Data disaggregation of SDG indicators by forced displacement

DECEMBER 2020

Contents

Acknowledgements

Acronyms

ANDA	National Data Archive, Colombia. Acronym in Spanish
CNPV	National Population and Housing Census, Colombia. Acronym in Spanish
CPR	Central Population Register, Norway
CRRF	Comprehensive Refugee Response Framework
DANE	National Administrative Department of Statistics, Colombia. Acronym in Spanish
DHS	Demographic and Health Survey
EGRIS	Expert Group on Refugee and IDP Statistics
EMIS	Education Management Information System
FDP	Forcibly displaced person
FDS	Forced displacement status
GEIH	Great Integrated Household Survey, Colombia. Acronym in Spanish
GCR	Global Compact on Refugees
HFS	High Frequency Survey
HMIS	Health Management Information System
IAEG-SDGs	Inter-agency Expert Group on SDG Indicators
ICLS	International Conference of Labour Statisticians
IDP	Internally displaced person
ILO	International Labour Organization
IOM	International Organization for Migration
IRIS	International Recommendations on IDP Statistics
IRRS	International Recommendations on Refugee Statistics
JDC	World Bank-UNHCR Joint Data Center on Forced Displacement
JIPS	Joint IDP Profiling Service
LSMS	Living Standards Measurement Study
MICS	Multiple Indicator Cluster Surveys
NSO	National Statistics Office
NSS	National Statistical System
OCHA	United Nations Office for the Coordination of Humanitarian Affairs
ODK	Open Data Kit
PIN	Personal identification number, Norway
SDGs	Sustainable Development Goals
UARIV	Unit for the Attention and Integral Reparation to the Victims, Colombia. Acronym in Spanish
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNSC	United Nations Statistical Commission
UNSD	United Nations Statistics Division
WASH	Water, Sanitation and Hygiene
WHO	World Health Organization

1. Introduction

The 2030 Agenda for Sustainable Development makes an ambitious commitment to "leave no one behind" in its implementation. To ensure that this commitment is effectively met, several vulnerable population groups are identified for attention in the Agenda including migrants, refugees, and other forcibly displaced persons (FDPs). The New York Declaration for Refugees and Migrants[1], the Comprehensive Refugee Response Framework (CRRF)[2] and the Global Compact on Refugees (GCR)[3] established a solid international commitment to supporting FDPs and the countries where they reside. The work of the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs)[4] to develop and implement the global indicator framework for monitoring the Goals and Targets of the 2030 Agenda, and to define and compile the standards and tools for disaggregating data, provided the basis for and systematic support required to meet the ambitious commitments made by member states.

One such commitment is to make forcibly displaced persons visible in the 2030 Agenda by enhancing the availability of better data and evidence. Towards this goal, UNHCR and JIPS have worked within the Expert Group on Refugee and IDP Statistics (EGRIS)[5] to develop a common set of indicators to measure refugees' inclusion in host communities. These indicators were selected pragmatically as the most salient, those for which data is likely to become available, and make reference whenever possible, to the SDG Indicators framework. This list of indicators can be found in *Chapter 5* of the International Recommendations on Refugee Statistics (IRRS)[6], published by EGRIS in 2018. Drawing from this work, in 2019, EGRIS identified three key policy areas that are most relevant for forcibly displaced

populations and mapped 12 top priority SDG indicators to these areas[7] [see Box 1]. The submission of the Compilers' Manual on Displacement Statistics[8] and the International Recommendations on IDP Statistics (IRIS) to the UN Statistical Commission in March 2020 by EGRIS[9] provided additional technical guidance for the production of reliable forced displacement disaggregated data.

The purpose of this report is to provide a review of the availability of published disaggregated SDG indicators and to assess the feasibility of estimating them based on existing data. It is hoped that this document will encourage National Statistics Offices (NSOs) and custodian agencies to improve the availability of data on forced displacement aligned to the SDGs.

The set of indicators explored are the 12 indicators identified as priority in consultation with the EGRIS [see Box 1]. The availability of published disaggregated SDG indicators was reviewed and when not available, we assessed the feasibility of estimating them based on the data and metadata available. This review also aimed at identifying and documenting best practices. Countries were invited to showcase their experiences and a series of interviews with experts [see Appendix 1] involved in the work of national statistics systems and international agencies that support the production of SDG indicators was carried out to help understand how the selected SDG indicators are (or are not) produced for FDPs.

This report is organized in four parts. Section II describes the datasets explored, assessing the feasibility of estimating disaggregated indicators based on the data and metadata available, and detailing the main technical challenges encountered

---

1 https://www.unhcr.org/new-york-declaration-for-refugees-and-migrants.html

2 https://www.unhcr.org/comprehensive-refugee-response-framework-crrf.html

3 https://www.unhcr.org/en-us/the-global-compact-on-refugees.html

4 https://unstats.un.org/sdgs/iaeg-sdgs/

5 https://ec.europa.eu/eurostat/web/expert-group-on-refugee-statistics

6 https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/International-Migration/2018_1746_EN_08-E.pdf

7 This proposal was included in a background document on data disaggregation which was prepared by UNSD for the 50th session of the UN Statistical Commission. See: https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/

8 https://unstats.un.org/unsd/statcom/51st-session/documents/BG-item-3n-compilers-manual-E.pdf

9 https://unstats.un.org/unsd/statcom/51st-session/documents/BG-item-3n-international-recommendations-on-IDP-statistics-E.pdf

in performing such disaggregation. Section III reviews the main data sources that may be used to improve the availability of statistics for FDPs and for the generation of disaggregated indicators, including opportunities related to integration among different sources. Section IV showcases two successful national experiences which illustrate the way in which countries have been able to incorporate FDPs in their national statistical systems, and therefore can generate disaggregated information and indicators about this vulnerable group. Section V concludes with a number of recommendations to strengthen this area of work and addresses some of the challenges associated with producing SDG indicators disaggregated by forced displacement status.

Box 1 | The 12 priority SDG indicators recommended to be disaggregated by forced displacement, by priority policy areas[10]

POLICY AREA 1: BASIC NEEDS AND LIVING CONDITIONS

2.2.1
Prevalence of stunting (height for age < - 2 standard deviation from the median of the World Health Organization (WHO) Child Growth Standards) among children under 5 years of age

3.1.2
Proportion of births attended by skilled health personnel

6.1.1
Proportion of population using safely managed drinking water services

11.1.1
Proportion of urban population living in slums, informal settlements, or inadequate housing

POLICY AREA 2: LIVELIHOODS AND ECONOMIC SELF-RELIANCE

1.2.1
Proportion of population living below the national poverty line, by sex and age

4.1.1
Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex

7.1.1
Proportion of population with access to electricity

8.3.1
Proportion of informal employment in total employment, by sector and sex

8.5.2
Unemployment rate, by sex, age and persons with disabilities

POLICY AREA 3: CIVIL, POLITICAL AND LEGAL RIGHTS

1.4.2
Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure, by sex and type of tenure

16.1.4
Proportion of population that feel safe walking alone around the area they live

16.9.1
Proportion of children under 5 years of age whose births have been registered with a civil authority, by age

---

10 This recommendation was included in a background document, developed by the IAEG-SDGs and submitted to the 50th session of UNSC, that included a compilation of all of the categories and dimensions of data disaggregation currently in place and planned by custodian agencies and the policy priorities concerning the most vulnerable population groups (including, for instance, people with disability, migrants, refugees and older people). See: https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Data-Disaggregation-E.pdf

Box 2 | Definition of population groups included in this report

'Forcibly displaced persons' (FDPs) in this report includes refugees, refugee-related populations, and internally displaced people (IDPs) as defined by the concepts and definitions of the statistical framework developed by EGRIS, to be made applicable for national statistical systems. Because the normative definitions of FDPs serve as the basis for collecting FDP statistics, this section will recapitulate these definitions.

The obligations of countries towards persons in need of international protection are contained in the international refugee protection regime, with the 1951 Convention Relating to the Status of Refugees[11] and the 1967 Protocol Relating to the Status of Refugees[12] at its heart, which establishes a specific rights system that protects refugees. Because of the complexity of the legal definitions of persons in need of international protection, simplification is needed to arrive at a workable classification. The IRRS identifies three distinct populations: (a) the population in the country needing international protection (prospective asylum-seekers; asylum-seekers; admitted as refugees; admitted for complementary or subsidiary forms of protection; admitted with temporary protection status; and others in refugee-like situations admitted for other international protection reasons); (b) persons with a refugee background (naturalized former refugees; children and descendants of refugees; family member reunification; and others); and (c) persons who have returned to their home country after seeking international protection abroad.

The Guiding Principles on Internal Displacement[13] states that IDPs are "persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights, or natural or human-made disasters, and who have not crossed an internationally recognized state border." This notion of an IDP is based on two components: 1) that the movement is coerced or involuntary (to distinguish from economic and other voluntary migrants), and 2) that the movement stays within internationally recognized state borders (to distinguish from refugees).

