

# Towards a statistical measure for durable solutions to internal displacement: a simulation to assess indicator and metric choice

Draft version for internal discussion in EGRIS subgroup 3 - version 2

April 9, 2021

# Towards a statistical measure for durable solutions to internal displacement: a simulation to assess indicator and metric choice

Draft version for internal discussion in EGRIS subgroup 3 (version 2)

## Contents

<b>Introduction</b>	<b>3</b>
<b>The IRIS “solutions measure”</b>	<b>4</b>
Remaining methodological challenges . . . . .	5
Ways forward for indicator and metric choice . . . . .	7
<b>Methodology</b>	<b>13</b>
Step 1: Identifying the universe of potential indicators . . . . .	13
Step 2: Selecting empirical datasets on IDP vulnerabilities & solutions . . . . .	13
Step 3: Preparing the indicators and indices for simulation . . . . .	15
Step 4: iteratively simulating the indicator and metric choice . . . . .	17
<b>Findings</b>	<b>17</b>
Simulation results for Hargeisa (UNHCR 2015) . . . . .	17
Simulation results for Nigeria (World Bank 2018) . . . . .	26
Simulation results for El Fasher (Sudan 2019) . . . . .	31
Simulation results for Colombia (DANE 2019) . . . . .	32
Discussion of housing, land, and property rights indicators . . . . .	37
Limitations and methodological concerns . . . . .	39
<b>Conclusions and key recommendations</b>	<b>41</b>
Informing the selection of indicators . . . . .	41
Informing the selection of a metric . . . . .	42
<b>Annex: full mapping of all indicators available to measure the IASC criteria</b>	<b>45</b>

## Introduction

The *International Recommendations on IDP Statistics* (IRIS), developed by the *Expert Group on Refugee and IDP Statistics* (EGRIS), provide a framework for capturing a country's stock of IDPs for statistical purposes. In order to correctly calculate the stock, it is crucial to define when an individual enters the stock and when they exit the stock. This note relates to the latter, and specifically to the way by which individuals exit the stock by overcoming key displacement-related vulnerabilities - sometimes referred to colloquially as the "solutions measure".

It is important to note that the purpose of a statistical definition for the "end of displacement" is explicitly **not** to identify specific households or individuals that are no longer displaced (e.g. for programming or assistance purposes) but only to enable aggregate statistics on the amount of internal displacement in each country that can be compared globally. Separate from the "solutions measure", IRIS also proposes a distinct but related framework to assess progress towards the achievement of durable solutions - colloquially referred to as the "progress measure". The progress measure aims to allow decision-makers "to understand at a glance in which aspects of vulnerability IDPs are struggling compared to others [...] and in which they are doing relatively well" (IRIS, page 55). While progress and solution can likely be measured with similar statistical indicators, their aims and application are very different: informing decisions on policy areas that need attention as well as more granular assistance-related decisions (through the IRIS progress measure) vs informing aggregate statistics on IDP stocks (IRIS solutions measure).

This note focuses on the IRIS solutions measure and how this measure can be implemented in practice. The IRIS have advanced the debate on a statistical measure on solutions considerably, including in the following ways:

- Specification of 5 priority criteria and 10 priority sub-criteria that should form the basis of a statistical measure<sup>1</sup>
- Recommendation of quality criteria to assess the suitability of indicators to measure each sub-criterion
- Move to a context-dependent assessment of durable solutions rather than an absolute approach whilst maintaining a level of comparison
- Narrowing of benchmark options (to national averages or host community averages)

Nevertheless, as stated in the recommendations themselves, the above achievements do not result in a final and applicable statistical measure for solutions. The IRIS falls short of 1) specifying the selection of statistical indicators for each of the priority criteria/ sub-criteria and 2) identifying a fully workable methodology of combining these indicators into an overall statistical measure that can be compared to a national or host community benchmark. While other methodological issues may also need further attention (i.e. the final definition of a comparator group, addressing statistical uncertainty and missing data points), the aim of this technical note is to inform solutions to these two outstanding methodological challenges.

This methodological note first outlines the solutions measure as suggested by the IRIS and demonstrates the remaining methodological challenges. The main body of the note then simulates how the selection of statistical indicators and the choice of a statistical framework to combine them affect the results of a solutions measure, i.e. the magnitude of exit from IDP stocks. To conduct these simulations, four empirical datasets from Hargeisa, Nigeria, Colombia and Sudan are used. The note compares how choosing different indicators and metrics to combine them increase or decrease the exit from the IDP stock in those countries. The note concludes with a summary of findings to improve the solutions metrics.

---

<sup>1</sup>The definition of priority criteria and sub-criteria is relevant to the IRIS recommended progress and solutions measure as they should be based on a similar or the same set of indicators.

## The IRIS “solutions measure”

Based on the *Framework for Durable Solutions by the Inter-Agency Standing Committee* (IASC), the IRIS specifies a set of 5 priority criteria and 10 sub-criteria that need to be measured in every displacement context to take IDPs out of the national stock as having overcome their displacement-related vulnerabilities (see Table 1). The indicators for each of the 10 sub-criteria should be collected at the household or individual level.

Table 1: IASC durable solution criteria and identified sub-criteria

Criteria	Subcriteria
1. Safety and security	1.1 Victims of violence 1.2 Freedom of movement
2. Adequate standard of living	2.1 Food security 2.2 Shelter and housing 2.3 Medical services 2.4 Education
3. Access to livelihoods	3.1 Employment and livelihoods 3.2 Economic security
4. Restoration of housing, land and property	4.1 Property restitution and compensation
5. Documentation	5.1 Documentation

To then assess how many displaced households in a specific context have progressed towards a durable solution, the situation of IDPs must be compared with that of the comparator population (national average or “host community” average). The logic is that IDPs that perform similar or better than the comparator population are no longer considered displaced for the purpose of statistics.<sup>2</sup> Regarding the methodology to decide whether a household has ended its displacement, IRIS writes:

*“The methodology used for calculating the composite measure is a simple pass/no pass (or binary) scenario at sub-criteria level that is then accumulated to produce a score at criteria level, and ultimately an overall score for the measure. [...] To determine if a sub-criterion has been overcome or not, for each household, a target needs to be set [...] It is foreseen that target setting will be more complex with categorical or binary indicators [...] If all sub-criteria receive a ‘pass’ mark, then that criterion [...] has been overcome. For criteria with multiple sub-criteria, all sub-criteria would need to receive a ‘pass’ mark for the criteria to be overcome. All of the five key-displacement related vulnerabilities (the five criteria) need to achieve a ‘pass’ mark for the composite measure to be fulfilled.[...] It is recommended to use the general/national population as the comparison group when deciding on the targets or thresholds for scoring each sub-criteria [...] In specific circumstances, thresholds can be set through a comparison with the average situation of a subset of the general population.” (IRIS, page 58-61)*

Table 2 provides an example to understand the framework. Household A passes the criteria 1. *Safety & security*, 4. *Restoration of HLP rights*, and 5. *Documentation*. However, because it has not passed the two sub-criteria in 3. *Access to livelihood*, it does not pass this criterion. The household also does not pass the criterion 2. *Adequate standard of living* because it has not passed the sub-criterion 2.4 *Education* and the framework requires to pass all sub-criteria. Overall, no durable solution has been achieved in this example and the household should not exit the IDP stock.

<sup>2</sup>Note that this framework is used to count the aggregate number of IDPs in a given context. It is not used to assess the allocation of aid to an individual or household.

Table 2: Example household in the IRIS framework

	Sub-criterion (compared to a benchmark)		Result on criterion level
Household A	1.1 Victims of violence	Pass	2/2 (Passed)
	1.2 Freedom of movement	Pass	
	2.1 Food security	Pass	3/4 (Not passed)
	2.2 Shelter and housing	Pass	
	2.3 Medical services	Pass	
	2.4 Education	No pass	
	3.1 Employment & livelihoods	No pass	0/2 (Not passed)
	3.2 Economic security	No pass	
	4.1 Property restitution & compensation	Pass	1/1 (Passed)
	5.1 Documentation	Pass	1/1 (Passed)
<b>Overall:</b> No durable solution achieved: displacement not ended			

### Remaining methodological challenges

While this framework appears intuitive, there are several challenges and methodological gaps when implementing it in practice.

The first methodological challenge is that **no set of statistical indicators has been agreed on to capture the 10 sub-criteria**. The lack of an agreed, measurable set of indicators presents a real challenge in operationalizing the IRIS framework. While this methodological assessment aims to shed some light on how specific indicators affect the exit from the IDP stock, further empirical work will be needed to assess how the indicator choice affects the aggregate stock of IDPs in a displacement context. Eventually, more suitable datasets will need to be produced to facilitate this.

The second methodological challenge arises from the above-cited assumption that it is possible to set a comparator target/benchmark for each sub-criterion which a household needs to match or surpass. The IRIS states that “target setting will be more complex with categorical or binary indicators”; however, **such benchmark comparison is not just complex but indeed technically meaningless for binary and categorical indicators. Binary and categorical household-level data points cannot be directly compared with average/distribution values in the comparator population.** To illustrate this problem, imagine a binary indicator “household has access to medical services” was selected to operationalize sub-criterion 2.3 *Medical services*. The value for any given household can only be yes or no, 1 or 0. The target value derived from the comparator population, in contrast, would be a percentage on a scale, let us assume 5% of the comparator population have access to medical services. Achieving a pass for any given household in comparison with the target value can only be achieved by achieving a “yes” in this indicator – despite the fact that access to medical services is extremely low in the comparator population. In effect, the value of the comparator population is irrelevant, because the condition for achieving a pass would not have changed if the comparator value had been, say, 1% or 90%. In other words, the entire idea of a contextualized target/benchmark that underlies the IRIS approach is devalued for binary or categorical indicators. Unfortunately, such types of indicators appear the pragmatic and thematically most valid choice of metric under the vast majority of sub-criteria.

A third methodological challenge not yet addressed by IRIS arises if the **comparator value comes with a level of statistical uncertainty**. Assume the benchmark value for the national/host population was produced using a sample survey – it would thereby come with a confidence interval around the comparator point estimate. It is not clear yet if a displaced household would have to perform as good as or better than the benchmark’s point estimate or whether displaced households should “just” perform as

good as or better than the lower bound of the confidence interval in the benchmark population. Deciding on an approach towards statistical uncertainty is crucial as the IDP stock estimates in most countries are not derived from a full mapping of the complete IDP population but from sampled surveys of displaced and non-displaced households. How to deal with levels of statistical uncertainty is relevant for binary but also metric indicators.

Even if a set of statistical indicators can be agreed on to operationalize the 10 sub-criteria, it is unlikely that data will be available for all selected indicators in all situations. The IRIS specifies that if data is missing an assessment of whether displacement has been overcome cannot be made. Further empirical work provides an opportunity to assess and further explore the feasibility of this approach and gain greater insight on the aggregate **effects of missing data points** (and potentially of imputation-based methods for dealing with them).

A fifth area left open by IRIS is the **definition of the comparator population** itself – national or “host”. If a host community is to be used as comparator, a clear definition of the term will need to be developed. While it may be preferable for IRIS to *not* provide a standardized recommendation on this and leave flexibility to data producers, further empirical work is recommended to at least assess how the choice of the comparator population affects the aggregate results across different context, and facilitate an evidence-informed decision.

Related to the definition of a comparator population is the question how to deal with **changing benchmark values over time**. If for example unemployment in the host community/ national population drops from one stock assessment to the next due to improving economic conditions in a country, this increases the benchmark value, and IDPs that have previously been taken out of the stock as having overcome their vulnerabilities may fall under the comparator benchmark again, re-entering and thereby enlarging the stock without the occurrence of new displacement-causing events.

Finally, the IRIS solution measure should address to what extent assistance received should be “factored out / imputed out” before an assessment is made of how many IDPs exit the stock. For example, if IDPs overcome key-displacement related vulnerabilities because their shelter and housing is provided through humanitarian assistance, they may exit the stock without actually having overcome their housing-related vulnerabilities. While the solutions measure only produces an aggregate number, the measure should ensure that the overall exit from the IDP stock is not based on the **overcoming of vulnerabilities through humanitarian assistance** but through sustainable solutions.

Box 1 summarizes the methodological challenges that need to be addressed to establish a workable solutions metrics. This note aims to advance the debate on the challenges 1 and 2.

- Challenge 1:** Specifying one or several statistical indicators for each of the 10 sub-criteria
- Challenge 2:** Find a way to make the “contextualized” approach (comparing against target/benchmark population) statistically applicable
- Challenge 3:** Address how to deal with statistical uncertainty in the target/benchmark values
- Challenge 4:** Specify how missing data points should be dealt with
- Challenge 5:** Define the comparator population
- Challenge 6:** Address changing benchmark values over time
- Challenge 7:** Specify how assistance is factored into the solutions measure

Box 1: Remaining challenges for a workable solutions metrics

## Ways forward for indicator and metric choice

One next step to advance the solutions measure is to identify the indicators that should be used to measure the 10 sub-criteria for a durable solution (“challenge 1” above). The choice of the right indicators is complex. Political sensitivities, contextual differences between displacement situations, comparability across different contexts and comparator populations as well as availability of data to measure the indicators may play a role in the final decision to operationalize the 10 sub-criteria. According to IRIS, possible indicators should be **commonly used** (e.g. the indicator is required for SDG reporting); an indicator should be **tested and evaluated** to satisfy quality criteria and comply with international standards; and an indicator should **cover the population in question** (i.e. school enrolment is widely recognized as a good indicator for measuring access to education, but it would not give any information on IDP households which do not have children).

While recommending the ‘best’ indicators to measure the 10 sub-criteria is beyond the scope of this note, this note takes an empirical approach to simulate how the choice of indicators affects the outcome of the solutions measure, that is the number of IDPs that would leave a national stock in different displacement contexts. Importantly, this is a technical assessment that generates practical information on 1) the extent to which the indicator choice for each sub-criterion matters and 2) the statistical characteristics or quality standards that an indicator should fulfil to be suitable from a statistical perspective. This can inform dialogue on higher political levels on the right indicator choice, but it does not lead to a final set of indicators that should be used.

Another necessary next step towards a full metric is to address how the different indicators are combined and then compared to the national population or the host community in a statistically applicable way (“challenge 2” in the above list). Different metrics, or combinations of indicators and comparisons with a benchmark population, are possible to implement the current IRIS solution measure. The overall aim is to be able to compare an IDP household to a benchmark on all 10 sub-criteria to decide whether a household performs the same or even better than the national average or the host community and should hence no longer be counted as IDP household in official statistics. As this comparability is currently not technically applicable in the IRIS solutions metric, different options are available to revise the proposed measure:

### *Pass/fail measure: Implementing the IRIS without a comparator population<sup>3</sup>*

The current measure described in IRIS does not allow for a meaningful comparison between IDPs and host communities/national averages. One could make the decision to fully focus on a pass/fail decision on the sub-criterion level (as demonstrated in Table 2 above) by only assessing whether a household achieves the sub-criterion. This is not a desirable option because no comparison takes place, and it thereby fails to comply with the fundamental logic set out in IRIS for a solutions measure, which is that it should be assessed relative to a host/national benchmark. A consequence of a context-blind pass/fail approach is that this option is likely to produce very low numbers of IDPs overcoming their vulnerabilities because achieving a ‘one’ for all 10 indicators is a hard threshold in many displacement contexts.

### *Option 1: An actual composite metric across all criteria*

The approach currently outlined in IRIS is described therein as a “composite” measure, which strictly speaking it is not, considering that a single indicator per sub-criterion is implicitly assumed and a pass

---

<sup>3</sup>For completeness and to allow comparisons with other metric options, this note also demonstrates simulations based on this approach but will not discuss the suitability of a pass/fail measure as option going forward.

needs to be achieved in each sub-criterion/indicator. One option to explore that aligns with the contextualized approach would be to redefine the overall framework as an actual composite index. A composite index groups together indicators across all criteria to one score (see an illustration in Table 3). This household level index could then be more readily compared to the average value of the same composite index in the comparator population. The shortcoming of this approach is that a household may be taken out of the IDP stock despite underperformance on a specific key criterion or sub-criterion (which is a deviation from the IRIS requirement that a pass needs to be achieved at the sub-criterion level).

Table 3: Option 1: A full composite metric for all sub-criteria

	Sub-criteria	Indicator	
Household A	1.1 Victims of violence	Indicator 1.1.a	1
	1.2 Freedom of movement	Indicator 1.2.a	1
	2.1 Food security	Indicator 2.1.a	1
	2.2 Shelter and housing	Indicator 2.2.a	1
	2.3 Medical services	Indicator 2.3.a	1
	2.4 Education	Indicator 2.4.a	0
	3.1 Employment and livelihoods	Indicator 3.1.a	0
	3.2 Economic security	Indicator 3.2.a	0
	4.1 Property restitution and compensation	Indicator 4.1.a	1
	5.1 Documentation	Indicator 5.1.a	1
	Total composite index for the household		7/10
	Benchmark average		8.6/10
<b>Decision:</b> The household has not overcome displacement as the composite index is below the benchmark.			

