IDP contexts, considering both the national or general population and the concept of the host community. Whilst this key ambition has been achieved, a number of additional observations that deserve attention have also been identified:

    - The lack of a conceptually standardised definition of the host community in policy discourse imposes challenges for the discussion of a statistical definition of the comparator population.
    - For statistical application, except for the general or national population, more research is needed on the standardisation of each definition and the details on selecting a definition that is best suited to the specific context.
    - Lastly, once the conceptual definition of the host community and the statistical definition of comparator populations as suggested become available, in-depth discussion is needed on their implementation to ensure their optimal application for various purposes, such as official statistics and operational data.

[8] This is the topic of a separate methodological workstream under EGRISS' Technical Subgroup 2.

References

African Union. 2009. “Kampala Convention: African Union convention for the protection and assistance of internally displaced persons in Africa.” Accessed July 16, 2025. https://au.int/sites/default/files/treaties/36846-treaty-kampala_convention.pdf.

Angrist, Joshua D. and Adriana D. Kugler. 2003. “Protective or Counter-productive? Labor market institutions and the effect of immigration on EU natives.” *Economic Journal* 113.

Betts, Alexander, Maria Flinder Stierna, Naohiko Omata, and Olivier Sterck. 2023. “Refugees Welcome? Inter-group interaction and host community attitude formation.” *World Development* 161.

Brookings Institute - University of Bern Project on Internal Displacement. 2010. “IASC Framework on Durable Solutions for Internally Displaced Persons.” Accessed July 16, 2025. https://interagencystandingcommittee.org/other/iasc-framework-durable-solutions-internally-displaced-persons.

Carrington, William J. and Pedro J.F. de Lima. 1996. “The Impact of 1970s Repatriates from Africa on the Portuguese Labor Market.” *Industrial and Labor Relations Review* 50 (2).

Ceritoglu, Evren, H. Burcu Gurcihan Yunculer, Huzeyfe Torun, and Semih Tumen. 2017. “The Impact of Syrian Refugees on Natives’ Labor Market Outcomes in Turkey: Evidence from a quasi-experimental design.” *Journal of Labor Policy* 6 (5).

Coates, Jeniffer, Anne Swindale, and Paula Bilinsky. 2007. “Household Food Insecurity Access Scale for Measurement of Food Access: Indicator guide.” Accessed July 16, 2025. https://www.fantaproject.org/monitoring-and-evaluation/household-food-insecurity-access-scale-hfias.

Coulibaly, Mohamed, Johannes G. Hoogeveen, Robert Benjamin Hopper, and Aboudrahyme Savadogo. 2024. “Forced Displacement in Niger: Why Donors Should Invest More in Refugee and IDP Endowments - The Cost of Displacement Under a ‘Perfect’ Refugee Hosting Regime.” Accessed July 16, 2025. https://documents.worldbank.org/pt/publication/documents-reports/documentdetail/099061224082030219.

Departamento Administrativo Nacional de Estadística. 2019. “Encuesta Nacional de Calidad de Vida (ECV) 2019.” Accessed July 16, 2025. https://microdatos.dane.gov.co/index.php/catalog/678/study-description.

Department of Statistics and ICF. 2019. “Jordan Population and Family and Health Survey 2017-18.” Accessed July 16, 2025. https://dhsprogram.com/methodology/survey/survey-display-500.cfm.

EGRISS. 2018. “International Recommendations on Refugee Statistics (IRRS).” Accessed July 16, 2025. https://egrisstats.org/recommendations/international-recommendations-on-refugee-statistics-irrs/.

EGRISS. 2020. “International Recommendations on Internally Displaced Persons Statistics (IRIS).” Accessed July 16, 2025. https://egrisstats.org/recommendations/international-recommendations-on-idp-statistics-iris/.

EGRISS. 2023. "Towards a Standardized Approach to Identify IDPs, Refugees and Related Populations in Household Surveys." Accessed July 16, 2025. https://egrisstats.org/resource/methodological-paper-towards-a-standardized-approach-to-identify-idps-refugees-and-related-populations-in-household-surveys/.

IDA. n.d. Window for Host Communities and Refugees. Accessed July 16, 2025. https://ida.worldbank.org/en/replenishments/ida19-replenishment/windows-host-communities-refugees.

IOM. 2024. "Population Groups." Accessed July 16, 2025. https://dtm.iom.int/sites/g/files/tmzbdl1461/files/tools/Population%20groups%20-%20Operational%20Definition%20-%20DTM%20March%202024.pdf.

Joint Data Center on Forced Displacement, World Bank Group, and UNHCR. 2020. "Compounding Misfortunes: Changes in poverty since the onset of COVID-19 on Syrian refugees and host communities in Jordan, the Kurdistan region of Iraq and Lebanon." Accessed July 16, 2025. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/878321608148278305/compounding-misfortunes-changes-in-poverty-since-the-onset-of-covid-19-on-syrian-refugees-and-host-communities-in-jordan-the-kurdistan-region-of-iraq-and-le.