These concepts and classifications are translated into core standard questions that can enable the identification of refugee, refugee-related populations, and internally displaced persons in data sources, in a way that is both practical and cost effective, in the IRRS and IRIS developed by EGRIS.

---

11 https://www.refworld.org/docid/3be01b964.html

12 https://www.refworld.org/docid/3ae6b3ae4.html

13 E/CN.4/1998/53/Add.2, 11 February 1998, UN Commission on Human Rights, Report of the Representative of the Secretary-General, Mr. Francis M. Deng, submitted pursuant to Commission resolution 1997/39. Addendum: Guiding Principles on Internal Displacement, available at: https://www.refworld.org/docid/3d4f95e11.html

BANGLADESH. *Rohingya refugee boys attending lessons at the Pin Gou Njan youth club.*
© UNHCR/VINCENT TREMEAU

2. Exploration of datasets and related technical challenges

An exploration of publicly available datasets was undertaken to determine the existence of variables that would allow for the disaggregation of the prioritized SDG indicators by forced displacement.

Fifty-seven publicly available datasets and reports were examined to establish the options for identification of FDPs and the possibility of estimating disaggregated SDG indicators. Datasets came from surveys and other data collection exercises carried out in 26 countries, namely Afghanistan, Australia, Canada, Colombia, Ecuador, Egypt, Germany, Greece, Iraq, Italy, Jordan, Kenya, Lebanon, Myanmar, Nigeria, Norway, Pakistan, Peru, Somalia, South Sudan, State of Palestine, Sweden, Syrian Arab Republic, Turkey, the United Kingdom, and the United States of America. Appendix 2 presents the detailed list of the datasets for which the calculation of disaggregated SDG indicators was possible.

All the datasets refer to countries with forced displacement situations at the time the survey was conducted. The datasets came from large-scale surveys including the Demographic and Health Survey (DHS), Living Standards Measurement Study (LSMS), Multiple Indicator Cluster Surveys (MICS), and others, as well as from surveys conducted on/with FDPs, and to compare FDPs and host populations.

For each dataset, we attempted to answer the following questions:

1. Was there systematic inclusion of FDPs in the sampling universe?

2. Can forcibly displaced persons be (directly or indirectly) identified in the datasets?

3. What sub-sample sizes are available for FDPs, and are they sufficient for statistically significant disaggregated inferences?

4. Can SDG indicators be calculated from the data available? If so, which are available?

2.1 Limitations and challenges

The datasets selected for this report are not a representative subset of datasets in all countries with FDPs. They were explored because the data producers made their data available for public use. This is consistent with the goal of this report, which is to identify available SDG indicators disaggregated by forced displacement using currently available data. The exploration did not include datasets that have not been made available to the public. We are aware that statistical organizations and agencies working with FDPs will have access to more datasets and full details about each of them.

When working with the datasets themselves, we found that there were some issues that made accessing them difficult. While significant progress has been made with data archives and public access to the same, we found that "online availability" did not always mean that a dataset could be easily used. There were several instances where metadata or supplementary information required to understand the data was unavailable. This included the questionnaire or parts of the questionnaire that would have made it possible to interpret the structure of the data.

This experience highlights the challenges of creating appropriate data archives, and the need to continue efforts to support data collection activities across countries and institutions. If this challenge is to be met, appropriate resources must be allocated to ensure that archiving and curation of datasets is effective and timely, with the provision of appropriate metadata, and following agreed principles of data responsibility. Capacity building for data archiving should be a priority to ensure that data archives can be used for future analyses.

2.2 Question 1: Was there systematic inclusion of FDPs in the sampling universe?

Our exploration shows that there is no systematic exclusion of FDPs from the target populations[14] in the datasets examined. However, the challenge is not the deliberate exclusion of FDPs, but rather the fact that unless a national survey has an explicit objective to generate estimates for FDPs, the survey is not designed to collect a sub-sample of FDPs in a way that allows unbiased estimation of any indicators for this subpopulation.

On the other end of the scale, when a study focuses specifically on FDPs, a challenge comes from its inability to produce estimates that can be used to make inferences about FDPs in the whole country. This is because the objective of the study is generally not to produce countrywide estimates. The latter is relevant given that SDG indicators normally apply to country level, while the estimates obtained from those studies are, by design, representative of segments of the overall forcibly displaced population in a country.

As seen in Table 1, of the 57 studies explored, 29 have national coverage, but only 5 studies were explicitly designed to include FDPs, mainly through the allocation of strata to refugee camps or geographical areas with large presence of FDPs. These five studies were undertaken in Jordan, Iraq and Somalia between 2016 and 2018. A subset of 16 studies with national coverage did not exclude FDPs but were not specifically designed to capture information about them. Consequently, it is not possible to provide unbiased estimates for those subpopulations. There were also three datasets with national coverage and where refugees are the target population. These correspond to surveys of refugees in Germany, the United Kingdom and the United States. The remaining datasets are studies of migrants in Canada, Norway and Sweden which rely on sampling frames or administrative data that enable national coverage.

Table 1 | Number of datasets explored by coverage and form of inclusion of FDPs (n = 57)

Form of inclusion of FDPs in the studies	National coverage?	
	Yes	No
FDPs included by design	5	0
FDPs not excluded, but not included by design either	16	0
FDPs are target population	3	23
Refugees are compared to host population	0	1
Migrants are target population	5	4
Total	**29**	**28**

We found 23 studies **with no national coverage** that had FDPs as their target population. While the issue of inclusion of FDPs is totally solved here, the studies do not allow inferences at national level. Of the remaining datasets, one comes from a study that attempts a comparison between host and refugee populations, and the last four are focused on migrants. While some of these migrants are FDPs, the studies were not designed to focus on them.

---

14 The target population for a survey is the entire set of units for which the survey data are to be used to make inferences. Thus, the target population defines those units the findings of the survey are meant to generalize to.

2.3 Question 2: Can forcibly displaced persons be identified in the datasets?

To answer this question, we explored the documentation that accompanies the datasets, including the questionnaires when available, and the datasets themselves. Out of the 20 studies where the target population was the general population:

- Four of them allow identification of individual FDPs, as they were included by design. These include surveys in Jordan (DHS VII), Iraq, and immigrant-focused studies in Canada and Norway. In the case of Jordan DHS VII, the survey includes a stratification for refugees, but the supplementary information available did not allow to identify the records in the dataset which correspond to the FDP stratum.

- Ten datasets contain variables from questions that would make it possible to infer that the respondent has been forcibly displaced. However, these do not meet standard definitions, or the question was included with a different purpose (i.e. not to identify FDPs). This type of proxy identifiers included variables resulting from questions about:

  › Current migration status

  › Whether the respondent had moved due to a specific event (e.g. a recent civil war, climatic disaster).

  › Push/pull factors for previous migration, including security and safety.

  › Six datasets did not allow the identification of FDPs.

2.4 Question 3: What sub-sample sizes are available for FDPs, and are they sufficient for statistically significant disaggregated inferences?

This is a relevant question for surveys that did not have FDPs as their target population - and there is no single answer. Sample sizes vary from a handful of households, where no attempt was made to establish appropriate sample sizes at the design stage, to several thousand households, when strata were established for FDPs.

The question about whether sample sizes were "sufficient" depends on the levels of precision required. We attempted to produce estimates based on the information available, but in none of the cases explored were we able to use sampling weights or full definitions of the sampling design to calculate unbiased estimates with appropriate margins of error. This is an area where work would be desirable to improve description of sampling designs, design effects and sampling weights in data archives.

Furthermore, in the case of large-scale surveys that are part of well-established processes like LSMS, MICS, DHS and others, engaging with organizations and teams that support those surveys could provide an opportunity to include, when relevant, the objective of disaggregation by FDPs at the stage of survey design. This would make production of estimates for FDPs not only possible but also part of the original analysis effort.

We found that of the 20 studies where the target population was the general population, disaggregation and estimation (not considering the sampling design and weights) were only possible in six of them, as explained in Table 2.