#### *Option 2: A set of composite indices at the criterion level*

Rather than construct a single index across all criteria, one could define composite indices for each criterion (see Table 4). For example, the composite index for the criterion on 2. *Adequate standard of living* could consist of four or more equally weighted indicators corresponding to the four related sub-criteria (2.1. *Food security*, 2.2 *Shelter and housing*, 2.3 *Medical services*, 2. 4 *Education*). A household could then score values of 0-4 on this sub-criterion index (or 0%, 25%, 50%, 75%, 100%). This score can be compared with a distribution average in the comparator population for the same composite index. As with option 1 above, this would constitute a deviation from IRIS, which explicitly states that a pass needs to be achieved at the sub-criterion level; under this option a household may be taken out of the IDP stock despite underperformance on a specific sub-criterion. In order to construct an index in each criterion, at least 2 binary indicators per criterion would be required, but more would be recommended. A small number of indicators per criterion would result in a cruder measure, thereby barely overcoming challenge 2 (as we will still be comparing discrete values – e.g. 0;1;2 in the case of 2 indicators for a given criterion - with a continuous distribution average; so for a given household to perform on par or better than the benchmark, it would be irrelevant whether the benchmark valuer was, say, 0.1 or 0.9 – in either case the household in question would need to score a 1 to achieve a pass on this criterion). Assuming at least 3 indicators per each of the 5 criteria (and 4 indicators in the case of criterion 2, in alignment with its 4 sub-criteria), this would result in an overall computation based on at least 16 indicators.



Table 4: Option 2: A composite metric at the criterion level

	Criterion	Sub-criteria	Indicator	Composite	Comparison to benchmark
Household A	1. Safety and security	1.1 Victims of violence	Indicator 1.1.a	2/3	1.9 (Pass)
		1.2 Freedom of movement	Indicator 1.1.b		
			Indicator 1.2.a		
	2. Adequate standard of living	2.1 Food security	Indicator 2.1.a	3/4	3.5 (No pass)
		2.2 Shelter and housing	Indicator 2.2.a		
		2.3 Medical services	Indicator 2.3.a		
		2.4 Education	Indicator 2.4.a		
	3. Access to livelihoods	3.1 Employment and livelihoods	Indicator 3.1.a	2/3	1.9 (Pass)
		3.2 Economic security	Indicator 3.1.b		
			Indicator 3.2.a		
	4. Restoration of housing, land and property	4.1 Property restitution and compensation	Indicator 4.1.a	1/3	0.9 (Pass)
			Indicator 4.1.b		
			Indicator 4.1.c		
	5. Access to documentation	5.1 Documentation	Indicator 5.1.a	0/3	0.9 (No pass)
			Indicator 5.1.b		
Indicator 5.1.c					
<b>Decision:</b> The household has not overcome displacement as the composite index for some criteria is lower than the benchmark.					

*Option 3: A set of composite sub-indices at the sub-criterion level*

Very similar to the second approach, one could define composite indices within each sub-criterion (see Table 5). For example, the composite index for sub-criterion 2.1 *Food security* could be associated with 3 equally weighted binary indicators. A household could then score values of 0-3 on this sub-criterion index (or 0%, 33%, 66%, 100%). This score can be compared with a distribution average in the comparator population for the same composite index. The major difference between this option and options 1&2 above is that this option would not constitute a methodological deviation from IRIS, in that it would ensure that a pass is achieved at the sub-criterion level for a household to be taken out of the IDP stock. However, there is a serious feasibility issue with this option. As with the preceding option, at least 2 binary indicators are required per sub-criterion to construct an index, but more would be recommended. A small number of indicators per sub-criterion would result in a cruder measure, thereby barely overcoming challenge 2 (as we will still be comparing discrete values – e.g. 0;1;2 in the case of 2 indicators for a given sub-criterion - with a continuous distribution average; so for a given household to perform on par or better than the benchmark, it would be irrelevant whether the benchmark valuer was, say, 0.1 or 0.9 – in either case the household in question would need to score a 1 to achieve a pass on this criterion). Assuming at least 3 indicators per each of the 10 sub-criteria, this would result in an overall computation based on at least 30 indicators – which appears overly burdensome for widespread application.

Table 5: Option 3: A composite metric at the sub-criterion level

Sub-criteria	Indicators	Composite	Comparison to benchmark
Household A	1.1 Victims of violence	Indicator 1.1.a Indicator 1.1.b Indicator 1.1.c	1 3 1 1 2.8 (Pass)
	1.2 Freedom of movement	Indicator 1.2.a Indicator 1.2.b Indicator 1.2.c	1 3 1 1 2.5 (Pass)
	2.1 Food security	Indicator 2.1.a Indicator 2.1.b Indicator 2.1.c	1 2 0 1 1.1 (Pass)
	2.2 Shelter and housing	Indicator 2.2.a Indicator 2.2.b Indicator 2.2.c	1 2 0 1 1.9 (Pass)
	2.3 Medical services	Indicator 2.3.a Indicator 2.3.b Indicator 2.3.c	1 3 1 1 3 (Pass)
	2.4 Education	Indicator 2.4.a Indicator 2.4.b Indicator 2.4.c	1 1 0 0 2.5 (No pass)
	3.1 Employment and livelihoods	Indicator 3.1.a Indicator 3.1.b Indicator 3.1.c	0 0 0 0 1.5 (No pass)
	3.2 Economic security	Indicator 3.2.a Indicator 3.2.b Indicator 3.2.c	0 1 0 1 2.9 (No pass)
	4.1 Property restitution and compensation	Indicator 4.1.a Indicator 4.1.b Indicator 4.1.c	1 3 1 1 2.5 (Pass)
	5.1 Documentation	Indicator 5.1.a Indicator 5.1.b Indicator 5.1.c	0 2 1 1 1.8 (Pass)
	<b>Decision:</b> The household has not overcome displacement as some composite indices on the sub-criterion level are lower than the benchmark.		

#### Option 4: Comparison of homogeneous cells:

Rather than trying to create a continuous indicator value at the household level through composite indices as in the first three options, an alternative approach could be to divide the IDP population into small homogeneous “cells” – for example by location of displacement, area of habitual residence and year of arrival (see Table 6). Even if selecting just one indicator per sub-criterion (i.e. 10 indicators for the measure overall), for each cell an average value could be calculated which subsequently can be compared against the distribution average in the comparator benchmark. The decision whether displacement has ended is no longer made on the individual household level but on the cell level. If an IDP cell has achieved a durable solution in comparison to the benchmark, all IDPs in this group are taken out of the stock of IDPs. If the cell has not achieved a durable solution, all IDPs remain as IDPs in the stock. The advantage of this option is that it is the only one to truly overcome challenge 2 and fully deliver to the contextualized approach proposed by IRIS, in that it would be comparing continuous cell values with benchmark continuous values. The shortcoming of this option is that the aggregate results may depend heavily on the criteria chosen for dividing IDP populations into cells. A strong assumption of homogeneity in the cells is necessary.

Table 6: Option 4: Comparison of homogeneous cells

	Sub-criteria	Indicator	Percentages in cell Z	Benchmark comparison
Household A is classified into cell Z (e.g. current location, area of origin and year of arrival)	1.1 Victims of violence	Indicator 1.1.a	29%	28% (Pass)
	1.2 Freedom of movement	Indicator 1.2.a	26%	25% (Pass)
	2.1 Food security	Indicator 2.1.a	2.2%	2% (Pass)
	2.2 Shelter and housing	Indicator 2.2.a	19%	19% (Pass)
	2.3 Medical services	Indicator 2.3.a	3.5%	3% (Pass)
	2.4 Education	Indicator 2.4.a	1%	3.5% (No pass)
	3.1 Employment and livelihoods	Indicator 3.1.a	12%	15% (No pass)
	3.2 Economic security	Indicator 3.2.a	25%	29% (No pass)
	4.1 Property restitution and compensation	Indicator 4.1.a	31%	25% (Pass)
	5.1 Documentation	Indicator 5.1.a	35%	18% (Pass)
<b>Decision:</b> All households in cell Z (that includes household A) have not overcome displacement as their average distribution in some sub-criteria is lower than the average in the benchmark.				

#### Option 5: Classifier/regression-based approach

Another option for the “solutions measure” could be to take a regression-based approach in which the 10 sub-criteria are used as covariates to predict whether a household should still be classified as an IDP household or not (see Table 7). A probabilistic classifier, such as a logistic regression, would estimate whether an IDP household is distinct from the host community (i.e. high probability to be an IDP) or whether an IDP household is very similar to households in the host community (i.e. low probability to be an IDP). Important decisions to make before implementing this option are which classifier to select, how to select probability cut-off points to determine whether a specific household is similar to IDPs or to hosts, and how to deal with sample imbalance in the data. While this approach would overcome the challenge of comparing IDP households with host community households, this approach needs further clarifications and is not sensible to underperformance on specific indicators.

Box 2 summarizes the main advantages and disadvantages of the different metrics. The note going forward aims to provide empirical information how these metrics perform (that is how many IDPs they count as having overcome their displacement-related vulnerabilities).

Table 7: Option 5: A classifier/regression-based approach

	Sub-criteria	Indicator		Regression weights
Household A	1.1 Victims of violence	Indicator 1.1.a	1	-1.58
	1.2 Freedom of movement	Indicator 1.2.a	1	-0.04
	2.1 Food security	Indicator 2.1.a	1	1.59
	2.2 Shelter and housing	Indicator 2.2.a	1	-2.24
	2.3 Medical services	Indicator 2.3.a	1	0.48
	2.4 Education	Indicator 2.4.a	0	4.48
	3.1 Employment and livelihoods	Indicator 3.1.a	0	6.72
	3.2 Economic security	Indicator 3.2.a	0	2.13
	4.1 Property restitution and compensation	Indicator 4.1.a	1	4.32
	5.1 Documentation	Indicator 5.1.a	1	-0.22
Predicted probability for household (e.g. logit transformed)				0.439
Cut-off point				0.5
<b>Decision:</b> The household has not overcome displacement as the predicted probability of being similar to the host community is below the cut-off point.				

Pass/fail measure (No comparison!)	<ul style="list-style-type: none"> <li>✓ Sensitivity to underperformance on the sub-criteria level</li> <li>✗ No comparison between IDP and comparator households; potentially restrictive</li> </ul>
1. Composite metrics across all criteria	<ul style="list-style-type: none"> <li>✓ Comparability of scores with few indicators needed</li> <li>✗ No sensibility to underperformance on specific (sub-criteria)</li> </ul>
2. Composite metrics at the criterion level	<ul style="list-style-type: none"> <li>✓ Comparability of scores</li> <li>✗ No sensibility to underperformance on specific sub-criteria; higher data needs (at least 16 indicators)</li> </ul>
3. Composite metrics at the sub-criterion level	<ul style="list-style-type: none"> <li>✓ Comparability of scores and sensibility to underperformance on sub-criterion level</li> <li>✗ Very high data needs (many indicators needed)</li> </ul>
4. Homogeneous cells	<ul style="list-style-type: none"> <li>✓ Full contextualization, comparison of continuous values</li> <li>✗ Depends on assumption of homogeneous IDPs within cells, may be highly sensitive to boundaries of cells</li> </ul>
5. Classifier/regression-based	<ul style="list-style-type: none"> <li>✓ Full contextualization possible</li> <li>✗ Depends on regression assumptions; no sensibility to underperformance on specific sub-criteria</li> </ul>

Box 2: Main advantage and disadvantage of different metrics

## Methodology

This note follows a purely technical and empirical approach to assess how different indicators and different metrics to combine these indicators perform and result in IDPs overcoming their displacement-related vulnerabilities, thereby exiting the IDP stock in statistical terms. Overall, the assessment follows four steps: The first step was to identify the universe of potential indicators that could be used to measure the 10 sub-criteria. The second step was to identify suitable empirical datasets that contain enough statistical indicators from this theoretical universe to compare their performance against each other. In a third step, the relevant indicators in the datasets were prepared for the analysis by turning all indicators into binary variables and by creating composite indices. Finally, the selected datasets were used to assess the indicators and the different metric options in simulations. This final simulation step means that for each metric option, it was assessed how many IDPs exit the dataset for each possible indicator combination. The simulated results provide information on the performance of indicators and metrics that can inform the process of developing a fully applicable measure for durable solutions. The following sections outline the details of the methodology.

### Step 1: Identifying the universe of potential indicators

The first methodological step was to map out the universe of potential indicators that could be used to operationalize the 10 sub-criteria on the basis of the Durable Solutions Library.<sup>4</sup> For each sub-criterion, multiple statistical indicators exist but there is no comprehensive list of all theoretically possible indicators. However, this is important to guide the selection of suitable datasets for this assessment. The universe of potential indicators should be a realistic list of indicators that are often used in surveys while also covering multiple potential operationalizations of the 10 sub-criteria.

As one of the most comprehensive collection of often-used statistical indicators for durable solutions, the list of indicators contained in the Durable Solutions Library was mapped on to the 10 sub-criteria. To demonstrate, the library specifies the indicator “Target population residing in durable housing structures” which was mapped on to the sub-criterion 2.2 *Shelter and housing*. Overall, 52 possible indicators were mapped on to the 10 sub-criteria. While the library only contains one possible indicator for some of the sub-criteria, up to 10 possible operationalizations for 3.2 *Economic Security* are covered.<sup>5</sup> See Table 9 for a break down of the possible indicators per sub-criterion.

This universe of potential indicators forms a theoretical baseline to identify empirical datasets that cover many of these indicators in practice. It should also be noted that the Durable Solutions Library identifies general indicators defined in broad terms. Empirical datasets can contain a plurality of more concrete, specific and narrower indicators that can be associated with one of the statistical indicators in the Durable Solutions Library.

### Step 2: Selecting empirical datasets on IDP vulnerabilities & solutions

In a second step, the aim was to identify and select empirical datasets that cover many potential indicators to measure the 10 sub-criteria defined in the IRIS and that are contained in the mapped-out universe of statistical indicators from the Durable Solutions Library. The first dataset selected for this simulation

---

<sup>4</sup>The Durable Solutions Library is an inter-agency project led by the Special Rapporteur on the Human Rights of IDPs, coordinated by JIPS, and technically steered by a broad group of partners engaged in durable solutions work.

<sup>5</sup>The DS library also includes indicators on distributions (e.g. Target population who experienced moderate or severe food insecurity in the last year, *by main obstacle to obtaining sufficient food*.). In the case of these indicators it is not clear what constitutes a solution to displacement (Which obstacle to obtaining food is tolerable for IDPs and which one is not?) and how hosts and IDPs can be compared. These disaggregated indicators were dropped from the mapping.

exercise is the *Internal Displacement Profiling in Hargeisa*<sup>6</sup> that was conducted by UNHCR with support of JIPS and the Protection Cluster's Profiling Task Force in 2015. The second dataset stems from the World Bank *Profile of Internally Displaced Persons in North-East Nigeria*<sup>7</sup> in 2018. Both datasets cover IDP households displaced by conflict and violence but the dataset in Hargeisa is focused on urban displacement while the Nigerian dataset allows an assessment of in-camp IDPs and IDPs in host communities. The two datasets are complemented with another dataset covering IDPs in camps on the household level: The third dataset comes from the *Durable Solutions Profiling in El Fasher*<sup>8</sup> which was conducted in 2019 in a cooperation between the Government of Sudan, the IDP communities of Abu Shouk and El Salam, the World Bank, and the UN Country Team. These three datasets were chosen because they cover a wide range of possible indicators for durable solutions and because they cover IDPs and host communities.<sup>9</sup> Finally, the case selection was complemented with the *Living Standards Measurement Study in Colombia 2019* (LSMS)<sup>10</sup>. The LSMS is a useful addition to the case selection as it allows to compare a national representation of displaced households to a national average, as opposed to a comparisons to the host community only. The identifying question whether displacement has occurred or not has been asked on the individual-level in this dataset. Hence, we briefly discuss how a household level assessment of the solutions measure performs in comparison to an individual-level assessment. The datasets - summarised in Table 8 - were readily available in the Humanitarian Data Exchange, the World Bank Microdata Library, the data library of the Colombian National Statistical Office or were provided by JIPS.<sup>11</sup>

Table 8: Selection of empirical datasets

Dataset	Year	Assessment level	Benchmark	Displacement context
Hargeisa (Somaliland)	2015	Household	Host community	Urban; conflict, insecurity & drought
North-East Nigeria	2018	Household	Host community	In camps & host communities; conflict & violence
El Fasher (Sudan)	2019	Household	Host community	In camps; conflict & violence
Colombia	2019	Household & Individual	National average	National representation; conflict & violence, some disaster-related

The datasets were then checked to identify whether they contained the relevant 52 indicators from the Durable Solutions Library. Table 9 summarizes how many indicators for each sub-criterion were present in the empirical datasets. As previously mentioned, the datasets can cover multiple more detailed indicators than the mapping from the Durable Solutions Library which lists general indicators. See the appendix for a full table of all indicators in the Durable Solutions library mapped to a sub-criterion and complemented with the indicators available in the empirical datasets.<sup>12</sup>

This identification of indicators is crucial as it builds the baseline for the following simulations. In Hargeisa, 30 indicators could be identified. For example, the dataset contains 3 possible indicators for

<sup>6</sup>UNHCR, 2015, "Internal Displacement Profiling in Hargeisa", <https://reliefweb.int/report/somalia/internal-displacement-profiling-hargeisa-december-2015>

<sup>7</sup>World Bank, 2018, "Profile of Internally Displaced Persons in North-East Nigeria 2018", <https://microdata.worldbank.org/index.php/catalog/3410>

<sup>8</sup>JIPS, 2019, "Progress towards durable solutions in Abu Shouk and El Salam IDP camps", <https://www.jips.org/uploads/2019/12/JIPS-Sudan-profilingreport-2019.pdf>

<sup>9</sup>Covering hosts and IDPs is crucial for this methodological assessment to facilitate the task of finding a comparator population.

