

Expert Group on Refugee, IDP and Statelessness Statistics

Towards a harmonized statistical measure for exits from the stock of internally displaced persons

Refining the composite measure for overcoming key displacement-related vulnerabilities proposed in the International Recommendations on IDP Statistics (IRIS)



August 2023
Methodological Paper Series
Paper No. 2

The Expert Group on Refugee, IDP and Statelessness Statistics (EGRISS) is a UN Statistical Commission (UNSC) mandated, multi-stakeholder group.

Since its establishment in 2016, it works with national statistical offices, international organizations and civil society to develop and support implementation of international standards and guidance to improve official statistics on forcibly displaced and stateless persons.

With 57 country members and 36 institutional members, EGRISS welcomes collaboration with all relevant partners from the statistical community.

EGRISS implements activities with and through its members, organized in Technical Subgroups (TSG). This paper was developed by TSG 2 'Methodological Research and Guidance Development'.

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

EGRISS would like to express particular appreciation for leading the development of Methodological Paper Series No. 2 to Sigrid Weber (Immigration Policy Lab, Stanford University), Hisham Galal (UNHCR), Felix Schmieding (World Bank–UNHCR Joint Data Center on Forced Displacement) and Kristine Vegard (Joint IDP Profiling Service). Thanks are extended also to Martina Caterina (UNHCR) for her contributions and expertise as the paper was developed



Contents

CONTENTS	1
INTRODUCTION.....	3
STATE OF PLAY ON THE “IRIS EXIT MEASURE”	4
The conceptual framework: ten sub-criteria and two principles	4
Nine remaining methodological challenges.....	5
METHODOLOGY	8
OVERCOMING CHALLENGE 1: METRIC CHOICE FOR BENCHMARKING AGAINST COMPARATOR POPULATION. 9	
Metric options.....	9
Pass/fail measure: Implementing the IRIS without a comparator population	9
Option 1: A composite metric across all criteria.....	9
Option 2: A set of composite indices at the criterion level	10
Option 3: A set of composite sub-indices at the sub-criterion level.....	10
Option 4: Comparison of homogeneous cells:	11
Option 5: Classifier/regression-based approach	12
Option 6: Empirical cumulative distribution approach	13
Simulation results for Hargeisa (UNHCR 2015).....	15
Simulation results for Nigeria (World Bank 2018)	17
Simulation results for El Fasher (Sudan 2019)	20
Simulation results for Colombia (DANE 2019)	22
Overall feasibility and application of metric options	24
Conclusions on challenge 1	26
Summary of findings	26
Recommendation	28
OVERCOMING CHALLENGE 2: INDICATORS FOR THE 10 SUB-CRITERIA	30
Simulation results for Hargeisa (UNHCR 2015).....	30
Simulation results for Nigeria (World Bank 2018)	32
Simulations results for El Fasher (Sudan 2019)	33
Simulation results for Colombia (DANE 2019)	35
Selecting indicators for the exit measure	37
Mapping and comparing possible indicators for each sub-criterion.....	38
Empirical performance of the recommended indicators	46
Conclusions on challenge 2	48
Summary of findings	48
Recommendation	49
OVERCOMING CHALLENGE 3: SUB-CRITERION ON PROPERTY RESTITUTION AND COMPENSATION.....	50

Simulations with and without the HLP sub-criterion	50
Addressing housing, land and property issues: shifting focus to security of tenure	52
Conclusions on challenge 3	53
Summary of findings	53
Recommendation	53
OVERCOMING CHALLENGE 4: DEALING WITH MISSING DATA POINTS	55
Data missingness due to design and non-response	55
Options for dealing with missing data	55
Simulations on the problem of missing data	56
Conclusion on challenge 4	57
Summary of findings	57
Recommendation	57
OVERCOMING CHALLENGE 5: AGGREGATION OF INDIVIDUAL-LEVEL INDICATORS TO HOUSEHOLD LEVEL ..	59
Available data aggregation rules	59
Simulations on the data aggregation from individuals to the household.....	60
Summary of findings	62
Recommendation	62
OVERCOMING CHALLENGES 6-9: THE ROAD TOWARDS A COMPARABLE EXIT MEASURE	63
Overcoming challenge 6: Addressing statistical uncertainty in benchmark values.....	63
Overcoming challenge 7: Defining the comparator population	64
Overcoming challenge 8: Vulnerabilities overcome merely by assistance received	65
Overcoming challenge 9: Addressing changing benchmark values over time.....	65
GOING “FULL CIRCLE”: A FINAL ROUND OF ANALYSIS IMPLEMENTING ALL OF THE ABOVE RECOMMENDATIONS	66
ANNEXES	68
Annex I: Detailed simulations methodology.....	68
Step 1: Identifying the universe of potential indicators.....	68
Step 2: Selecting empirical datasets on IDP vulnerabilities & solutions	68
Step 3: Preparing the indicators and indices for simulation	70
Step 4: Iteratively simulating the indicator and metric choice	71
Annex II: Limitations and methodological concerns	72
Annex III: Full mapping of all available indicators for the exit measure	73
Annex IV: Presence and implementation of proposed indicators for the IRIS exit measure	79
Annex V: Difficulty of passing proposed indicators under challenge 2	80

Introduction

1. The *International Recommendations on IDP Statistics* (IRIS), developed by the *Expert Group on Refugee, IDP and Statelessness Statistics* (EGRISS), provide a framework for capturing a country's stock of IDPs for statistical purposes. To correctly calculate the stock, it is crucial to define when individuals enter the stock and when they exit the stock. This paper relates to the latter, and specifically to the way by which individuals exit the stock by overcoming key displacement-related vulnerabilities - hereinafter referred to as the "IRIS exit measure".
2. It is important to note that the purpose of a statistical definition for the exit from the IDP stock 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. This exit measure "should be used for all IDPs (i.e. IDPs in locations of displacement, IDPs in locations of return and IDPs in other settlement locations)" (IRIS, par 160).
3. Separate from the exit measure, IRIS also proposes a distinct but related framework to assess progress towards the achievement of durable solutions - referred to in short 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 both progress towards solutions and exit from the stock 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 exit measure).
4. This note focuses on the IRIS exit measure and how it can be implemented in practice. With the introduction of the exit measure, the IRIS have advanced the debate on measuring exits from the IDP stock 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.
 - Move to a context-dependent assessment rather than an absolute approach.
 - Narrowing of benchmark options (to national averages or host community averages).
 - Recommendation of quality criteria to assess the suitability of indicators to measure each sub-criterion.
5. Nevertheless, as stated in the recommendations themselves, the above achievements do not result in a final and applicable statistical measure which can be applied directly in the work of statistical producers. This methodological paper first outlines the conceptual framework for the exit measure as suggested by the IRIS and identifies the remaining methodological challenges that persist in turning this into a fully applicable statistical measure. The main body of the paper then makes concrete suggestions and provides decision points for EGRISS on how to overcome these challenges. The suggestions in this paper are backed up by empirical data from Hargeisa, Nigeria, Colombia and Sudan. Considering the absence of a concrete statistical measure at this stage, a simulation approach - cycling through all foreseeable metric and indicator choices for the exit measure - is applied to illustrate exit from the IDP stock.

State of play on the “IRIS exit measure”

The conceptual framework: ten sub-criteria and two principles

6. 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

7. To then assess how many displaced households in a specific context have progressed, 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 in each sub-criterion are no longer considered displaced for the purpose of statistics.¹ The IRIS promulgates this as follows:

*“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)

8. To sum it up, the IRIS established that the exit measure should be based on an assessment of the ten sub-criteria, and should follow two key principles, namely (a) benchmarking against a comparator, and (b) achieving a pass in *each* of the sub-criteria. While following these two key technical principles, the measure must be globally relevant to a wide range of displacement contexts and must realistically minimise potential biases in reported IDP stocks resulting from the fact that IDP inflows are easier to measure than outflows (IRIS, par 160).
9. Table 2 provides an example to understand this framework. Household A passes the criteria *1. Safety & security*, *4. Restoration of HLP rights*, and *5. Documentation*. However, because it has not

¹ 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.

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, the household should not exit the IDP stock because it has not overcome the key displacement-related vulnerabilities measured in the exit measure.

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 and livelihoods	No pass	0/2 (Not passed)	
	3.2 Economic security	No pass		
	4.1 Property restitution and compensation	Pass	1/1 (Passed)	
	5.1 Documentation	Pass	1/1 (Passed)	
	Overall: Vulnerabilities not overcome; displacement not ended			

Nine remaining methodological challenges

10. While this framework appears intuitive, there are several challenges and methodological gaps when implementing it in practice.
11. The first 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.
12. The second 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.
13. A third challenge relates to one of the ten sub-criteria of the exit measure, namely 4.1 *Property restitution and compensation*. An implicit assumption in examining access to restitution or compensation mechanisms is that such mechanism are in place in displacement contexts. Experts on housing, land and property confirm that this is only the case in a limited number of cases across

the world. This effectively bars household from exiting the IDP stock in the vast majority of countries. Moreover, this sub-criterion cannot be easily benchmarked against a comparator, thereby violating one of the two technical principles of the exit measure as listed above. A decision on the **operationalization or the reframing of this sub-criterion** may be needed to develop a globally relevant exit measure.

14. 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. This results in a fourth methodological challenge where data points are missing by design (e.g. households without children cannot be assessed for whether children currently attend school). 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**.
15. A fifth methodological challenge is the unclear guidance on the **aggregation of individual-level indicators to the household-level**. Some of the sub-criteria promoted for the exit measure are more naturally measured on the individual-level than the household-level. For example, employment is easiest to assess on the individual-level. There is hence a need to develop clear rules on how different indicators should be aggregated to the household-level at which the assessment and comparison with national/host community averages takes place.
16. The sixth not addressed methodological challenge 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.
17. A seventh 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.
18. As the eighth challenge, the IRIS exit 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 exit measure only produces an aggregate number, the measure should ensure that the overall exit from the IDP stock is not merely achieved through **reliance on humanitarian assistance**.
19. Related to the definition of a comparator population is the ninth challenge, on 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 any new displacement-causing events.

Box 1: Remaining methodological challenges for a workable exit measure

Challenge 1: Metric choice for benchmarking against comparator population

Challenge 2: Indicators for the 10 sub-criteria

Challenge 3: Sub-criterion on property restitution and compensation

Challenge 4: Dealing with missing data points

Challenge 5: Aggregation of individual-level indicators to household level

Challenge 6: Addressing statistical uncertainty in benchmark values

Challenge 7: Defining the comparator population

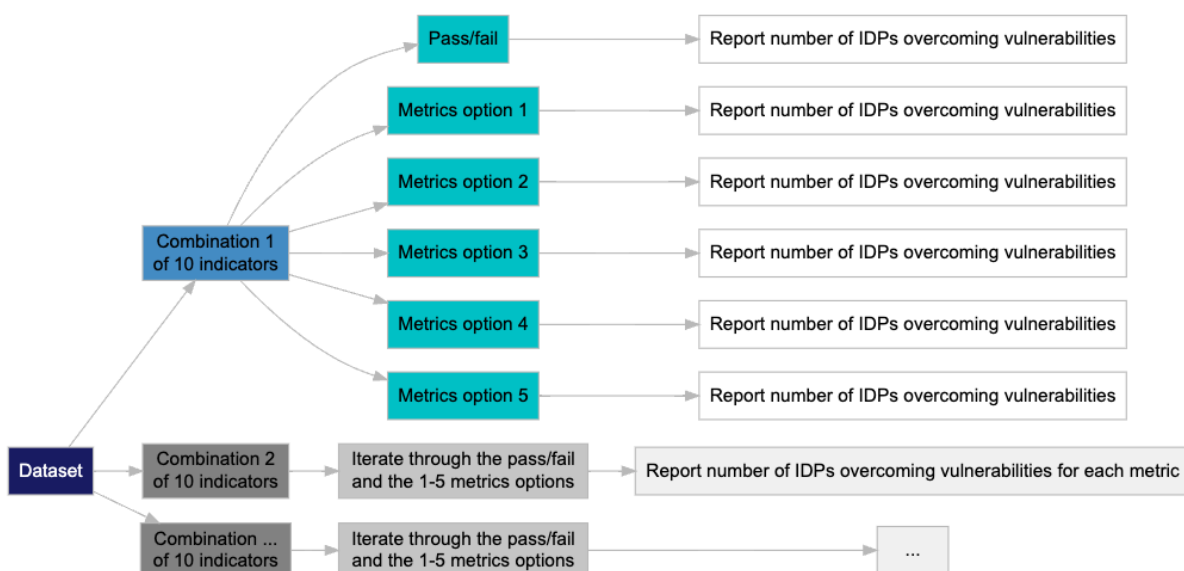
Challenge 8: Vulnerabilities overcome merely by assistance received

Challenge 9: Addressing changing benchmark values over time

Methodology

20. This paper provides actionable suggestions for overcoming the above challenges and aims to substantiate all suggestions with empirical evidence. Given that - for now - the exit measure is not yet statistically workable, the empirical evidence on effects of any proposal on the aggregate IDP stock cannot be based on a single “exit measure estimate” because this does not yet exist. To overcome this, we took stock of all potential metric choices and all potential indicator choices and ran iterative simulations of all combinations thereof (see Figure 1). We selected four empirical contexts (Hargeisa, Nigeria, Colombia, Sudan). In each dataset, 1000 possible combinations of 10 indicators (one for each of the sub-criteria) were assessed using all six metric options outlined further below. The results - both in terms of the mean and distribution of exits from the stock across combinations - are highly indicative and insightful for understanding the effect of a proposed solution on the aggregate exits from the IDP stock, without having to decide for one or the other option. Annex I provides further details on the simulation methodology.

Figure 1: Simulation approach



Overcoming challenge 1: Metric choice for benchmarking against comparator population

21. A necessary next step towards an applicable metric is to address how the various sub-criteria are combined and then compared to the national population or the host community in a statistically applicable way (“challenge 1” in the above list). 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 exit measure, different metric options - approaches to combine indicators and compare them with a benchmark population - are available to develop the IRIS exit measure.

Metric options

Pass/fail measure: Implementing the IRIS without a comparator population ²

22. 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 an exit measure, which is that it should be assessed relative to a host/national benchmark.

Option 1: A composite metric across all criteria

23. One option to explore that aligns with the contextualized approach would be to redefine the overall framework as an actual composite index. This would mean adding up all 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).³

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

	Sub-criterion	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

Overall: The household has not overcome displacement as the composite index is below the benchmark.

² 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 an option going forward.

³ One way to address this shortcoming could be to create a “hybrid” composite measure. For example, one could imagine a full composite index for the criteria 2, 3, and 5 while the more rights-based criteria 1 (freedom to move) and 4 (property restitution) are scored a pass or a no pass.