Table 2 | Sample sizes for the six datasets where disaggregation was possible

Survey name	Disaggregation	Comments
High Frequency Survey Wave 2, South Sudan, 2016	Disaggregation was possible by IDP status using two questions. One question was whether they had fled since the December 2013 conflict; if yes, would they have moved regardless? If no, this means their move was directly/indirectly a result of the conflict and therefore they could be considered an IDP. The other question was about the push reasons for their most recent move, including conflict and security concerns. The sample of IDPs included only 122 households.	No appropriate sampling weights were available, and the estimates computed are not unbiased estimates for all IDPs in the country.
High Frequency Survey Wave 3, South Sudan, 2016-2017	Disaggregation was possible by IDP status using two questions. One question was whether they had fled since the December 2013 conflict; if yes, would they have moved regardless? If no, this means their move was directly/indirectly a result of the conflict and therefore they could be considered an IDP. The other question was about the push reasons for their most recent move, including conflict and security concerns. Filtering by these two variables to define IDPs created a subsample of 1,099 household members in 193 households, 555 of which were of working age, with 260 in the labour force.	No appropriate sampling weights were available, and the estimates computed are not unbiased estimates for all IDPs in the country.
High Frequency Survey Wave 4 & Crisis Recovery Survey, South Sudan, 2017	Disaggregation was possible by IDP status using three questions. One question was whether they had fled since the December 2013 conflict; if yes, would they have moved regardless? If no, this means their move was directly/indirectly a result of the conflict and they therefore could be considered an IDP. The second question was the push reasons for their most recent move including conflict and security concerns. The third was a variable defining which dataset the observation was from (HFS or CRS), as all IDPs surveyed lived in camps/settlements covered by the CRS. Filtering by these three variables to define IDPs created a subsample of 13,217 household members in 2,771 households, 7,005 of which were of working age, with 3,711 in the labour force.	No appropriate sampling weights were available, and the estimates computed are not unbiased estimates for all IDPs in the country.
Multiple Indicator Cluster Survey, Iraq, 2018	This dataset could be disaggregated as it included a question concerning the motivation behind an individual's most recent displacement. One answer to this question concerned armed conflict or security issues. This can be used as a proxy to define a refugee or IDP household. 1,740 households were recorded and used for subsequent analysis. This encompassed 11,661 household members, 1,321 children and 2,762 women with birth history, and corresponding information on 6,156 births.	No appropriate sampling weights were available, and the estimates computed are not unbiased estimates for all IDPs in the country.

Survey name	Disaggregation	Comments
High Frequency Survey Wave 1, Somalia, 2016	This dataset could be disaggregated as displacement was a key focus of the survey, with 431 IDP households specifically sampled from settlements. Another 154 IDP households were identified outside of camps using details on push reasons behind the households' most recent displacement including security and conflict issues.	No appropriate weights available and estimates are not representative of IDPs in Somalia.
High Frequency Survey Wave 2, Somalia, 2017-2018	This dataset could be disaggregated as displacement was a key focus of the survey, with 468 IDP households specifically sampled from settlements. Another 665 IDP households were identified outside of camps using details on push reasons behind the households' most recent displacement including security and conflict issues.	No appropriate sampling weights were available, and the estimates computed are not unbiased estimates for all IDPs in the country.

2.5 Question 4: Can SDG indicators be calculated from the data available? If so, which are available?

To estimate an SDG indicator from a published dataset, we identified the following conditions that need to be met:

1. The dataset and its metadata are made available by the data producer or data guardian.

2. Access to microdata is granted by the data producer or data guardian.

3. Technical information about variables contained in the dataset is available, and it is possible to map these variables to the original data collection instrument.

4. A variable or set of variables allow FDPs to be identified and extracted from the dataset (except in the cases of datasets that only included FDPs).

5. Variables required for the computation of the SDG indicator are available.

A total of 121 estimates for the prioritized SDG indicators were obtained. Given that we explored 57 datasets, and that up to 12 prioritized SDG Indicators could have been calculated, in theory the total number of estimates could have been as high as 684 (57x12). Instead, 121 estimates were generated. The difference (563 fewer estimates) is due to the estimation process failing to comply with any of the five conditions set above. For the indicators which could be estimated, an additional limitation was created by the difficulty to calculate sampling weights, which were generally not available, and would make it burdensome to calculate the precision of the estimates. Table 3 presents the number of times each indicator could be calculated for the 57 datasets explored.

Table 3 | SDG indicators computed from the explored datasets

SDG Indicator		Description	# of estimates generated
	1.2.1	Proportion of population living below the national poverty line, by sex and age	6
	1.4.2	Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure, by sex and type of tenure	7
	2.2.1	Prevalence of stunting (height for age < - 2 standard deviation from the median of the World Health Organization (WHO) Child Growth Standards) among children under 5 years of age	6
	3.1.2	Proportion of births attended by skilled health personnel	6
	4.1.1	Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex	8
	6.1.1	Proportion of population using safely managed drinking water services	18
	7.1.1	Proportion of population with access to electricity	17
	8.3.1	Proportion of informal employment in total employment, by sector and sex	7
	8.5.2	Unemployment rate, by sex, age and persons with disabilities	13
	11.1.1	Proportion of urban population living in slums, informal settlements, or inadequate housing	12
	16.1.4	Proportion of population that feel safe walking alone around the area they live	14
	16.9.1	Proportion of children under 5 years of age whose births have been registered with a civil authority, by age	7
	Total		**121**

2.6 Compilation of SDG indicators about FDPs

Compiling SDG indicator estimates for FDPs across time and geography would provide information to better understand the situation of these vulnerable groups and support evidence-informed policy choices and decision making. This is an effort that would need to include multiple countries and international agencies over a certain period of time to assess progress and allow multi-country comparisons.

Reaching this goal would require that the objective of disaggregation by forced displacement status be placed higher in the agenda for data collection processes. It would seek to ensure that the resulting data includes FDPs, makes it possible to identify FDPs (when safe to do so), and includes the variables that allow for the computation of the priority SDG indicators. There is, however, one further challenge: the need to compile the resulting estimates of the SDG indicators for FDPs over time and across geography. This coincides with the point made in the International Recommendations on Refugee Statistics (IRRS), published by EGRIS in 2018[15], which states:

> "Lack of comparability between statistics on refugees and asylum-seekers produced by different countries, and across displacement situations within countries. This arises due to the lack of consistency of terminology, concepts, definitions and classifications, as well as variation in the methods of data collection, compilation and presentation at national and international levels. Data on refugees and asylum-seekers are typically collected for administrative not statistical purposes, and the definitions employed reflect country-specific legislation, policies and practices."[16]

Some of the more general aspects of harmonization, coordination and support required are discussed later in this report, but the exploration of datasets and computation of SDG indicators for FDPs highlighted some challenges:

- None of the priority SDG indicators for FDPs that we found or computed are estimates that represent the whole population of FDPs in a country. This is either because they come from special studies that targeted specific segments of the forcibly displaced population or because they are the result of unweighted estimates from national surveys.

- The variables that were useful to distinguish FDPs in datasets were inconsistent across time and geographies.

- The variables that allowed for the identification of FDPs in a dataset are often practical interpretations of formal definitions for IDPs or refugees or the result of self-reporting, and rarely

coincide with recent recommendations such as those in the IRRS or the IRIS.

- Estimates of FDPs' satisfaction of immediate and ongoing needs were found and are regularly produced by agencies and government bodies that work with FDPs but they do not necessarily align with SDG indicators.

Given these challenges, the resulting collection of SDG indicator estimates may be useful but need to be interpreted with care.

Future efforts by researchers to compile SDG indicators across time and geographies need to be aware that:

1. Compiling the prioritized SDG indicators for FDPs is likely to generate a patchwork of non-harmonized estimates.

2. The interpretation of those estimates requires extra information, in particular the specific definition of the study's target population, the

---

15 And similarly, in the International Recommendations on IDP Statistics, published in 2020.
16 IRRS paragraph 17a

exact definition of variables used to identify FDPs, and a description of known or potential sources of bias for the estimates.

3. A collaborative process of compilation, rather than a centralized initiative, is more likely to be successful because those involved in the original production of the estimates are better informed about the specific characteristics and limitations of the estimates they produced.

4. Involving national partners and international agencies in the process of compiling SDG indicators for FDPs may contribute to enabling disaggregation and making progress towards obtaining estimates that are more harmonized and representative through the inclusion of FDPs in the objectives and design of studies.

As part of the work for this report, we built a repository of the 12 priority SDG indicators for FDPs that we found or were able to calculate from multiple sources. It is planned to make available the resulting set of non-harmonized statistics for a small number of countries on a micro website.

Given the lack of harmonization in the estimates compiled, each of them needs to be accompanied by additional information to allow appropriate interpretation. Making estimates publicly available may help involve the actors generating disaggregated SDG indicators, including in-country partners and international agencies, in the process of compiling the estimates and would contribute towards improving the availability of statistics for SDG indicators disaggregated by forced displacement status. It could also provide an additional opportunity to discuss and make progress on how to improve harmonization, use of common tools, and capacity building.

SUDAN. *Former Sudanese refugees who have been living in Chad are forced to return to West Darfur due to outbreak of violence in Chad. UNHCR and partners are providing urgently needed assistance while an assessment of the evolving situation is underway.*
© UNHCR/ANA BIURRUN RUIZ

3. Main data sources

3.1 Review of main data sources

The International Recommendations on Refugee Statistics (IRRS), published by EGRIS in 2018, and the International Recommendations on IDP Statistics (IRIS), endorsed by the UNSC in March 2020, present a comprehensive discussion of the technical issues, potential uses of data sources, and recommendations to improve the availability of statistics on refugees and IDPs. The Compilers' Manual on Displacement Statistics, also prepared by EGRIS, and currently under revision, takes steps towards supporting the implementation of both recommendations. These recommendations on methodology (design and implementation) coincide with our findings about the main issues that affect the possibility of generating SDG indicators disaggregated by FDPs.