<sup>10</sup>DANE, 2019, <http://microdatos.dane.gov.co/index.php/catalog/678/study-description>

<sup>11</sup>If survey weights were available (not the case in Hargeisa), they are considered in the simulations to ensure representativeness.

<sup>12</sup>When identifying available indicators in the datasets, the aim was to be as inclusive as possible and to use as much information present in the data as possible. However, some indicators had to be dropped because of too many missing values or no variance at all.

Table 9: Potential and empirically existent indicators for durable solutions

Subcriteria in IRIS	Indicators from	Of which, available in ...			
	Durable Solutions Library	IDP Profiling in Hargeisa	IDP Profiling in Nigeria	IDP Profiling in El Fasher	LSMS in Colombia
1.1 Victims of violence	7	3	8	4	2
1.2 Freedom of movement	1	1	1	0	0
2.1 Food security	1	2	1	2	1
2.2 Shelter and housing	8	6	9	4	4
2.3 Medical services	4	3	2	2	2
2.4 Education	3	5	2	3	3
3.1 Employment and livelihoods	8	1	1	2	6
3.2 Economic security	10	3	7	4	5
4.1 Property restitution and compensation	6	3	2	3	0
5.1 Documentation	4	3	1	2	1
Number of possible combinations		14580	4032	9216	1440

measuring 2.3 *Medical services* as the survey asked respondents for access to child vaccinations and basic services, as well as for attendance of births by medical personnel. For the two sub-criteria 1.2 *Freedom of movement* and 3.1 *Employment and livelihoods*, the data only contained one possible indicator and an indicator choice cannot be evaluated for these two dimensions. As the indicators are not used individually but in a combination of at least 10 (for the 10 sub-criteria), this leads to a total of 14580 possible indicator combinations that can be assessed in simulations.

In Nigeria, 34 indicators could be identified. However, since the indicators in Nigeria are more concentrated on two sub-criteria, only 4032 possible combinations for simulations exist. The dimensions 1.2 *Freedom of movement*, 2.1 *Food security*, 3.1 *Employment and livelihoods*, and 5.1 *Documentation* are only covered with one possible indicator. The dataset in Nigeria is particularly rich covering whether IDPs were victims of violence, their shelter and housing conditions as well as their economic security.

In Sudan, the profiling in El Fasher yielded 26 indicators that can be combined to 9216 possible combinations for simulations. It should be noted that the IDP profiling in El Fasher did not contain a suitable indicator to measure whether IDPs enjoy freedom of movement. For all other sub-criteria, the profiling contains 2 to 4 possible indicators. However, some of these indicators had to be aggregated from the household member surveys that were conducted alongside the household assessments.<sup>13</sup>

In Colombia, the LSMS contains a total of 24 indicators that could be used to measure solutions to displacement. The dataset does not contain any indicators to measure 1.2 *Freedom of movement* and 4.1 *Property restitution and compensation*. These sub-criteria are hence omitted in the analysis. For 2.1 *Food security* and 5.1 *Documentation*, the data only contained one indicator, which allows to include the sub-criteria in the analysis but does not allow for an assessment of how the indicator selection may affect the results. However, the data is particularly rich assessing the economic security and the livelihoods of IDPs and the national average.

### Step 3: Preparing the indicators and indices for simulation

After having identified all relevant indicators in the data, the aim is to apply the above outlined metric options to the datasets to identify how many IDPs exit the IDP stock in each of the metrics. Additionally,

<sup>13</sup>More specifically, to make an assessment whether the household has access to education, employment, and documentation, the variables are aggregated from the household member surveys by assuming that at least one child should go to school, at least one household member should be employed, and at least one household member should have documentation.

this should not only be done for one set of 10 indicators but for at least 10000 combinations (if available) in the data. To iterate through at least 10000 indicator combinations for each metric option, the indicators were first prepared for these simulations. Importantly, all indicators were coded as binary indicators that are coded as 1 if a displacement-related vulnerability was overcome and coded as 0 if the vulnerability persists for a specific household. For example, not reporting a security incident is a 1 as this is a positive achievement for durable solutions. For a simple pass/fail measure and the different composite metrics, those binary indicators can be directly combined to the composites at the relevant total, criterion and sub-criterion level.<sup>14</sup>

Metric option 4 is based on homogeneous IDP population cells that first have to be defined before they can be assessed as a group average. In El Fasher (Sudan), the IDP population was grouped into subgroups based on their type of housing. This was chosen because the El Fasher questionnaire did not contain many other variables to group the IDPs into logical groups. In Colombia, we did not divide the IDP households into different cells. In Nigeria and Hargeisa, the IDP population was split into different subgroups based on core demographics, their timing of displacement and their places of displacement and origin.

To do so, the assessment did not pick out one variable to group the population but iterated through different ways to split the IDP population to be able to make a judgement to what extent the different groupings affect the overall results (i.e. how many IDPs exit the stock). In Hargeisa, 10 different groupings were tested based on using three of the grouping variables below. For example, one population split was based on the gender, the clan, and the departure period of a household (e.g. one grouping consists of female-headed households that belong to the Isa clan and were displaced after 2010). In Nigeria, 4 different groupings were tested based on picking three of the grouping variables below. The following cell variables were used to split the IDP population in Hargeisa and Nigeria into subgroups:

#### Hargeisa:

- **Gender of the household head** (female, male)
- **Clan of the household** (Haw, Isa, Rah, Dar, Dir, Gab, Other, Unknown)
- **Origin district** (29 origin districts)
- **Region of origin** (Displaced from South Central, from Somaliland, from South Central OoS),
- **Departure period** (Before 1990, Between 1990 and 200, Between 2000 and 2010, After 2010)

#### Nigeria:

- **Year of displacement** (ranging from 1998 to 2018)
- **Year of arrival** (ranging from 1998 to 2018)
- **Region of displacement** (Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe)
- **Region of origin** (Different state in Nigeria, Outside Nigeria, Same local government area, Same state, Same ward)

Finally, a regression-based approach requires a methodological decision how to classify an IDP household as still vulnerable or as able to exit the IDP stock. In this assessment a logistic classifier was fitted to determine whether a household is similar to an IDP household or a host community household. After fitting a regression, we can predict the probability that a specific IDP household is closer to the other IDPs or to the host community but this requires a threshold value (e.g. 0.5) to make the decision. In this demonstration, we have not picked a threshold but optimized this threshold from the data by choosing the threshold that best separates the IDP and host population. If a regression-based approach is used in the future, this cut-off point requires further refinement and a final methodological decision.

<sup>14</sup> Regarding metric option 2 and 3, the composite indices were built by combining three indicators at the (sub-)criterion level (where possible). In the case of having only one or two empirical indicators for a sub-criterion or a criterion and not enough empirical indicators to build an index in the dataset, the single indicator was used for this specific (sub-)criterion while the other (sub-)criteria while more indicators were turned into indices. In a final metric development, it would be necessary to have enough indicators to build a complete composite index with 3 or more elements.



#### Step 4: iteratively simulating the indicator and metric choice

After identifying the indicators for this assessment in each survey and preparing the indicators for analysis, iterative simulations following the scheme in Figure 1 were run. For each dataset, 10 000 possible combinations of 10 indicators (one for each of the sub-criteria) were assessed using all five metric options outlined above. To be able to compare the five possible metric options to a simple pass/fail approach, the simulations also included this option. In each simulation (given a set of indicators and the chosen metric), we report the number of IDPs overcoming their displacement-related vulnerabilities. After all combinations of indicators and metrics were analyzed, it is possible to assess to what extent a single indicator affects the number of IDPs exiting the stock in each metric type by running linear regressions.

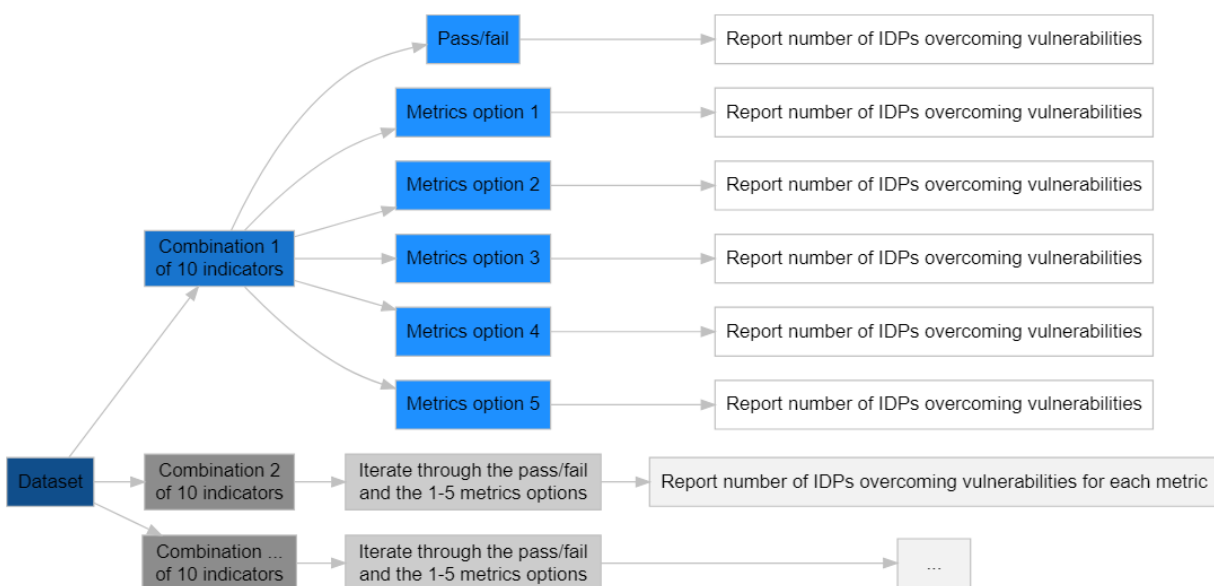


Figure 1: Simulation approach

## Findings

### Simulation results for Hargeisa (UNHCR 2015)

The IDP profiling in Hargeisa covered a total of 4780 IDP households that could be compared directly to their hosts. On average across all simulated indicators and metrics, 211.07 IDP households have overcome their vulnerabilities and exit the IDP stock. According to this assessment, an average of 4.42% of the sampled IDP population in Hargeisa could exit the stock for statistical purposes because they have overcome key displacement-related vulnerabilities. However, some IDP households could not be assessed as the IRIS recommends that no assessment should take place if data is missing on at least one of the 10 sub-criteria. Depending on the different indicator combination, an average of 16.76% of IDP households could not be assessed due to missing data on at least one of the key indicators. In those instances, the IDP household remains in the stock. This demonstrates that a lot of the variation in the number of IDPs exiting the stock does not come from the selection of indicators but rather the missingness for some of those possible operationalizations of the 10 sub-criteria.

### Importance of indicator selection

The first question this assessment aims to address is how much difference there is between using different operationalizations of the 10 sub-criteria. To address the question of selecting statistical indicators, Table 10 summarizes the relevance of the indicator choice in Hargeisa. The table lists all indicators that are present in the data and provides information on the average effect of choosing one indicator over the other across all metrics.

Table 10: Average relevance of indicator choice across metrics (Hargeisa)

Subcriteria IRIS	Indicators in Hargeisa UNHCR 2015	Average effect size on stock	Description of variation
1.1 Victims of violence	Experience of security incident (baseline), Feeling of safety, Worries about security	-0.41 to 0.1 percentage points	Hardly any discernable difference between different indicators
1.2 Freedom of movement	Only available indicator was whether IDPs are free to visit public places.	No other indicator	
2.1 Food security	More meals than average (baseline), Ability to pay food	-0.61 to -0.61 percentage points	Hardly any discernable difference between different indicators
2.2 Shelter and housing	Bathroom access (baseline), Electricity, Adequate housing, No overcrowded housing, Toilet facilities, Water access	-0.08 to 1.32 percentage points	Hardly any discernable difference between different indicators
2.3 Medical services	Access to medical service if needed (baseline), Birth attendance, Child vaccination	2.47 to 2.87 percentage points	Some difference between different indicators
2.4 Education	Child ever attended school (baseline), Child able to read/write, Child in school, Mobile phone, Secondary school	-0.58 to 0.58 percentage points	Hardly any discernable difference between different indicators
3.1 Employment and livelihoods	Only available indicator was whether IDP household has a breadwinner or not.	No other indicator	
3.2 Economic security	Durable assets (baseline), Paying rent, Cover unexpected expenses	1.22 to 3.38 percentage points	Some difference between different indicators
4.1 Property restitution and compensation	Recompensation mechanisms accessible (baseline), Documentation for HLP, Access to HLP restored	16.68 to 16.97 percentage points	Notable difference between different indicators for HLP rights
5.1 Documentation	Possession of birth certificate (baseline), Possession of ID, Ability to replace ID	0.02 to 2.7 percentage points	Some difference between different indicators

For example, choosing an indicator that covers the ability to pay for food as opposed to an indicator that counts whether an IDP household had more meals than the average IDP household decreases the percentage of IDPs exiting the stock by an average of -0.61 percentage points across all metrics in this simulation. Expressed differently, the ability to pay for one's own food is a slightly higher bar to pass than consuming above-average meals, which means that less IDPs will overcome their vulnerabilities. The last column of the table gives a verbal summary of the average effect of indicator selections across all metrics.

It should be noted that for most of the 10 sub-criteria, the exact indicator choice does not matter strongly across the different metric option. There is no discernible difference in the indicators for 1.1. *Victims of violence*, 2.1. *Food security*, 2.2 *Shelter and housing*, and 2.4. *Education* as the percentage point change is on average smaller than 1. For the sub-criteria 2.3. *Medical services*, as an example, there is some differences between using access to medical services, birth attendance by skilled health personnel and child vaccinations with access to medical services being the most restrictive indicator. However, in substantive terms, the difference between these indicators increases or decreases the exit from the IDP stock by only around 2.47 to 2.87 percentage points.

Most importantly, there is a strong and notable difference between the different indicators for 4.1 *Property restitution and compensation*. It is observable that having access to documentation for housing, land and property (HLP) or having HLP access restored is more permissive, allowing more IDPs to exit the stock than the presence of compensation mechanisms. Compensation mechanisms seem to be hardly present for IDPs in the Hargeisa case, leading to a decrease in IDP exits from the stock by 16.68 to 16.97 percentage points, if such an indicator is chosen.

To visualize the results of this indicator assessment, Figure 2 displays the average effect of choosing one indicator over the other averaged across all assessed metric options. The shown effect sizes must be understood in comparison to the baseline indicator (see baselines in Table 10). The effect plot displays clearly that HLP indicators for access to property or documentation for property are crucial indicators that drive a lot of variation in the simulations compared to other indicators for sub-criteria. This means, that if a decision is made on which indicator to choose for measuring 4.1 *Property restitution and compensation*, the selection of other indicators is changing the amount of IDPs exiting the stock significantly less.

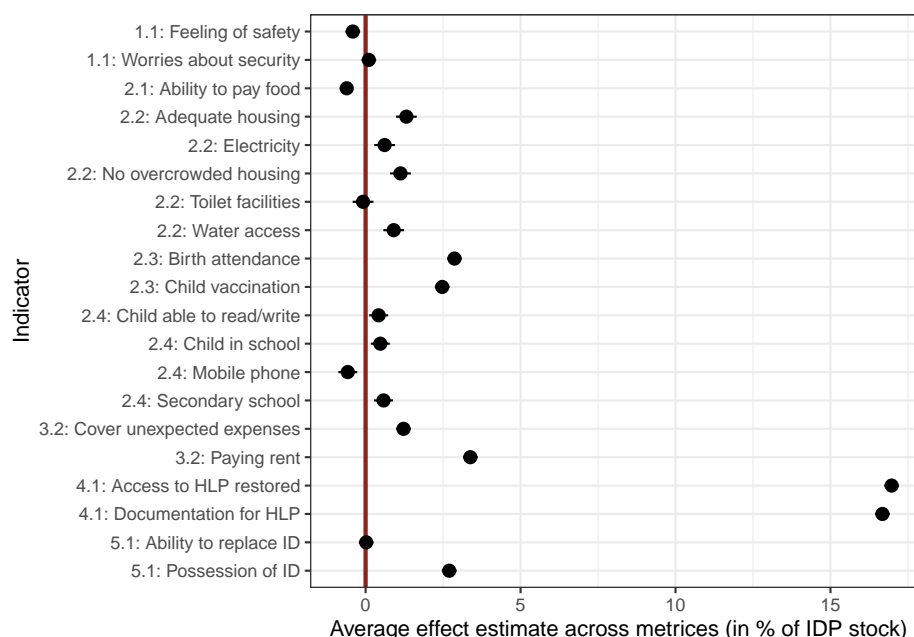


Figure 2: Average effect of indicator selection across metrics on IDPs exiting the stock (Hargeisa)

Overall, even if there are some differences in the operationalization of the 10 sub-criteria, and the indicator selection tends to matter more, the simulations in the context of Hargeisa suggests that the average effect of choosing one indicator over the other is often neglectable in practice. In this indicator assessment, the most relevant choice is clearly related to HLP rights.