Loschmann, Craig, Özge Bilgili, and Melissa Siegel. 2017. "Considering the Benefits of Hosting Refugees: Evidence of refugee camps influencing local labour market activity and economic welfare in Rwanda." *IZA Journal of Develop Migration* 9 (5).

Mäkelä, Erik. 2017. "The Effect of Mass Influx on Labor Markets: Portuguese 1974 evidence revisited." *European Economic Review* 98.

Mayda, Anna Maria, Chris Parsons, Giovanni Peri, and Mathis Wagner. 2017. "The Labor Market Impact of Refugees: Evidence from the U.S. resettlement program." Accessed July 16, 2025. https://www.state.gov/wp-content/uploads/2018/12/The-Labor-Market-Impact-on-Refugees-Evidence-form-the-U.S.-Resettlement-Program-1.pdf.

Morales, Juan S., 2018. "The Impact of Internal Displacement on Destination Communities: Evidence from the Colombian conflict." *Journal of Development Economics* 131.

Uganda Bureau of Statistics. 2023. "Demographic and Health Survey 2022." Accessed July 16, 2025. https://www.ubos.org/wp-content/uploads/publications/UDHS-2022-Report.pdf.

UNHCR. 2019. "Global Compact on Refugees: Defining the boundaries of a host community for the purposes of needs assessment, targeting and programme monitoring." Unpublished document.

UNHCR. 2023. "Socio-Economic Survey in Grand Kasai, 2022-23." Accessed July 16, 2025. https://microdata.unhcr.org/index.php/catalog/1027.

UNHCR. 2024. "Forced Displacement Survey: South Sudan 2023." Accessed July 16, 2025. https://microdata.unhcr.org/index.php/catalog/1175.

UNHCR. n.d.-a. Shelter, Camp and Settlement, Emergency Handbook. Accessed July 16, 2025. https://emergency.unhcr.org/emergency-assistance/shelter-camp-and-settlement.

UNHCR. n.d.-b. UNHCR Master Glossary of Terms. Accessed July 16, 2025. https://www.unhcr.org/glossary.

UNHCR and JIPS. 2021a. "Durable Solutions Analysis Survey: North Darfur State." Accessed July 16, 2025. https://microdata.unhcr.org/index.php/catalog/698.

UNHCR and JIPS. 2021b. "Durable Solutions Analysis Survey: South Darfur State." Accessed July 16, 2025. https://microdata.unhcr.org/index.php/catalog/697.

UNHCR and WFP. 2022. "UNHCR-WFP Joint Assessment Mission in Maratane 2020-2021." Accessed July 16, 2025. https://microdata.unhcr.org/index.php/catalog/680.

United Nations. 1998. Guiding Principles on Internal Displacement. Accessed July 25, 2025. http://un-documents.net/gpid.htm.

United Nations. 2016. New York Declaration for Refugees and Migrants. Accessed July 28, 2025. https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_71_1.pdf.

United Nations. 2018. "Global Compact on Refugees." Accessed July 16, 2025. https://www.unhcr.org/media/global-compact-refugees-booklet.

United Nations. 2021. "Shining a Light on Internal Displacement: A Vision for the Future." Accessed July 25, 2025. https://internaldisplacement-panel.org/.

United Nations. 2022. "The United Nations Secretary-General's Action Agenda on Internal Displacement: Follow-up on the report of the UN Secretary-General's high-level panel on internal displacement." Accessed July 16, 2025. https://www.un.org/en/content/action-agenda-on-internal-displacement/assets/pdf/Action-Agenda-on-Internal-Displacement_EN.pdf.

Verme, Paolo, and Kirsten Schuettler. 2021. "The Impact of Forced Displacement on Host Communities: A review of the empirical literature in economics." *Journal of Development Economics* 150.

World Bank. 2018. "Refugee and Host Communities Household Survey in Uganda." Accessed July 16, 2025. https://microdata.worldbank.org/index.php/catalog/3867.

World Bank. 2019. "Survey of Syrian Refugees and Host Communities 2015-16." Accessed July 16, 2025. https://microdata.unhcr.org/index.php/catalog/406.

World Bank. 2024. "World Bank Group Scorecard FY24-FY30 Methodology Note: World Bank Group results indicator." Accessed July 16, 2025. https://wbgscorecardcdn.worldbank.org/scorecard-methodology/indicator/CSC_RES_DISP_COMM_SERVICES/14.1.[RESULTS]_FDPs%20and%20Host%20Communities_20241119064635173.pdf.