This section describes how different data sources may be used to improve the availability of statistics for FDPs and for the generation of SDG indicators, that are used by national statistics offices.

3.1.1 Population Census

Censuses can provide reliable information about stock figures of FDPs if appropriate questions are included. While population censuses do not collect data that directly allows for the calculation of SDG indicators, there is the potential to use census figures in the preparation of sampling frames for surveys and the derivation of weights for FDPs. If such uses are attempted, it is important to bear in mind that events that affect FDPs between censuses may reduce the relevance of census figures. This implies that appropriate ways to update those figures would be needed; research on approaches to address this issue may be required. We did not find examples of the use of census data for these

purposes, except the confirmation that national governments are not always willing to include FDPs (in particular refugees) as part of the usually resident population[17], and FDPs are therefore excluded from some national census figures.

An international effort to produce a set of questions, suitable for a census form, that would allow the identification of different types of FDPs has already been led by EGRIS. This effort should raise the profile of the need for statistics for these subpopulations and generate tools that could make it easier for countries to use their census for that purpose. The IRRS (Chapter 4, section C.1.b) presents a set of relevant international recommendations and the Compilers' Manual contains sample questions and guidance for implementation.

3.1.2 Administrative data sources

Some countries have well-developed administrative data systems that can support the production of statistics for subpopulations of interest, including FDPs. The "Country in Focus" cases included in this report, Colombia and Norway, illustrate an example of how such systems can work. Those systems have been developed over time, driven by national information needs, accompanied by legislation that provides the mandate and resources for their establishment and maintenance. In addition to the technical and resource requirements to establish such systems, it must be highlighted that unless integration across datasets from different sources is possible in the country, the data available may not be suitable for estimation of statistics about SDG indicators.

Even where effective administrative data sources are available, it is important to be aware that they are useful to provide data for information needs that

---

17 Principles and Recommendations for Population and Housing Censuses, Revision 3, UN Department of Economic and Social Affairs Statistics Division. https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf

do not change substantially over time. If information needs change rapidly, administrative systems tend to be too slow at adapting. In consequence, crisis situations involving FDPs are unlikely to be appropriately captured by routine administrative processes.

Administrative data sources that are used to generate statistics about the general population, including SDG indicators, do not always include the ability to identify FDPs. For example, sources of education indicators rely on administrative sources of data for their estimation but rarely allow the identification of FDPs. Work would be required with national governments and international agencies to adapt the education management information systems (EMIS) to include the possibility of FDP identification to enable the generation of those disaggregated statistics, for example for the estimation of SDG indicator 4.1.1 for FDPs. This, however, should be based on a context-specific assessment of the protection risks created by the disaggregation of education data.

As with censuses, administrative data sources can also provide sampling frames for surveys that can be used to collect information for the calculation of disaggregated indicators.

3.1.3 Large-scale surveys

The design of surveys for populations that are rare, unevenly distributed, or difficult to find[18], as is often the case with FDPs, is a well-known challenge for sampling specialists. Its solution starts with the ability to identify the members of those populations. In the case of FDPs, this implies including questions for FDP identification in survey questionnaires.

Our exploration of datasets and questionnaires from surveys in countries with FDPs showed that little progress has been made in the inclusion of questions that allow FDP disaggregation based on internationally accepted, harmonized definitions. The basis for harmonization has been set by the IRRS and IRIS, and practical steps towards making it easier to use those definitions have already started among international agencies working with FDPs. However, in our exploration of surveys and datasets,

we found that FDPs tend not to be a priority, and the sampling of national surveys is rarely designed to allow estimation of statistics for FDPs. This could be due to lack of capacity, political sensitivities, or protection concerns. More details on specific issues are presented below.

Issues with sample designs

Large-scale surveys, designed to represent the population of a country, generate data suitable to estimate statistics, including SDG indicators and other indicators, for the national population. They are the approach of choice for many SDG indicators. However, they are normally designed with objectives that do not prioritize estimation for forcibly displaced populations. As a consequence, the resulting sample sizes for FDPs tend to be small and the statistics produced have low precision and may be biased.

Under some conditions, a case can be made for these surveys to attempt over-sampling of FDPs. However, while the resulting sample size will be larger, it can be difficult to derive sampling weights in cases where reliable information about the overall size of the forcibly displaced population in the country is not available.

Sampling weights

The data that allows for the derivation of appropriate weights for FDPs is rarely available. In our exploration of datasets, these weights were not available at all. As a consequence, the extent to which estimates of disaggregated SDG indicators for FDPs can be considered representative of all FDPs in the country remains difficult to establish. While in some cases estimates can be calculated for FDPs, they are likely to be biased and cannot be interpreted without caveats about who they apply to. In the case of SDG indicators, this is risky (and not recommended), as the general expectation is that SDG indicators are representative at country level.

Harmonized questions

Most of the surveys explored for the purposes of this report did not use harmonized questions for the identification of forcibly displaced respondents,

---

18 A rare population is sometimes defined as one with a low number of individuals. However, even a large population can appear to be rare either because of the elusive behaviour of its members or because the population is sparsely distributed over large geographic areas.

3. MAIN DATA SOURCES

even when these sub-populations were included by design in the survey. To help change this situation, and following the recommendation from the IRRS and IRIS, a battery of easy-to-use questions, based on harmonized definitions, should be made widely available, and standard surveys should be encouraged to include them. The Compilers' Manual has already made progress in this area, as part of EGRIS's phase 3 of work focusing on refinement and implementation.

It is worth bearing in mind that large-scale surveys strive to keep questionnaires and interview times as short as possible. Adding questions that allow the production of SDG indicators disaggregated by FDP subpopulations is likely to face resistance unless these are easy to implement and do not represent a significant increase in the time of the interview. It is important to avoid unnecessary burden on respondents (and potential harm), especially when dealing with vulnerable groups and/or conducting surveys in humanitarian or fragile contexts.

Examples: The South Sudan HFS Wave 4 (2017) and the Somalia HFS Wave 2 (2017-2018)

South Sudan HFS Wave 4

This is likely the best example of a dataset that allows calculation of SDG indicators for IDPs.

The crisis recovery survey portion of this dataset is one of the most complete sources of data for IDPs, as the depth of the survey allowed for more indicators to be calculated than in any other survey, and the sampling focus on IDP camps meant that the sample size was large, not only for households but also household members. Hence, it was possible to calculate indicators for education as enough children were recorded in the data. This is compared to the DHS, for example, where after filtering, there were no children in the households born within the past five years to allow calculation of birth registration.

South Sudan HFS Wave 2

This survey included questions and sampling methods designed to capture IDPs in the data, both in and out of camps. This is an example of how a large-scale survey can be designed to allow calculating SDGs for FDPs.

The questionnaire provided online does not include the module with one of the questions required to calculate SDG indicator 16.1.4. Fortunately, the dataset contains the data and through careful reconstruction it was eventually possible to calculate it.

The question on whether the household would have moved regardless of the 2013 conflict was particularly useful. This type of question could be useful in contextualizing specific situations when trying to capture IDPs and refugees. For instance, while we could make assumptions based on a person's migration history, birthplace etc., if we do not know the reasons for their move then we would not be certain about their FDP status.

3.1.4 Specialized studies

As data sources, specialized studies have the main advantage that they target specific FDPs, identified and delimited for the purpose of the study. Well-designed studies of this type yield estimates that are representative of the 'target population' as

defined by the study's objectives, and at times also at country level.

The important question to address here is whether, in the case of SDG indicators for FDPs, users of statistics should be presented with statistics about FDP populations that are not unbiased estimates of

DATA DISAGGREGATION OF SDG INDICATORS BY FORCED DISPLACEMENT   20

all FDPs in the country. This means that estimates can only be used if the user is fully aware of the specific definition of the study's "target population". This will be the population that the estimates represent. On that basis, it may be possible to assess potential biases when making more general inferences. While in theory this is possible, in practice there are many opportunities for unintended misuse of this type of statistics. This is further complicated by the fact that SDG indicators are normally produced at country level, and there is an implicit expectation that SDG estimates for FDPs would also be representative at country level.

More work would be needed on how to use estimates from specialized studies of FDPs, including discussions with users and producers of statistics at national and international level about how to address the issues highlighted. This problem is not unique to the case of FDPs; progress has been made with the estimation of SDG indicators at sub-national level and it is likely that the work on SDG indicators for FDPs can learn from this experience and approach.

Other concerns with specialized studies as a source of data include: these studies may not use standard indicators; they tend to be heavily influenced by a thematic interest; and they tend to be considered expensive.

3.2 Integration of data sources

The IRRS (Chapter 4, Section E) makes specific recommendations that provide the basis for the work on the integration of data sources. In the short-term, the most likely steps that will contribute towards this are related to making progress with harmonization, improving data accessibility, sharing of methodologies, aggregation of statistics, availability of metadata, and in particular, the use of the same Personal Identification Number across all datasets.