### *Importance of metric selection*

The simulation approach is also intended to answer a second question that is to what extent the different metric options produce varying estimates of the exit from the IDP stock. Table 11 summarizes the key findings related to the range of IDP households that exit the stock across all metrics (how much variation there is in results), and how important the choice of individual indicators is in each approach. The table also provides the mean number of IDP households exiting the stock. It can be generally said that the choice which indicators are used matters in some metrics more than in others as some metrics are more volatile than others.

Table 11: Simulation results for Hargeisa across all metrics (Total sample of IDPs: 4780 households)

	Pass/fail measure (no comparison!)	1: Full composite	2: Composite at criterion level	3: Composite at sub-criterion level	4: Comparison of homogeneous cells	5: Classifier/ regression- based
Mean of IDPs exiting the stock	24.85 IDPs (0.52% of stock)	1317.18 IDPs (27.56% of stock)	131.72 IDPs (2.76% of stock)	0 IDPs (0% of stock)	4.25 IDPs (0.09% of stock)	1442.93 IDPs (30.19% of stock)
Range of IDPs exiting the stock	0 to 942 IDPs (0 to 19.71% of stock)	42 to 2778 IDPs (0.88 to 58.12% of stock)	0 to 1440 IDPs (0 to 30.13% of stock)	0 to 0 IDPs (0 to 0% of stock)	0 to 692 IDPs (0 to 14.48% of stock)	92 to 3830 IDPs (1.92 to 80.13% of stock)
Variation in how many IDPs exit the stock	Low (SD: 76.96 )	High (SD: 642.74 )	Middle (SD: 239.7 )	No variation (SD: 0 )	Very low (SD: 22.85 )	Very high (SD: 1043.81 )

When using one composite metric across all criteria (all 10 indicators for each sub-criterion are summarized into one index) and when using a regression-based approach, the choice of indicators matters the most and the metric yields high variation. In the case of one composite metric, measuring the sub-criterion 4.1. *Property restitution and compensation* by assessing the documentation and restoration of property rather than the access to compensation mechanisms increases the percentage of IDPs overcoming their vulnerabilities by around 22.5 percentage points. In the case of a regression-based assessment whether an IDP household is more likely to be similar to other IDPs vs. the host communities, the same indicator can even increase the percentage of IDPs overcoming their vulnerabilities by 41.61 percentage points. Albeit less strongly, the choice which indicators are used is also more severe in the case of a composite metric at the criterion level (as up to 6.21 percentage point difference can be between different indicators). For options 1, 2 and 5, it is hence important to note that these metric options create varying estimates of exits from the stock and are very sensitive to the indicator selection.

For other metric options, the importance of choosing the right indicator is very diminished as the metrics produce very similar estimates of the exit from IDP stocks across all indicators. As multiple

indicators are combined at the sub-criterion level, option 3 indicates that none of the IDPs exit the stock. The option performs very similar to a pass/fail scenario in which no comparison with the host community takes place. Finally, when IDPs are divided into subgroups and then compared with the host population, the simulation also revealed that very few IDP subgroups exit the IDP stock. In the simulations, 10 different groupings for IDPs were tried based on the region of origin, the clan, gender, and departure period of IDPs in Hargeisa. However, none of the groupings yielded strongly different results compared to other groupings. See the additional Box 3 for more details on the effect of different ways of grouping the IDP stock on the performance of the homogeneous cells approach.

Figure 3 displays the density of the conducted simulations for the 5 identified metric options. To allow comparisons, a simple pass/fail measure on the sub-criterion level has been added but readers should note that this does not enable any comparisons with hosts. The graph displays on the x-axis what percentage of the IDP population in the dataset has overcome vulnerabilities and exits the stock. The y-axis displays the density of simulations or how many of the simulations shared the same outcome (number of IDPs exiting the stock). Note that the y-axis and the x-axis in the different graphs have their own scaling to make visualization easier.

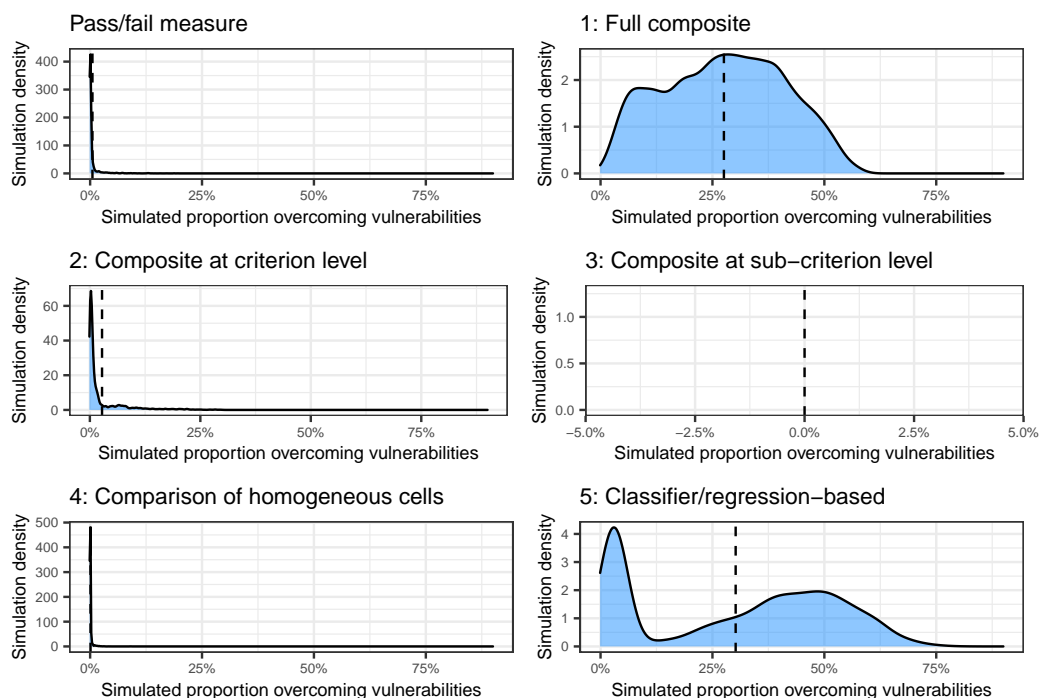


Figure 3: Density of simulations for all possible metrics (Hargeisa)

The first observation from these density plots is that the pass/fail measure, the composite measure at criterion level, the composite measure at sub-criterion level, and the comparison of cells all yield similar results: most of the simulations based on these metric options result in 0 IDPs overcoming their vulnerabilities and exiting the stock. Simulations based on a composite measures at the criterion level has a clear peak around 0 percentage of IDPs - that means that most simulations show that no IDPs exit the stock - while the average simulation suggests that 132 IDP households overcome their vulnerabilities (2.76% of the stock). For a composite measure at the sub-criterion level, there is no variation and all simulations indicate that no IDP exits the stock. This means that both composite metric options are not strongly dependent on the selected indicators. It also indicates that the metrics are relatively restrictive

and classify few IDPs as having overcome their key vulnerabilities. Because of their high threshold to exit IDPs from the stock, they behave like a pass/fail measure that does not compare IDPs with a comparator population. Similarly, the cell-based metric option showed low volatility to the selection of indicators and the selection of grouping variables in the case of Hargeisa.

Making a regression-based assessment whether a specific household is predicted to be an IDP household or a host community household is an option that has the highest variability, which means that depending on the indicator combination chosen, the resulting number of IDPs exiting the stock can be considerably higher or lower. Keeping all other indicators constant, using an indicator whether IDPs have restored their access to property can classify around 1989 more IDP households as exiting the stock as opposed to measuring this sub-criterion as access to restoration mechanisms. The simulations reveal a bimodal distribution: many simulations cluster around 0-5% of the IDP population overcoming their vulnerabilities. At the same time, many simulations center around 55% of IDPs overcoming their vulnerabilities. From this distribution, it is hence difficult to assess what the true percentage of IDPs having achieved a solution in statistical terms might be. The split into bimodal distribution is strongly linked to the selection of the HLP indicators that separate the IDP population strictly into those that do not have HLP rights realised and those without. If a restrictive HLP indicator is chosen the simulations peak around 0, if a more lenient indicator is chosen, the average simulations move closer to 50%.

Constructing one composite measure across all criterion (option 1) is a metric option that has the second highest variability. Keeping all other indicators constant, using an indicator whether IDPs have documentation to prove their property rights can classify around 1075 more IDP households as exiting the stock as opposed to measuring this sub-criterion as access to restoration mechanisms. The conclusion is that if data is combined into one index across all criteria and if a regression-based approach is used, it is strongly advisable to carefully determine a final set of indicators to measure solutions to ensure that data across displacement contexts is comparable.

#### *Identification of solutions on household level*

We also investigated if the different simulations always identify *the same* IDP households as candidates to exit the stock or if the different metric options pick up different vulnerable groups. To do this, we calculate the Cohen's kappa coefficient. Cohen's kappa is a statistic that is normally used to measure "inter-rater reliability" or the idea that if a coder A identifies an IDP household as household that overcomes all vulnerabilities, an independent coder B should also identify this IDP household as no longer vulnerable. In the case of these simulations, the "coder" is essentially the different metric options we have available. For example, to be a robust measure, the homogeneous cell approach should pick up more or less the same few households as vulnerable as the composite measures at the criterion level.

Figure 4 displays the results of the Cohen's kappa coefficient for the different metric options. The interpretation works as follows: Brighter blue suggests that the two metrics options on the x and y axis pick up the same households as vulnerable. Darker blue suggest that Cohen's kappa is low and they hardly pick up the same households as vulnerable. Overall, because many metrics are very identify few exits from the stock, Cohen's kappa is rather low. However, we can for example see that grouping IDPs into cells based on the gender, origin district and broader region yields similar results to groupings based on gender, origin district and clan or gender, origin district and departure year. In contrast, groupings based on broader region, gender and departure year do not compare well to other groupings.

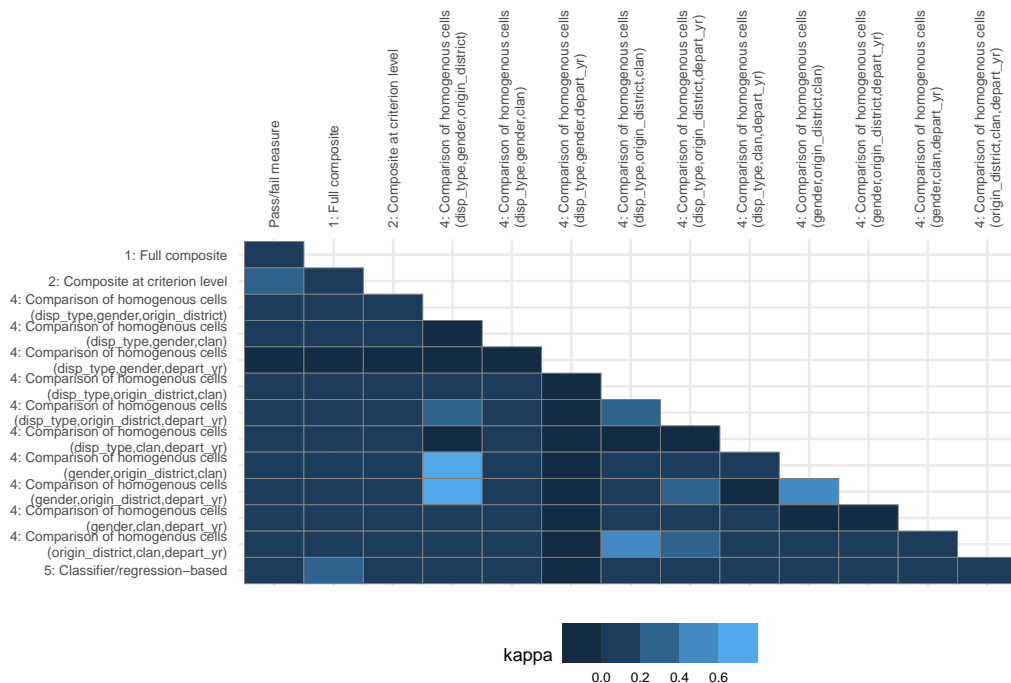


Figure 4: Inter-rater agreement between different metric options for Hargeisa. Note that metric option 3 is excluded as no variation in IDP exits is achieved.

### Overall feasibility and application of metric options

Some metric options turned out to be easier to implement than others in the dataset on IDPs in Hargeisa. In terms of feasibility, the metric option 1, which showed high sensitivity to indicator selection, was easy to implement as the construction of one index is straightforward. However, a problem with this metric option is not only its variability but also that elements of the index may be collinear.

Regarding the metric options 2 + 3, that both require multiple indices on the criterion and the sub-criterion level, the feasibility of this approach is much lower. First, they require more indicators than the other metrics to be able to construct additive indices. For example, in Hargeisa there were simply not more than one indicator available to measure 1.2 *Freedom of movement* and 3.1 *Employment and livelihood*. It is expected that producers of official statistics may also have challenges to generate the data necessary for these more data hungry indices. Second, the theoretical conceptualization is not fully advanced yet. An “ideal” composite index is always made of theoretically distinct elements (i.e. we measure different dimensions of displacement-related vulnerabilities). However, when multiple indicators that measure very much the same theoretical concept are combined to one index at the sub-criterion level (e.g. similar indicators that all measure whether IDPs’ shelter and housing conditions are satisfactory), the different indicators are likely very much correlated. An example is that individual living in an overcrowded space very often also live in a makeshift space. This multicollinearity of elements of an index raises questions whether strongly correlated elements of an index should be weighted differently than other elements. The problem of developing a theoretically sound index also exists when an index is built on the criterion level. So far, the IRIS already identifies four distinct sub-criteria for the adequate standard of living of IDPs. However, it is not clear if we can theoretically identify more than one dimension to empirically measure the criterion 5. *Documentation*. These questions of theoretical conceptualization, correlation

between elements in the index, and the high data demands reduce the feasibility to implement the metric options 2 and 3 in practice.

Dividing the IDP population in Hargeisa into sub-groups to assess homogeneous cells that can then be compared to the host community is relatively easy to implement for National Statistical Offices. However, it goes beyond the scope of this methodological note to assess precisely how the boundaries between cells (i.e. the criteria to divide the IDP populations) affect the results. In Hargeisa, the available variables to build cells did not indicate high volatility to the grouping. However, a final selection of this metric option would need to assess 1) which delimitations should be used, 2) if the number of variables to divide the IDP population matters, 3) if the size of the cells matters, and if the 4) homogeneity or heterogeneity in a cell matter. While this option is initially easy to implement, it raises many new methodological challenges.

Finally, using a regression-based approach has some disadvantages and advantages. First, further methodological work is needed to identify an easy-to-implement but also successful classifier that predicts whether an IDP household is more similar to other IDPs or to the host community. A logistic regression or linear regression seems plausible here. Furthermore, it has to be decided what the probability cut-off is, or at what predicted probability an IDP household should be no longer considered displaced. Another practical problem in this approach is that comparisons across contexts might be difficult. While a regression in one country might assign a lot of weight to indicators for the freedom of movement, in another country it might be much more important whether food security has been achieved. From the point of a national statistical system, it might be difficult to identify why the regression coefficients vary across contexts and sub-criteria and how different assessments can be compared between displacement situations.



### How should the IDP population be split into homogeneous cells?

To fully assess the suitability of the homogeneous cell approach, we have to assess different ways in which the IDP population can be grouped in cells. Theoretically, the expectation is that some ways of grouping the overall population will lead to larger groups; other groupings could lead to smaller groups. The group size, the homogeneity within IDP cells and the heterogeneity across IDP cells could affect how many IDP households exit the stock.

**Grouping variables in the data to split the IDP population:** First, we selected three grouping variables from the data - for example the gender of the household head, the clan of the household and their origin district - and then grouped the IDPs into subgroups based on these grouping variables and calculated averages to compare to the host community average. In the simulations, we can do this iteratively for multiple possible grouping variables. In total, in Hargeisa, we tried 10 different groupings (see table). The key findings are summarised here:

- The grouping leading to the fewest exits from the IDP stock was based on region of origin, gender and departure period with no IDPs exiting the stock.
- The grouping leading to the most exits from the IDP stocks was based on the origin district, the clan, and the departure period with over 10 IDPs exiting the stock on average across indicator combinations.
- The overall impression is that more fine-grained groupings (e.g. by clan, by departure period, and by district) can potentially lead to more exits from the stock as some small groups might outperform the host population.
- In contrast, broader and more coarse groupings (e.g. by gender or region of origin) are more restrictive as broader IDP groups do not outperform the host population.
- In total, the groupings did not affect the results strongly.

**Algorithmic clustering:** Second, we employed automatic hierarchical clustering algorithms to group the IDPs into groups based on their performance on the different indicators. That means, an algorithm identifies IDP households in the data that are similar to each other in the sense that they perform equally well on certain sub-criteria while equally bad on others. These automatic groupings yield IDP cells that are most homogeneous within the cells and most heterogeneous in comparison to other cells. This is a data-driven approach that can provide some key lessons for finding good grouping variables. The hierarchical clustering revealed a data-driven cell approach could potentially increase the exit from the IDP stock as similarities between IDP households are exploited and the "best-performing" households are grouped together and taken out of the stock. While there are easily applicable software packages for hierarchical clustering, further exploration is needed if National Statistical Offices can implement such approaches.