Appendices

Appendix 1. Mapping of the definitions of host communities from UNHCR Microdata Library

ID	Dataset	Definition type	Definition details	Comments
FAO_UGA_2017_RIMA_v01_EN_M_v01_A_OCS	Resilience Index Measurement and Analysis 2017	Physical proximity	The sample of the household survey is composed of 3 034 households, including both the refugee population as well as host communities. The focus is on the seven settlements where the South Sudanese refugees are living, as well as, nearby host communities.	
FAO_UGA_2018_RIMA_v01_EN_M_v01_A_OCS	Resilience Index Measurement and Analysis 2018	Physical proximity	This study defines host communities as those households that live in neighbouring villages to the refugee camps.	
HARVARD_DATAVERSE_BGD_2019_CBPS_v2.1	Cox's Bazar Panel Survey Baseline, 2019	Physical proximity	The sample is representative of three strata: i) residents of the refugee camps, ii) host communities within 15km of refugee camps, iii) host communities further than 15km from refugee camps.	
HDX_JIPS_SOM_2016_IDPProfiling_vEXT	Internal Displacement Profiling in Hargeisa, 2015	Population in affected administrative areas	In order to capture representative samples of the different target populations outlined above, the survey used two different sampling approaches. A probabilistic sampling approach was used for IDPs from Somaliland, refugee returnees, economic migrants and host community living in settlements. The sample in the settlements was stratified by population group and was representative at the total population level. The distribution of each population group was proportionate to their size in each settlement and was considered when distributing the sample across the settlements.	

HDX_reach-car-2022-msna-dataset_vEXT	Multi-Sectoral Needs Assessment, 2022	Population in affected administrative areas	Includes refugees and non-displaced households. It does not specifically say host communities but it is implied. People are sampled to be representative at the sub-prefecture and prefecture levels.	
UNCHR_UGA_MSNA_2018_v2	Joint Multi-Sector Needs Assessment - 2018	Unclear	Unclear.	The universe is just refugee settlements but the description of the purpose of the study describes collecting information from hosts as well.
UNHCR_AFG_2021_MSRNA_v2.1	Multi-Sectorial Rapid Assessments - 2021	National population	This study surveys host communities alongside refugees across all provinces in Afghanistan.	
UNHCR_AFG_2021_PARR_December_KII_v2.1	Evaluation of Community-Based Protection and Solutions Programme in Priority Areas of Return and Re-integration in Afghanistan, Key Informants Interview - December 2021	Population in affected administrative areas	The host communities are defined generally as the administrative borders from the PARR locations that refugees live in.	
UNHCR_AFG_2021_PARR_March_HH_v2.1	Evaluation of Community-Based Protection and Solutions Programme in Priority Areas of Return and Re-integration in Afghanistan, Household Survey - March 2021	Population in affected administrative areas	The host communities are defined generally as the administrative borders from the PARR locations that refugees live in.	
UNHCR_AFG_2021_PARR_March_KII_v2.1	Evaluation of Community-Based Protection and Solutions Programme in Priority Areas of Return and Re-integration in Afghanistan, Key Informants Interviews - March 2021	Population in affected administrative areas	The host communities are defined generally as the administrative borders from the PARR locations that refugees live in.	
UNHCR_AFG_2022_RHAF_anon_data_v2.1	Rapid Household Assessments 2022	Physical proximity	Focuses on refugees and IDPs but also includes vulnerable host community members. All provinces were sampled. There was no explicit sampling strategy. The summary implies	

			that host community members were those living near (and thus were sampled) PoCs.	
UNHCR_BGD_2020_JMSNA_Host Community_v2.1	Joint Multi Sector Needs Assessment: Cox's Bazar, Rohingya Refugee Response – August 2020, Host Community	Physical proximity	The universe for host communities is all households within 6 km of UNHCR camps in Ukhiya and Teknaf upazilas	
UNHCR_BGD_2021_jmsna_host_anon_data_v2.1	Joint Multi Sector Needs Assessment: Cox's Bazar, Rohingya Refugee Response – 2021, Host Community	Physical proximity	Households were sampled from UNHCR's host community database covering host community populations living within 6 km from UNHCR camps. UNHCR, WFP and IOM beneficiary databases were used to sample households in wards outside this radius, or with limited UNHCR host community database coverage. The share of the sample drawn from each database can be found in annex 1. When interpreting the findings, a bias towards beneficiary populations has to be considered for areas outside the UNHCR host community database coverage.	
UNHCR_BGD_2023_jmsna_host_anon_data_v2.1	Joint Multi Sector Needs Assessment: Cox's Bazar, Rohingya Refugee Response – 2023, Host Community	Population in affected administrative areas	Examines host communities living in Cox's Bazar which has 34 refugee camps.	Unclear exactly how they are grouped, but it is clear hosts are defined as those impacted by nearby refugee camps.
UNHCR_BGD_KAP_2018_v2.1	WASH KAP Survey Rohingya Cox's Bazar, 2018	Unclear	Unclear.	It says that it does measure host communities but unclear where they are being sampled from.
UNHCR_CIV_2021_CBI_PDM_Metadata	Post-Distribution Monitoring of Cash-Based Interventions - Feb 2021	Other	Infrastructure service/catchment areas. Host communities are defined as those that are not refugees or IDPs that received the cash based intervention. This seemed to leave only returnees in this category.	
UNHCR_CMR_2016_SEA_v2-1	Socio-economic assessment of Central African refugees in Cameroon's Adamanou,	Population in affected administrative areas	Areas hosting Central African refugees in 5 subsistence zones under study within Cameroon's Adamanou, Eastern and	