Disaggregation of SDG indicators requires the ability to differentiate at the micro-data level those individuals that qualify as FDPs, whatever the source of data. This requires a high level of harmonization not yet achieved across datasets and geographical areas explored.

The exploration of datasets highlighted that the progress on harmonization of definitions and questions is limited, that communication across

international agencies can still be improved and that support to national statistical systems to enable harmonized data collection needs to be stepped up. However, the relatively recent development of the IRRS, IRIS and the Compilers' Manual as well as UNSC's endorsement in March 2020 of the Phase 3 of EGRIS on the implementation of the recommendations, is expected to help improve this situation.

3.3 Non-technical challenges related to the availability of disaggregated indicators

3.3.1 Relevance of and mandate for FDP statistics

The relevance of and interest in statistics for FDPs from national and international stakeholders vary over time and across geographies. This affects the priority and allocation of national or international resources invested towards the generation of statistics for FDPs. Overall, with some exceptions, at the national level the interest in the disaggregation of SDG indicators by FDPs is generally still low. In countries with refugees and IDPs, there is some emphasis on the estimation of stock and flow figures, but SDG indicators about FDPs seem not yet to be considered priority information. If the availability of SDG indicators disaggregated by FDPs is to improve, increased buy-in at the national level is needed. A strategy for improving the availability of disaggregated data must acknowledge the variation in the importance of forced displacement issues across countries, and improve collaboration and responsible information sharing.

International agencies have mandates to support and collect thematic data about the general population or subpopulations linked to their mandate, disaggregated by age, sex, or other relevant diversity characteristics, but they do not have explicit obligations to disaggregate indicators by forced displacement. This was consistently found in the interviews conducted and confirmed in datasets made available by those agencies. The tension between the thematic mandate and the transversal disaggregation will remain until agencies explicitly adopt FDP disaggregation as part of their own, or shared, information needs, bearing in mind the potential protection risks in certain contexts.

The new SDG indicator 10.7.4 ("Proportion of the population who are refugees, by country of origin")[19], which was added to the global indicator framework following the 2020 Comprehensive Review, explicitly recognizes the relevance of forced displacement for the achievement of the 2030 Agenda, and provides an excellent platform to further the discussion about including disaggregation by forced displacement across SDG indicators. Increased awareness and communication strategies should be pursued to build on this achievement that made refugees explicitly visible in the 2030 Agenda, and to make sure it becomes a first step towards the disaggregation of SDG indicators and other statistics for FDPs.

3.3.2 Country context and information needs

Government information needs and the relevance of FDP statistics in a country, at a specific time, determine the efforts to gather data and generate statistics for FDPs by country agencies. Demand for information needed for decision-making, together with the willingness of governments to allocate resources to the production of relevant indicators, creates conditions for the generation of statistics for FDPs. This goes hand in hand with the creation of a legal framework that enables the national statistical system to allocate resources and organize data collection activities to meet the demand for this type of information.

The case of Colombia[20] illustrates how a complex system of IDP statistics evolved from the information needs associated with the reparations to the victims of the long internal armed conflict in the country. The information system for IDPs in Colombia responds directly to the government's need for statistics on the number of IDPs, and indicators on the extent to which needs are met. The existence, functions, and operations of each element of the information system is governed by legislation at all relevant levels. The information needs led to the creation of a system of integrated data sources from institutions that gather data and are custodians of data (including, but not

exclusively, data on IDPs), and to the inclusion of questions into the national census and large-scale surveys that allow production of statistics and analysis for IDPs.

A different example is Norway[21], where a culture of transparency and the existence of an effective national data infrastructure has led to the integration of data sources, particularly administrative data, at country level. The availability of statistics for sub-groups of special interest was a result of an initiative by Statistics Norway, which has worked to make it possible to generate statistics for those groups. While Norway represents a successful case of integration of data sources, its replicability may be limited given that only few countries have the necessary legal framework, culture of transparency, and the ability to link administrative data across data sources using unique personal identifiers.

In both examples above, administrative sources of data made important contributions, as these are contexts that change slowly and which remain unchanged for the long term. In Colombia, providing services and reparations for IDPs has been high on the political agenda and will continue to be so for the central and local governments. In Norway, the administrative data systems are part of a well-established system of information for the government at all levels and across sectors, which covers most aspects of life of the resident population, including the refugee population.

These two systems would be less able to cope in situations of crisis where conditions, and therefore information needs, change fast, mainly because administrative data sources are not built to be easily adaptable. This has been observed in Colombia with the influx of Venezuelan refugees and migrants, who are not covered by the same systems as the internally displaced population. It should be highlighted that Colombia is working towards integrating data sources and has introduced changes to their "Gran Encuesta Integrada de Hogares" (GEIH), a National Household Survey, to better capture information about refugees and migrants.

19 See UNHCR Global Trends 2019, p.26 (https://www.unhcr.org/5ee200e37.pdf)
20 See Country in Focus: Colombia
21 See Country in Focus: Norway

While administrative data sources may not be best suited to adapt to fast changing conditions or may not provide data to estimate the 12 SDG priority indicators for FDPs, the institutional infrastructure that exists in countries like Colombia and Norway is an important enabling factor for FDP disaggregation of SDG indicators.

3.3.3 Diversity of objectives

Statistics available from governments and international agencies about sub-populations of interest include, but are not limited to, SDG indicators. Indicators that are useful for agencies to assess situations, manage programmes or monitor activities do not necessarily map exactly to SDG indicators. For example, UNICEF's Water, Sanitation and Hygiene (WASH) survey collects and publishes data for indicators that are relevant to the agency's work but that do not correspond exactly to SDG indicators. They also do not include all 12 SDG priority indicators identified as priority by EGRIS for FDPs. This, of course, is due to the different purposes of the indicators (they measure a different 'what', at different units of measurement, and at different moments in the programme cycle), but it does nonetheless create challenges for reporting on SDG indicators by forced displacement status.

There is also the issue of the overlap in sub-populations of interest, and how the responsibility and incentives for generating information about them affect the statistics produced by agencies, either at national or international level. This generates related, but not uniform, sets of statistical indicators for overlapping populations that are covered by the mandate of those agencies. While the collective term Forcibly Displaced Persons is useful at a high level, in the process of generating statistics, the composition and distribution of populations for whom statistics are produced varies based on the mandate of the agency. This poses an extra challenge when attempting to bring together statistics generated by different institutions. For example, in a conflict setting where IOM, UNHCR, OCHA and UNICEF may be working with FDPs, IOM would focus on migrants or IDPs, UNHCR on refugees and IDPs, OCHA on IDPs, and UNICEF on children, including FDP children. These agencies produce data and statistics that reflect their respective mandate and information needs, which is of course important in terms of deploying their mandate. However, it creates challenges in combining datasets due to differences in population definitions and sets of relevant indicators.

At the national level, this issue can be illustrated by the complex set of indicators used in Colombia for assessing the satisfaction of needs of IDPs. The system is tailored to meet the information needs of government programmes. Those indicators overlap with - but do not focus on - SDGs. Despite the general agreement about the importance of SDGs and the consequent international harmonization they have brought about, we encountered a tension between contextual purposes and information needs and the data requirements for reporting on SDGs, which will continue to exist.

While efforts to support the production of the prioritized SDG indicators are important for international purposes, information about FDPs is determined by the demand for statistics that inform evidence-based policy and decision making in specific contexts. As a consequence, the resulting statistical sets do not necessarily map with the 12 priority SDG indicators.

3.4 A case for inclusion of FDP statistics into the official mandates of international agencies

The emphasis and resources that international agencies allocate to disaggregate their data by forced displacement status is determined by their mandates and operational needs. While this is obvious, even when there is a genuine intention to improve the availability of FDP statistics, the absence of a specific mandate on forced displacement affects the ability of agencies to collaborate in this area. The following points illustrate this case. Interviews with international agency personnel indicated that:

- Disaggregation by FDPs was not found to be part of the mandate of custodian agencies other than those directly working with specific groups of FDPs, and as such, they do not see it as their responsibility. This produces a patchwork of coverage and representativity of the indicators available. Even when there are overlaps with issues related to forced displacement, as in the case of IOM, their mandate is for migrants, which entails a different definition and focus for data disaggregation. This is one of the reasons why

international agencies may not collect data that allows disaggregation of SDG indicators by FDPs.

- Staff working in organizations that are custodians of SDG indicators highlighted that their duty did not include disaggregation of the indicators by FDPs. We were told that they are aware of discussions and efforts to produce disaggregation by FDPs, and while they expressed their interest in contributing to the efforts, interviewees indicated that as long as the disaggregation remains outside their mandate, it is unlikely they would be able to allocate resources to the relevant data collection or production of statistics.

- There are some exceptions that occur in geographical areas where forced displacement is an especially relevant issue, as for instance in the Middle East and North Africa. In these cases, the agencies' mandates include FDPs and, consequently data availability, as well as efforts for cooperation and harmonization to enable the disaggregation of data by FDPs. However, definitions and metrics are not always fully harmonized or interoperable, and comparison of the data across geographies, agencies and over time is therefore not straightforward.