Group variable 1	Group variable 2	Group variable 3	Average number of IDPs exiting the stock	Average percentage of IDPs exiting the stock	Range and mean of group size
Region of origin	Clan	Departure period	0.59	0.01	1 to 1042 (~ 65 )
Region of origin	Gender	Clan	2.76	0.06	1 to 1338 (~ 111 )
Region of origin	Gender	Departure period	0.00	0.00	1 to 774 (~ 184 )
Region of origin	Gender	Origin district	4.82	0.10	1 to 1715 (~ 34 )
Region of origin	Origin district	Clan	4.88	0.10	1 to 953 (~ 46 )
Region of origin	Origin district	Departure period	4.55	0.10	1 to 1288 (~ 65 )
Gender	Clan	Departure period	0.50	0.01	1 to 922 (~ 26 )
Gender	Origin district	Clan	5.94	0.12	1 to 513 (~ 31 )
Gender	Origin district	Departure period	8.09	0.17	1 to 673 (~ 67 )
Origin district	Clan	Departure period	10.32	0.22	1 to 811 (~ 21 )
Algorithmic clustering			30.32	0.63	

Box 3: Identifying IDP subgroups to implement the homogeneous cell approach (Hargeisa)

## Simulation results for Nigeria (World Bank 2018)

The IDP profiling in Nigeria covered a total of 1437 IDP households. Because the dataset includes survey weights, these IDPs represent an overall amount of over 129.41 thousand IDP households. On average across all simulated indicators and metrics, 5067.24 IDP households exit the IDP stock, which corresponds to an average of 3.92% of the IDPs represented by the profiling. For an average of 14.76% of the IDP households represented by the profiling in Nigeria, IDP households could not be assessed due to missing data on at least one dimension. Similar to the assessment in Hargeisa, the question remains if it is acceptable that around a seventh of the IDPs remain in the stock because of missing data.

### Importance of indicator selection

Table 12 summarizes how much the choice of one indicator over others can affect the exit from the IDP stock on average across all simulations and all metric options. In the dataset from Nigeria, it is only possible to assess the relevance of the indicator selection for six sub-criteria of the total of 10 sub-criteria. For the remaining four sub-criteria, the data only contains one possible indicator. The table is quickly summarized by concluding that only the indicator choice in 4.1. *Property restitution and compensation* matters strongly. For the majority of the other indicator choices, the difference between indicators only increases or decreases the IDP stock by less or around 1 percentage points.

However, for 4.1 *Property restitution and compensation*, choosing an indicator that measures legally owning the dwelling in comparison to having access to recommendation mechanisms for lost HLP is a more permissive indicator, allowing more IDPs to exit the stock by 13.4 percentage points. Compensation mechanisms seem to be hardly present for IDPs in Nigeria, and checking legal ownership of the house in the pre-flight location is a more feasible indicator.

To visualize the results, Figure 5 displays the average effect of choosing one indicator over the other averaged across all assessed metric options. Again, the difference between the effect size of the HLP indicators in comparison to the other indicators is striking. While the indicator selection does not matter strongly for the sub-criteria 1-8 and 10, 4.1 *Property restitution and compensation* requires a careful assessment to determine how to best measure this for the purpose of the “solutions measure”.

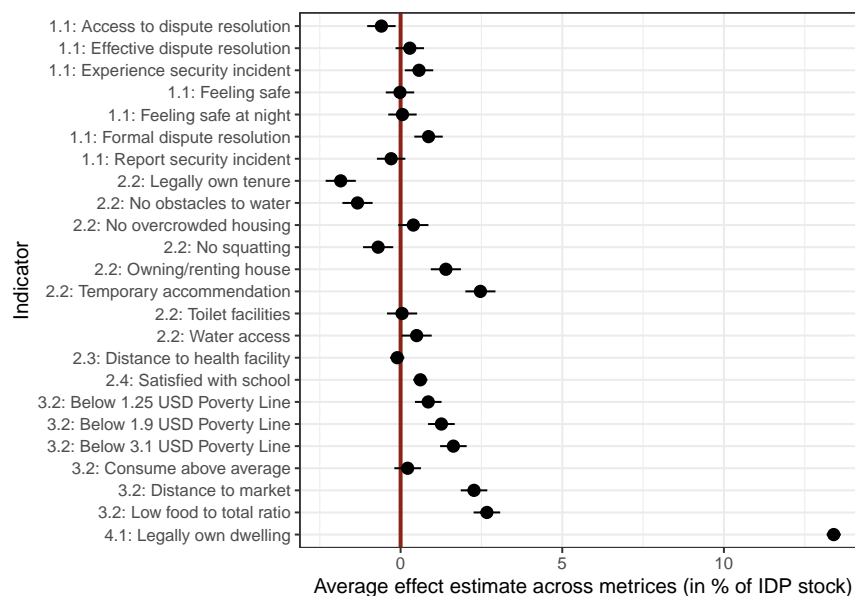


Figure 5: Average effect of indicator selection across metrics on IDPs exiting the stock (Nigeria)

Table 12: Average relevance of indicator choice across metrics (Nigeria)

Subcriteria IRIS	Indicators in Nigeria World Bank 2018	Average effect size on stock	Description of variation
1.1 Victims of violence	Feeling of safety at day (baseline), Access to dispute resolution, Effective dispute resolution, Formal dispute resolution, Experience security incident, Feeling safe at night, Report security incident, Feeling safe	-0.59 to 0.86 percentage points	Hardly any discernable difference between different indicators
1.2 Freedom of movement	Only available indicator was whether IDPs feel free to move	No other indicator	
2.1 Food security	Only available indicator was whether IDPs score higher than average on Food Security Scale.	No other indicator	
2.2 Shelter and housing	Non-durable living arrangement (baseline), No overcrowded housing, Owning/renting house, No squatting, Temporary accommodation, Toilet facilities, No obstacles to water, Water access, Legally own tenure	-1.85 to 2.47 percentage points	Some difference between different indicators
2.3 Medical services	Access to medical services as needed (baseline), Distance to health facility	-0.1 to -0.1 percentage points	Hardly any discernable difference between different indicators
2.4 Education	Duration to next education facility (baseline), Satisfied with school	0.61 to 0.61 percentage points	Hardly any discernable difference between different indicators
3.1 Employment and livelihoods	Only available indicator was income generation.	No other indicator	
3.2 Economic security	Having a bank account (baseline), Consume above average, Low food to total ratio, Distance to market, Below 1.9 USD Poverty Line, Below 1.25 USD Poverty Line, Below 3.1 USD Poverty Line	0.22 to 2.67 percentage points	Some difference between different indicators
4.1 Property restitution and compensation	Access to recompensating mechanism for lost HLP (baseline), Legally own dwelling	13.4 to 13.4 percentage points	Notable difference between different indicators for HLP rights
5.1 Documentation	Only available indicator was if IDPs had documents/access to replace missing documents if lost.	No other indicator	

### Importance of metric selection

While the indicator selection matters most for HLP indicators in Nigeria, the different metric options also produce varying estimates for the exit from the IDP stock. As already seen in the data from Hargeisa, using a regression-based approach or one composite metric across all criteria are the most sensitive to the exact indicators used while other approaches show little variation.

In the case of a full composite metric the selection of an indicator can in some instances make up to 19.33 percentage point difference in the outcome of a simulation. That means that over 25.01 thousand IDP households may be classified as having overcome their vulnerabilities when one indicator to measure 4.1. *Property restitution and compensation* is used while those households do not exit the IDP stock when another indicator is used. In the case of the regression-based classification of IDPs, using an indicator whether a household is the legally recognized owner of their property can on average, when all other indicators stay the same, increase the number of IDPs achieving a durable solution by 33.4 percentage points compared to an indicator capturing the access to compensation mechanisms.

However, for all other metric options presented in this note, the choice of indicators seems to be of very minor relevance for the Nigerian case as less than 1 percentage point difference is made by selection one indicator over the other. Table 13 summarizes the key findings related to variation across indicators and metrics. Most importantly, we find no variation in the number of IDPs exiting the stock for composite measure at the sub-criterion level and very little variation for homogeneous cells.

Table 13: Simulation results for Nigeria across all metrics (Weighted IDP households: 129.41 thousand)

	Pass/fail measure (no comparison!)	1: Full composite	2: Composite at criterion level	3: Composite at sub-criterion level	4: Comparison of homoge- neous cells	5: Classifier/ regression- based
Mean of IDPs exiting the stock	61.1 IDPs (0.05% of stock)	28047.2 IDPs (21.67% of stock)	985.25 IDPs (0.76% of stock)	0 IDPs (0% of stock)	3.5 IDPs (0% of stock)	23997.92 IDPs (18.54% of stock)
Range of IDPs exiting the stock	0 to 883 IDPs (0 to 0.68% of stock)	1101 to 73623 IDPs (0.85 to 56.89% of stock)	0 to 7109 IDPs (0 to 5.49% of stock)	0 to 0 IDPs (0 to 0% of stock)	0 to 470 IDPs (0 to 0.36% of stock)	1662 to 76193 IDPs (1.28 to 58.88% of stock)
Variation in how many IDPs exit the stock	Low (SD: 139.85 )	Very high (SD: 17430.41 )	Low (SD: 1592.3 )	No variation (SD: 0 )	Very low variation (SD: 40.43 )	Very high (SD: 22869.46 )

The simulations also show that the groupings of the IDP population into different homogeneous cells to then make a group-level assessment yields little volatility regarding the variables used for the IDP sub-groups. It should be noted that the grouping in Nigeria was based on date of arrival, date of displacement and the origin and displacement location of IDPs as the dataset contained few other variables useful for classification of IDPs into sub-groups. It is realistic that other groupings could be more sensitive and could produce more variation in the number of subgroups exiting the stock of IDPs. A more detailed discussion of different delimitations for IDPs in cells can be found in Box 4.

#### Splitting the Nigerian IDPs into homogeneous cell

In Nigeria, the simulations iterated through 4 possible groupings based on arrival or displacement year, and region of origin or displacement. Most of these groupings result in very few or no IDPs exiting the stock. This may be a result of too small or too big cells of IDPs. Overall, the groupings based purely on displacement timing and location of IDPs is not optimal. Further work is needed to define better grouping variables. Using core demographics (such as gender of household head, etc.) seems more promising. Using hierarchical clustering, we find can confirm that the homogeneous cell approach in Nigeria yields few exits from the overall IDP stock.

Group variable 1	Group variable 2	Group variable 3	Average number of IDPs exiting the stock	Average percentage of IDPs exiting the stock	Range and mean of group size
Year of arrival	Region of origin	Region of displacement	0	0.00	3 to 46270 (~ 1221 )
Year of displacement	Year of arrival	Region of origin	0	0.00	12 to 45057 (~ 1294 )
Year of displacement	Year of arrival	Region of displacement	0	0.00	3 to 49797 (~ 1407 )
Year of displacement	Region of origin	Region of displacement	14	0.01	3 to 43390 (~ 1362 )
Algorithmic clustering			12.32	0.01	

#### Box 4: Identifying IDP subgroups to implement the homogeneous cell approach (Nigeria)

Figure 6 displays the simulation density for the 5 different metric options that could be used to implement the IRIS solutions measure in practice. The five options are compared to a simple pass/fail measure which does not implement a comparison to host communities. As a reminder, the graphs show which outcome (i.e. number of IDPs exiting the stock) is particularly common in the simulations across indicator combinations. The same pattern as in Hargeisa can be detected in Nigeria. The composite measure at the criterion level, the composite measure at the sub-criterion level and a comparison of homogeneous cells yield an average number of IDPs exiting the stock close to 0. In fact, the metric option 3 results in 0 IDPs overcoming their key displacement-related vulnerabilities across all indicator simulations. These three metric options essentially generate similar results to a simple pass/fail measure without comparisons to hosts. In the case of the full composite measure (option 1), the density plots show higher variability and up to 21.67 percent of the IDP population assessed in Nigeria may have overcome their displacement-related vulnerabilities according to this measure. For a regression-based approach, the assessment again yielded a bimodal distribution. The full composite measure and the regression-based approach are characterized by a much higher variability and the indicator selection is more sensitive in this context.

#### Identification of solutions on household level

Figure 7 displays the results of the Cohen's kappa coefficient for the different metric options. Brighter colours suggest that the two metric options on the x and y axis pick up the same households as able to exit the IDP stock. The assessment in Nigeria shows that the composite measure at the criterion level identifies mostly the same IDPs as exiting the stock as a pass/fail measure without comparison to hosts. There is also big overlap between the full composite index (option 1) and a regression-based approach or a composite measure at the criterion level.

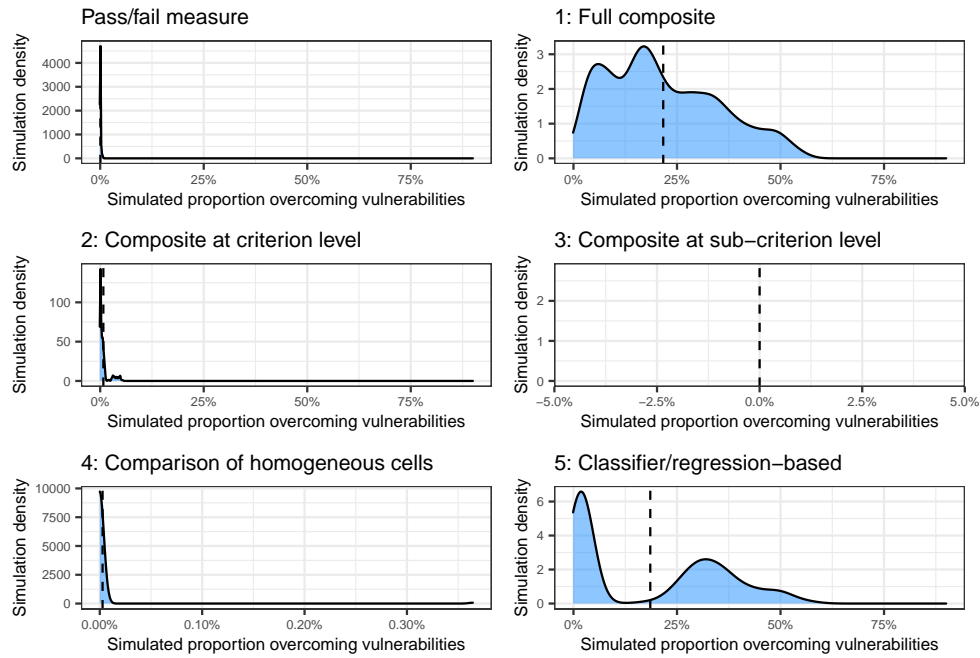


Figure 6: Density of simulations for all possible metrics (Nigeria)

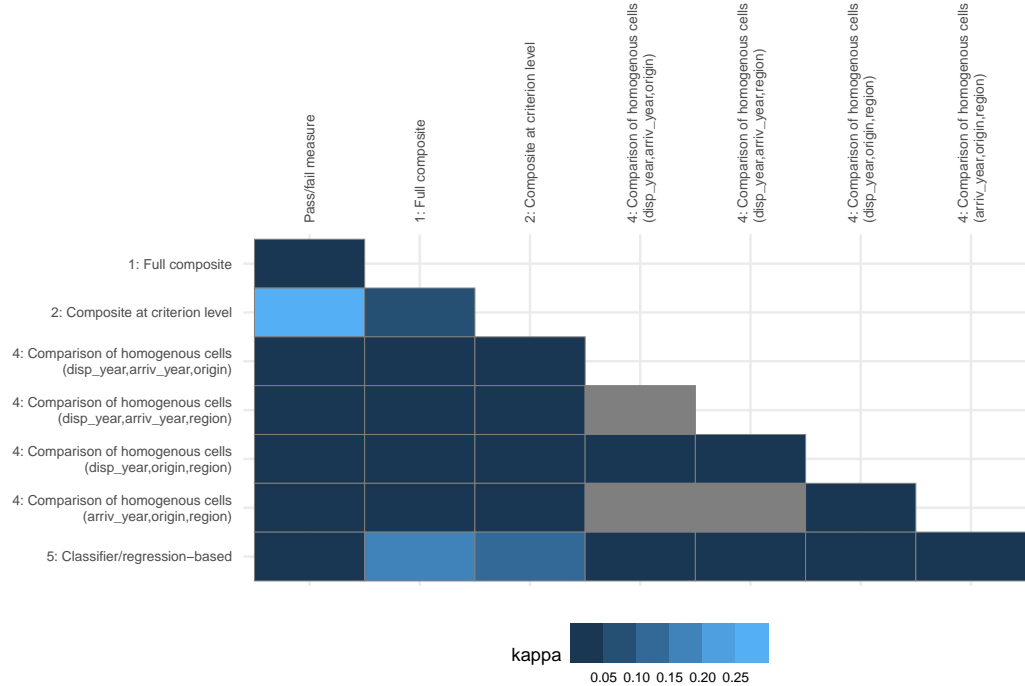


Figure 7: Inter-rater agreement between different metric options for Nigeria. Metric option 3 is excluded as no IDPs were classified as exiting the stock. A gray color suggests that kappa is undefined because the two metric options evaluated only output that no IDP exits the stock under either metric.