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

24. 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.
25. 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 1 (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-criterion	Indicator	Composite	Comparison to benchmark (population average)	
Household A	1. Safety and security	1.1 Victims of violence	Indicator 1.1.a	2/3	1.9 (Pass because 2 > 1.9)	
			Indicator 1.1.b			
		1.2 Freedom of movement	Indicator 1.2.a			
	2. Adequate standard of living	2.1 Food security	2.2 Shelter and housing	Indicator 2.1.a	3/4	3.5 (No pass because 3 < 3.5)
				Indicator 2.1.b		
				Indicator 2.2.a		
				Indicator 2.3.a		
	2.3 Medical services	2.4 Education	Indicator 2.4.a			
			Indicator 2.4.a			
	3. Access to livelihoods	3.1 Employment and livelihoods	3.2 Economic security	Indicator 3.1.a	2/3	1.9 (Pass because 2 > 1.9)
				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 because 1 > 0.9)
				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 because 0 < 0.9)	
			Indicator 5.1.b			
			Indicator 5.1.c			

Overall: The household has not overcome displacement-related vulnerabilities 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

26. 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.

27. 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 1 (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 value was, say, 0.1 or 0.9 – in either case the household in question would need to score at least a composite value of 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 relatively burdensome for widespread application.

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

	Sub-criterion	Indicator	Composite		Comparison to benchmark
Household A	1.1 Victims of violence	Indicator 1.1.a	1	3	2.8 (Pass)
		Indicator 1.1.b	1		
		Indicator 1.1.c	1		
	1.2 Freedom of movement	Indicator 1.2.a	1	3	2.5 (Pass)
		Indicator 1.2.b	1		
		Indicator 1.2.c	1		
	2.1 Food security	Indicator 2.1.a	1	2	1.1. (Pass)
		Indicator 2.1.b	0		
		Indicator 2.1.c	1		
	2.2 Shelter and housing	Indicator 2.2.a	1	2	1.9 (Pass)
		Indicator 2.2.b	0		
		Indicator 2.2.c	1		
	2.3 Medical services	Indicator 2.3.a	1	3	3 (Pass)
		Indicator 2.3.b	1		
		Indicator 2.3.c	1		
	2.4 Education	Indicator 2.4.a	1	1	2.5 (Pass)
		Indicator 2.4.b	0		
		Indicator 2.4.c	0		
	3.1 Employment and livelihoods	Indicator 3.1.a	0	0	1.5 (Pass)
		Indicator 3.1.b	0		
		Indicator 3.1.c	0		
	3.2 Economic security	Indicator 3.2.a	0	1	2.9 (Pass)
		Indicator 3.2.b	0		
		Indicator 3.2.c	1		
4.1 Property restitution and compensation	Indicator 4.1.a	1	3	2.5 (Pass)	
	Indicator 4.1.b	1			
	Indicator 4.1.c	1			
5.1 Documentation	Indicator 5.1.a	0	2	1.8 (Pass)	
	Indicator 5.1.b	1			
	Indicator 5.1.c	1			
Overall: The household has not overcome displacement-related vulnerabilities as some composite indices on the sub-criterion level are lower than the benchmark.					

Option 4: Comparison of homogeneous cells:

28. 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 proportion achieving a pass 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 outperforms the benchmark value, all IDPs in this group are taken out of the stock of IDPs. If the cell does not outperform the benchmark value, all IDPs remain as IDPs in the stock.

29. The advantage of this option is that it is the only one to truly overcome challenge 1 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 homogenous cells

	Sub-criterion	Indicator	Percentages in cell Z	Benchmark comparison
Cell Z composed of different IDP households (e.g. based on current location, area of origin and year of arrival)	1.1 Victims of violence	Indicator 1.1.a	29% of cell Z have positive outcome	28% of population have positive outcome
	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)
Overall: 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.				

Box 2: The role of qualitative data collections in metric option 4

Using key-informant interviews and qualitative data collections for the exit measure

Many data collections in contexts of forced displacement use key informant interviews, focus groups and other qualitative strategies. These data collections, such as the Displacement Tracking Matrix by the International Organization for Migration, are crucial data sources for humanitarians and rapid response teams. However, for the production of official statistics on IDP exits from the stock a representative sample of IDP households is required that can be compared with national or host community averages. Nevertheless, data on the community-level, as often generated through key informants, can be a valuable step in the transition to a fully comparable micro-level exit measure. In the homogeneous cell approach (Option 4), interviews with community leaders and other informants could in theory be used to identify if particular IDP subgroups and communities perform well on the 10 sub-criteria if no household-level or individual-level is available or data collections are infeasible. For example, information on the freedom to move may be informative on the community-level until household- and individual-level data is produced.

Option 5: Classifier/regression-based approach

30. Another option for the exit 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).

31. 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.

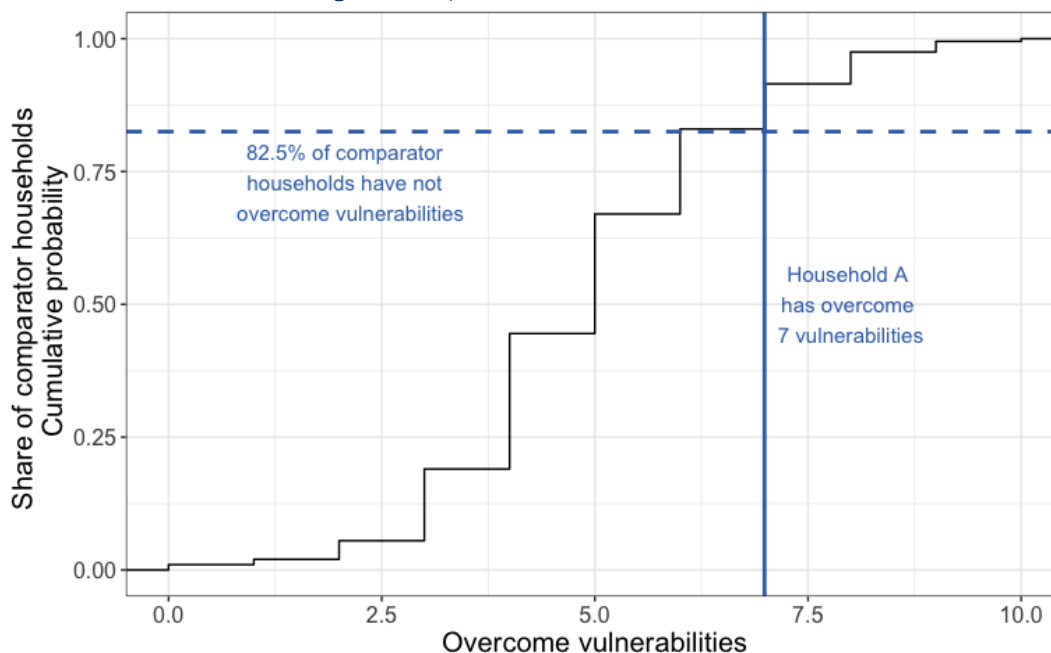
Table 7: Option 5: A classifier/regression-based approach

Sub-criterion	Indicator	Regression weights		
Household A		1 Pass		
	1.1 Victims of violence	Indicator 1.1.a		
		1 Pass	-1.58	
	1.2 Freedom of movement	Indicator 1.2.a	1 Pass	-0.04
	2.1 Food security	Indicator 2.1.a	1 Pass	1.59
	2.2 Shelter and housing	Indicator 2.2.a	1 Pass	-2.24
	2.3 Medical services	Indicator 2.3.a	0 No	0.48
	2.4 Education	Indicator 2.4.a	pass	4.48
	3.1 Employment and livelihoods	Indicator 3.1.a	0 No	6.72
	3.2 Economic security	Indicator 3.2.a	pass	2.13
	4.1 Property restitution and compensation	Indicator 4.1.a	0 No	4.32
	5.1 Documentation	Indicator 5.1.a	pass	-0.22
			1 Pass	
			1 Pass	
	Predicted probability for household (e.g. logit transformed)		0.439	
Cut-off point		0.5		
Overall: The household has not overcome displacement as the predicted probability of being similar to the host community is below the cut-off point.				

Option 6: Empirical cumulative distribution approach

32. Another approach is to focus on the multidimensional empirical cumulative distribution (eCDF). In line with IRIS, the general idea remains that an IDP household A is less vulnerable than another household if household A has overcome all vulnerabilities that the comparison household has overcome. The empirical cumulative distribution then describes the share of households that are more vulnerable than a given household. For example (also conceptually displayed in Figure 2), imagine that a household A has overcome 7 of the 10 displacement-related vulnerabilities. Given the distribution of vulnerabilities in the comparator population, we can say that around 82.5% of the comparator households have not yet overcome all of these vulnerabilities. The 82.5% are hence the share of host community households that are more vulnerable than the example IDP household A. This share also describes the probability that the IDP household A is less vulnerable than a “randomly selected” or average household. In the example, the probability that household A is less vulnerable than the average comparator household is high, with over 80%.
33. While previous metric options provide a clear-cut decision whether a specific household exits the stock or not, this metric generates a probability that a given IDP household is less vulnerable than the average comparator household. To then calculate exits from the stock for aggregated statistics, one can calculate the average (weighted) probability of overcoming displacement-related vulnerabilities across all IDP households in comparison to the host community. Although conceptually more complex, the metric option incorporates the logic of IRIS by comparing IDPs to the comparator population while the requirement of achieving a pass on all indicators is implicitly incorporated into the definition of the multivariate cumulative distribution function.

Figure 2: Empirical cumulative distribution



34. Box 3 summarizes the main advantages and disadvantages of the different metrics options. The paper 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).

Box 3: Main advantages and disadvantages of different metric options

1. Composite metric across all criteria	✗	Ignores principle that a pass is required in each sub-criterion
	✓	Low amount of indicators required
2. Composite metric at the criterion level	●	Relaxes principle that a pass is required to the criterion level
	●	Medium amount of indicators required
3. Composite metrics at the sub-criterion level	✓	Complies with principle that a pass is required in each sub-criterion
	✗	Very high amount of indicators required
4. Homogenous cells	✓	Complies with principle that a pass is required in each sub-criterion
	✓	Low amount of indicators required
	✗	Adds methodological complexity to define homogenous cells
5. Classifier/ regression-based	✗	Ignores principle that a pass is required in each sub-criterion
	✓	Low amount of indicators required
	✗	Analysis requires familiarity with regression-based analysis
6. Empirical cumulative distribution	✓	Complies with principle that a pass is required in each sub-criterion
	●	Implements probability rather than strict decision
	✗	Complexity, requires familiarity with eCDF

Simulation results for Hargeisa (UNHCR 2015)

35. The IDP profiling in Hargeisa covered a total of 939 IDP households that could be compared directly to their hosts. In the following sections, we simulate how many IDP households would exit the stock if we apply the different metric options 1-6 on different indicators of vulnerability. As a reference, a pass/fail measure that does not compare IDPs with comparator households exits on average 0.5 IDP households from the stock in Hargeisa. On average across all simulated indicators and 6 metric options, 4.99 IDP households overcome their vulnerabilities and exit the IDP stock. This corresponds to an average of 0.53% of the sampled IDP population in Hargeisa. The low exit numbers are largely a result of the fact that IDP households could not be assessed as the IRIS recommends in paragraph 168 that no assessment should take place if data is missing on at least one of the 10 sub-criteria (on average this applies to 90.79% of households across the indicator combinations). This report will discuss the challenge of missing data in subsequent sections.
36. Table 8 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. Overall, very few exits from the stock are possible in the Hargeisa context. As elaborated later, this is to some extent an effect of data missingness and the available indicators for *4.1 Property restitution and compensation*, which are hard to pass for IDP households. It should be noted that a composite measure at the sub-criterion level yields no exits from the stock, independent of the chosen indicators. A full composite index across all 10 sub-criteria and a regression-based approach generate the highest possible exits from the IDP stock, dependent on the indicator set chosen.

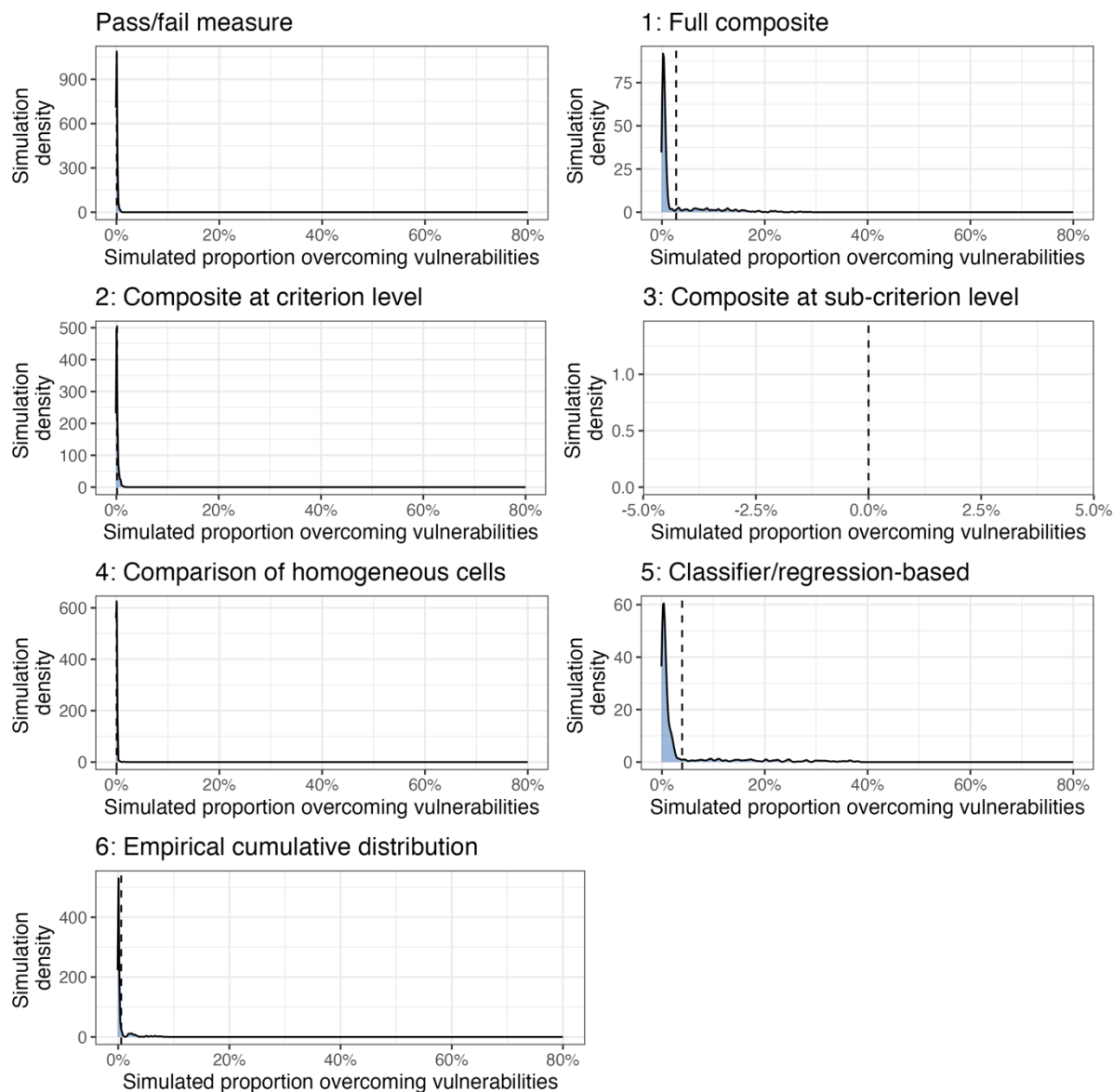
Table 8: Simulation results for Hargeisa across all metric (Total sample of IDPs: 939 households)