KENYA. *UNHCR distributes hygiene kits and firewood during the COVID-19 crisis.*
© UNHCR/SAMUEL OTIENO

4. Countries in Focus

This section presents two case studies which illustrate examples of successful ways in which countries have been able to incorporate FDPs in their national statistical systems, and therefore can generate disaggregated information and indicators about this vulnerable group. While these indicators do not always align to SDGs, this section shows the way in which NSOs have been able to approach the issue and highlights some enabling factors that may make it possible for other national statistical systems to generate this type of information in the future.

4.1 Colombia: Disaggregating statistics by IDP status[22]

4.1.1 The road to IDP statistics

The first legislation to address internal displacement in Colombia was Law 387/1997.[23] This law provided a precise definition of an IDP and adopted measures to prevent forced displacement, and consolidate the socioeconomic situation of internally displaced persons due to armed conflict.

In 2004, by Ruling T-025, the Supreme Court declared an Unconstitutional State of Affairs, to acknowledge the failure of both the Legislative and Executive branches of government in enforcing public policies against widespread and systemic violation of human rights. The Court highlighted the condition of extreme vulnerability of the displaced population, not only because of the displacement itself, but also because in most cases they are people under special protection by the Constitution, such as women heads of household, minors, ethnic minorities and senior citizens.

With Ruling T-025, the Government was due to establish specific indicators to detect if the intended goals of the legislation were met. Two years later, the National Government adopted what is known as *The Effective Enjoyment of Rights Indicators Library*. This framework is intended to measure the progress towards the fulfilment of the effective enjoyment of rights of the internally displaced population.

Current legislation on armed conflict in Colombia, The Victims Law (Law 1448/2011), endorses the need to measure this progress through a Vulnerability Assessment. The Victims Law allocates that responsibility to the Victims Unit,[24] the Government Agency responsible for evaluating displacement-related vulnerabilities of IDPs (see below). This assessment is carried out through administrative records and the existing mechanisms to monitor IDP households. With nearly 9 million victims registered as of December 2019, 85 per cent of which are IDPs, the Effective Enjoyment of Rights Framework has led the process of allocating resources to humanitarian response and the decision-making process to end displacement.

Law 1448/2011 created the Unit for the Attention and Integral Reparation of the Victims (UARIV, Spanish acronym), the government agency responsible for coordinating the National System of Attention and Reparation for Victims regarding the implementation of public policy to assist and provide comprehensive reparation to victims. The Victims Unit has within its mandate the task of guaranteeing the operation of the National Information Network for assistance and reparation to Victims, including the interoperability of the various relevant information systems. The Unit also manages the Single Victims Registry, overseeing the integrity of all existing records about displacement and human rights violations due to armed conflict.

As coordinator of the National System of Attention and Reparation for Victims, the Victims Unit gathers official data and produces official statistics on internal conflict. These statistical operations are reported to the National Administrative Department of Statistics (DANE, Spanish acronym) to be included into the National Statistical System, particularly those related to the Single Victims Registry and the Vulnerability Assessment for IDPs.

4.1.2 Sets of statistics that are available

As the core entity of the national statistics system, DANE produces the socioeconomic and demographic information of the country (Decree 262 of January 28 (2004) Article 1), and within its mission emphasizes the importance of compliance with international standards and the use of innovation and technology to support the comprehension and solution of the country's social, economic and environmental problems. Two additional statistical operations managed by DANE can either be a source of information or be matched with existing data for producing IDPs statistics: the Great Integrated Household Survey (GEIH, Spanish acronym) and the National Population and Housing Census (CNPV, Spanish acronym).

The Great Integrated Household Survey (GEIH, Spanish acronym) is a survey and the official source of information for labour market statistics and monetary poverty of the resident population in Colombia. This survey also contains information related to the socio-demographic characteristics of the population. The GEIH is statistically representative at the national level and for some subnational domains: urban and rural areas, the 23 biggest cities in Colombia and at the regional level. The questionnaire includes a migratory module, which allows for the collection of information on the migratory dynamics at the national level of the last five years and the last twelve months from the date of the survey. It allows characterization of the IDP population by a variety of factors, has national coverage, and has temporal and geographical levels of disaggregation.

Since 2012, the GEIH migration module has been expanded and modified to allow better characterization of the population (DANE, 2019). As of mid-2020, the GEIH was being redesigned, and two more questions were expected to be added to the migration module.

The National Population and Housing Census (CNPV, Spanish acronym) has the purpose of counting and characterizing the universe of persons and households residing in Colombia, and it is the statistical operation with the greatest possibilities to obtain disaggregated data by population groups, based on their location, gender, migration status, and other characteristics. Thus, the CNPV allows the generation of statistical information that supports decision-making in a variety of areas relevant to the interests for FDPs and it is useful to evaluate the scope of development policies targeting specific groups of the population (DANE, 2019).

The official methodological document of CNPV 2018 considers that the census, as a statistical operation that reaches all parts of the country, may be quite appropriate for collecting recent information on internal mobility, which cannot be captured through other surveys. The variables incorporated into the CNPV 2018 allowed for the characterization of the internally displaced population and international migrants (DANE, 2018).

In 2018, the implementation of digital technology strengthened DANE's data collection processes and helped facilitate census processing, especially with the first electronic census (eCensus). Other innovations implemented in this census included the process of consultation and concertation with ethnic groups, the monitoring and control of the operation, and the use of administrative records as input to evaluate the census.

In addition to the CNPV and the GEIH, the following statistical operations capture information on IDPs:

- Production and analysis of annual migration statistics
- Population projections and demographic studies
- Statistics on the Single Victims' Registry
- Statistics on reestablishment of rights

Currently, information derived from statistical operations conducted by DANE, including anonymized metadata and microdata, can be found

in the National Data Archive (ANDA).[25] Statistical Operations are grouped in three categories: society, territory and economy.

When it was first proposed in 2007, the Effective Enjoyment of Rights Framework was composed of indicators per right. Each right would be attached to a main indicator, complementary indicators and sectorial indicators. Main indicators aimed to measure the overall progress, whereas complementary indicators aimed at breaking down main indicators into relevant disaggregation. Sectorial indicators were designed to help policy makers oversee progress regarding government sectors in order to guide budget allocation decisions and address institutional challenges.

The first battery of indicators was composed of the following main indicators: housing, health, education, food, income generation and identification documents. A composite indicator of socioeconomic stabilization was defined with the insertion of displaced households into the Social Protection System, plus the percentage of families that gradually met the stabilization criteria.

A year later, new indicators were added to measure preservation of life, integrity and freedom, coverage of minimum subsistence-related needs, and support received for family reunification, as well as sectorial indicators on participation and coordination.

In 2014, taking advantage of the Victims Law and of the need to periodically measure displacement-related vulnerabilities at an individual level, the National Government adopted Decree 2569 by which a set of socioeconomic indicators were established as the criteria to measure progress towards overcoming the situation of vulnerability of displaced persons. The criteria were based on the Durable Solutions framework adopted by the United Nations in 2010.[26] The Victims Unit is the government agency responsible for conducting the assessment.

Along these lines, the UARIV (Unidad para la Atención y Reparación Integral a las Víctimas), through the National Information Network for the

Assistance and Reparation to Victims, made progress to include more than 100 official administrative records from all government agencies responsible for providing response to displacement-related vulnerabilities. As of 2020, the vulnerability assessment is conducted twice a year, using only administrative records regarding the rights to identification, health, education, income generation, housing, family reunification and food (the income generation indicator is used as a proxy for the food indicator, for which no data is available). Even though administrative records still do not reach the entire IDP population, it is the closest available option to periodically and feasibly cover most of it.

4.1.3 Uses of the data

Displacement statistics are used both for operational purposes and public policy making and data is used by government agencies part of the National System of Attention and Reparation for Victims. The Single Victims Registry, which counts the stock figure of victims since 1985 and disaggregates by year and place of displacement, other victimizations, age, gender, ethnicity and current geographic location, is used by these entities to allocate resources nationwide to assist victims. These data are also available online.[27]

The results of the vulnerability assessment at an individual level are sent to the relevant government agencies in order to target social assistance, mainly in income generation, housing and social subsidies. The aggregate statistics are also reported to the Constitutional Court as the official mechanisms to measure progress to durable solutions to displacement and adjust public policy if needed.

Based on Colombia's technical leadership, both statistically and conceptually, the aim is to unite efforts and knowledge between the UARIV and the DANE to increase the regular statistical production of the most relevant IDP indicators, considering the public policy priorities. This initiative is directly related with the central, transformative promise of the 2030 Agenda of "Leave No One Behind", that

---

25 https://www.dane.gov.co/

26 https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/15%20UN%20framework%20ending%20displacement.pdf

27 www.unidadvictimas.gov.co (click on "Reportes by Red Nacional de Información").