	Eastern and Northern regions 2016		Northern regions. This included 7 refugee camps (Gado-Badzere, Lolo, Bile, Timangolo, Borgop, Ngam and Ngarisingo) as well as various non-camp sites.	
UNHCR_CMR_2020_SEI_CAR_COVID19_v2.1	Assessment on the socioeconomic impact of COVID-19 on Central African Refugees - East (May 2020)	Population in affected administrative areas	Samples from refugees and hosts in specific villages.	
UNHCR_CMR_2020_SEI_COVID19_v2.1	Assessment on the socioeconomic impact of COVID-19 on refugees, IDPs and host communities - June 2020, Extreme North	Population in affected administrative areas	Samples from refugees and hosts in specific villages.	
UNHCR_CMR_2021_JAM_v2.1	Far North Joint Assessment of Nigerian Refugees - Nov 2021	Physical proximity	Examines host communities as non-refugees that live in or near and around the Minawao settlement.	
UNHCR_COD_2023_RMS	Results Monitoring Survey (RMS) - 2023 Congo (Democratic Republic of the), 2023	Population in affected administrative areas	Refugees and asylum-seekers were sampled from UNHCR's registration database (proGres), stratified by areas. Refugee returnees, IDPs, and host communities were sampled using lists provided by the operation in each area, with systematic sampling conducted if sampled households were unreachable.	Unclear. Says Refugees are stratified into areas, but not sure if hosts and others are also sampled from these areas or other areas.
UNHCR_IRQ_2016_PROFILING_v2.1	Profiling of South and Central Governorates - 2016	Physical proximity	Each of the assessments in the 9 Governorates (merged in the present dataset) surveyed approximately 400 host community households and 400 IDP households in 30 different sites. The 30 sites were selected using a random stratified sample based on data from the Displacement Tracking Matrix of the International Organization for Migration (IOM). The sites were stratified based on urban or peri-urban characteristics, the number of IDP families present, and shelter typology of the IDP households. The household assessment attempts to survey as representative a population as possible given the available data on location of IDP	

			households across the nine assessed governorates.
UNHCR_IRQ_2021_MSNA_v2.1	Multi-Sector Needs Assessment, 2021	Physical proximity	For the host community cluster geo-sampling (Table 2), all three KR-I Governorates were divided into hexagons of 2km and filtered to contain at least 500 persons per hexagon (on the basis of population concentration figures applicable down to the sub-district level) to ensure sufficient host community households for data collection, as a means to filter out empty, industrial, or inaccessible hexagons from selection.
UNHCR_JIPS_SDN_2021_CERF_BlueNile_v2.1	Durable Solutions Analysis Survey: Blue Nile State, 2021	Population in affected administrative areas	The focus was set on profiling of IDPs (in camp settlements and out of camps), IDP returnees, refugee returnees, and non-displaced. This survey specifically referred to the host communities as those that are not displaced in the localities of interest.
UNHCR_JIPS_SDN_2021_CERF_NorthDarfur_v2.1	Durable Solutions Analysis Survey: North Darfur State, 2021	Population in affected administrative areas	The focus was set on profiling of IDPs (in camp settlements and out of camps), IDP returnees, refugee returnees, and non-displaced. This survey specifically referred to the host communities as those that are not displaced in the localities of interest.
UNHCR_JIPS_SDN_2021_CERF_SouthDarfur_v2.1	Durable Solutions Analysis Survey: South Darfur State, 2021	Population in affected administrative areas	The focus was set on profiling of IDPs (in camp settlements and out of camps), IDP returnees, refugee returnees, and non-displaced. This survey specifically referred to the host communities as those that are not displaced in Kaas.
UNHCR_JIPS_SDN_2021_CERF_SouthKordofan_v2.1	Durable Solutions Analysis Survey: South Kordofan State, 2021	Population in affected administrative areas	The focus was set on profiling of IDPs (in camp settlements and out of camps), IDP returnees, refugee returnees, and non-displaced. This survey specifically referred to the host communities as those that are not displaced in the localities of interest.

UNHCR_JIPS_SDN_2021_CERF_WestDarfur_v2.1	Durable Solutions Analysis Survey: West Darfur State, 2021	Population in affected administrative areas	The focus was set on profiling of IDPs (in camp settlements and out of camps), IDP returnees, refugee returnees, and non-displaced. This survey specifically referred to the host communities as those that are not displaced in the localities of interest.	
UNHCR_JIPS_SDN_2021_PBFProfiling_v2.1	Peacebuilding Fund Durable Solution Analysis Survey - Darfur, 2020-2021, IDP returnees, IDPs, non-displaced and nomad profiling in Darfur	Population in affected administrative areas	Host communities are defined as non-displaced populations living in the same localities as refugees.	
UNHCR_MLI_2020_CBI_PDM_v2.1	Post-Distribution Monitoring of Cash-Based Interventions – Apr 2020	Other	Infrastructure service/catchment areas. Host communities are defined as non-refugee populations living around the refugee camp that are part of the Cash-based intervention.	
UNHCR_MMR_2023_RMS_anon_data_v2.1	Results Monitoring Survey 2023	Other	Infrastructure service/catchment areas. Unclear details of how hosts were identified and sampled.	The survey is about seeing how services provided by UNHCR are impacting people, including the hosts.
UNHCR_MOZ_2021_JAM_v2.1	UNHCR-WFP Joint Assessment Mission in Maratane - 2020-2021	Physical proximity	Host communities: A sample frame was constructed based on detected buildings from satellite images because of the lack of complete list. JAM Coordinator confirmed the interaction between the camp and the host community within 6 km radius; however, those live in the West of the river, located in the West from the camp, are excluded since the excluded residents have limited interaction with the host community. In addition, due to the limited scope of the most recent satellite image from January 2021, old images complemented to cover the circle of 6 km radius.	
UNHCR_MWI_2016_SENS_v2.1	Standardised Expanded Nutrition Survey (Dzaleka and Luwani Refugee Camps and	Physical proximity	Interviewed host communities in the district around the refugee camps.	