	Mean of IDPs exiting the stock	Range of IDPs exiting the stock	Variation in how many IDPs exit the stock
Pass/fail measure (no comparison!)	0.5 IDPs (0.05% of stock)	0 to 8 IDPs (0 to 0.85% of stock)	Very low (SD: 1.1)
1: Full composite	25.86 IDPs (2.75% of stock)	0 to 270 IDPs (0 to 28.75% of stock)	Low (SD 48.38)
2: Composite at criterion level	1.29 IDPs (0.14% of stock)	0 to 17 IDPs (0 to 1.81% of stock)	Very low (SD: 2.13)
3: Composite of homogenous cells	0 IDPs (0% of stock)	0 to 0 IDPs (0 to 0% of stock)	No variation (SD: 0)
4: Comparison of homogenous cells	0.24 IDPs (0.03% of stock)	0 to 107 IDPs (0 to 11.4% of stock)	Very low (SD: 1.93)
5: Classifier/regression-based	36.85 IDPs (3.92% of stock)	0 to 358 IDPs (0 to 38.13% of stock)	Low (SD: 74.81)
6: Empirical cumulative distribution	5.08 IDPs (0.54% of stock)	0 to 82 IDPs (0 to 8.68% of stock)	Low (SD: 12.76)

37. Figure 3 depicts the distribution of the simulation results. The figure displays the density of the conducted simulations for the 6 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)



38. These density plots corroborate what was discussed above. Most simulations - independent of the metric or indicators chosen - result in 0 IDP households overcoming their vulnerabilities and exiting the stock. Because of their high threshold to exit IDPs from the stock, they effectively behave like the pass/fail measure that does not compare IDPs with a comparator population. Making a regression-based assessment (option 5) 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 higher or lower but the difference is marginal in the case of the Hargeisa data.
39. The finding that very few IDP households can exit the stock can be traced back to three challenges that will be discussed in later sections: *Data missingness* for individuals and households on relevant indicators can lead to an inability to assess large proportions of the IDP population because IRIS recommends that no assessment should take place and they should remain in the stock. Second,

the *aggregation of individual-level data* to the household level (e.g., employment) can exacerbate the problem of missing data and creates ambiguity as to how to do the aggregation in a reasonable way. Third, the assessment of all 10 sub-criteria and the comparison to hosts is a *very high benchmark* to pass for IDPs in Hargeisa, in particular for the sub-criterion *4.1 Property restitution and compensation*.

Simulation results for Nigeria (World Bank 2018)

40. 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, 10.18 thousand IDP households exit the IDP stock, which corresponds to an average of 7.87% of the IDPs represented by the profiling. For an average of 13.42% 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.
41. In Nigeria, the different metric options produce varying estimates for the exit from the IDP stock. Using a regression-based approach or one composite metric across all criteria produce the highest number of exits from the IDP stock and are the most sensitive to the indicators used while other approaches show little variation. Assessing exits from the IDP stock through the empirical cumulative distribution also generates a higher exit from the IDP stock. Table 9 summarizes the key findings related to variation across indicators and metrics. Most importantly, we find very little variation in the number of IDPs exiting the stock for the composite measure at the sub-criterion level and for homogeneous cells.
42. Regarding metric option 4, different groupings into cells in Nigeria were tried based on date of arrival, date of displacement, and the origin and displacement location of IDPs. The simulations for both Hargeisa and Nigeria show that the variable used to group the IDP population into homogeneous cells to then make a group-level assessment against the comparator yields little volatility regarding the variables used for defining the cells. Details on the different groupings of IDPs into cells can be found in Box 4.
43. Figure 4 corroborates the above findings and displays the density of simulation outcomes for the 6 different metric options that could be used to implement the IRIS exit measure in practice. The options are compared to a simple pass/fail measure which does not implement a comparison to host communities. 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. 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 35.09 percent of the IDP population assessed in Nigeria may have overcome their displacement-related vulnerabilities according to this measure. For a regression-based approach or the cumulative distribution, the assessment yielded a bimodal distribution.

Table 9: Simulation results for Nigeria across all metrics (Weighted IDP households: 129.41 thousand)

	Mean of IDPs exiting the stock	Range of IDPs exiting the stock	Variation in how many IDPs exit the stock
Pass/fail measure (no comparison!)	609 IDPs (0.47% of stock)	0 to 7501 IDPs (0 to 5.8% of stock)	Low (SD: 1184.25)
1: Full composite	45406 IDPs (35.09% of stock)	35241 to 53582 IDPs (27.23 to 41.41% of stock)	High (SD: 3454.31)
2: Composite at criterion level	1238 IDPs (0.96% of stock)	0 to 6259 IDPs (0 to 4.84% of stock)	Low (SD: 1706.53)
3: Composite of homogenous cells	518 IDPs (0.4% of stock)	0 to 1200 IDPs (0 to 0.93% of stock)	Very low (SD: 435.52)
4: Comparison of homogenous cells	60 IDPs (0.05% of stock)	0 to 923 IDPs (0 to 0.71% of stock)	Very low (SD: 145.45)
5: Classifier/ regression-based	34149 IDPs (26.39% of stock)	1251 to 72925 IDPs (0.97 to 56.35% of stock)	Very high (SD: 24919.7)
6: Empirical cumulative distribution	10083 IDPs (7.79% of stock)	86 to 24630 IDPs (0.07 to 19.03% of stock)	High (SD: 7639.75)

Box 4: Identifying IDP subgroups to implement the homogenous cell approach

How should the IDP population be split into homogenous 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 as the size of cells, the homogeneity within IDP cells and the heterogeneity across IDP cells could affect how many IDP households exit the stock.

Approach I: Grouping variables in the data to split the IDP population: First, we selected three grouping variables from each of the datasets based on household characteristics - for example the gender of the household head, the location of origin, or the time of displacement - and then we grouped the IDPs into subgroups based on these variables to then calculate the average scores that can be compared to the host community. In the simulations, we can do this iteratively for multiple possible grouping variables (see Annex I for details) in all four displacement contexts. The results suggest:

- **The grouping of the IDP population into cells did not strongly affect the how many IDPs exit the stock.**
- The reason for this is that **achieving a pass on all group averages for all sub-criterion is a hard benchmark to pass.**
- 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 while this is unlikely for bigger groups (e.g., by gender or region of origin).
- It is recommended that grouping variables yield similarly sized groups.

Example differences across groupings in Nigeria:

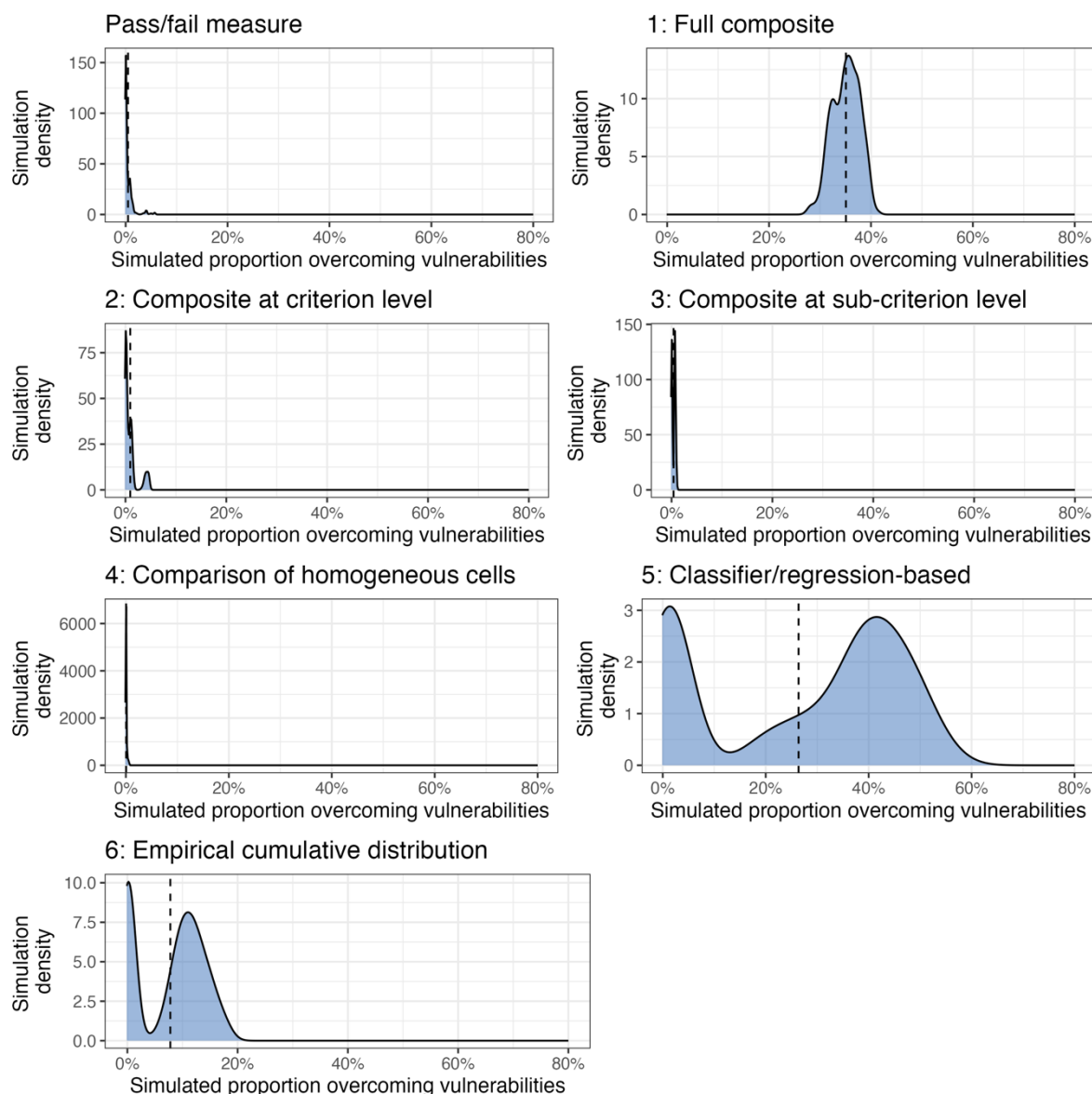
Group variable 1	Group variable 2	Group variable 3	Average number of IDPs exiting the stock	Average percentage of IDPs exiting the stock	Range and mean of group size
Year of arrival	Region of origin	Region of displacement	43.99	0.03	3 to 46270 (~1221)
Year of displacement	Year of arrival	Region of origin	72.29	0.06	12 to 45057 (~1294)
Year of displacement	Year of arrival	Region of displacement	73.37	0.06	3 to 49797 (~1407)
Year of displacement	Region of origin	Region of displacement	49.59	0.04	3 to 43390 (~1362)
Algorithmic clustering			328.35	0.25	

Approach II: Algorithmic clustering: Second, we employed automatic hierarchical clustering algorithms to group the IDPs into groups based on their performance on the different indicators. 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 data-driven groupings yield IDP cells that are most homogeneous within the cells and most heterogeneous in comparison to other cells. The results suggest:

- **A data-driven cell approach could potentially increase the exit from the IDP stock.** 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.**

In the example case of Nigeria, algorithmic clustering leads to an average of 328.25 IDPs exiting the stock (0.25%). This is higher than the mean exit from the IDP stock when using core demographics as groupings (only around 59.81 IDPs or 0.05% of the stock exit).

Figure 4: Density of simulations for all possible metrics (Nigeria)



Simulation results for El Fasher (Sudan 2019)

44. 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 survey weights suggesting that the sample is representative of 18.13 thousand IDP households. On average across all simulated indicators and metrics, 145.49 IDP households exit the IDP stock (0.8% of the IDPs represented by the profiling).⁴ An average of 82.18% of the IDP households could not be assessed due to missing data on at least one dimension.
45. Table 10 summarizes the key findings related to variation across indicators and metrics. The overall pattern is the same as in other country examples: While a composite index at the sub-criterion is a high standard to pass and no IDPs exit the stock, option 1 and 5 lead to the most substantial exits from the stock. Very few IDPs exit the stock in Sudan if a homogeneous cell approach is used (0.02% of the stock). Around 0.85% of the IDP stock exit if a composite is built at the criterion level,

⁴ Note that the indicator for sub-criterion 4.1. used in Sudan is security of tenure as other HLP indicators have not been asked to the host community which makes comparisons not applicable.

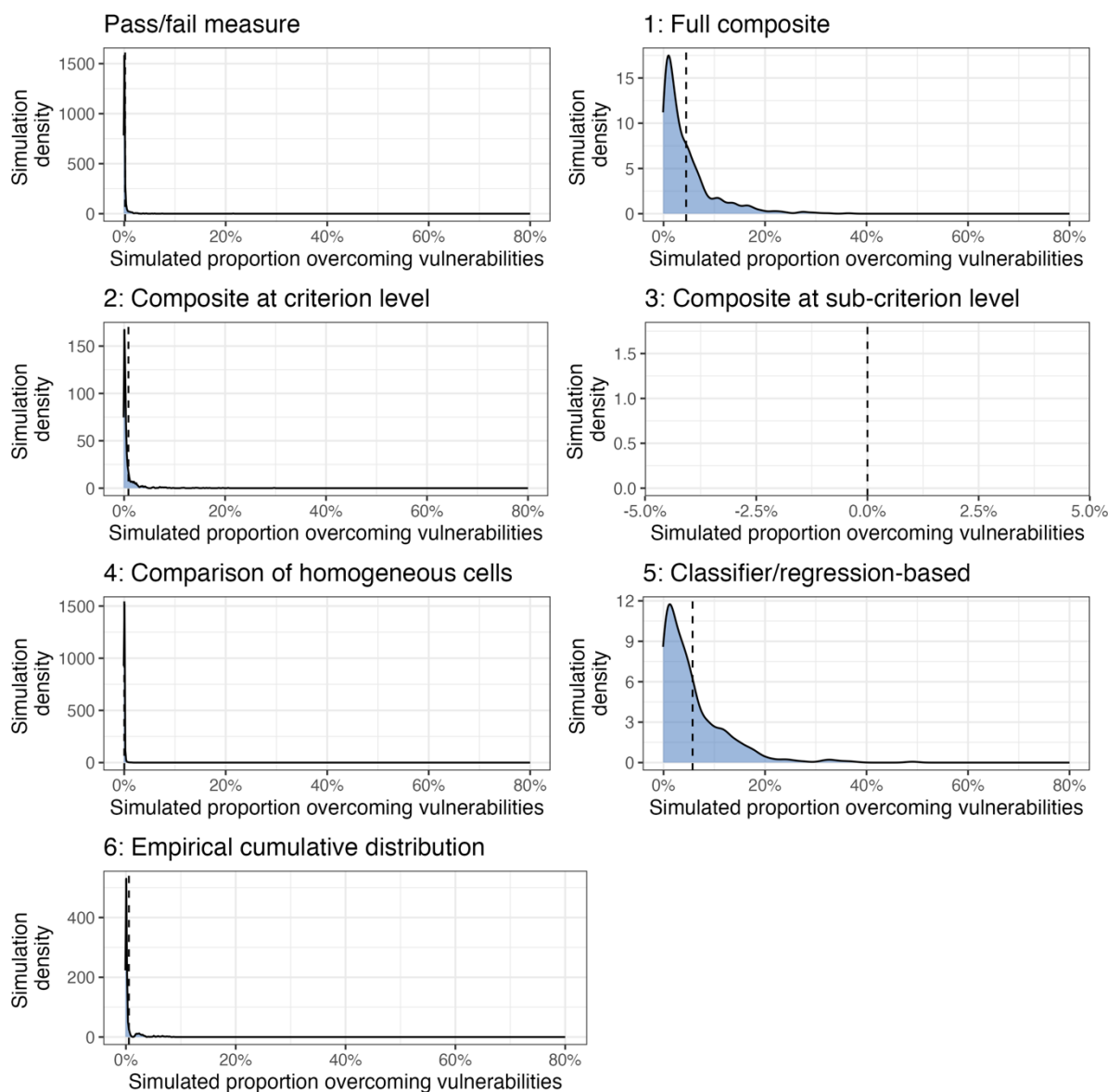
relaxing the IRIS requirement. 0.83% of the IDP stock exit according to the empirical cumulative distribution.