JORDAN. *Syrian refugee children are playing with old tires just before sunset in Azraq camp.*
© UNHCR/CHRISTOPHER HERWIG

implies, among others, a quantitative monitoring of the progress of IDPs towards the targets, highlighting measurement and information needs in the generation of data for decision-making and public policy design.

This initiative seeks to take advantage of available data from both entities and the cross-checking of information that may be used to generate the most complete characterization for IDPs and refugees, also allowing to identify information and data gaps. The alignment with the 2030 Agenda enhances the generation of statistics for different population groups to conduct comparisons and to design better targeted policies. Furthermore, this initiative aims to guide previous studies that the UARIV is carrying out for the evaluation of the renewal of the Victims Law.

4.2 Norway: Disaggregating statistics by refugee status

4.2.1 Refugees in Norway[28]

Forcibly Displaced Persons in Norway fall under four main categories:

1. Asylum-seekers
2. Resettled refugees
3. Other refugees
4. Family members of refugees

Refugees are given a temporary or permanent residence permit, and a unique national personal identification number (PIN), like all other legal residents of Norway.

Most refugees apply for citizenship and become citizens after the required period of residence, usually seven years. Stateless refugees can apply for citizenship after three years of residence in Norway.[29] Asylum-seekers whose applications were rejected but who remained in Norway are not considered residents of Norway and most of them do not have a PIN.

---

28 Section contributed by Helge Brunborg, Statistics Norway. Vebjørn Aalandslid, Lars Østby, and Kåre Vassenden, also from Statistics Norway, provided useful comments.
29 See https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/statelessness-many-worldwide-few-in-norway.

4.2.2 Population registration in Norway

Norway has an extensive and well-developed system of administrative registers with data on the population since the 1960s. These registers can be linked through the PIN to produce statistics from different sources.

All current and previous residents of Norway are registered in the Central Population Register (CPR) of the Norwegian Tax Administration. The CPR includes variables such as country of birth and date of immigration for people not born in Norway,

but there is no variable indicating refugee status. However, Statistics Norway publishes annual statistics on 'persons with refugee background' based on information from the Directorate of Immigration and the CPR, which includes the table below.[30] This variable includes persons who were refugees when they immigrated to Norway for the first time, as well as their family members who came later, but not children born in Norway with one or two refugees as parents.

Table 4 | Persons with refugee background in Norway, as of 1 January 2020[31]

	1 January 2020	Change 2019-2020, per cent	Persons with refugee background in per cent of:	
			All immigrants	Total population
Total	238,281	1.9	30.1	4.4
Principal applicants	**173,524**	**2.0**	**22.0**	**3.2**
Asylum-seekers	115,080	0.9	14.6	2.1
Resettled refugees	40,676	6.6	5.1	0.8
Other refugees	9,606	-0.9	1.2	0.2
Unspecified	8,162	-0.7	1.0	0.2
Family connection to refugee	**64,757**	**1.6**	**8.3**	**1.2**
Family enlargement	17,656	2.1	2.2	0.3
Family reunification	47,055	1.5	6.0	0.9
Family unspecified	46	-2.1	0.0	0.0

Statistics Norway has published several analyses of refugees in Norway, most recently in a special issue of the journal *Samfunnsspeilet* in 2016, which was translated into English in 2017.[32]

---

[30] See https://www.ssb.no/en/befolkning/statistikker/flyktninger/aar. See IRRS at https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/International-Migration/2018_1746_EN_08-E.pdf for a definition of "persons with refugee background."

[31] Source: https://www.ssb.no/en/befolkning/statistikker/flyktninger/aar

[32] See https://www.ssb.no/en/statistical-analysis/2017

4.2.3 Disaggregation of statistics on refugees

Because of the system of administrative registers including unique PINs, statistics on the flow of refugees to Norway and the stock of people with refugee background in Norway can be disaggregated in the same way as all other register-based statistics on the population of Norway. This applies to basic demographic variables such as age, sex, country of birth, citizenship, educational attainment, fertility, mortality, emigration, etc., but also to sustainable development indicators such as infant and maternal mortality rates, poverty indicators, etc.

Concerning refugees in Norway, the SDGs can be grouped into different categories based on a partial list of estimates of SDG indicators published on 1 April 2020 in Norwegian at https://www.ssb.no/sdg (English version forthcoming). In general, since refugees have PINs like all other residents, statistics on SDG indicators that focus on individuals and that can be produced by administrative data, such as 1.2.2, 1.3.1, 3.2.1, 3.2.2, 3.6.1, 3.7.2, 4.1.1, 4.2.2, 4.3.3, 5.5.1, 5.5.2, 5.b.2, 8.6.1, 8.8.1 and 17.8.1, may be produced for refugees as well - some quite easily, while others require more work on record linkage and perhaps also for confidentiality issues. Other SDG indicators that focus on individuals, but which cannot be estimated from administrative sources, can sometimes be estimated from sample surveys. Statistics Norway has conducted several sample surveys of immigrants, including on living conditions. The last national survey was carried out in 2016.[33] As the sample is quite large (4,435 interviewed) and persons with refugee background are about one third of all immigrants, it should be possible to estimate some of the SDG indicators for refugees from these surveys. The Labour Force Survey may perhaps also be used for this, although the sample of refugees is considerably smaller. For other indicators, special surveys would probably be required, for example, on 3.a.1, 5.4.1, 16.1.4 and 16.b.1.

SUDAN. *South Sudanese refugees and their host community received assistance following floods in Bahri, Al Droshab, Khartoum.*
© UNHCR/ROLAND SCHÖNBAUER

---

33 See https://www.ssb.no/308564/living-conditions-among-imigrants-in-norway-2016

5. Recommendations

Through the exploration of the 57 datasets, this report has drawn some important conclusions regarding the challenges associated with producing SDG indicators disaggregated by forced displacement status. The exploration also allows us to formulate a number of recommendations to strengthen this area of work.

1. **Develop an awareness raising strategy**

Significant progress has been achieved with the publication of the International Recommendations on Refugee Statistics, the International Recommendations on IDP Statistics, and the Compilers' Manual on Displacement Statistics. Together with the work of UNHCR, JIPS and other actors, the approval of EGRIS' Phase III of work - which will prominently focus on capacity building for forced displacement statistics - and other initiatives like the World Bank-UNHCR Joint Data Center on Forced Displacement, these documents set a solid basis on which to move forward. To amplify their use and impact, there is a need for complementary efforts to raise awareness among producers of statistics about the guidelines, definitions, and other technical aspects that will help them generate SDG indicators disaggregated by forced displacement. With this aim, a strategy should be developed to raise awareness about the international recommendations, definitions, manuals, tools, and support options for the disaggregation of SDG indicators by FDPs. This strategy should be tailored to two specific audiences:

   a) International agencies working with FDPs or supporting national statistical systems;

   b) Producers of statistics in countries where forced displacement is or could become a relevant issue.

2. **Make it easier to generate micro-data that is suitable for disaggregation**

In order to address the current challenge of reconciling the different questions used by different data collection processes to establish whether the individual/household can be classified as a FDP, custodian agencies and agencies in charge of large-scale surveys should continue with the development and promotion of guides and tools that make it easy to use harmonized definitions and survey questions in data collection processes. It is suggested that a survey module for the identification of refugees and internally displaced persons in national multi-topic household surveys be developed, as well as modules for the collection and processing of data for estimation of each of the prioritized SDG indicators. This should be done in partnership with national governments, and any other actor supporting the implementation of large-scale surveys. These modules should be available in multiple formats (electronic and non-electronic), together with documentation that enables users to take them off-the-shelf and include them in their data collection processes. Specifically, this work includes:

   a. Providing questions in a variety of formats (for example, written form, ODK, CSPro, etc.) to encourage their inclusion in data collection instruments.

   b. Providing data processing scripts to make it easy to process the micro-data. This should result in datasets which include variables that allow for the calculation of the prioritized SDG indicators, in addition to the required disaggregation variables. These scripts should be made available for commonly used statistical packages.

   c. Providing supporting documentation for the integration and implementation of those questions into data collection processes, differentiating modules for surveys and questions for census.

3. **Step up collaboration efforts**

There is interest and goodwill across international agencies to improve the availability of statistics on FDPs. However, there is also competition for resources arising from the difficult financing environment, multiple mandates, and specific information needs. Without increased collaboration and awareness of FDP-related issues, international

agencies and other actors that are already working towards ensuring that data is collected to estimate SDG indicators are unlikely to take the extra steps needed for the disaggregation of those indicators by forced displacement. Priority may be given to a discussion among agencies that are custodians of SDGs, but also to identifying and selecting some of the actors who support primary data collection through administrative systems and large-scale surveys. UNHCR, the World Bank-UNHCR Joint Data Centre, JIPS and the EGRIS have mandates, incentives and opportunities to increase coordination and synergies in this area.

It is also recommended to target key stages in the review of data collection processes of international agencies with the aim of including the disaggregation of SDG indicators by FDPs in their policies, guidance and/or information systems, especially in countries with large forcibly displaced populations. This should include agencies and institutions supporting large-scale international surveys such as LSMS, MICS, DHS, and data derived from administrative records like PRIMERO (UNICEF) and HMIS (WHO).