	Host Communities) – 2016			
UNHCR_MWI_2020_CBI_PDM_v2.1	Dzaleka refugee camp - Post Distribution Monitoring for Cash-Based Intervention - 2020	Other	Infrastructure service/catchment areas. Host communities are defined as non refugee populations living around the refugee camp that are part of the Cash-based intervention.	
UNHCR_NGA_2020_SEA_COVID19_v2.1	Socio-economic impact assessment of COVID-19 pandemic among persons of concern in Nigeria (July 2020)	Unclear	Unclear	The description says it includes community members hosting refugees. It is unclear if this is hosting in homes or just living in the same area.
UNHCR_PHL_2016_Zamboanga_HB_IDP_Profiling	Zamboanga Home Based IDP Re-Profiling 2016	Cohabitation or criteria-based	Under the concept of "home based IDPs" host communities are those households that directly host refugees in their own homes.	
UNHCR_POL_2022_MSNA_v2.1	Multi-Sector Needs Assessment - 2022	Cohabitation or criteria-based	The survey does not sample those that are not refugees but does distinguish between types of refugees: those that live among hosts and those that live in collective sites. Hosts in this case are defined as those that allow refugees to live with them for free or rent a space to them.	
UNHCR_SENS_TCD_2019_v1	SENS in Refugee Camps and Host Villages in South Chad and the Lac Region 2019	Physical proximity	Collects host community information by looking at some villages hosting refugees near to the camps but not all. Unclear how these cities were selected.	
UNHCR_SVK_2022_ABA_v2.1	Area-Based Assessment of Refugees and Host Communities - 2022	Physical proximity	Refugees in 4 cities in Slovakia were used to sample refugees (Bratislava, Kosice, Nitra, Zilina). Host community members were also interviewed in each of those four cities. Used a purposive sample to find these people, not representative.	
UNHCR_SVK_2022_MSNA_v2.1	Multi-Sector Needs Assessment - 2022	Cohabitation or criteria-based	The survey does not sample those that are not refugees but does distinguish between types of refugees: those that live among hosts and those that live in collective sites. Hosts in this case are defined as those that	

			allow refugees to live with them for free or rent a space to them.	
UNHCR_SYR_2021_PA_R_v2.1	Participatory Assessment for Syrians, 2021	National population	Covers host households at a national level.	
UNHCR_TAJ_2022_soec_anon_data_v2.1	Socio-Economic Survey 2022	Unclear	Unclear.	The quantitative survey does not include host communities, but the qualitative study conducts a Focus Group that concerns refugees and hosts, but unclear if hosts are interviewed and how they were selected.
UNHCR_TCD_2021_SENS_EAST_SOUTH_LAKE_REGION_v2.1	Standardised Expanded Nutrition Survey (East, South and Lake Region) - 2021	Physical proximity	The data collection took place between June 17 and July 22, 2021, and covered 25 camps of Sudanese, Central African and Nigerian refugees in the East, South, and Lake Regions of Chad, including host villages living 25 kilometres around the camps. Unclear if this was only refugees, but it clearly considers host villages those living close to the refugees.	
UNHCR_TJK_2021_CBI_PDM	Post-Distribution Monitoring of Cash-Based Intervention, April 2021	Other	Infrastructure service/catchment areas. Host communities are defined as those that are not refugees or IDPs that received the cash-based intervention. This seemed to leave only returnees in this category. Note: This survey only asked refugees, but its universe of what it defined as a host community would have been those that receive the benefit.	
UNHCR_UGA_2019_CBI_PDM_v2.1	Post-Distribution Monitoring of Cash-Based Intervention, April 2019	Other	Infrastructure service/catchment areas. Host communities are defined as non-refugee populations living around the refugee camp that are part of the Cash-based intervention.	
UNHCR_UGA_2019_KAP_Palabek_v2.1	WASH KAP Survey Palabek Settlement (Refugees & Host	Other	Infrastructure service/catchment areas. The KAP Survey was conducted in	