### *Overall feasibility and application of metric options*

Regarding the feasibility of implementing the different metric options in practice and with the capacities of a National Statistical Office in mind, it can be said that the patterns are similar to Hargeisa: The composite measure across all indicators is the easiest to implement. Dividing the IDP population into cells that can then be compared is also fairly easy to implement but may require clearer guidance on the criteria to distinguish different IDP sub-groups. Similarly, a regression-based approach is feasible to implement but requires clarification on probability cut-off, classifier used, and comparisons across contexts. Implementing indices on the criterion or sub-criterion level was more difficult to implement. This problem was even more severe in the dataset used in Nigeria compared to the data in Hargeisa because the dataset did not contain enough possible indicators to form indices for 6 out of 10 sub-criteria. Future assessments of how the IRIS solutions measure can be implemented should use datasets that provide a solid number of different indicators.

### **Simulation results for El Fasher (Sudan 2019)**

The IDP profiling in El Fasher (Sudan) covered a total of 1979 IDP households. As with the data from Nigeria, the results presented here are weighted by the survey weights included in the El Fasher sampling. Absolute and relative numbers should hence be compared to the number of 18.13 thousand IDP households.

When running the different metric options, the simulations all resulted in 0 IDP households exiting the stock as they compare worse than the host community on the 10 sub-criteria. This finding occurred across all metric options and across all indicators. This is because on average across the different possible indicators, 37.75 of the data cannot be assessed due to missingness on one of the 10 dimensions. Three of the indicators that miss the most in the data from Sudan is information on the ownership and documentation of property in the habitual residence and on access to compensation mechanisms. The HLP indicators are missing for many IDP households. If they are present and an assessment can be made, the HLP indicators are very restrictive: most IDP households do not have access to compensation or a written documentation to prove ownership. This explains why 0 IDP households exit the stock across all metrics. In the following sections, the results of the simulations are presented when excluding the HLP dimension (sub-criterion 4.1) in a briefer summary.<sup>15</sup>

First, it should be noted that the effect of indicator selection on the overall result is much more diminished when excluding the indicators for restitution and compensation. The biggest change between one indicator and the other indicator can be observed on the dimension of economic security: when using a measurement for whether the household is above or below the poverty 3.2 USD poverty line the exit from the IDP stock increases by 3.87 percentage points as compared to an indicator that measures the access to a mobile bank account.

Second, the importance of the indicator selection differs across metrics options and the overall pattern is the same as in Nigeria and Hargeisa: A composite measure at the criterion or sub-criterion level is very restrictive, allowing few IDP households to exit the stock. Similarly, a comparison of homogeneous cells is restrictive and classifies few IDPs as accessing the stock. In Hargeisa, IDPs were grouped into cells based on the housing structure they live in. This resulted in some larger groups, such as 12575 IDP households that live in a Tukul and only around 8 households that live in a house constructed of woods. Across those cells, very few IDP households could exit the stock. However, with other groupings - for which no suitable indicators were available in the dataset - this result could look differently.

---

<sup>15</sup> A later section in this note compares more systematically how the sub-criterion 4.1 *Property restitution and compensation* affects the results for Hargeisa and Nigeria.

We again find that the full composite index and a regression-based approach are more lenient, identifying more exits from the IDP stock - but they are also more prone to variation depending on the indicators selected and may classify too many IDPs as no longer subjected to displacement-related vulnerabilities when in fact they are. See Figure 8 for the density plots of the resulting simulations.

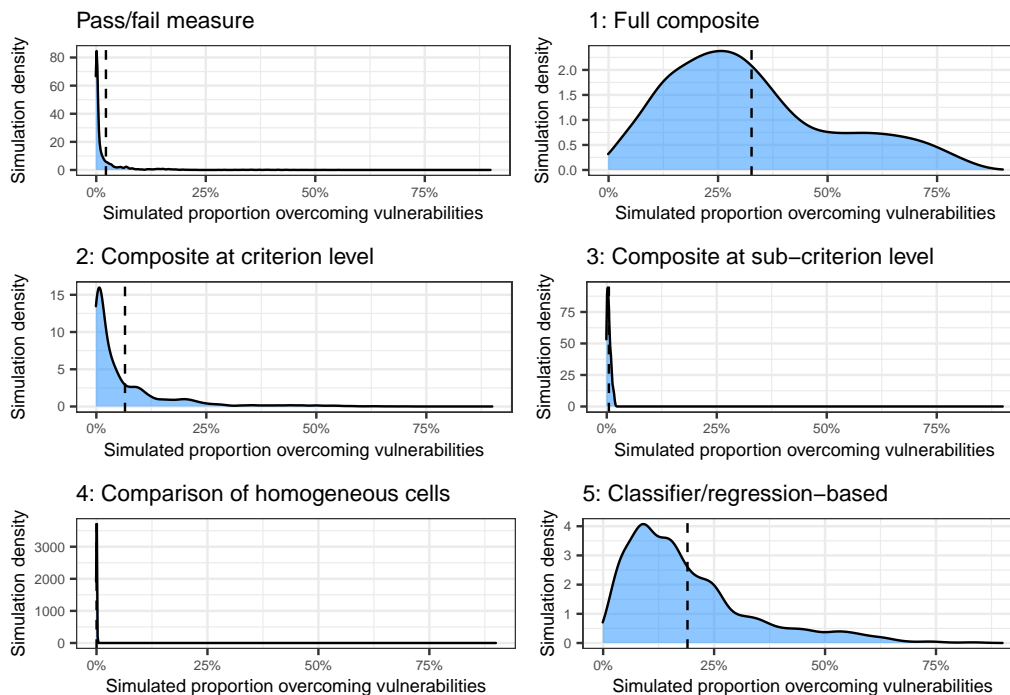


Figure 8: Density of simulations for all possible metrics (Sudan)

A final word on the feasibility of the different metric options in El Fasher (Sudan): To fully implement a composite measure at the criterion or sub-criterion level, there are not enough possible indicators on the household level. To implement the solutions simulation, the data from individual household members regarding their educational status, employment and documentation had to be first aggregated, making the process not impossible but more complicated. The IRIS suggests that the solutions measure should be assessed at the household level. If a survey only includes information on the individual level for key individuals, the guidelines developed by EGRIS for implementing the final solutions measure should give a clear indication how this aggregation from the individual to the household should occur: For example, should a binary household-level indicator for employment take the value 1 if one member of the household is employed or if all adult members of the household are employed? The dataset also did not include suitable grouping variables to assess the homogeneous cells approach in different ways. The full composite measure and the regression-based assessment were hence the easiest to implement but they are conceptually more difficult to allow comparisons across displacement contexts and they make the task of selecting the right indicators more acute.

### Simulation results for Colombia (DANE 2019)

The LSMS in Colombia covered a total of 4096 households that have fled from violence or natural disaster (weighted number of households: 744.25 thousand IDP households). Recognizing the survey weights included in the survey, around 14.02% of the IDPs in Colombia exit the stock on average across all indi-



cators and metrics. On average, 62.27% of the IDPs cannot be assessed due to missing data on at least one dimension. This number is so high because 4 variables are often not applicable/missing: attendance of a recognized educational institution, contractual terms of employment and working contract as well as presence of a written lease. For the remaining variables, data is almost complete. This high missingness is likely to also stem from the fact that some indicators have to be aggregated from individual-level survey questions to the household level.

### Importance of indicator selection

In the Colombian case, it was only possible to assess the importance of selecting the right indicators for six dimensions. For two sub-criteria, no indicators were available at all (1.2 *Freedom of movement*, 4.1 *Property restitution and compensation*). For two additional sub-criteria, there was only one indicator available (2.1 *Food security*, 5.1 *Documentation*). Table 14 summarizes how much the choice of one indicator over others can affect the exit from the IDP stock on average across all simulations and all metric options in Colombia for the six possible dimensions that could be assessed.

Overall, the selection of indicators creates less differences in the exit from the IDP stock for 2.2 *Shelter and housing*, 1.1 *Victims of violence* and 2.3 *Medical services*. For the remaining 3 sub-criteria, the indicator selection matters much more as choosing one indicator over the other can increase or decrease the exit from the stock by easily over 5 percentage points. However, it should be kept in mind that many assessments are not possible due to missing data. In fact, the reason why we find notable differences across indicators is entirely driven by missing data: For 2.4 *Education*, it seems as if measuring this dimension with the question whether respondents attend an officially recognized educational institution decreases the exit from the stock by -14.79 percentage points. However, in reality, missingness for this indicator is by almost 80% and this explains why so many IDPs no longer exit the IDP stock. This effect of data missingness also explains the notable differences that can be found between the indicators for 3.1 *Employment and livelihoods* and 3.2 *Economic security*.

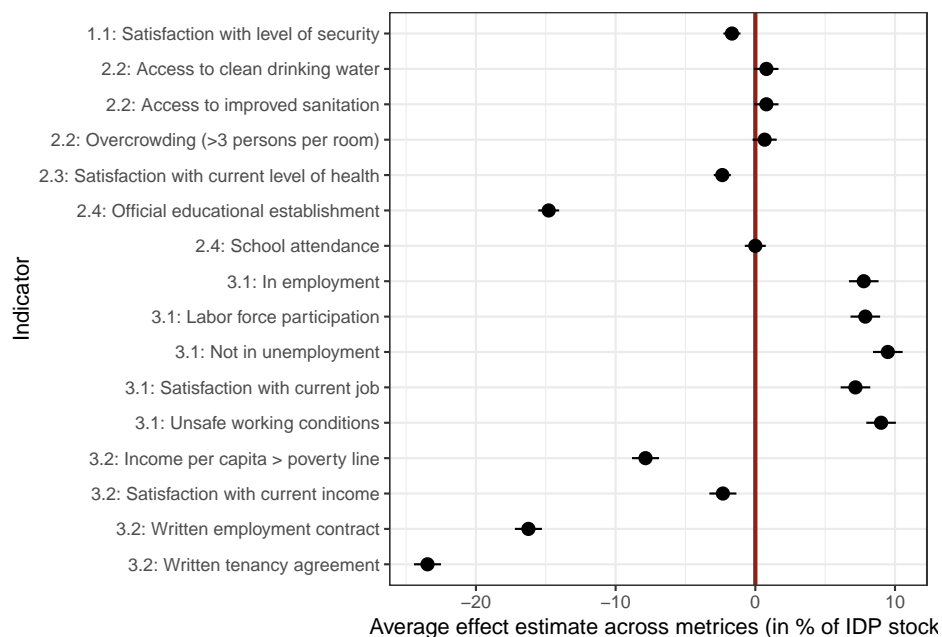


Figure 9: Average effect of indicator selection across metrics on IDPs exiting the stock (Colombia)

Table 14: Average relevance of indicator choice across metrics (Colombia)

Subcriteria IRIS	Indicators in Nigeria World Bank 2018	Average effect size on stock	Description of variation
1.1 Victims of violence	Experience of a natural disaster (baseline), Satisfaction with level of security	-1.67 to -1.67 percentage points	Some difference between different indicators
1.2 Freedom of movement	No indicator available	No other indicator	
2.1 Food security	Only available indicator was whether IDPs have an income per capita greater than the food security line	No other indicator	
2.2 Shelter and housing	Legal occupation of dwelling (baseline), Overcrowding (>3 persons per room), Access to improved sanitation, Access to clean drinking water	0.67 to 0.79 percentage points	Hardly any discernable difference between different indicators
2.3 Medical services	Possession of health insurance (baseline), Satisfaction with current level of health	-2.36 to -2.36 percentage points	Some difference between different indicators
2.4 Education	Literacy (baseline), School attendance, Official educational establishment	-14.79 to 0 percentage points	Notable difference between different indicators
3.1 Employment and livelihoods	Employment contract duration (baseline), In employment, Satisfaction with current job, Labor force participation, Not in unemployment, Unsafe working conditions	7.17 to 9.49 percentage points	Notable difference between different indicators
3.2 Economic security	Defaulting on utility bills (baseline) Satisfaction with current income, Income per capita > poverty line, Written employment contract, Written tenancy agreement	-23.47 to -2.32 percentage points	Notable difference between different indicators
4.1 Property restitution and compensation	No indicator available	No other indicator	
5.1 Documentation	Only available indicator was if IDPs had identification documents	No other indicator	

The lesson learnt from this is that indicators should cover the whole population and should minimise missingness to be able to make a reliable assessment. Figure 9 summarises the effect of the indicator selection again but readers should keep in mind that these differences are strongly determined by data availability. Comparing an indicator that covers the whole IDP households to an indicator that is missing in almost all of the IDP households will automatically result in a high difference between the two indicators. In practice, as the solutions measure is further developed by EGRIS, attention should be paid to selecting indicators that usually result in a high response rate and few missing data points.

### Importance of metric selection

The importance of selecting the right metric to combine the different indicators is summarised in Table 15 and the density plots in Figure 10<sup>16</sup>. Overall, the variation and standard deviation in how many IDPs exit the stock is higher in the case of Colombia but follows a similar pattern as in the other three cases:

The most permissive metric options are again a full composite index and a regression-based approach. Both approaches are the most sensitive to the chosen indicators and produce a higher estimate of IDPs exiting the stock on average across simulations. However, in the case of Colombia, we find that the composite measure at the criterion level generates the second-highest exit from the IDP stock with around 22.7 % of IDPs leaving the stock on average. This result is largely an artefact of the fact that few different indicators are available for each sub-criterion. As a result, the composite indices on the criterion-level are made of very few components and essentially behave like a composite index across all 10 criteria (similar to option 1).

Table 15: Simulation results for Colombia across metrics (Weighted IDP households: 7454.25 thousand)

	Pass/fail measure (no comparison!)	1: Full composite	2: Composite at criterion level	3: Composite at sub-criterion level	4: Comparison of homoge- neous cells	5: Classifier/ regression- based
Mean of IDPs exiting the stock	4916 IDPs (0.66% of stock)	173296 IDPs (23.28% of stock)	168961 IDPs (22.7% of stock)	21253 IDPs (2.86% of stock)	4997 IDPs (0.67% of stock)	215849 IDPs (29% of stock)
Range of IDPs exiting the stock	0 to 217234 IDPs (0 to 29.19% of stock)	20054 to 437423 IDPs (2.69 to 58.77% of stock)	20054 to 434418 IDPs (2.69 to 58.37% of stock)	0 to 76873 IDPs (0 to 10.33% of stock)	0 to 210128 IDPs (0 to 28.23% of stock)	0 to 627187 IDPs (0 to 84.27% of stock)
Variation in how many IDPs exit the stock	Middle (SD: 26112.29 )	Very high (SD: 93981.06 )	Very high (SD: 97640.42 )	Low (SD: 19761.57 )	Middle (SD: 25613.22 )	Very high (SD: 208771.21 )

The density plots also reflect this pattern and demonstrate that in the case of Colombia, a pass/fail measure, a composite at sub-criterion level and a comparison of cells lead to similar and restrictive results. In contrast, there is more variation and a higher estimate of IDPs leaving the stock in the case of a full composite index, an index on the criterion level and a regression-based approach.

### Identification of solutions on household level

Plotting the inter-rater agreement between the different metric options in Colombia in Figure 11 shows that on the household level the full composite measure behaves very similar to the composite measure at the criterion level (bright blue color). Again, this is likely due to the fact that the components of the criterion-level index are very few.

<sup>16</sup>Note that in Colombia, the homogeneous cell approach was not fully implemented: IDP households were all together assessed in comparison to the national average and not sub-divided into cells.

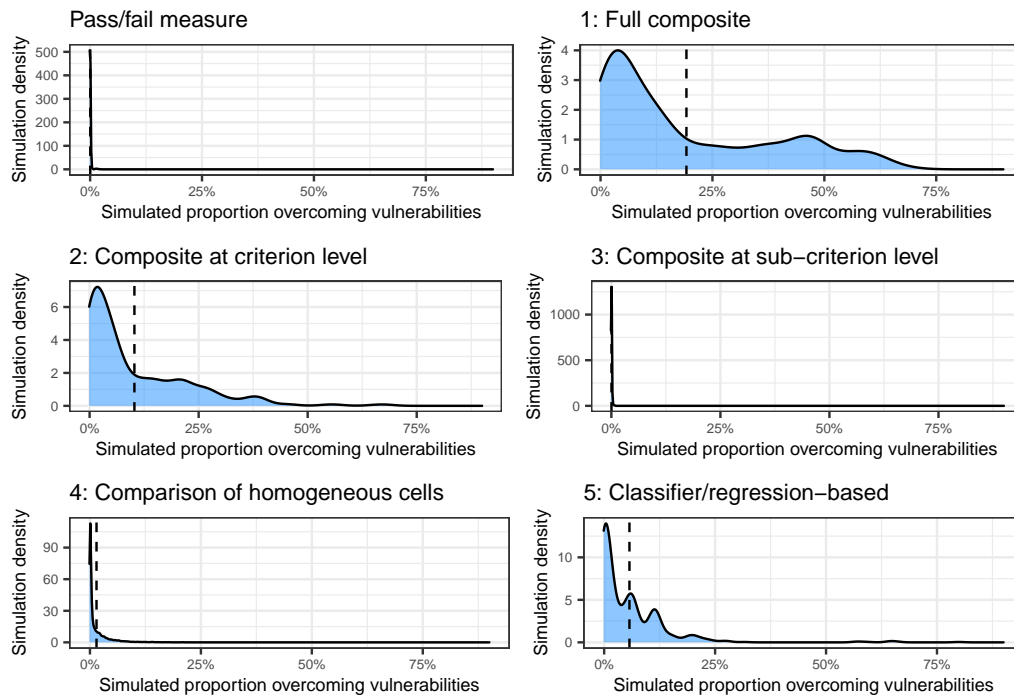


Figure 10: Density of simulations for all possible metrics (Colombia)

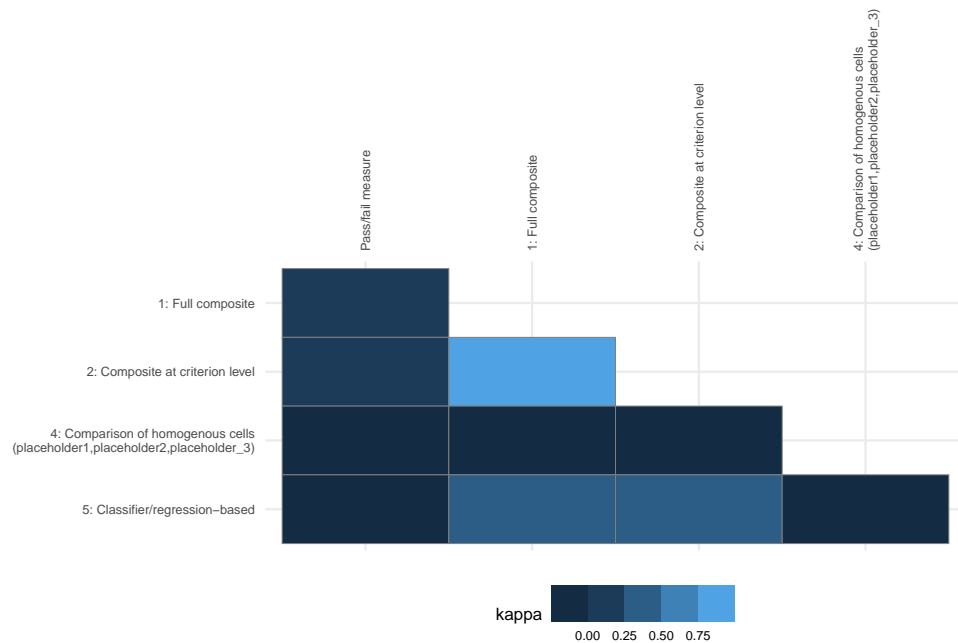


Figure 11: Inter-rater agreement between different metric options for Colombia. Metric option 3 is excluded as no IDPs were classified as exiting the stock.