4. Identify easy gains

When working to implement and test initiatives, capacity building efforts should be concentrated in geographical areas where forced displacement is a current issue in order to increase visibility. This may help to elevate the issue of disaggregating SDG indicators by FDPs on the agenda of governments, international agencies, and others interested in addressing information needs in displacement situations. Actions to make it easier to adopt harmonized approaches for data collection, data processing, and responsible sharing of information will help to achieve this.

5. Identify, systematize, and aggregate available data

Data dissemination systems should be strengthened to make estimates about SDG indicators disaggregated by FDPs more easily available to users. This will require collaboration between agencies and will also highlight current strengths and weaknesses in the indicators themselves and generating them. This should be an opportunity to develop joint efforts to improve the availability and quality of the statistics of interest. This work would involve:

a. Identification of data producers that are interested in contributing to the production of SDG indicators disaggregated by FDPs.

b. Provision of targeted support to generate and contribute those indicators to a common repository, including the global SDG portal.

c. Accumulation of estimates of SDG indicators disaggregated by FDPs as well as the information needed to make proper use of these statistics.

Appendix 1. List of interviews

Organization	Interviewee	Interview date
UNHCR	Erica Cristina Aiazzi	Oct/2019
Joint IDP Profiling Service	Natalia Baal	Nov/2019
UN-Habitat	Dennis Mwaniki	Nov/2019
UNHCR	Eva Barrenberg	Nov/2019
DANE SDG Indicators Working Group (Colombia)	Anggie Katizza, Carvajal Arciniegas	Dec/2019
ILO	Rafael Diez de Medina	Dec/2019
ILO	Hakki Ozel	Dec/2019
Statistics Norway	Live Margrethe Rognerud	Dec/2019
Statistics Sweden	Viveka Palm	Dec/2019
UNICEF	Yanhong Zhang	Dec/2019
Unidad Victimas (Colombia)	Oscar Ivan Rico Valencia	Dec/2019
UNSD	Yongyi Min	Dec/2019
World Bank	Umar Serajuddin	Dec/2019
UNHCR	Petra Nahmias	Jan/2020
World Bank-UNHCR Joint Data Centre on Forced Displacement	Felix Schmieding	Feb/2020

We are grateful to the persons above for their willingness to share their time and for the information they provided. We have made our best effort to reflect their views while supporting the confidentiality that was agreed upon. If we have made errors or misinterpreted their contributions, we apologize in advance.

Appendix 2.
Data sources used to compute disaggregated SDG indicators

Survey type	Country	Survey Name	Date	National coverage	Notes	Indicators extracted or computed
DHS	**Colombia**	DHS VII	2015/ 2016	Yes	6.1% of internal migrants reported that the reason for migration was armed conflict and violence. » A new migration module, not present in older surveys, was included in the most recent DHS.	2.2.1 │ 3.2.1 │ 6.1.1 │ 11.1.1 │ 8.5.2 │ 8.3.1 │ 16.9.1 7.1.1 │ 4.1.1
URL	Colombia: Standard DHS, 2015, USAID					
Other	**South Sudan**	High Frequency Survey: Wave 1	2015	Yes		6.1.1 │ 4.1.1 │ 7.1.1 │ 8.3.1 │ 8.5.2 │ 1.4.2 │ 16.1.4
URL	South Sudan: High Frequency Survey 2015, Wave 1, DFID					
Other	**South Sudan**	High Frequency Survey	2012/ 2014	Yes		6.1.1 │ 11.1.1 │ 1.2.1 │ 7.1.1 │ 8.3.1 │ 8.5.2 │ 16.1.4
URL	South Sudan: High Frequency Survey 2012-2014, Panel Data, DFID					
MICS	**Iraq**	MICS 6	2018	Yes	Respondents asked if they have faced discrimination in the past 12 months and why. One response is due to displacement or immigration - 5.1% selected this option » 1567 people	16.9.1 │ 3.2.1 │ 2.2.1 │ 7.1.1 │ 16.1.4 │ 6.1.1
URL	Iraq: Multiple Indicator Cluster Survey 2018, Central Statistical Organization (CSO), Kurdistan Region Statistics Office (KRSO), Ministry of Health, United Nations Children's Fund (UNICEF)					
MICS	**Lebanon**	MICS 4	2011	No	Entire sample are Palestinian refugees	16.9.1 │ 3.2.1 │ 2.2.1 │ 7.1.1 │ 16.1.4 │ 6.1.1
URL	Lebanon: Multiple Indicator Cluster Survey 2011, Palestinian Camps, Palestinian Central Bureau of Statistics, United Nations Children's Fund					
MICS	**Lebanon**	MICS 3	2005/ 2006	No	Entire sample are Palestinian refugees	16.9.1 │ 3.2.1 │ 2.2.1 │ 7.1.1 │ 16.1.4 │ 6.1.1
URL	Lebanon - Multiple Indicator Cluster Survey 2006, Palestinian Refugee Camps, United Nations Children's Fund					
MICS	**Lebanon**	MICS 2	2001	No	Entire sample are Palestinian refugees	16.9.1 │ 3.2.1 │ 2.2.1 │ 7.1.1 │ 16.1.4 │ 6.1.1
URL	Lebanon: Multiple Indicator Cluster Survey 2000					
MICS	**Syrian Arab Rep.**	MICS 3	2006	No	Entire sample are Palestinian refugees	3.2.1 │ 2.2.1 │ 6.1.1 │ 4.1.1 (primary/completion) │ 16.9.1
URL	Syria: Multiple Indicator Cluster Survey					
Other	**Iraq**	Rapid Welfare Monitoring Survey	2017	Yes		6.1.1 │ 11.1.1 │ 7.1.1 │ 8.5.2
URL	Iraq: Rapid Welfare Monitoring Survey 2017					

Other	**Iraq**	Survey of Syrian Refugees and Host Communities in Kurdistan	2015/2016	No		6.1.1	11.1.1	7.1.1	8.5.2	16.1.4
URL	Iraq: Survey of Syrian Refugees and Host Communities in Kurdistan, 2015-2016, The World Bank					
Other	**Jordan**	Survey of Syrian Refugees and Host Communities	2015/2016	No		6.1.1	11.1.1	7.1.1	8.5.2	16.1.4
URL	Jordan: Survey of Syrian Refugees and Host Communities, 2015-2016,					
Other	**Jordan**	Comprehensive Baseline Study on Digital Remittances	2016	No		8.5.2
URL	Jordan: Comprehensive Baseline Study on Digital Remittances 2016, Demand-side Survey of Low-income Jordanians and Syrian Refugees in Jordan, IPSOS Public Affairs, IPSOS Jordan					
Other	**Lebanon**	Survey of Syrian Refugees and Host Communities	2015/2016	No		6.1.1	11.1.1	7.1.1	8.5.2	16.1.4
URL	Lebanon: Survey of Syrian Refugees and Host Communities, 2016-2016, The World Bank					
Other	**Nigeria**	Profile of Internally Displaced Persons in North-East Nigeria	2018	No		6.1.1	11.1.1	1.2.1	4.1.1	7.1.1	8.5.2	1.4.2	16.1.4
URL	Nigeria: Profile of Internally Displaced Persons in North-East Nigeria, 2018, IBRD, IOM.					
Other	**Somalia**	Somali High Frequency Survey	2017/2018	Yes		6.1.1	11.1.1	1.2.1	4.1.1	7.1.1	8.3.1	8.5.2	1.4.2	16.1.4
URL	Somalia: Somali High Frequency Survey December 2017 Wave 2, The World Bank					
Other	**Somalia**	Somali High Frequency Survey	2016	Yes		6.1.1	11.1.1	1.2.1	7.1.1	8.3.1	8.5.2	1.4.2	16.1.4
URL	Somalia: Somali High Frequency Survey 2016 Wave 1, The World Bank					
Other	**South Sudan**	High Frequency Survey: Wave 4 and Crisis Recovery Survey	2017	Yes		6.1.1	11.1.1	1.2.1	4.1.1	7.1.1	8.3.1	8.5.2	1.4.2	16.1.4
URL	South Sudan: High Frequency Survey Wave 4, 2017, South Sudan National Bureau of Statistics					
Other	**South Sudan**	High Frequency Survey: Wave 3	2016/2017	Yes		6.1.1	11.1.1	4.1.1	7.1.1	1.4.2
URL	South Sudan: High Frequency Survey 2016 Wave 3, The World Bank					
Other	**South Sudan**	High Frequency Survey: Wave 2	2016	Yes		6.1.1	11.1.1	1.2.1	4.1.1	8.3.1	8.5.2	1.4.2
URL	South Sudan: High Frequency Survey 2016 Wave 2, The World Bank					

Data disaggregation of SDG indicators by forced displacement

UNHCR
The UN Refugee Agency

JIPS
Joint IDP Profiling Service

SUSTAINABLE
DEVELOPMENT
GOALS