	Community), October 2019		Palabek Settlement Zones and four (4) host sub counties of Plabek Ogili, Palabek Kal, Plabek Gem and Lamwo Town council where the UNHCR Multi-Sector Humanitarian Assistance for Refugees and Host Community project is being implemented. Those that received WASH services were considered host communities.
UNHCR_ZAM_2020_COVID19_v2.1	COVID-19 Impact Assessment on Refugee Livelihoods: Multistakeholder Rapid Assessment – 2020	Population in affected administrative areas	Samples from refugee camps and nearby villages.
UNHCR_ZMB_2024_JICA	One Meheba Local Area Plan - 2024	Population in affected administrative areas	Details are unclear, but the abstract says that all non-refugees in the Meheba region are considered as hosts.
UNHCR-CMR-2016-AUB-2.1	Analysis and Refinement of Targeting Mechanisms for Food and Multipurpose Cash Assistance to Central African Republic Refugees - 2016	Population in affected administrative areas	Areas hosting Central African refugees in 5 subsistence zones under study within Cameroon's Adamanou, Eastern and Northern regions. This included 7 refugee camps (Gado-Badzere, Lolo, Bile, Timangolo, Borgop, Ngam and Ngarisingo) as well as various non-camp sites.
UNHCR-RWA-2016-SEA-2.1	Socio-economic assessment of refugees in Rwanda's Gihembe, Kigeme and Kiziba camps - 2016	Physical proximity	This survey was conducted to examine the influence of Congolese refugees on host communities in Rwanda, with a focus on labour market activity and economic welfare. The survey covered three refugee camps as well as their surrounding host communities.
WBG_ETH_2017_SPS_v01_M	Skills Profile Survey 2017, A Refugee and Host Community Survey	Physical Proximity	The Skills Profile Survey 2017 covered refugees in camps, and surrounding host communities, in Ethiopia. Refugees of four nationalities were surveyed: Eritrean, Somali, South Sudanese, and Sudanese. Only the refugees living in camps were surveyed, because tracing households outside the camps was not feasible. However, 66 percent of all refugees in Ethiopia live in camps, while

			those that live outside camps are largely Eritrean. Host communities, defined as Ethiopian non-displaced households living within a 5km radius of a camp, were also surveyed.
WBG_IRQ_2015_SRHCS_v01_M	Survey of Syrian Refugees and Host Communities in Kurdistan, 2015-2016	Physical proximity	Pseudo-enumeration areas were constructed via geographical and population distributions and all the households were listed as either Syrian or non-Syrian. Tried to sample 40 households with an even 50/50 split between hosts and refugees.
WBG_JOR_2015_SRHCS_v01_M	Survey of Syrian Refugees and Host Communities, 2015-2016	Physical proximity	Jordan: Sampled refugees within specific camps, but then also sampled in governorates that were close to these camps to capture the host community population that should be similar to the refugees.
WBG_LBN_2015-2016_SRHCS_v01_M	Survey of Syrian Refugees and Host Communities, 2015-2016	National Population	Lebanon: Pseudo-enumeration areas were constructed via geographical and population distributions and all the households were listed as either Syrian or non-Syrian. Tried to sample 40 households with an even 50/50 split between hosts and refugees.
WBG_NGA_2018_IDP_v01_M	Profile of Internally Displaced Persons in North-East Nigeria 2018	Population in affected administrative areas	The sample had 6 strata with each state as a stratum. Each stratum was divided into grids of 150 by 150 meters using GIS technology defining the enumeration areas (EAs). The number of EAs to be selected from each stratum was obtained proportional to the IDP population sizes in each state. For the IDP sample, the EAs consisted of two types of IDP settlements, which were host community settings and camp settings. A total of 111 EAs across the 6 states were randomly selected proportional to size. Out of these, 77 EAs were host community type EAs and 48 were camp-like EAs. For the host community sample of 1,500 households, EAs were

			restricted to the 77 host community type EAs that were selected based on IDPs population numbers, in the absence of residential population estimates at this level. The survey is representative for IDPs and host communities, defining host communities as the non-displaced population living in the EAs with displaced populations.	
WBG_UGA_2018_RHCS_v01_M	Refugee and Host Communities Household Survey 2018	Population in affected administrative areas	The host population is defined as the native population in districts where refugee settlements are situated.	The first domain is the host population in the regions of West Nile and South West. The third, the refugee and host population in Kampala. For the host communities and Kampala, before the selection of the PSUs, district EAs were sorted by residence type (urban/rural), district sub-county, parish, village and EAs. For Kampala, only EAs that contained more than ten refugee households according to the 2014 Census were considered.

Appendix 2. Mapping of the definitions of host communities from Verme and Schuettler (2021)