Table 10: Simulation results for Sudan across all metrics (Weighted IDP households: 18.13 thousand)

	Mean of IDPs exiting the stock	Range of IDPs exiting the stock	Variation in how many IDPs exit the stock
Pass/fail measure (no comparison!)	32 IDPs (0.18% of stock)	0 to 3891 IDPs (0 to 21.47% of stock)	Low (SD: 157.55)
1: Full composite	807 IDPs (4.45% of stock)	0 to 6604 IDPs (0 to 36.44% of stock)	Very high (SD: 944.37)
2: Composite at criterion level	155 IDPs (0.85% of stock)	0 to 5386 IDPs (0 to 29.72% of stock)	Low (SD: 421.4)
3: Composite of homogenous cells	0 IDPs (0% of stock)	0 to 0 IDPs (0 to 0% of stock)	No variation (SD: 0)
4: Comparison of homogenous cells	3 IDPs (0.02% of stock)	0 to 652 IDPs (0 to 3.82% of stock)	Very low (SD: 18.96)
5: Classifier/regression-based	1036 IDPs (5.72% of stock)	0 to 8898 IDPs (0 to 49.09% of stock)	Very high (SD: 1159.04)
6: Empirical cumulative distribution	151 IDPs (0.83% of stock)	0 to 4713 IDPs (0 to 26% of stock)	Low (SD: 370.69)

46. Figure 5 displays the performance of the different metric options in density plots. The plots make clear that - in the case of Sudan - many metric options hardly yield any exits from the IDP stock because the benchmark to pass all sub-criteria is very high. Only full composite indices and regression-based approaches classify a substantial proportion of the IDP households as having overcome displacement-related vulnerabilities.

Figure 5: Density of simulations for all possible metrics (Sudan)



Simulation results for Colombia (DANE 2019)

47. The LSMS in Colombia covered a total of 4096 households that have fled from violence or natural disaster (weighted number of households: 1145.42 thousand IDP households). Recognizing the survey weights included in the survey, around 1.19% of the IDPs in Colombia exit the stock on average across all indicators and metrics. On average, 94.98% of the IDPs cannot be assessed due to missing data on at least one dimension.⁵
48. The importance of selecting the right metric to combine the different indicators is summarised in Table 11 and the density plots in Figure 6. Overall, the variation and standard deviation in how many IDPs exit the stock is higher in the case of Colombia. However, there are less differences in the distribution of the simulations across all possible metric options. The metric options follow a similar

⁵ 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.

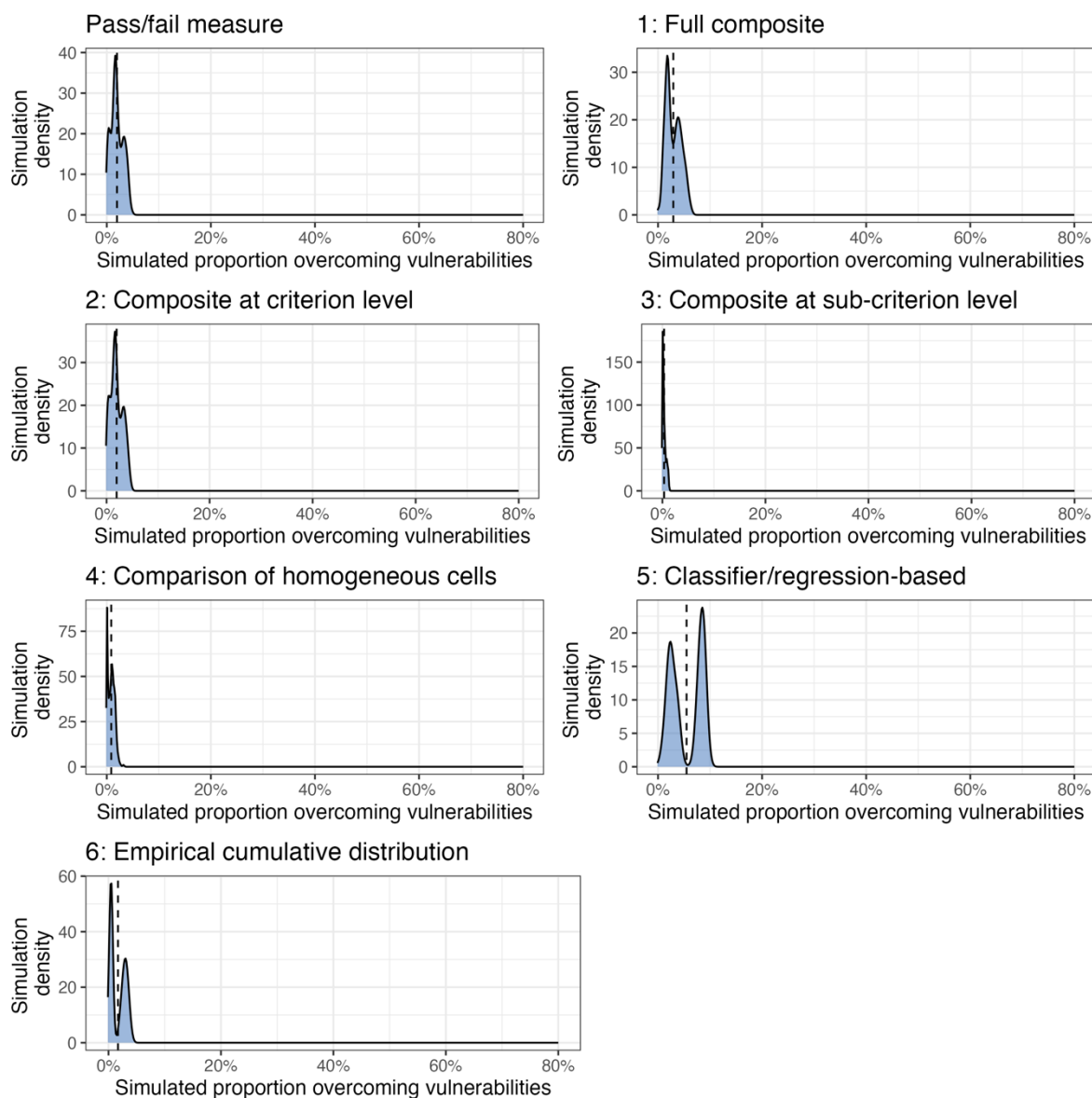
pattern as in the other three cases: The most permissive metric options are again a full composite index, a regression-based approach and a cumulative distribution approach. These approaches are the most sensitive to the chosen indicators and can produce a higher estimate of IDPs exiting the stock on 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 1.95 % 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 indices for each sub-criterion are often only binary. IDP households then only have to achieve a binary pass on each criterion (in total: 5 criteria) as opposed to a pass on at least 10 numerical values.

Table 11: Simulation results for Colombia across all metrics (Weighted IDP households: 7454.25 thousand)

	Mean of IDPs exiting the stock	Range of IDPs exiting the stock	Variation in how many IDPs exit the stock
Pass/fail measure (no comparison!)	55278 IDPs (1.96% of stock)	2246 to 141227 IDPs (0.08 to 5.01% of stock)	High (SD: 33228.07)
1: Full composite	82974 IDPs (2.94% of stock)	2246 to 182633 IDPs (0.08 to 6.47% of stock)	High (SD: 40404.27)
2: Composite at criterion level	55079 IDPs (1.95% of stock)	2246 to 135403 IDPs (0.08 to 4.8% of stock)	High (SD: 33649.81)
3: Composite of homogenous cells	9267 IDPs (0.33% of stock)	0 to 33624 IDPs (0 to 1.19% of stock)	Low (SD: 9410.01)
4: Comparison of homogenous cells	24059 IDPs (0.85% of stock)	0 to 98417 IDPs (0 to 3.49% of stock)	Low (SD: 18160.61)
5: Classifier/ regression-based	154791 IDPs (5.49% of stock)	32045 to 247877 IDPs (1.14 to 8.79% of stock)	Very high (SD: 83968.7)
6: Empirical cumulative distribution	42135 IDPs (1.69% of stock)	2382 to 110142 IDPs (0.1 to 4.42% of stock)	High (SD: 32751.39)

49. The density plots reflect this pattern and demonstrate that in the case of Colombia, a pass/fail measure 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. Overall, IDP exits are low.

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



Overall feasibility and application of metric options

50. In terms of practical application, some metric options turned out to be easier to implement than others. In most datasets, metric option 1 was the easiest 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.
51. Regarding the metric options 2 + 3, which both require multiple indices on the criterion and the sub-criterion level, the feasibility of doing this in practice is much lower. First, these metric options 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 3 of the 10 sub-criteria. For Nigeria, the dataset did not contain sufficient indicators to form indices for 7 out of 10 sub-criteria. For Sudan and Colombia, 6 sub-criteria were covered with only one or two possible indicators. It is to be expected that producers of official statistics may have challenges to generate the data necessary for these more “data hungry” metric options.
52. Secondly, 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 within such a composite are likely highly correlated. An example is that individuals living in overcrowded spaces very often also live in makeshift spaces. This multicollinearity of elements of an index raises the question 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 identify four distinct sub-criteria for the adequate standard of living of IDPs. It is not clear if we can theoretically justify more than one dimension to empirically measure other criteria, such as documentation. The questions around theoretical conceptualization, multicollinearity, and high data demands reduce the feasibility to implement metric options 2 and 3 in practice.

53. Dividing the IDP population in Hargeisa into homogeneous cells that can then be compared to the host community is also relatively easy to implement and is not expected to pose substantial difficulties for National Statistical Offices in terms of the analytical process. Results indicate that defining the boundaries between cells (i.e. the grouping variables to divide the IDP population) have only limited effect on the aggregated result (see Box 4). This is, on the one hand, encouraging, as EGRISS can suggest a degree of international standardization for the cell boundaries without this being complicated by heightened volatility in the aggregate. On the other hand, the exit from the IDP stock is minimal under this metric option. Because it is a hard bar to pass all 10 sub-criteria as a homogeneous IDP cells, the boundary issue is of secondary importance. If the criteria to assess passing are relaxed or reduced, the issue of drawing boundaries between IDP subgroups may re-emerge.
54. Using a regression-based approach has some disadvantaged 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 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 no longer be considered as part of the IDP stock. Another challenge in this approach is that comparisons across contexts will be difficult. While a regression in one country might assign a lot of weight to an indicator for 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. Finally, the implementation of this option can be more computationally complex than other options.
55. The last possible metric option is the empirical cumulative distribution approach. This approach is more complex in its theoretical conceptualisation, requiring some knowledge about probabilities and empirical distributions. The approach is slightly more computationally expensive than composite indices. On the other hand, focusing on the empirical cumulative distribution has no high data demands (i.e. not more than 10 indicators are needed) and the multidimensional empirical cumulative distribution implicitly fulfils the requirement of achieving a pass on all sub-criteria while upholding the IRIS principle of a comparison. Clear guidelines on the usage for National Statistical Offices could facilitate the practical implementation of this metric.
56. The results in Colombia demonstrate that the exit 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 exit measure can either compare IDP households to host communities or to national averages. The results presented here showcase that with a careful inclusion of all sub-criteria of the exit measure in the questionnaires of nationally representative datasets, and the inclusion of IDPs in their sampling frames, one could easily conduct an exit measure assessment without separate IDP profiling surveys. This puts further emphasis on the importance of recommending commonly used indicators for the exit measure - such as the

Sustainable Development Goals indicators - to increase the chances that those indicators being included in broader surveys to generate official statistics. Questions around HLP rights in the place of habitual residence and questions around the freedom to move, which are not commonly covered in large-scale national surveys, would need to be added.

Conclusions on challenge 1

Summary of findings

57. The assessment showed 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 12 summarizes these insights. The simulations demonstrate that the selection of a metric to implement the comparison between IDPs to a national/host community average is crucial and is the first and main problem to address as the 6 proposed metric options perform very differently. 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 (indicator choice will be discussed in more detail in the subsequent chapter).⁶
58. 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. This is largely because a composite metrics across all criteria violates the IRIS principle that all sub-criteria must be met and cannot guarantee that IDP households overcome all key displacement-related vulnerabilities.
59. Option 2, a composite measure at the criterion level, 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. Furthermore, EGRISS would have to provide further details on the weighting of different indicators and criteria. More evidence on the performance across different displacement contexts may also be needed.
60. Option 3, a composite measure at the sub-criterion level, 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. Option 3 is a very hard bar to pass for IDP households to be able to exit the stock.

⁶ We also investigated if the different simulations always identify IDP households as candidates to exit the stock or if the different metric options pick up different vulnerable groups. We calculate the Cohen's kappa coefficient as a statistic that measure "inter-rater reliability". Overall, we can observe that the full composite index and the index at the criterion level pick up similar IDP households as exits from the stock but most metric options differ substantially.

Table 12: Summary of metric advantages and disadvantages

	1: Composite metric across all criteria	2: Composite metric at criterion level	3: Composite metric at sub-criterion level	4: Homogenous cells	5: Classifier/regression-based approach	6: Empirical cumulative distribution
Aggregation of indicators to indices	Aggregation of indicators to indices	High-level aggregation	Medium-level aggregation	Low-level aggregation	No aggregation	No aggregation
Crudeness of (sub-)indices	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	Not applicable, no index
Importance of indicator choice	High to medium	Medium to low	No variation	Very low to no importance	High to very high	High to medium
Characteristics of the metric	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	Probability of exit rather than decision
Exit from IDP stock	Dependent on indicators	Low by design	Very low by design	Very low by design	Dependent on indicators	Dependent on indicators

61. The simulations also 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 (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 EGRISS for producers of IDP statistics on how to segment the population under analysis into smaller cells while maintaining cross-context comparability. Another disadvantage of the homogeneous cells approach is that the bar for IDP cells to exit the stock is very high by design, making it difficult to envision a practical implementation.
62. A regression-based approach classifying IDP households as being more similar or dissimilar with other IDP households and the host community 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. Application of this option might be challenging for some National Statistical Offices.