### *Overall feasibility and application of metric options*

The results in Colombia demonstrate that the solutions measure and the different proposed metric options can also be applied to a nationally representative sample that includes IDPs. This demonstration is important as the IRIS solutions measure can either compare IDP households to host communities or to national averages. Nationally representative datasets have the clear advantage to cover the IDP population more thoroughly and not just in specific regions. The solution measure aims to determine the exit from the IDP stock for the purpose of official and aggregated statistics; the here presented results show promisingly that with a careful inclusion of all sub-criteria of the solutions measure in nationally representative datasets, one could easily conduct a solutions assessment without separate IDP profiling surveys. This increases the incentives for EGRIS to use commonly used indicators, such as the Sustainable Development Goals indicators, in the final selection of indicators for the solutions measure to increase the chances that those indicators are already included in broader surveys to generate official statistics.

Compared to targeted IDP profilings, as used in the other country examples, a large-scale nation-wide survey has fewer indicators available to conduct the simulations. Hence, the results of the Colombian case are difficult to interpret. As there are fewer indicators available, the composite indices on the criterion and sub-criterion level contain not enough elements. Furthermore, two of the 10 sub-criteria had to be ignored to be able to run the simulations. This indicates that some of the metric options, in particular option 2 and 3, might have too high data demands and require too many additional questions than can be expected to add to existing survey rounds by National Statistical Offices. A composite measure across all sub-criteria, a homogeneous cells approach, or a regression-based approach seems more plausible in this context.

Another lesson learnt from this is that any nationally representative survey that wants to implement the IRIS solutions measure should carefully include at least one indicator for each sub-criterion in the household questionnaire. This is particularly important for questions around HLP rights in the place of habitual residence and questions around the freedom to move that are *not* commonly covered in large-scale national surveys. If those topics could be included more prominently in surveys, implementing the solutions measure could become more feasible.

Finally, it should be added that we ran the simulations on the household and the individual level. Some of the indicators are more easily measured on the individual level (like education and employment). Others are most appropriate to measure on the household level. The IRIS prescribes that an assessment should take place on the household level and we present the main results on this level. However, it should be noted briefly that if an individual-level assessment is made, a smaller share of IDPs exits the stock. Fewer individuals tick off all the boxes. Aggregating the results at the household level may mask some inequalities in access to solutions within the household. While it might be acceptable that a household exits the stock if one person has access to employment, this might not be the case for other indicators: For example, access to identification documentation should be universal for all members of a displaced household. A definition of the final set of indicators should hence consider carefully how indicators on the individual level translate into household-level indicators.

### **Discussion of housing, land, and property rights indicators**

Amongst other results, the simulations have shown that the choice of indicators matters the most regarding the sub-criterion 4.1 *Property restitution and compensation*. This was the case in Hargeisa and Nigeria, where most indicator selections did not affect more than 1 percentage point difference in the number of IDPs exiting the stock. However, we have seen that choosing a more permissive HLP indicator can affect large percentages of the IDP stock. In the extreme case of Sudan, the property restitution indicators prevented all exits from the IDP stock. In Colombia, the dataset did not include questions on HLP

rights in the habitual residence, indicating the need to cover those modules in large data collections more systematically. To further investigate how the HLP indicators affect the simulation results, the figures 12 and 13 plot the simulation densities for an assessment of all 10 sub-criteria and for an assessment based on 9 sub-criteria (excluding the HLP indicators). The blue shaded curves exclude the HLP indicators; the grey shaded curves are the original assessment.

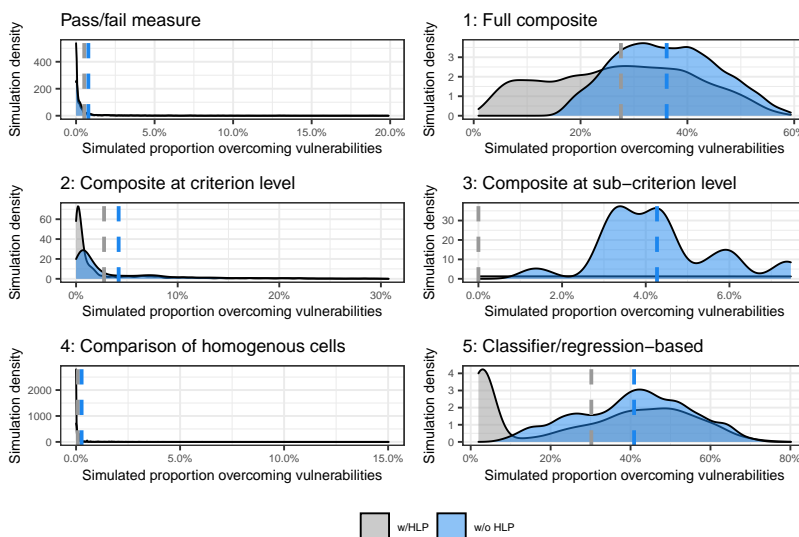


Figure 12: Density of simulations when including and excluding HLP indicators (Hargeisa)

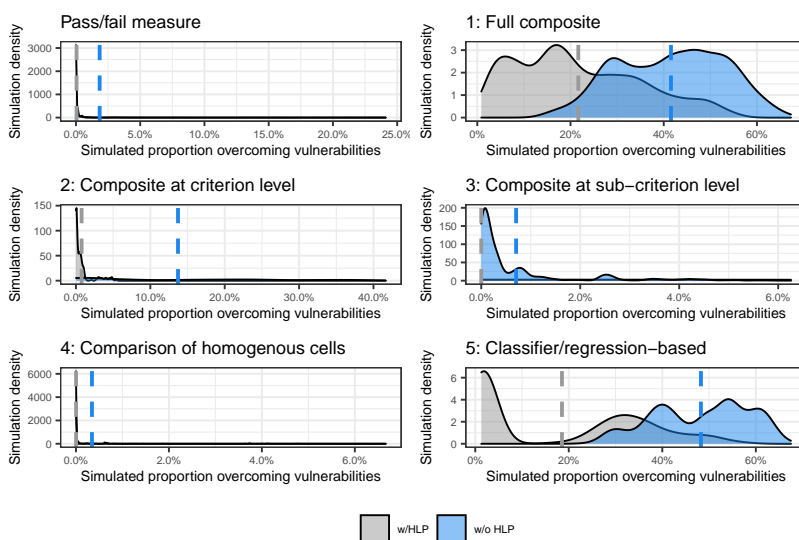


Figure 13: Density of simulations when including and excluding HLP indicators (Nigeria)

There are several important insights to gain from these plots: First, the dashed vertical lines represent the mean percentage of IDPs exiting the stock. Across both displacement contexts, one can observe that the mean exit from the IDP stock is higher if the HLP indicators are excluded from the assessment. This means that, overall, there seems to be little access to property rights restitution or compensation

mechanisms in the assessed displacement contexts. If HLP indicators are then included in the assessment, very few IDPs can pass the bar and can be considered as no longer suffering from key-displacement related vulnerabilities. Only assessing the other 9 dimensions is more permissive.

Related to this, excluding HLP indicators also has an effect on the different composite measures assessed in this study (options 1, 2, and 3). From the plots, we can see that the distribution of simulations moves towards the right (more IDPs exiting the stock) and we can observe more variation across indicators. This is the clearest in the composite metric at the sub-criterion level. While the original assessment based on all 10 sub-criteria did not yield any variation, we find more variation if the HLP indicators are excluded as too restrictive. In Hargeisa and Nigeria, removing the HLP indicator also “solves the problem” of a bimodal distribution for the regression-based approach. If the HLP dimension is ignored, the distribution resemble more a uniform distribution.

For the purpose of providing guidance on the further development of the IRIS “solutions measure”, the lesson learnt from this assessment is two-folded: First, the policy community should be aware that HLP rights are critical and are often lacking for IDPs. They are crucial and must be monitored to give IDPs the possibility to overcome their vulnerabilities and return to their own property and housing. However, for a statistical assessment how many IDPs overcome displacement-related vulnerabilities, the presence of restitution and compensation mechanisms is a very high bar that effectively hinders that IDPs can be taken out of the stock. One consideration could be to assess the IDP stock without the sub-criterion 4.1. *Property restitution and compensation* and to engage in separate but important policy work to ensure that IDPs can regain or prove their property after displacement.

The second lesson learnt is that if 4.1. *Property restitution and compensation* is to remain part of the “solutions measure”, an indicator should be chosen that can plausibly be achieved by IDP households. For example, secured ownership over the current housing or having proof for housing, land, and property in the form of documentation are more easily achievable indicators than access to formal restitution processes.

## Limitations and methodological concerns

This note set out to provide evidence to what extent the choice of a set of statistical indicators to measure solutions (“challenge 1”) and the choice of a method to combine these indicators into one statistical metric (“challenge 2”) affect a “solutions measure” for internal displacement. While the next sections will summarize the key insights from this assessment, it is important to acknowledge some key limitation in this simulation study:

1. **Sample:** Only four datasets were used to test the solutions measure due to time constraints. The coverage includes three household-level IDP profiling exercises that compare IDPs to the host community. This was complemented with a nation-wide survey that compares to the national overage. Given limited data and time, the surveys were not specifically selected to diversify context or policy relevance. Main issues with the selection are that no data from IDP contexts with a natural disaster are used, most datasets are not necessarily nationally representative or do not cover a large proportion of the IDP population, and most data collections were not part of an exercise to generate official statistics or feed into SDG reporting. Future assessments should identify other displacement contexts and samples that may be useful to analyze with the IRIS solution measure in mind.
2. **Data quality:** While the datasets were useful because they included multiple indicators that could be used to operationalize the IRIS solutions measure, there are some concerns about data quality, in particular in terms of comparability across contexts. Additionally, the survey in Hargeisa included more significant data gaps related to the employment of respondents. In Sudan, no indicator for

the freedom to move was included, the assessment did not include grouping variables that could be relevant for this simulation exercise and some indicators had to be aggregated from the household member level to the overall household. The Colombia dataset had a range of sub-criteria that were either not covered (e.g. freedom to move and HLP rights in habitual residence) or only covered with few indicators. Many indicators also had to be aggregated from the individual to the household level without clarity how this aggregation should be done. Hence, after some methodological decisions are made, reassessing how the improved solutions measure behaves in those contexts seems crucial.

3. **Compliance with standards:** Another concern is that indicators coded from any of the three datasets to measure the 10 sub-criteria are not necessarily in full compliance with SDG indicators. For the solutions measure, it would be beneficial to streamline the elements needed to form a solutions measure with SDG indicators, to increase the likelihood that indicators are already available in various displacement contexts, to increase comparability across contexts, and to reduce the needs to collect new data in order to implement the IRIS solutions measure. Future analyses could define a set of potential indicators guided by more statistical standards and good practices.
4. **Analysis:** Ideally, the work on operationalizing the relevant sub-criteria from the IRIS solutions metric would be done in close cooperation with different experts and organizations to ensure that the work aligns with other efforts, to be guided by statistical standards and common practices, to allow more context-specific knowledge and to cross-compare simulation results as the way statistical indicators are defined may vary depending on the coder. Additionally, the analysis had to retrofit potential indicators for the solutions measure from existing datasets. This is problematic as the indicators are hence not necessarily standardized and there are not always enough indicators to compare thoroughly. Options for future analyses include to fully simulate theoretical data on how much indicators matter in different metric options or to use datasets that are pre-designed for this methodological assessment.

Gathering the data and running these simulations has reiterated and revealed a few important methodological and technical difficulties that should be addressed going forward:

- **Missing data:** EGRIS should select indicators for the solutions measure that usually generate high response rate and have few missing values. In particular the Colombian case has shown that too much missingness on key indicators renders the solutions assessment very difficult. The Expert Group could also consider approaches to impute missing data
- **Data aggregation from the individual to the household level:** While the IRIS suggest a household-level assessment of solutions to displacement, several indicators are more naturally asked at the individual level (e.g. employment, education). Clear guidance is necessary how data should then be aggregated from the individual to the household level. Should this follow a rule that at least one household member is employed/in school, etc. or should this be based on an average?
- **Non binary indicators:** The simulations only covered binary indicators. Clear guidance is necessary if the final indicator selection will also include categorical or numeric variables. In the case of categorical variables, it is important to formulate them as “solution-oriented” indicators and to specify which categories are desirable (e.g. working in formal employment) or not.
- **Accounting for assistance, uncertainty and differences in the comparator population:** As outlined in the beginning of the note, it is still necessary to answer the question how assistance is factored out in the solutions measure, how uncertainty is considered, and if there are differences between host and national comparator populations.
- **Distinction between 3.1 *Employment & livelihoods* and 3.2 *Economic security*:** In some instances, it is not easy to discern if a potential indicator falls into 3.1 or 3.2. Clearer distinctions of the two dimensions could be achieved by developing the final set of statistical indicators.



## Conclusions and key recommendations

### Informing the selection of indicators

The first insight from this simulation study is that the **choice of a set of indicators to measure the 10 sub-criteria may matter less than the data quality and the metric chosen to combine these indicators**. On average across all simulations and metric options, very few indicators had a large effect on the exit of IDPs from the stock (except for HLP indicators). This finding may facilitate and de-politicize the debate as EGRIS moves towards defining concrete statistical indicators against the IASC (sub-)criteria of the solutions measure. However, as discussed in a separate section, the selection of suitable HLP indicators is crucial to develop a robust “solutions measure”.

The process of running the simulations re-emphasized the quality standards IRIS already promotes for the statistical indicators for a solutions measure: It is indeed crucial that indicators are commonly used, tested and evaluated and cover the population in question. It is important that the final “solutions measure” is constructed of **indicators that maximize coverage in the IDP population**. While it may be permissible to measure access to education in IDP populations with school attendance rates of children, the indicator does not provide information about IDP households that do not have children. As such, this may not be a problem because one can argue that they did not suffer from this vulnerability in the first place. However, if a large amount of the 10 indicators is only applicable to subsets of the IDP population, one risks that the count of IDPs exiting a national stock is based on only a small proportion of “real assessments” while a huge part of the IDP population is just not seen as vulnerable on the dimensions measured.

Another key conclusion concerns missing values. In particular in the case of the assessment in Sudan and Colombia, but also in the other cases, a big proportion of IDPs could not be assessed because they lack information on at least one indicator. Hence, it is important that a **clear list of common and standardized indicators on the household level** is identified that is already covered in many surveys. This list of indicators should be as much standardized as possible, for example by using the SDG indicators rather than other indicators. Finally, a preference should be given to indicators that commonly have high response rates and few missing values.

**Many potential indicators found in the Durable Solutions library are unclear regarding the question what constitutes a durable solution and what not.** For example, in which sectors households are employed (e.g. agriculture) does not provide any information whether this household overcomes displacement-related vulnerabilities or not. Nevertheless, the Durable Solutions library suggest to analyse the proportion of the IDP population working in self-employment or in work that requires seasonal movements. Another indicator suggests to look at access to agricultural land for agricultural IDP groups - but those indicators do not give a clear indication what a solution looks like for IDPs that are not self-employed, do not move for work or do not need agricultural land. It would hence be crucial to formulate the indicators to implement the “solutions measure” in a **solution-oriented** way (e.g. ‘Target population experiencing no security incidents’ as opposed to ‘Target population who think it likely they will experience serious consequences due to hazard.’) and to focus on a clear concept how a solution should look like for the final indicator selection. This is particularly important for potential categorical indicators.