Paper	Definition type	Definition details	Comments
Akgunduz and Torun (2018)	Population in affected administrative areas	Admin 2 level (NUTS-2)	Syrian refugees in Turkey. Analysis at admin 2 level (NUTS-2).
Akgunduz et al. (2015)	Population in affected administrative areas	Admin 3 level (NUTS-3)	Syrian refugees in Turkey; analysis at admin 3 level (Provinces, NUTS-3)
Aksu et al (2018)	Population in affected administrative areas	Admin 3 level (NUTS-3)	Syrian refugees in Turkey; Assess local economics in the affected regions as a whole; Difference-in-difference; analysis at admin 3 level (Provinces, NUTS-3)
Alhawarin et al (2018)	Population in affected administrative areas	Admin 2 level (district), and admin 1 level for time trends (governorate)	Syrians in Jordan, difference-in-difference
Alix-Garcia and Bartlett (2015)	Population in affected administrative areas	Nyala, Darfur and el Obeid, Kordofan	Impact of displacement in labour market (IDPs), Darfur, Sudan
Alix-Garcia and Saah (2010)	Physical proximity	Impact by distance to camps and impact in Kagera and Kigoma	Refugee inflows from Burundi and Rwanda in 1993 and 1994 on host populations in western Tanzania;
Alix-Garcia et al (2012)	Population in affected administrative areas	Dynamics between residents, aid workers and IDPs in Nyala, Sudan	IDPs in Nyala caused by Darfur crisis, Sudan
Alix-Garcia et al (2017)	Physical proximity	10km distance of the camp	Impact of long-term refugee camps on host populations in northern Kenya. Nighttime lights data, official statistics and new household survey data.
Angrist and Kugler (2003)	National		Immigrants in EU, 18 countries
Bahar et al. (2019)	Population in affected administrative areas	Mainly at admin 1 level (municipalities)	Venezuelans in Colombia, socioeconomic impacts of the PEP program.

Balkan and Tumen (2016)	Population in affected administrative areas	Admin 2 level (NUTS-2)	Syrians in Turkey
Balkan et al. (2018)	Population in affected administrative areas	Admin 1 level (NUTS-3)	Syrians in Turkey
Bodvarsson et al. (2008)	Population in affected administrative areas	Assess the impact in Miami, compared to Atlanta, Tampa, Houston and Los Angeles.	Mariel Boatlift (Cubans in Miami)
Borjas (2017)	Population in affected administrative areas	Miami	Marielitos (Cubans in Miami)
Borjas and Monras (2017)	National/Population in affected administrative areas	Assessment both at subnational and national levels	The influx of Marielitos into Miami in 1980; the influx of French repatriates and Algerian nationals into France at the end of the Algerian Independence War in 1962; the influx of Jewish emigrants into Israel after the collapse of the Soviet Union in the early 1990s; and the exodus of refugees from the former Yugoslavia during the long series of Balkan wars between 1991 and 2001.
Bozzoli et al. (2013)	Population in affected administrative areas	Grouping municipalities to not displaced/displaced/stayers in expulsing communities	IDPs in Colombia, labour market
Braun and Kvasnicka (2014)	Population in affected administrative areas	District level	Germens expellees to West Germany after World War II
Braun and Mahmoud (2014)	Population in affected administrative areas	State level	German expellees in postwar Germany
Calderon-Mejia and Ibanez (2016)	Population in affected administrative areas	City level	IDPs in Colombia, labour market
Card (1990)	Population in affected administrative areas	Miami	Mariel Boatlift (Cubans in Miami)

Carrington and de Lima (1996)	National/Population in affected administrative areas	Portugal comparing to Spain and France; Comparisons between districts within Portugal	Angolans and Mozambicans in Portugal
Caruso et al. (2019)	Population in affected administrative areas	Admin 1 (departments)	Venezuelans in Colombia, impacts of the recent labour supply shock-driven by Venezuelans
Cengiz and Tekguc (2018)	Population in affected administrative areas	Admin 2 (NUTS-2)	Syrians in Turkey
Ceritoglu et al. (2017)	Population in affected administrative areas	Admin 2 (NUTS-2)	Syrians in Turkey
Clemens and Hunt (2019)	Population in affected administrative areas	Miami	Mariel Boatlift (Cubans in Miami)
Cohen-Goldner and Paserman (2011)	National	The paper looks at the overall impact in Israel. One of the labour market segmentations is district of residence interacting with occupation.	Immigrants from the Former Soviet Union to Israel, natives' wages and employment
del Carpio and Wagner (2016)	Population in affected administrative areas	Admin 2 (NUTS-2)	Syrians in Turkey
Depetris-Chauvin and Santos (2017)	Population in affected administrative areas	Admin 2 (municipality level)	IDPs in Colombia
Depetris-Chauvin and Santos (2018)	Population in affected administrative areas	Admin 2 (municipality level)	IDPs in Colombia
Esen and Binatli (2017)	Population in affected administrative areas	Admin 2 (NUTS-2)	Syrians in Turkey
Fakih and Ibrahim (2015)	Population in affected administrative areas	Admin 1 (governorate)	Syrians in Jordan
Fallah et al. (2018)	Population in affected administrative areas	Admin 4 (locality)	Syrians in Jordan
Foged and Peri (2015)	Population in affected administrative areas	Admin 2 (municipality level)	Immigrants in Denmark in 1991-2008