63. Option 6 refers to an assessment of the empirical cumulative distribution. This approach generates a probability of how likely the IDP household is less vulnerable than the average comparator household (and hence likely to exit the stock). This can easily be aggregated to the average probability of IDPs to be less vulnerable or the average share of IDPs to exit the stock for aggregated statistics. The application of this option might initially be challenging for some National Statistical Offices to work with the concept of an empirical cumulative distribution. Nevertheless, the option has demonstrated in the simulations that a reasonable number of IDPs exit the stock (although the metric is also sensitive to indicator choices) and that it fulfils both criteria of IRIS to ensure that all sub-criteria are met while a true comparison to the comparator population takes place.

Recommendation

64. It is suggested that EGRISS does not further pursue the option of a composite index at the sub-criterion level, which is not feasible in practical application and also - by its very design - allows almost no exits from the IDP stock. Likewise, a regression-based approach is not recommended (high volatility, capacity challenges for producers of official statistics) for further development of the exit measure. A full composite measure cannot currently be recommended as it is too far removed from the key principle that seeks to guarantee that IDPs overcome their vulnerabilities across all relevant dimensions before being taken out of the statistical stock.
65. A composite measure at the criterion level, while slightly relaxing one of the measure's key principles from the sub-criterion to the criterion level, constitutes a feasible way forward. The validity of this metric option will benefit greatly from standardizing the number and specification of the index elements (i.e. indicators) for each criterion. A homogeneous cell approach also constitutes a feasible way forward although the current implementation allows almost no exits from the stock by design. In addition to standardizing the specification of indicators, a degree of standardization in the methodology for grouping IDPs into such cells would also be required. Finally, focusing on a metric based on the empirical cumulative distribution could be a feasible way forward giving the metric's compliance with IRIS. Assessments whether this option is feasible for National Statistical Offices will be necessary as well as a standardization of the indicator selection.

Box 5: Validity and strictness of the IRIS exit measure

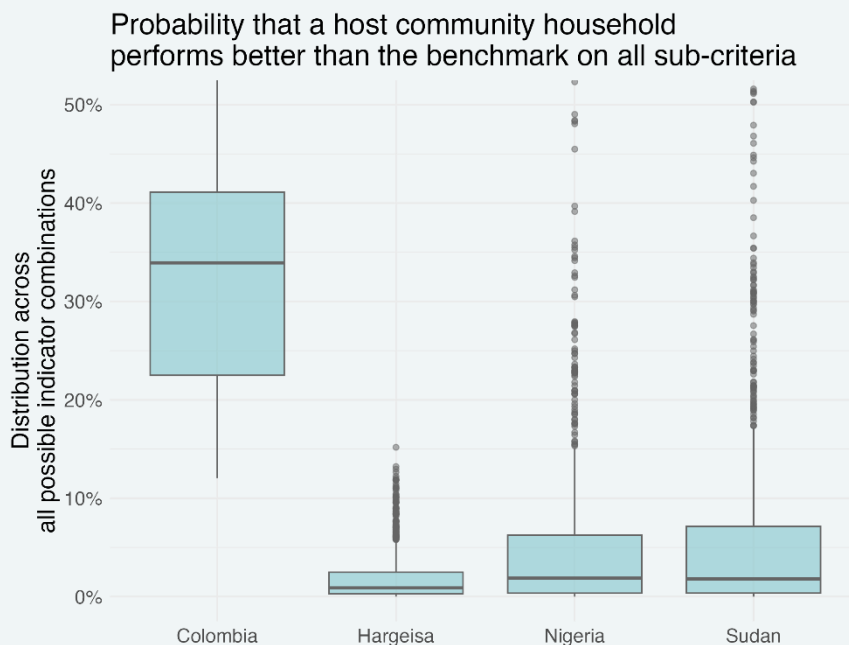
The IRIS exit measure as a "strict" measure of stock exits

In interpreting the results on challenge 1 and especially the very low number of exits from the stock under the metric options 3 and 4 - which are in full compliance with both the IRIS key principle for the exit measure - it is useful to remind oneself that one of the two key principles of the IRIS exit measure, namely having to perform equally or better than the average host *in each* of the ten sub-criteria to exit the stock, is a significantly stricter requirement than merely having to perform equally or better than the average host *across* the ten sub-criteria. In other words, the IRIS key requirement of achieving a pass in each sub-criterion **effectively means that IDPs need to significantly outperform the median host if they are to exit the IDP stock.**

A simple probabilistic argument can be used to illustrate this point. Assume, for example, that the host benchmark values for the first four indicators in the assessment framework were all 50%. Then the probability of a randomly selected host community household performing better than the host benchmark on these four indicators would be only 6.25% (= 50%* 50% * 50% * 50%). Extend the calculation to ten indicators and the probabilities become diminishingly small.

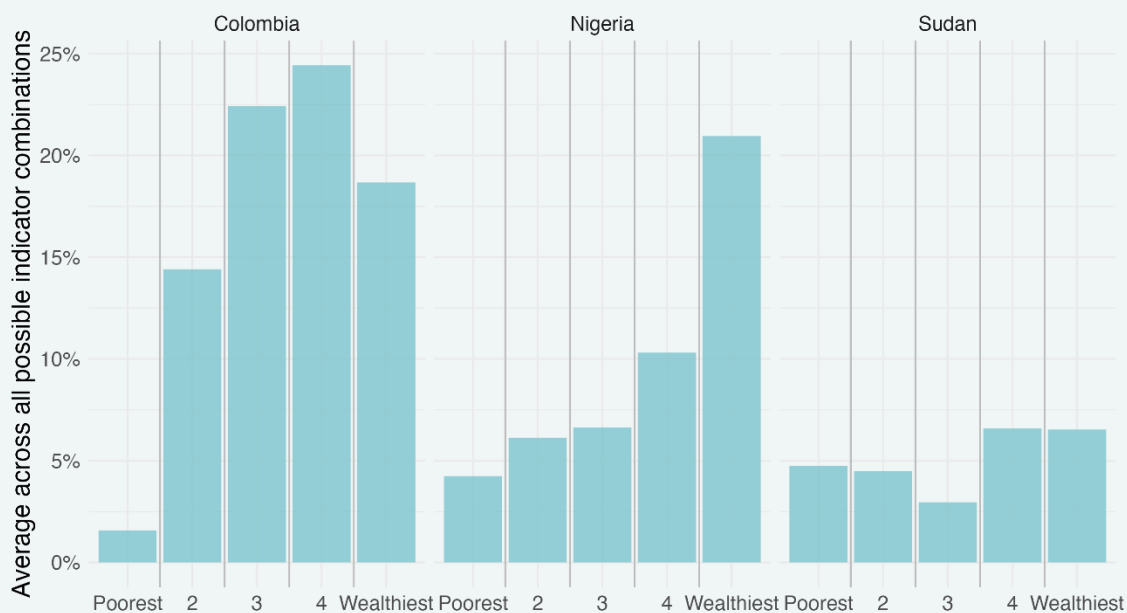
For the sake of further illustrating this point, we have applied a *role reversal scenario* in the available datasets, in which we test how many host community households are actually able to perform similarly or better than the host community benchmark in each sub-criterion. The graph on the left displays the probability that a host community household performs better than the benchmark on all sub-criteria for each country example. While the mean probability is at almost 40% for host community households in Colombia (see horizontal line in each box), the majority of households in Hargeisa, Nigeria, and Sudan have almost zero probability to outperform the benchmark.

We complement this assessment with information on households' wealth quintiles (for Colombia, Nigeria and Sudan). The figure on the right displays the share of the host communities that perform better than the benchmark for each country and for each wealth quintile for the metric option 2 (Composite measure at criterion level). For a valid exit measure, ideally the host communities in the wealthiest quintiles should exit the stock to a high percentage. However, less than 30% of the richest households outperform the average on all dimensions.



Share of host community households that perform better than the benchmark

By wealth quintile, for Option #2: Composite measure at the criterion level



To sum up, the low number of exits under the metric options compliant with the IRIS requirement of achieving a pass on every single sub-criterion is because the requirement is **difficult to pass by design**, not due to any inherent vulnerabilities among IDP households. While the need to outperform hosts in each sub-stock is clearly justified from a protection angle, the empirical implications in practice are that even IDPs that substantially outperform hosts in their living conditions do not exit the IDP stock.

Overcoming challenge 2: Indicators for the 10 sub-criteria

66. Another essential step to advance the exit measure into a fully workable tool is to identify the indicators that should be used to measure the 10 sub-criteria (“challenge 2” above). The choice of the right indicators to capture progress towards durable solutions can already be complex at the country level, let alone for EGRISS to make a recommendation at the global level. 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.
67. Nevertheless, a standardization proposed by EGRISS is key to ensure that IDP stock numbers are harmonized and comparable across displacement contexts. As the volatility of the exit measure under metric options 1 and 5 (full index and regression approach) has shown: the selection of the right indicators to compare IDPs and a benchmark population across the 10 sub-criteria can yield different results. Albeit being a challenging task, the way forward to make exits from the IDP stock comparable in different countries can hence only be to provide guidance on standardized indicators to measure the 10 sub-criteria.
68. To advance this debate on indicator selection, this paper follows two approaches. First, the paper uses the empirical simulations and demonstrates to what extent the choice of available indicators in the four country examples matters for the extent of IDP exit from the stock. This assessment is technical in nature and generates general information on a) the extent to which the indicator choice matters; and b) the statistical characteristics or quality standards that an indicator should fulfil to be suitable from a statistical perspective. Second, this paper theoretically reviews the *Durable Solutions Library* indicators and the *Sustainable Development Goals* indicators to identify if any of these indicators are suitable for the exit measure. The section ends with a recommended set of indicators to EGRISS.

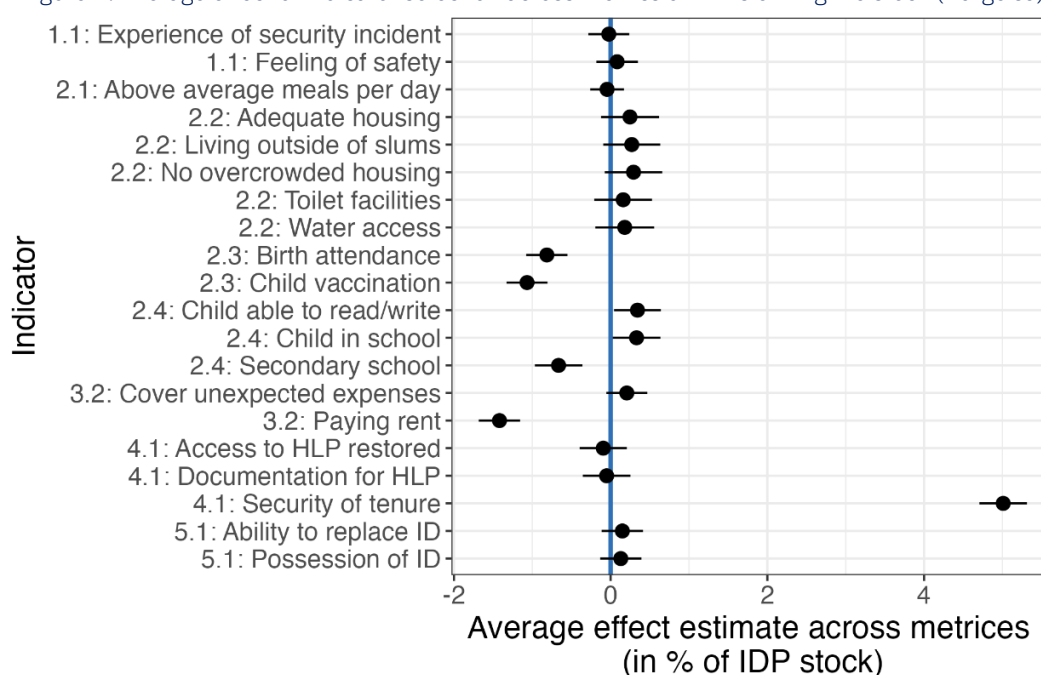
Simulation results for Hargeisa (UNHCR 2015)

69. Table 13 summarizes the effect of the indicator choice on aggregate outcomes, i.e. exits from the IDP stock for 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.
70. For example, choosing an indicator that counts whether an IDP household had more meals than the average as opposed to an indicator that covers the ability to pay for food as decreases the percentage of IDPs exiting the stock by an average of -0.05 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. Overall, the table is quickly summarized by the statement that none of the indicator choices make a discernible or substantial difference in the outcome (i.e. the number of IDP exits from the stock), except for the indicator choice regarding *4.1 Property restitution and compensation*. Using tenure security as opposed to other measures of property rights (e.g., restoration of access to property) yields higher estimates of the exit from the IDP stock.
71. To visualize the results of this indicator assessment, Figure 7 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 13). Overall, the simulations in the context of Hargeisa suggest that the average effect of choosing one indicator over the other is often negligible in practice with the exception of the indicator choice for measuring property restitution and compensation.