Finally, the indicator choice across all simulations seemed to matter the most for the sub-criterion 4.1. *Property restitution and compensation*. This was the case across displacement contexts (Hargeisa, Nigeria, Sudan) and across metric options. Given that 4.1. is a rights-based indicator, it seems crucial to decide whether these indicators should be measured based on factual compensation or only on access to compensation mechanism or only documentation of property rights. An alternative is to take out HLP rights from the durable solutions assessment and monitor them in separate work streams. Once this decision is made, the variation found across metrics may be drastically diminished. The assessment

suggests as lessons learnt that the assessment of restitution and compensation should either be separate of the 'solutions measure' (excluding this dimension from the 10 sub-criteria) or to identify realistic and permissive indicators to mitigate that the property indicators are a very high bar to pass.

## Informing the selection of a metric

Given that the individual indicators are on average less relevant for the final number of IDPs exiting the stock, the discussion around the solutions measure should first focus more deeply on the selection of a model/metric by which the indicators are then combined and analyzed in a way that complies with the wider IRIS framework and principles. How the 10 different sub-criteria are combined and compared to a comparator population drives more of the variation than the exact definition of each statistical indicator.

The assessment shows that the different available options to implement the contextualized approach that IRIS promotes and to find a workable measure have different strengths and weaknesses. Table 16 summarizes these insights.<sup>17</sup>

Table 16: Summary of metric advantages and disadvantages

	Pass/fail measure	1: Composite metric across all criteria	2: Composite metric at criterion level	3: Composite metric at sub-criterion level	4: Homogeneous cells	5: Classifier/regression-based approach
Aggregation of indicators to indices	No aggregation	High-level aggregation	Medium-level aggregation	Low-level aggregation	No aggregation	No aggregation
Crudeness of (sub-)indices	Not applicable, no index	Low (consists of 10 elements)	Medium to high (each criterion can be measured with multiple elements)	High (realistically each sub-criterion can only be measured with 3-4 elements)	Not applicable, no index	Not applicable, no index
Importance of indicator choice	Very low to low	High to medium	Medium to low	No variation	Very low to no importance	High to very high
Characteristics of the metric	Easy to implement; no comparison between IDPs and hosts	High variability of index likely reduces comparability across contexts	Need to define multiple indicators for each criterion	High data demands; potentially sensitive to crudeness of index	Potential challenges in defining the cells	Potential challenges in defining cut-off points and comparing across contexts

Option 1, a composite metric across all criteria, has shown high variability in the simulation. This means that this way of combining the indicators and comparing one score to the host population is very sensitive to the indicators chosen. It is hence not necessarily advisable to use one unified metric across all criteria, which also violates the IRIS principle that all sub-criteria must be met, because it is highly sensitive to the exact combination of indicators, and because it is hence likely that a comparison across displacement contexts is flawed.

Option 2 has merits as it is not overly sensitive to the choice of indicators in the composite metrics at the criterion level. However, it is necessary to define enough indicators for each criterion to ensure that the indices are not too crude. This increases the data demands to implement these options.

<sup>17</sup>For the sake of testing the validity of the simulation findings, we also compared non-displaced households to displaced households as the benchmark. The logic is that host communities should exit the stock as they perform significantly better. However, when doing this recoding, we still find very similar patterns across metric options with few "fictive exits from the stock". This suggests that our findings detect that certain metric options are difficult to pass *by design* (independent of the status as host or IDP household).

Option 3 is based on a similar approach as option 2 but is focused on indices on the sub-criterion level. This yielded no to very little variation in the results as very few IDPs were classified as exiting the stock. Moreover, option 3 will often be infeasible in practical terms as the datasets used simply do not contain enough indicators to fully implement indices at the sub-criterion level.

The recommendation is that further analysis and better data is needed to first fully implement these two composite metrics and to then assess how they behave comparatively. A first indication is that a composite metric at the criterion level may be more realistic although it also violates the IRIS principle that all sub-criteria must be met.

The simulations assessed whether a division of the IDP stock into sub-groups that can then be compared to the host population may be a feasible option. The analysis delivered generally favourable findings but is not fully conclusive on this option 4. On the one hand, the homogeneous cell approach is easy to implement, comes with manageable data needs, is relatively straightforward to communicate, and does not appear overly sensitive to indicator choice. On the other hand, it might be sensitive to how the IDP population is divided into cells. This assessment did not find any stark differences between different ways of dividing the IDP stock into cells. Nevertheless, more analysis is needed to be sure and clear guidance would need to be developed by EGRIS for producers of IDP statistics on how to segment the population under analysis into smaller cells while maintaining cross-context comparability. The datasets used in this assessment were not designed with grouping variables in mind which should be considered more theoretically before making a final decision.

Eventually, it may also be possible to use a regression-based approach classifying IDP households as being more similar or dissimilar with other IDP households and the host community. However, this approach is highly sensitive to the indicator selection, with the highest variability across indicator combinations. Similar to option 1, it is not necessarily advisable to use one regression-based approach because: (a) it violates the IRIS principle that all sub-criteria must be met, (b) it is sensitive to the indicator selection, and (c) it might be difficult to compare different displacement situations. This option requires further methodological work to assess how the choice of a probability cut-off point, and the choice of a classifying framework affect the results. Further work is also needed to investigate how the regression coefficients for the 10 different sub-criteria may vary across contexts and how this affects the comparability between countries and displacement situations. Another issue with this option is that it yields a multimodal distribution (i.e. two peaks) which raises the question what the true average population of IDPs that has overcome their vulnerabilities is. This could be mitigated by excluding the HLP indicators.

**If EGRIS wants a permissive solutions measure that exits a substantial proportion of IDPs from the overall stock,** the full composite measure and the regression-based assessments are possible options that make it easier to “reduce” the IDP stock in reporting. Since this can be seen as a risk, it may be advisable to select relatively strict indicators to measure the sub-criterion to counteract the permissive tendencies of those metric options. A simple pass/fail measure (without context comparison) or a composite measure at the sub-criterion are not good options due to the high data-demands and the restrictive nature of the metrics as almost no IDP household can meet the standard. **If EGRIS wants a more restrictive solutions measure that exits few IDPs from the stock,** a possibility would be to implement a composite measure at the criterion level or to identify homogeneous IDP cells. For the composite measure, a set of clear and sufficient indicators per criterion is necessary, which was not always given in the data used in this simulation. For a cell-based approach, better grouping variables have to be identified that generate small but meaningful IDP cells. Cells that are too big will make it hard for the IDP population to perform as well as the host community. Very small cells will make the assessment very crude and dependent on outliers.

Box 5 provides an overview of the key findings and recommendations. As stated in the introductory sections of this note, at least 7 challenges remain in turning the IRIS solutions measure into a fully workable and robust statistical metric. This note explored only the first two of these challenges. The other challenges will also need to be discussed.

#### Indicator selection

- Indicator selection may matter less for aggregate outcomes than data quality and metric selection
- Focus on indicators that maximize coverage of the IDP population and can be easily measured/aggregated to the household level
- Focus on indicators that have high response rates and few missing data
- Focus on indicators that are commonly available in already existing data collections
- Focus on clearly defined and standardized indicators that can be compared across contexts (e.g. SDG indicators)
- Define the indicators as solutions-oriented
- Address the rights-based sub-criterion 4.1. *Property restitution and compensation* that could be measured based on documentation for rights, access to compensation mechanisms/replacements or factual compensation/replacement: either take HLP indicator out of the assessment or define a lenient and achievable indicator

#### Metric selection

- **Permissive metric options:** Concerns about one full composite metric (option 1) and a regression-based approach (option 5) due to high variability and lack of alignment with IRIS principle of achieving a solution in each (sub-)criteria
- **Restrictive metric options:** Excessive data demands and too little variability to implement a composite metric at the sub-criterion level (option 3)
- **Potential options:** Further analysis needed to fully implement and assess a composite metric at the criterion level (option 2) or homogeneous cells (option 4) as the datasets used here do not fully allow a comparison

### Box 5: Key conclusions and recommendations

## Annex: full mapping of all indicators available to measure the IASC criteria

Sub-criteria (IRIS)	Potential indicators from the inter-agency Durable Solution Library	IDP Profiling in Hargeisa (UNHCR 2015)	IDP Profiling in Nigeria (World Bank 2015)	IDP Profiling in El Fasher (JIPS 2019)	LSMS in Colombia (DANE 2019)
1.1 Victims of violence	<p>Target population who think it likely they will experience serious consequences due to armed conflict and other situations of social instability or tension which are subject to international humanitarian law, human rights violations and national legislation.</p> <p>Target population who think it likely they will experience serious consequences due to a hazard.</p> <p>Target population who feel safe walking alone around the area they live (during day or night).</p>	<p>Target population is worried about being exposed to theft, crime or vandalism in their place of residence.</p> <p>Target population is feeling unsafe or insecure in their place of residence.</p>	<p>Target population feeling very or moderately safe.</p> <p>Target population feeling very or moderately safe walking at night.</p> <p>Target population feeling very or moderately safe walking during the day.</p> <p>Target population experiencing non-physical or physical harm in the last 12 months.</p>	<p>Target population feeling very or moderately safe walking at night.</p> <p>Target population feeling very or moderately safe walking during the day.</p> <p>Target population experiencing a robbery.</p>	<p>Target population satisfied with the current level of security.</p> <p>Target population who experienced a natural disaster in the past 12 months.</p>
	<p>Target population who were subjected to physical, psychological or sexual violence in the previous 12 months (or since time of displacement, if displaced for less than 12 months).</p> <p>Target population who have experienced other safety or security incidents in the previous 12 months (or since time of displacement, if displaced for less than 12 months).</p> <p>Target population who were affected by hazard in the previous 12 months (or since time of displacement, if displaced for less than 12 months).</p> <p>Target population who experienced violence in the previous 12 months, who reported their victimization to competent authorities or other officially recognized conflict resolution mechanism.</p>	<p>Target population who experienced victimizing events in their place of residence in the past 12 months.</p>	<p>Target population who have experienced harm and have reported it in the last 12 months.</p> <p>Target population who report thefts or disputes to formal conflict resolution mechanisms.</p> <p>Target population who find it very easy or somewhat easy to access dispute resolution mechanisms.</p> <p>Target population that find dispute resolution very or moderately effective.</p>	<p>Target population who have experienced harm and have reported it to the police.</p>	

1.2 Freedom of movement	Target population facing restrictions to their freedom of movement.	Target population who face legal or administrative restrictions of their freedom of movement (i.e. lack of documentation, restricted movements in living area).	Target population feeling free to move in and out of their area.		
2.1 Food security	Target population by prevalence of moderate or severe food insecurity in the past year, based on the Food Insecurity Experience Scale (FIES).	Target population unable to pay for food in the last 6 months.	Index of food insecurity from a combination of indicators.	Target population unable to pay for food in the last 7 days.	Target population with per capita income higher than the food security line.
		Number of meals eaten per day.		Target population having to borrow money for food.	
2.2 Shelter and housing	Target population with secure tenure rights to land, with legally recognized documentation, and who perceive their rights to land as secure. Target population having been forcibly evicted over the past 12 months. Target population, not being evicted in the past 12 months, but living in constant fear of eviction (population who do not perceive their current tenure rights as secure). Target population residing in sufficient living space.	Target population living in overcrowded housing/shelter (> X persons per room). Target population living in inadequate housing conditions (risk of landslide, near trash receptacles or industry). Target population with access to electricity or other modern energy services.	Target population having access to land and renting or owning it legally.	Target population legally owning the property.	Target population legally occupying their dwelling.
	Target population residing in durable housing structures.		Target population living in non-durable housing conditions (incomplete, not intended or makeshift housing). Target population squatting or living in temporary shelter. Target population owning or renting housing. Target population squatting.		
	Target urban population living in slums, informal settlements, or inadequate housing. Target population with access to basic drinking water services.	Target population with access to adequate source of drinking water (tanks)	Target population with access to safe drinking water.  Target population with no obstacles for water access.	Target population with access to improved drinking water sources.	Target population with access to clean drinking water.

	Target population with access to basic sanitation facilities including a hand-washing facility on premises with soap and water.	Target population with flushing toilet in household  Target population with bath/shower in household	Target population with improved sanitation facilities.	Target population with improved sanitation facilities.	Target population with improved sanitation facilities.
2.3 Medical services	Target population covered by essential health services. Target population who accessed essential health care services (including mental health care) the last time they needed it in the past 12 months. Births within target population attended by skilled health personnel within the past 12 months (% of total births taken place within the past 12 months). Children under the age of one in the target population covered by all vaccines included in their national programme (% of total child population under one in the target population).	Target population with access to essential health care when needed.  Target population with births or pregnancies attended by skilled health personnel.  Target population with immunized children.	Distance to health facilities in hours.  Target population that access essential health care when needed.	Distance to health facilities in hours.  Target population satisfied with health care.	Target population in possession of health insurance. Target population satisfied with current level of health.
2.4 Education	Primary school net attendance ratio in targeted population (% of children of primary school age in target population).      Secondary school net attendance ration in target population (% of children of secondary school age in target population). Target population who own a mobile phone.	Target population of school age that can read and write.  Target population of school age that has ever attended school (primary and secondary). Target population of school age that is currently attending school.  Target population of school age that is currently attending or have completed secondary school or university. Target population which owns a mobile phone.	Time to next education facility.  Target population being satisfied with primary education.	Time to next education facility.  Target population of school age that has ever attended school.  Target population of school age that is currently attending school.	Target population that is literate.  Target population attending school.  Target population attending an officially-recognized educational establishment.
3.1 Employment and livelihoods	Target population employed in formal and informal sector (employment rate).  Employed and self-employed target population in informal employment in non-agriculture employment (% of total employed target population). Employed and self-employed target population that is underemployed. Target population undertaking pendular or seasonal movement due to work. Target population aged 5-17 engaged in child labour (% of total child population 5-17 years of age).	Target population with a breadwinner in household.	Target population whose primary income are wages, salary, own businesses, or pension.	Target population engaged in paid job.	Target population in employment.

	Target population aged 15-24 years not in education, employment or training. Self-employed target population employing others (% of total self-employed target pop). Labor force population who are unemployed (unemployment rate).			Target population that is reporting unemployment.	Labor force participation.  Target population who are not unemployed. Target population in unsafe working conditions. Employment contract duration.
3.2 Economic security	Unemployed target population covered under social security schemes (public or private). Target population relying primarily on sustainable income sources over the last 30 days.  Target population who in the last 12 months was not able to pay for basic expenses. Target population who in the last 12 months obtained a loan to cover basic expenses. Target population's average expenditure against average total expenditures, per capita. Ratio of average food expenditures against average total expenditures, per capita. Target population below the poverty line.	Target population capable of managing unexpected expenses without borrowing money or receiving help from others.  Target population who was not able to pay house rental / services in the last 6 months.	Target population consuming more than average.  Ratio of food consumption against total consumption.  Target population below 1.9 USD PPP 2011 Poverty Line.  Target population below 1.25 USD PPP 2011 Poverty Line Target population below 3.1 USD PPP 2011 Poverty Line	Target population below 1.9 USD PPP 2011 Poverty Line.  Target population below 3.2 USD PPP 2011 Poverty Line.	Target population with written employment contract.  Target population in possession of a written tenancy agreement. Target population defaulting on utility bills.  Target population satisfied with current level of income.  Target population requiring child labor.  Target population with per capita income greater than national poverty line.
	Target population who own productive assets by type of assets. Target population with access to markets. Target population where at least one person in household has a bank account.	Average number of assets owned by target population.	Distance to market in hours. Target population with access to a bank account.	Distance to market in hours. Target population with mobile money account.	
4.1 Property restitution and compensation	Target population with documents to prove ownership/tenancy of housing, land and property left behind (% of total target population who left behind HLP).	Target population with documents to prove ownership of their lost HLP.	Target population that is legally recognized owner of dwelling.	Target population that is legally recognized owner of dwelling.  Target population that has documentation to prove ownership.	



	<p>Target population with housing, land and property left behind who successfully accessed restitution or compensation mechanism (% of target population with HLP left behind) – if relevant to context.</p> <p>Target population with housing, land and property left behind who have had their claims to assets (incl. land and property) resolved.</p> <p>Target population with housing, land and property left behind who have had their claims to assets (incl. land and property) enforced.</p> <p>Agricultural target population with ownership or secure rights over agricultural land.</p> <p>Agricultural target population with use rights to agricultural land.</p>	<p>Target population with lost HLP who have accessed restitution or compensation mechanisms.</p> <p>Target population with lost HLP who have had their claims resolved and enforced.</p>	Target population with access to compensation mechanisms.	Target population with access to compensation mechanisms.	
5.1. Documentation	<p>Target population currently in possession of valid birth certificates, national ID cards or other personal identification documents relevant to the context.</p> <p>Target population with other personal documentation necessary to accessing their rights.</p> <p>Target population registered by authorities as Internally Displaced Persons – if relevant to context.</p> <p>Children under 5 years of age in target population whose births have been registered with a civil authority.</p>	<p>Target population in possession of birth certificates.</p> <p>Target population with personal documentation or access to mechanisms to replace them if they are lost.</p> <p>Children in target population with a birth certificate or registration.</p>	Target population that have not lost their documents and know how to replace them if missing.	Target population in possession of birth certificates.	<p>Target population with personal documentation.</p> <p>Target population with personal documentation.</p>