Friedberg (2001)	National		Immigrants in Israel in 1990-1994
Gehrsitz and Ungerer (2018)	Population in affected administrative areas	Admin 3 (county level)	Refugees in Germany
Glitz (2012)	Population in affected administrative areas	Admin 1 (State level)	German expellees in postwar Germany
Hercowitz and Yashiv (2002)	National		Immigrants from the Former Soviet Union to Israel in the 1990s
Hunt (1992)	Population in affected administrative areas	Admin 2 (departments)	Repatriates from Algeria in 1962
Kreibaum (2016)	Population in affected administrative areas	Admin 2 (district)	Congolese in Uganda
Kurschner Rauck and Kvasnicka (2018)	Population in affected administrative areas	Admin 3 (county level)	Refugees in Germany in 2015
Lach (2007)	Population in affected administrative areas	City level	Immigrants from the Former Soviet Union to Israel in the 1990s
Loschmann et al. (2019)	Physical proximity	10km distance of the camp/above 20km around the camp	Congolese in Rwanda
Mäkelä (2017)	National		Angolans and Mozambicans in Portugal
Mansour (2010)	Population in affected administrative areas	City level in West Bank	Palestinians in Israel
Mayda et al. (2017c)	Population in affected administrative areas	US commuting zones	Resettled refugees in the US
Maystadt and Duranton (2018)	Population in affected administrative areas	Admin 4 (village); Construction of an refugee index by the size of camp and distance	Burundi (1993) and Rwanda (1994) influx into Tanzania, 1991-2010
Maystadt and Verwimp (2014)	Population in affected administrative areas	Admin 4 (village); weight of proximity to the 13 refugee camps	Burundi (1993) and Rwanda (1994) influx into Tanzania, 1991-2010

Morales (2018)	Population in affected administrative areas	Admin 2 (municipality level)	IDPs in Colombia
Murard and Sakalli (2018)	Population in affected administrative areas	Admin 4 (municipality level)	Greeks fled Turkey to Greece after the Greco-Turkish war of 1919-1922
Penaloza Pacheco (2019)	Population in affected administrative areas	Admin 1 (departments)	Venezuelans in Colombia
Peri and Yasenov (2019)	Population in affected administrative areas	Miami	Mariel Boatlift (Cubans in Miami)
Rozo and Sviastchi (2018)	Population in affected administrative areas	Admin 1 (governorate)	Syrians in Jordan
Ruiz and Vargas-Silva (2015)	Population in affected administrative areas	Kagera	Burundi (1993) and Rwanda (1994) influx into Tanzania, 1991-2010
Ruiz and Vargas-Silva (2016)	Population in affected administrative areas	Kagera	Burundi (1993) and Rwanda (1994) influx into Tanzania, 1991-2010
Ruiz and Vargas-Silva (2018)	Population in affected administrative areas	"exposure" to refugees	Burundi (1993) and Rwanda (1994) influx into Tanzania, 1991-2010
Saiz (2003)	Population in affected administrative areas	Miami	Mariel Boatlift (Cubans in Miami)
Taylor et al. (2016)	Physical proximity	10 km radius of 3 Congolese refugee camps	Congolese in Rwanda
Tsuda (2018)	Population in affected administrative areas	Admin 4 (village)	Burundi (1993) and Rwanda (1994) influx into Tanzania in the 1990s
Tumen (2016)	Population in affected administrative areas	Admin 2 level (NUTS-2)	Syrians in Turkey

Appendix 3. Additional sources of mapping of the definitions of host communities

Dataset/Paper	Definition type	Definition details	Comments
Uganda Demographics and Health Survey 2022	National/Population in affected administrative areas		National household survey including refugees
Jordan Demographics and Health Survey 2017-18	National/Population in affected administrative areas		National household survey including Syrians in settlements
Colombia Encuesta Nacional de Calidad de Vida 2019	National/Population in affected administrative areas		National household survey including IDPs
South Sudan Forced Displacement Survey 2023	Physical proximity/Criteria		Refugees in North and its host communities, and refugees in South
Alexander Betts, et al. (2022)	Physical proximity	The host nationals in the community living nearby	Refugees in Uganda, Kenya, and Ethiopia
Fajth et al. (2019)	Physical proximity	Based on their distance from each camp	Congolese refugees in Rwanda
Codjoe et al. (2013)	Physical proximity	Surrounding the camp, within a 5-km buffer zone	Liberian refugees in Ghana
Ansar and Khaled (2021)	Physical proximity/Criteria	All efforts were made to choose respondents from relatively lower-income households based on the hypothesis that the poorer among the local communities are most severely affected, and the opposition towards the refugees higher among them. Respondents, both living within the vicinity of the refugee camps and far from the camps, were purposively selected for the interviews.	Rohingya refugees in Bangladesh

Doocy et al. (2016)	National	100 cluster × 21 household (14 Syrian refugee households and seven host community households) design was used.	Syrian refugees in Lebanon
Mohammed et al. (2024)	Population in affected administrative areas	The sampling of refugee/internally displaced households is a two-stage stratified process, with the primary sampling unit being the Enumeration Area (EA), drawn from an exhaustive list of refugees and IDP sites obtained from UNHCR. A total of 165 EAs were randomly selected across all refugee/IDP sites, with approximately 20 percent comprising host community clusters (villages). In the second stage of sampling, 12 households were selected from each sampled EA following an enumeration operation that created a comprehensive list of households within the EA.	Refugees and IDPs in Niger