Table 13: Average relevance of indicator choice across metrics (Hargeisa)

Sub-criteria IRIS	Indicators in Hargeisa (UNHCR 2015)	Average effect size on stock	Description of variation
1.1 Victims of violence	Reporting of security incident (baseline), Experience of security incident, Feeling of safety	-0.02 to 0.08 percentage points	Hardly any discernible 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), Above average meals per day	-0.05 to -0.05 percentage points	Hardly any discernible difference between different indicators
2.2 Shelter and housing	Bathroom access (baseline), Adequate housing, No overcrowded housing, Toilet facilities, Water access, Living outside of slums	0.16 to 0.29 percentage points	Hardly any discernible difference between different indicators
2.3 Medical services	Access to medical services if needed (baseline), Birth attendance, Child vaccination	-1.07 to -0.81 percentage points	Hardly any discernible difference between different indicators
2.4 Education	Child ever attended school (baseline), Child able to read/write, Child in school, Secondary school	-0.66 to 0.34 percentage points	Hardly any discernible 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	-1.42 to 0.21 percentage points	Hardly any discernible difference between different indicators
4.1 Property restitution and compensation	Recompensation mechanisms accessible (baseline), Documentation for HLP, Access to HLP restored, Security of tenure	-0.09 to 5.01 percentage points	Some difference between different indicators for HLP rights
5.1 Documentation	Possession of birth certificate (baseline), Ability to replace ID, Possession of ID	0.13 to 0.15 percentage points	Hardly any discernible difference between different indicators

Figure 7: Average effect of indicator selection across metrics on IDPs exiting the stock (Hargeisa)



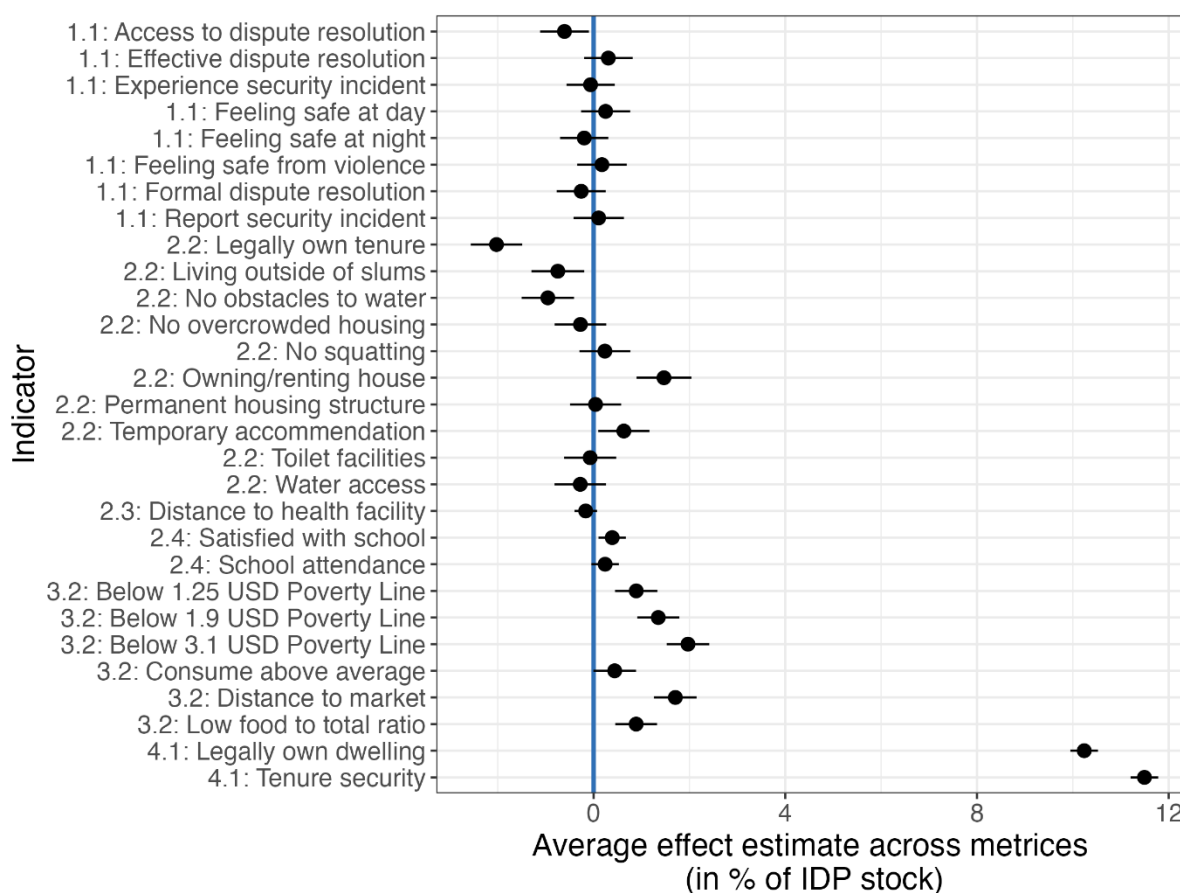
Simulation results for Nigeria (World Bank 2018)

72. 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 for Nigeria. In this dataset, 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 the indicator choice - similar to the Hargeisa case above - does not matter 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. The notable exception, however, is the indicator choice for *4.1. Property restitution and compensation*.
73. In the case of *4.1 Property restitution and compensation*, choosing an indicator that measures legally owning the dwelling in comparison to having access to compensation mechanisms for lost HLP is an indicator that sees more IDPs exiting the stock (10.2 percentage points more exits). The reason is that compensation mechanisms seem to be hardly present for IDPs in Nigeria but a substantial proportion seems to legally own dwelling in their place of origin.
74. To visualize the results, Figure 8 displays the average effect of choosing one indicator over the other averaged across all assessed metric options. 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 exit measure.

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

Sub-criteria 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), Feeling safe at day, Access to dispute resolution, Effective dispute resolution, Formal dispute resolution, Experience of security incident, Feeling safe at night, Reporting security incident, Feeling safe from violence	-0.61 to 0.31 percentage points	Hardly any discernible 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, Permanent housing structure, No squatting, Temporary accommodation, Toilet facilities, No obstacles to water, Water access, Legally own tenure, Living outside of slums	-2.03 to 1.47 percentage points	Some difference between different indicators
2.3 Medical services	Access to medical services if needed (baseline), Distance to health facility	-0.16 to -0.16 percentage points	Hardly any discernible difference between different indicators
2.4 Education	Duration to next education facility (baseline), Satisfied with school, School attendance	0.24 to 0.39 percentage points	Hardly any discernible 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 1.25 USD Poverty Line, Below 3.1 USD Poverty Line, Below 1.9 USD Poverty Line	0.44 to 1.97 percentage points	Some difference between different indicators
4.1 Property restitution and compensation	Access to recompensating mechanism for lost HLP (baseline), Legally own dwelling, Tenure security	10.24 to 11.40 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	

Figure 8: Average effect of indicator selection across metrics on IDPs exiting the stock (Nigeria)



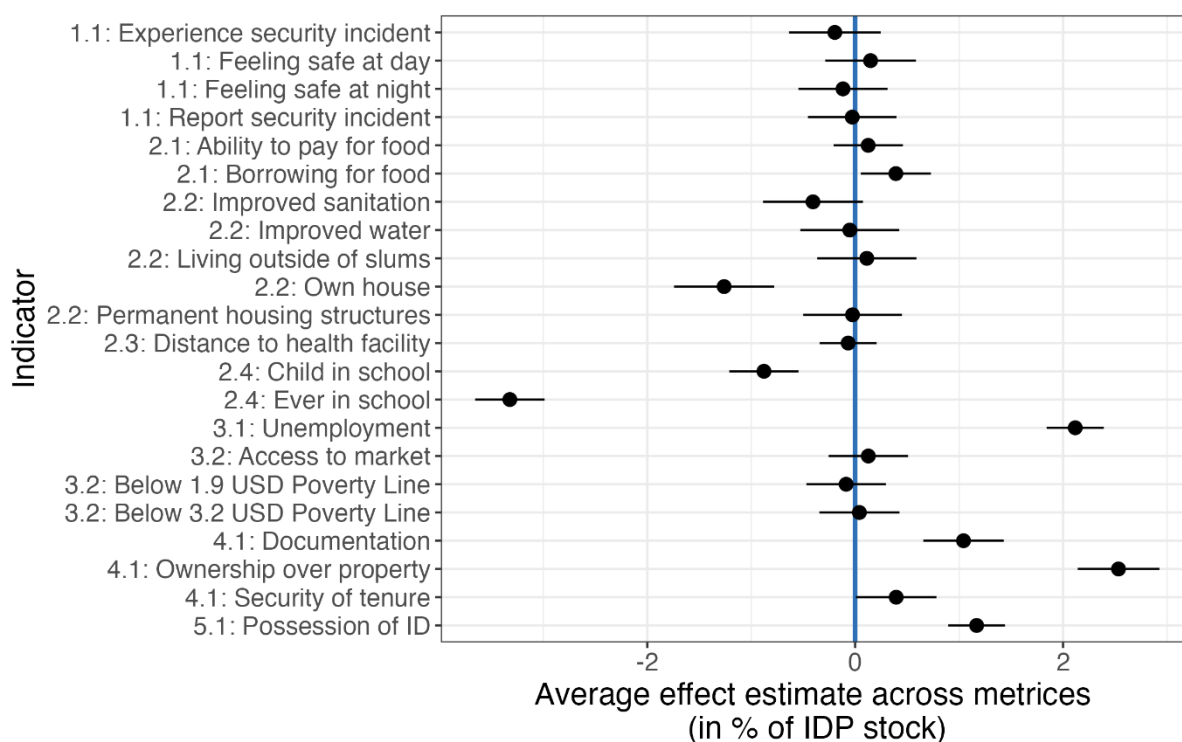
Simulations results for El Fasher (Sudan 2019)

75. Table 15 summarizes how much the choice of one indicator over the others affects the IDP stock across all simulations and all metric options for El Fasher (Sudan). The dataset in Sudan did not include an indicator for *1.2 Freedom of movement* but at least two indicators are available for any of the other sub-criteria. Overall, the finding is that the indicator choice does not affect the exits of IDPs from the stock very strongly across all sub-criteria. There are some percentage point changes in the range of 1 to 4 percentage points for the sub-criteria *2.2 Shelter and housing*, *2.4 Education*, *3.1 Employment and livelihoods*, *4.1 Property restitution and compensation*, and *5.1 Documentation* but the substantial difference is not strong.
76. For example, choosing to measure access to education with the question whether a child has ever been in school, reduces the exit from the IDP stock by -3.32 percentage points compared to measuring education with the distance to the next school. The second most relevant indicator in this selection is the question how to implement *4.1 Property restitution and compensation*. Again, Figure 9 displays the average effect sizes and gives an overview over the indicators that explain most variation in the data.

Table 15: Average relevance of indicator choice across metrics (Sudan)

Sub-criteria IRIS	Indicators in El Fasher (Sudan 2019)	Average effect size on stock	Description of variation
1.1 Victims of violence	Feeling of safety at day (baseline), Feeling safe at day, Experience security incident, Feeling safe at night, Report security incident	-0.2 to 0.15 percentage points	Hardly any discernible difference between different indicators
1.2 Freedom of movement	No available indicator	No indicator	
2.1 Food security	Borrowing for food (baseline), Borrowing for food, Ability to pay for food	0.13 to 0.39 percentage points	Hardly any discernible difference between different indicators
2.2 Shelter and housing	Overcrowded housing (baseline), Own house, Permanent housing structures, Improved sanitation, Improved water, Living outside of slums	-1.26 to 0.11 percentage points	Some difference between different indicators
2.3 Medical services	Distance to health facilities (baseline), Distance to health facility above average	-0.07 to -0.07 percentage points	Hardly any discernible difference between different indicators
2.4 Education	Distance to school (baseline), Child in school, Ever in school	-3.32 to -0.88 percentage points	Some difference between different indicators
3.1 Employment and livelihoods	Employment (baseline), Unemployment	2.12 to 2.12 percentage points	Some difference between different indicators
3.2 Economic security	Bank account (baseline), Access to market, Below 3.2 USD Poverty Line, Below 1.9 USD Poverty Line	-0.09 to 0.13 percentage points	Hardly any discernible difference between different indicators
4.1 Property restitution and compensation	Access to compensation (baseline), Documentation, Ownership over property, Security of tenure	0.39 to 2.53 percentage points	Some difference between different indicators
5.1 Documentation	Birth certificate (baseline), Possession of ID	1.17 to 1.17 percentage points	Some difference between different indicators

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



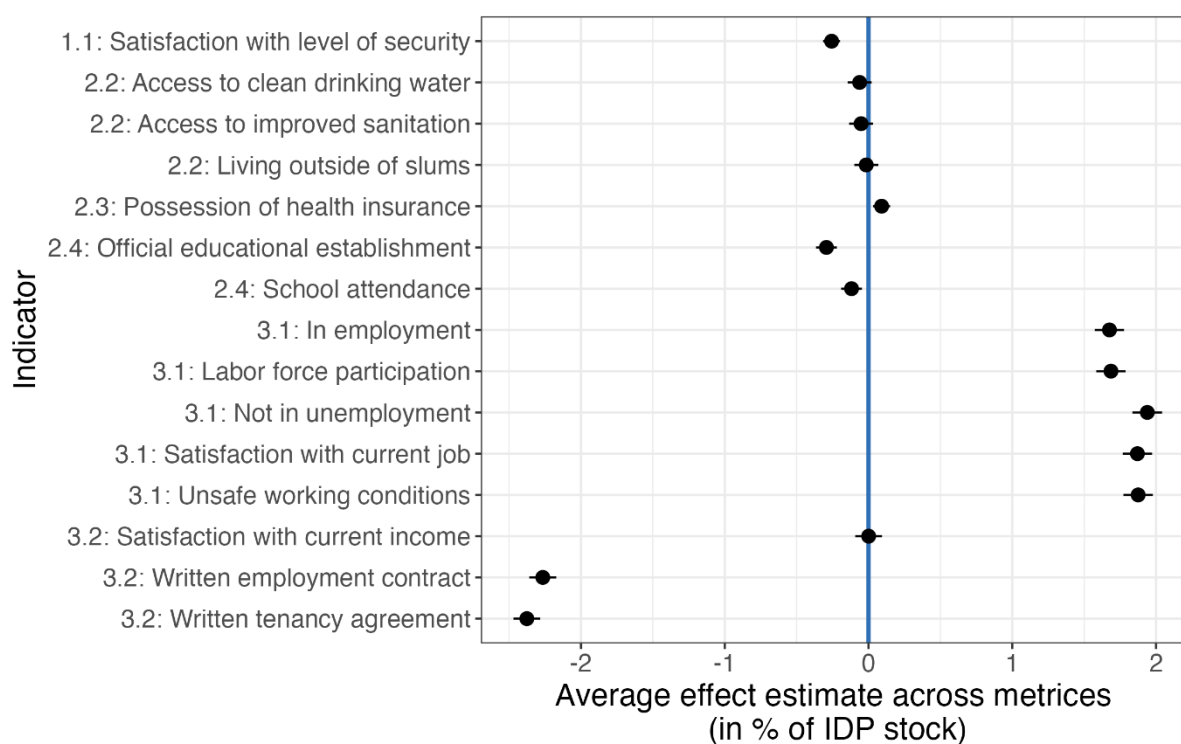
Simulation results for Colombia (DANE 2019)

77. 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 16 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.
78. Overall, the selection of indicators results in very limited differences in the exits from the IDP stock for *2.2 Shelter and housing, 1.1 Victims of violence, 2.3 Medical services* and *2.4 Education*. For the remaining 2 sub-criteria, the indicator selection matters a little bit more. However, this result should be interpreted with great caution, as it is driven by the large amount of missing data points for some indicators in this dataset rather than other properties of the respective indicators - and the fact that IRIS posits that IDP households cannot exit the stock if an assessment cannot be made in each sub-criterion (see the chapter on overcoming challenge 4 below for a more detailed discussion).
79. Figure 10 summarises the effect of the indicator selection again, but readers should keep in mind that these differences are strongly determined by data “missingness”. 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.

Table 16: Average relevance of indicator choice across metrics (Colombia)

Sub-criteria IRIS	Indicators in Colombia (DANE 2019)	Average effect size on stock	Description of variation
1.1 Victims of violence	Experience of a natural disaster (baseline), Satisfaction with level of security	-0.26 to /0.26 percentage points	Hardly any discernible difference between different indicators
1.2 Freedom of movement	No available indicator	No 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), Access to improved sanitation, Living outside of slums, Access to clean drinking water	-0.06 to -0.02 percentage points	Hardly any discernible difference between different indicators
2.3 Medical services	Possession of health insurance (baseline), Possession of health insurance	0.09 to 0.09	Hardly any discernible difference between different indicators
2.4 Education	Literacy (baseline), Official educational establishment, School attendance	-0.29 to -0.12 percentage points	Hardly any discernible 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	1.68 to 1.94 percentage points	Some difference between different indicators
3.2 Economic security	Defaulting on utility bills (baseline), Satisfaction with current income, Written employment contract, Written tenancy agreement	-2.38 to 0 percentage points	Some difference between different indicators
4.1 Property restitution and compensation	No indicator available	No indicator	
5.1 Documentation	Only available indicator was if IDPs had identification documents	No other indicator	

Figure 10: Average effect of indicator selection across metrics on IDPs exiting the stock (Colombia)



Selecting indicators for the exit measure

80. The simulations so far have shown that the indicator choice matters to a lesser extent than the selection of an appropriate metric. Nevertheless, the simulations have shown that data missingness, or non-applicability, can explain some of the variation in IDP exits from the stock. The subsequent sections propose an operationalization and indicator selection for each of the 10 sub-criteria, heavily relying on the *Durable Solutions Indicator Library* (DS Library) and the *Sustainable Development Goals* (SDG) indicators and discusses the implications of this choice. The IRIS postulate that “indicators selected in each national context should, as far as possible, be aligned with already tested and standardized indicators” (IRIS, p.45). The DS Library and the SDG indicators are referenced as useful resources in the IRIS because both collections include indicators that are “commonly used”, “tested and evaluated”, and “cover the population in question” (IRIS, p.45).⁷
81. The DS Library is one of the most comprehensive collections of often-used statistical indicators for durable solutions and 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. Given the involvement of many relevant partners, such as UNHCR, IOM, IDMC and JIPS, it serves as a useful starting point.
82. The SDG indicator framework has been developed with the involvement of many partners and agencies under the Inter-Agency and Expert Group on SDG Indicators and has been adopted as the global framework to monitor the implementation of the 2030 Agenda for Sustainable Development by the UN General Assembly. The SDG indicators cover a wide range of indicators to measure wellbeing and development with the advantage of providing standardized operationalizations of each indicator and a wide application across contexts. Using SDG indicators is also advantageous as benchmark data for hosts/national averages is widely available. Committed to making forcibly

⁷ For a review of available indicators for the exit measure, see “Statistical Measuring of Overcoming Internal Displacement-Related Vulnerabilities”, prepared for EGRISS IDP sub-group Solutions Working Group by Lauren Herby, 2019.

displaced persons visible in the reporting on sustainable development, EGRISS has identified 12 SDG indicators that should be disaggregated by displacement status as a priority.

Mapping and comparing possible indicators for each sub-criterion

1.1 Victims of violence:

83. In Hargeisa, Nigeria, Sudan, and Colombia the victimization of IDPs is measured with over 9 different indicators, ranging from capturing exposure to crime, harm, other victimizing events, satisfaction with the current security situation, to general feelings of safety. This wide range of potential indicators suggests a need to standardize across contexts. Two possible SDG indicators that also form a part of the DS Library plausibly measure sub-criterion 1.1.
84. **SDG indicator 16.1.4: Proportion of population that feels safe walking alone around the area they live** is one of the 12 priority SDG indicators recommended for disaggregation by forced displacement status. This indicator is broadly adopted in many household surveys across contexts and can be measured with a single question. This is a relevant advantage as it places a low data collection burden on NSOs and many questionnaires do not have to be adjusted to include this indicator. However, while the two can be expected to correlate, the SDG indicator measures the fear of violence rather than the actual experience of violence.
85. An alternative indicator could hence be **SDG indicator 16.1.3: Proportion of population subjected to (a) physical violence, (b) psychological violence and (c) sexual violence in the previous 12 months** as this indicator measures the reported experience of violence. Several methodological issues, however, arise from the fact that capturing psychological violence can be difficult in survey contexts. Additionally, questions on sexual violence are often too sensitive to include in a household survey that serves multiple purposes. For SDG reporting, this data is hence usually collected through specialized victimization surveys. The implementation of the exit measure is conceptualized for broad surveys by National Statistical Offices and including SDG indicator 16.1.3. appears infeasible in this context.
86. Beyond these two SDG indicators, the DS Library includes other operationalizations to measure the victimization of IDPs: This includes asking households how likely they will experience serious consequences due to armed conflict, other situations of instability or a hazard. However, this captures future expectations rather than actual experiences. Target populations could also be asked if they experienced certain types security incidents (e.g., abductions, child recruitments, attacks on property)/ hazards and if they reported relevant experiences to relevant authorities . However, a standardization based on the already specified SDG indicators seems appropriate to not have severe differences in the definition of security incidents or hazards. Reporting structures to different authorities also appear to be very context dependent.
87. **It is suggested that EGRISS recommends measuring sub-criterion 1.1. (“Victims of violence”) of the exit measure via the SDG Indicator 16.1.4.: Proportion of population that feels safe walking alone around the area they live.**

1.2 Freedom of movement:

88. The freedom to move is not commonly asked in many household/individual level surveys of displaced populations but rather inferred through legal analysis of relevant laws or through a community-level assessment. Only the questionnaires in Hargeisa and Nigeria included a question on the freedom to move. Additionally, the theoretical concept of what freedom to move entails can differ across contexts depending on whether IDPs live in camps (Are IDPs able to leave the camp or not) or out-of-camps (Are IDPs able to freely choose their place of residence).
89. In contrast to the first sub-criterion, there is no standardized indicator available in the SDG indicators that could capture the sub-criterion 1.2. The DS Library proposes to ask for any restrictions to the freedom to move at the individual, household and/or community level (**Indicator**

1.4.1 in the DS Library: Target population facing restrictions to their freedom to move). However, the exact formulation of this indicator is not standardized/implemented yet.

90. Nevertheless, some surveys on the household/individual level have tried to capture the freedom to move. The IDP Profiling in Hargeisa (UNHCR 2015), for example, uses these questions:
1. Do you have any problems visiting official and public places in Hargeisa?
 2. If yes, do you have any problems visiting ...
 - a. Religious places?
 - b. Locations for humanitarian assistance?
 - c. Municipal offices?
 - d. Social activities?
 - e. Market?
 - f. Other
 3. Please name the principal/secondary reasons why you have problems visiting.
91. The IDP Profiling in North-East Nigeria (World Bank, 2018) uses one question: Do you feel free to move in and out of this area whenever you choose?
92. **It is suggested that EGRISS recommends measuring sub-criterion 1.2. (“Freedom to move”) of the exit measure via the Durable Solutions Library Indicator 1.4.1 Target population facing restriction to their freedom. The indicator should be implemented with a simple household/individual level question (Do you feel free to move in and out of the area you currently live in whenever you choose?).**

Box 6: Community- and household-level variation in the freedom to move

Capturing the freedom to move in household surveys

The freedom to move is commonly collected through community-level assessments. However, for the exit measure it is recommended to move to the household-level. Ideally, a household-level question should demonstrate consistency in responses from people in the same area. This would imply that previous community-level analysis are valid and that a household-level data collection makes sense for the exit measure. In Nigeria, the recommended question to measure the freedom to move has already been implemented. The table below demonstrates how many households per community feel free to move and whether there is high deviation in responses. Overall, most households within a community give similar responses to the question whether they feel free to move.

Region of displacement	Feeling not free to move	Feeling free to move	Standard deviation
damawa	42	184	0.389
Bauchi	2	53	0.189
Borno	114	2312	0.212
Gombe	4	33	0.315
Taraba	2	41	0.213
Yobe	12	146	0.266

Food security:

93. Although food security is captured in the four empirical displacement data collections covered in this analysis, the exact operationalization varies with around 5 different indicators capturing the ability to pay for food, the number of meals eaten, or indices of food insecurity.
94. Four SDG indicators could be used to measure sub-criterion 2.1 in the exit measure. **SDG indicator 2.1.1: Prevalence of undernourishment** is ideally measured through a survey that includes individual-level dietary intakes, actual daily food consumptions, and heights and weights for each surveyed individual. If this is not feasible, a well-designed household survey should at minimum administer a full food consumption module. This data is usually collected through specialized demographic and health surveys rather than multi-purpose surveys. Additionally, the indicator is then computed at the population level through a parametric probability density function, which does

not serve the purpose of a feasible and easy to implement exit measure across displacement contexts.

95. On the other hand, **SDG indicator 2.1.2 Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)** can be readily captured in surveys as standard survey modules are available. The SDG indicator 2.1.2 also forms an integral part of the DS Library to capture any food insecurities and is the only indicator proposed for this sub-criterion. Nevertheless, the analysis of the FIES requires the application of the Rasch model, a more sophisticated statistical technique, that could raise challenges for low-capacity National Statistical Offices.⁸ Despite the growing popularity of the FIES, there is also recent methodological discussion whether the FIES accurately reflects the food-insecure population in a country.⁹
96. An alternative to the FIES - the recommended standard for measuring food security for the purpose of official statistics - could be to fall back to the **reduced Coping Strategies Index** by the World Food Program (rCSI). The rCSI is the de facto standard for measuring food security in humanitarian settings as it is very commonly used. The rCSI is based on five core questions that are administered with universally standardized weights to allow comparability across contexts. While the rCSI only captures a reduced set of indicators for food insecurity and may be sensitive to seasonal differences, it may serve as a reasonable alternative for contexts in which the full FIES is infeasible. An alternative to these two scales could be to revert to the proportion of the population living below the national food poverty line. The disadvantage of this operationalization of the sub-criterion is that it requires the collection and computation of full income/expenditure data which will not be feasible in most surveys aimed at capturing durable solutions.
97. Finally, the indicators **SDG indicator 2.2.1 Prevalence of stunting among children under 5 years of age** and **SDG indicator 2.2.2. Prevalence of malnutrition among children under 5 years of age** could be used to monitor performance under this sub-criterion. However, both SDG indicators require collecting anthropometrics, which is a burdensome and infeasible requirement in many displacement contexts.
98. **It is suggested that EGRISS recommends the SDG indicator 2.1.2: Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale as the standard measure for sub-criterion 2.1 (“Food security”). For displacement contexts in which this indicator is infeasible, the reduced Coping Strategies Index could be offered as second operationalization.**

2.3 Shelter and housing:

99. Housing and shelter is covered in all four displacement contexts assessed in this study with over 10 different indicator - such as a focus on overcrowding, durable housing structures, and water and sanitation facilities.
100. The 12 priority SDG indicators recommended for disaggregation by forced displacement status include **SDG indicator 11.1.1: Proportion of urban population living in slums, informal settlements or inadequate housing**. This indicator, which is also included in the DS Library, seems an appropriate choice to operationalize sub-criterion 2.2. Considering that the SDG indicator 11.1.1 is in fact a compilation of two related indicators, namely the proportion in slums or informal settlements and the proportion in inadequate housing (where the latter is a subset of the former), for the sake of consistency and methodological clarity, the exit measure should focus on the proportion of population in slums and informal settlements. The data to fully implement the SDG indicator is commonly collected in most household surveys and covers access to improved water, access to improved sanitation, sufficient living area, the structural quality/durability of dwelling

⁸ For the purpose of the exit measure, it might be feasible to use the raw scores as an ordinal measure of the severity of food insecurity. Further guidance: FOA, “The Food Insecurity Scale”, <http://www.fao.org/3/i7835e/i7835e.pdf>

⁹ World Bank Data Blog, 2020, “How should we measure food security during crises? The case of Nigeria”, <https://blogs.worldbank.org/opendata/how-should-we-measure-food-security-during-crises-case-nigeria>

and security of tenure . All five elements must be captured to implement the SDG indicator. Security of tenure is measured by collecting data on formal title deeds to land and residence, and agreements/ documentation as a proof for tenure arrangements (in full concurrence with SDG indicator 1.4.2). Although measuring this indicator requires multiple survey questions, the fact that this is already captured in many questionnaires and standard modules are available increases the feasibility to implement the indicator in most displacement contexts.

101. One problem with this indicator may constitute potential overlaps with the sub-criterion for property restitution and compensation that may be re-focused around tenure rights (see later discussions in the note). Other potential indicators in the DS Library heavily focus on the security of tenure rights as well (e.g. fear of eviction recognized documentation for land), and do not capture shelter and housing to the same multi-faceted extent as SDG indicator 11.1.1. It is technically possible that EGRISS focuses this sub-criterion 2.2. on just one of the sub-items of the SDG indicator, such as sufficient living space or durable housing structures, but it seems appropriate to focus on the SDG indicator as a whole which combines the different elements of shelter and housing to one standardized and widely used indicator.
102. **It is suggested that EGRISS recommends the SDG indicator 11.1.1 Proportion of urban population living in slums/informal settlements as a measure for sub-criterion 2.2 (“Shelter and housing”). Later discussion in this study may make a reconsideration necessary to exclude overlap with the sub-criterion 4.1.**

2.3 Medical services:

103. Medical services are captured in different ways in already existing surveys of displaced households. Over 6 possible operationalizations can be found in the four displacement contexts used in the simulations. They range from access to essential health care when needed (most often used), to birth attendance by skilled health personnel, distance to health facilities, health insurance status to satisfaction with current levels of health.
104. The sub-criterion 2.3 could be covered by the **SDG indicator 3.8.1: Coverage of essential health services**. However, the indicator is comprised of 14 sub-indicators of which 3 are derived from administrative data rather than household surveys. The resulting indicator is an aggregated index at the national level that is not applicable for household-level assessments and hence not to be recommended for the exit measure. The SDG indicator 3.8.1. also feeds into the reporting on **SDG indicator 1.4.1: Proportion of population living in households with access to basic services**, which has a health component but is – given its reliance on SDG indicator 3.8.1 – not applicable for the exit measure.
105. Another option could be **SDG indicator 3.8.2: Proportion of population with large household expenditures on health as a share of total household expenditure or income**. The indicator is a relatively straightforward proxy for the affordability of medical services but it requires the administration of a full income/expenditure module on the household level. This may make this indicator unsuitable for many data collections and displacement contexts. Additionally, low expenditure on health may be a result of unavailable or inaccessible medical services, in particular in displacement settings. The focus of this indicator is on capturing universal health coverage rather than differences in access to health care with are the key component for the exit measure.
106. A third SDG indicator related to medical services is **SDG indicator 16.6.2: Proportion of population satisfied with their last experiences of public services**. The indicator measures the availability and quality of services along accessibility, affordability, quality, equal treatment and courtesy of treatment - for three types of services: education, healthcare and “government services”.¹⁰ It can be administered in a reduced form as one question asking the respondents for their overall

¹⁰ Government services refer to services to obtain government-issued identification documents and services for the civil registration of life events such as births, marriages and deaths.

satisfaction with public health services – a relevant question that is included in many household surveys and is feasible in displacement contexts. Albeit not ideal, the SDG indicator 16.6.2 could be adjusted to only/explicitly focus on health-related services. However, the core logic of the indicator focuses on satisfaction with health services, which limits its usefulness for a solutions-oriented assessment. However, the metadata sheet for SDG indicator 16.6.2 does list some filter questions covering access to health services, which are highly relevant to the purpose of the exit measure - see further discussion below.

107. Alternatively, one could use the **SDG indicator 3.1.2: Births attended by skilled health personnel** or **SDG indicator 3.b.1: Proportion of the target population covered by all vaccines included in their national programme**. While both indicators are easy to implement, one should consider if they sufficiently capture access to medical services. Focusing on births attended by skilled health personnel only applies to a small subset of the IDP population that experiences births in a reference time. The indicator may not sufficiently capture the population in question. Access to vaccines only captures one limited part of health services and may hence similarly not be recommendable.
108. With the **DS Library indicator 2.1.7: Target population who accessed essential health care services when needed in the past 12 months**, the DS Library provides another alternative measure to easily capture access to health services for IDPs. This indicator inevitably only captures access to health services for IDPs that needed a health intervention in the last 12 months. On the other hand, the indicator is feasible for data collection in many displacement contexts and allows a straightforward comparison to host/national averages. In terms of questionnaire specification, the indicator could be operationalized through the above-mentioned filter questions developed for SDG indicator 16.6.2:
1. Was there any time during the past 12 months when you (or a child in your household) really needed a medical examination or treatment (Yes, No, Refuse to answer)
 2. Did you (or a child in your household) have a medical examination or treatment each time you (or a child in your household) really needed it? (Yes, No, Refuse to answer)
 3. What was the main reason for not having the medical examination or treatment?
109. It is suggested that EGRISS recommends capturing the sub-criterion 2.3 (“*Medical services*”) via the Durable Solutions Indicator 2.1.8 Target population who accessed essential health care services the last time they needed it in the past 12 months. This indicator could be operationalized in surveys by applying a question set originally developed as “denomination filters” for SDG indicator 16.6.2.

2.4 Education:

110. The last sub-criterion to capture adequate standards of living is 2.4. Across the empirical datasets in this analysis, access to education is captured by varying indicators from the ability to write and read, the proportion of school-age children in primary/secondary school, distance to the next education facility to school completion rates. More than seven different operationalizations can be found across the four datasets.
111. Access to education is the most closely covered by **SDG indicator 4.1.1: Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics**. Despite being one of the 12 priority SDG indicators recommended for disaggregation by forced displacement status, the implementation of this indicator in survey-based data collections is often not feasible. Moreover, while standard modules are available for children in grades 2/3 there is no standard survey module available for children at the end of primary or lower secondary. The difficulties of collecting this data means that the SDG reporting is usually based on large-scale national learning assessments in schools that cannot be linked back to individual IDP households.

112. An alternative indicator could be the **SDG indicator 4.1.2: Completion rate (primary education, lower secondary education, upper secondary education)**. The advantage of this SDG indicator is that there are minimal data requirements that are typically collected in survey rosters. On the downside, the indicator does not account for variation in the quality of schooling or learning outcome. Furthermore, children do not count as vulnerable until they reach an age in which they should have completed specific education levels and the indicator is hence lagged. An additional challenge stems from the fact that the completion rate is partially dependent on the pre-displacement access to education and may hence be less informative for the analysis at hand. As the SDG indicator focuses on three outcomes (primary, lower secondary and upper secondary education), it makes sense to limit this to primary school completion for the purpose of the solution measure.
113. Finally, the **SDG indicator 16.6.2: Proportion of population satisfied with their last experiences of public services** could be used here as well (as discussed in further detail for sub-criterion 2.3. Medical services). However, as discussed in the preceding section, the indicator focuses on satisfaction among those that had access, rather than access per se.
114. Indicators in the DS Library heavily focus on school attendance in primary and secondary schools to capture access to education (**DS indicator 2.1.11 Primary school net attendance** and **DS indicator 2.1.12 Secondary school net attendance**). The above described SDG indicators are not suggested in the DS Library. The advantage of indicators focusing on school enrollment is that the indicator is commonly collected, easy to implement, and applicable across displacement contexts.
115. **It is suggested that EGRISS recommends to measure sub-criterion 2.4 (“Access to education”) via the SDG indicator 4.1.2: Completion rate (primary education). Alternatively, EGRISS could recommend the DS Library Indicators 2.1.11: Primary school net attendance and 2.1.12: Secondary school net attendance.**

3.1 Employment and livelihoods

116. Three relevant SDG indicators could be used to measure sub-criterion 3.1. The most direct choice could be the **SDG indicator 8.5.2: Unemployment rate***, which is also included in the DS Library and prioritized as one amongst 12 SDG indicators to be disaggregated by forced displacement status. The indicator is commonly collected in most household surveys. At the same time, this operationalization could be criticized for not comprehensively capturing unmet demands for work (i.e. other forms of underutilization, in particular time-based underemployment and “discouragedness” of working-age persons that are no longer counted into the unemployment rate due to having given up their job search), and the quality of available work opportunities.
117. Alternative SDG indicators should be discussed briefly: **SDG indicator 8.3.1: Proportion of informal employment in total employment** is equally one of the 12 priority SDG indicators for disaggregation by migratory status and measures the quality of work to some extent. However, the statistical definition of formality may not be applicable in many forced displacement contexts. In many developing countries, informal employment by global definitions makes up the vast majority of employment. The next alternative SDG indicator is **SDG indicator 8.5.1: Average hourly earnings of employees**, which may come closest to measuring the quality of work conditions but requires the administration of a lengthy labor income module.
118. The DS Library also includes further indicators for employment centering around agricultural vs non-agricultural employment, self-employment, occupation types, underemployment, seasonal employment and child labor. However, describing what constitutes a solution and the overcoming of displacement-related vulnerabilities are not straightforward for a range of these indicators and some are not applicable to the whole IDP population but only to a subset.
119. **It is suggested that EGRISS recommends measuring sub-criterion 3.1 (“Employment and livelihoods”) via the SDG indicator 8.5.2: Unemployment rate.**

3.2 Economic security:

120. Similar to the sub-criterion on employment, sub-criterion 3.2 is covered by a range of SDG indicators. Firstly, the closely related **SDG indicator 1.1.1: Proportion of the population living below the international poverty line** and **SDG indicator 1.2.1: Proportion of population living below the national poverty line** could be used to cover economic security. The latter has been prioritized for disaggregation by forced displacement status. However, the indicator requires a full income/expenditure module or the application of statistical estimation procedures. This seems infeasible in many survey contexts.
121. A third relevant indicator choice could be **SDG indicator 1.2.2: Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions**. The implementation of this indicator requires to measure 10 sub-indicators as described in the metadata. The multi-dimensional poverty index is a full composite measure in its own right that overlaps with several other sub-criteria in the exit measure. Its scope ranges far beyond economic security, which is the clear focus of this sub-criterion. From a statistical perspective, having overlaps in indicators across sub-criteria can raise statistical challenges - such as collinearity - and theoretical challenges - if the exit measure essentially captures the same vulnerabilities multiple times - in the final exit measure. This indicator is hence not a recommended choice.
122. A fourth relevant SDG indicator comes closest to capturing the concept of economic security in the face of unexpected shocks: **SDG indicator 1.3.1: Proportion of population covered by social protection floors/systems** is included in the DS Library as well. However, the survey practice for implementing this indicator can vary in content and coverage across displacement contexts and the data typically stem from administrative registries that cannot be accessed and linked to household surveys.
123. In addition to these options, the DS Library contains indicators measuring sustainable income sources and problems to pay for basic expenses or cover loans. **DS indicator 3.2.4 Target population who in the last 12 months was not able to pay for basic expenses** could be an alternative for the exit measure due to the easier application. The indicator does not require the collection of a full consumption or poverty module in a data collection. The Durable Solutions Question Bank suggests the following precise operationalisation, which is based on the Zaragoza indicators for integration in the EU:
1. In the past 12 months, was your household at any point not able to pay any if the following:
 - a. to pay rent or utility bills
 - b. to keep home adequately warm/cold
 - c. to face unexpected expenses (of xx amount in x currency)
 - d. to eat meat, fish, or a protein equivalent every second day.
124. Similarly, **DS indicator 3.2.2 Target population relying primarily on sustainable income sources over the last 30 days** could be considered given the easier application. Capturing unsustainable income sources, such as donations, aid, loans, assets sales, etc., could also help to address the problem that humanitarian aid should be factored in when assessing whether - for example - camp-based IDP populations can exit the statistical stock.
125. **It is suggested that EGRIS recommends measuring sub-criterion 3.2. (“Economic security”) via the SDG indicator 1.2.1: Proportion of population living below the national poverty line. If the collection of full expenditure/poverty data is infeasible, EGRIS could recommend the Durable Solutions Indicator 3.2.4 Target population who in the last 12 months was not able to pay for basic expenses.**

4.1 Property restitution and compensation (or security of tenure):

126. As will be discussed in later sections of this note (“challenge 3”), there may be a need to redefine the sub-criterion 4.1 to focus on security of tenure rather than property restitution and

compensation due to the fact that mechanisms to reconstitute property or receive compensation are lacking in many IDP contexts. Nevertheless, this section reviews which indicators are potentially available to either measure property restitution or security of tenure. For the final indicator selection, one should consider to what extent there is overlap between the security of tenure (under sub-criterion 4.1) and shelter and housing (under sub-criterion 2.2) if this sub-criterion's focus is shifted away from property restitution and compensation to tenure security. To recap, the proposed indicator for sub-criterion 2.2 above is SDG indicator 11.1.1 which already explicitly includes the security of tenure as one component.

127. The tenure component within SDG indicator 11.1.1 also features as a stand-alone SDG indicator: **SDG indicator 1.4.2 : Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure**. This indicator has been highlighted as one of the 12 priority SDG indicators for disaggregation by forced displacement status, and standard survey modules exist as guidance for National Statistical Offices. In the context of the exit measure, the question whether this would focus on the security of tenure in the displacement location or in the place of origin may be left for the producer of official statistics to decide.
128. Relatedly, **SDG indicator 5.a.1 (a) Proportion of total agricultural population with ownership or secure rights over agricultural land** measures the prevalence of people in the agricultural population that own or have secure tenure rights. As such, the indicator is closely related to SDG indicator 1.4.2 but applies only to a subset of IDPs (only agricultural population). The disaggregation by sex may also be dropped for the purpose of the exit measure.
129. Finally, a related option is **SDG indicator 16.3.3: Proportion of the population who have experienced a dispute in the past two years and who accessed a formal or informal dispute resolution mechanism** as standard survey modules exist for this question and property restitution falls within the scope of the sub-criterion. The indicator is problematic as it does not capture how many IDPs have property disputes but only captures how many of those with property disputes have access to resolution mechanisms. At the same time, the reference period is limited to two years according to the metadata, which may be problematic given longer displacement situations. Additionally, these indicators and questions are usually included in dedicated victimization surveys rather than multi-purpose household surveys.
130. Beyond the above indicators, the DS Library includes a range of related indicators that focus on property prior to displacement only; as well as the resolve of claims to assets and the enforcement of such claims. Given the recommended shift of focus in the exit measure towards security of tenure in the displacement location (see later discussion), these are not further explored here.
131. **As discussed in greater detail in the chapter below on “challenge 3”, it is recommended that EGRISS moves away from restitution and compensation mechanisms and focuses on security of tenure for property and land. It is suggested that EGRISS hence recommends to measure sub-criterion 4.1 via the SDG indicator 1.4.2: Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure. Considering that SDG indicator 1.4.2 is already a (binding) sub-element of SDG indicator 11.1.1. which was recommended further above for operationalizing sub-criterion 2.2 of the exit measure, the option of dropping sub-criterion 4.1 altogether is discussed in the chapter “challenge 3”.**

5.1 Documentation:

132. The final sub-criterion 5.1 could be measured by using the **SDG indicator 16.9.1: Proportion of children under 5 years of age whose birth have been registered with a civil authority** in lack of an SDG indicator on adults' possession of documentation. While this is a prioritized SDG indicator on forced displacement, and which is applied in many household surveys already, the indicator is limited in scope and does not cover the full displaced population (i.e. households without children

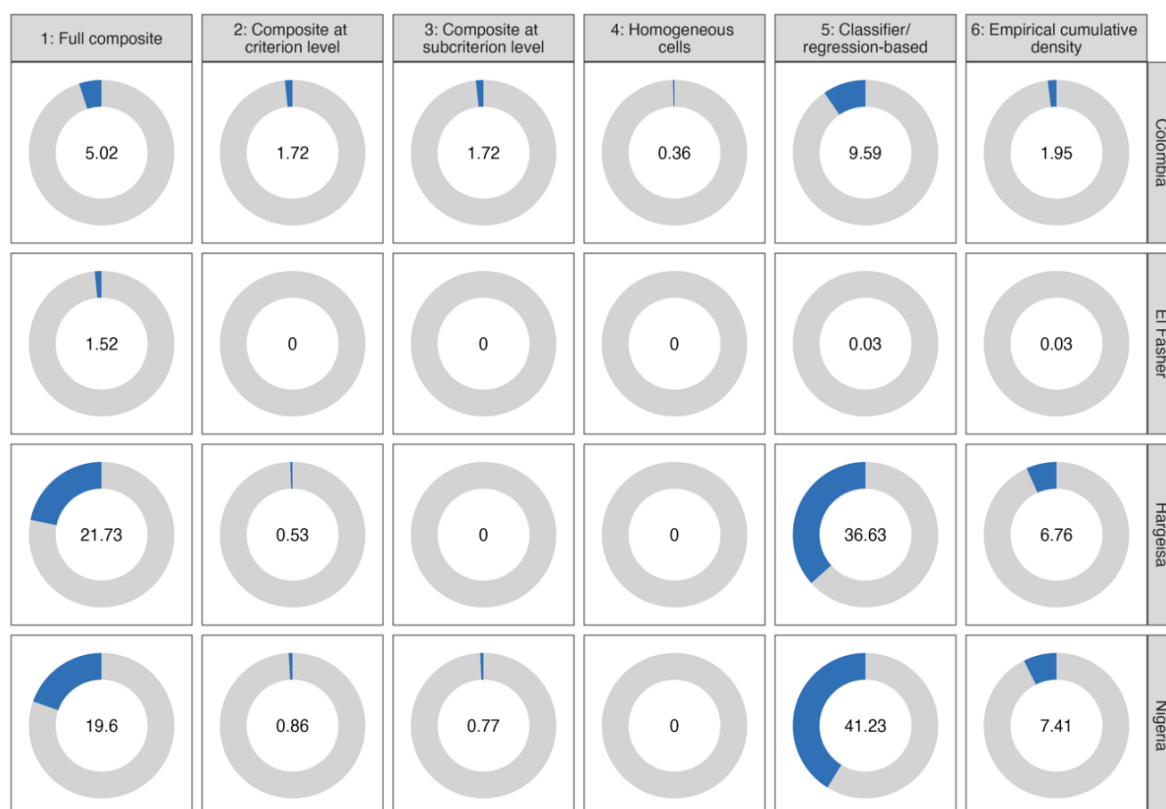
or households with younger children). Given those weaknesses, it might make sense to move away from SDG indicators.

- 133. The DS Library indicator 5.1.1 Target population currently in possession of valid birth certificates, national ID cards or other personal identification documents relevant to the context seems the most appropriate here. Other options in the DS Library, such as the ability to replace lost documentation, again may not be applicable to the whole IDP population. Indicators related to registration as IDPs also may not be applicable in many displacement contexts.
- 134. It is suggested that EGRISS recommends the DS Library Indicator 5.1.1: Target population currently in possession of valid birth certificates, national ID cards, or other personal identification documents to measure the sub-criterion 5.1 (“Documentation”).

Empirical performance of the recommended indicators

- 135. How do the proposed indicators for each sub-criterion perform empirically? This section summarises the IDP exit from the stock that is generated if we implement the proposed indicators for each country example and each metric option. Note that the datasets do not contain *all* proposed indicators. Hence, the approach was to implement the proposed statistical indicators *as closely as possible*.¹¹ Figure 11 presents the percentage of IDPs exiting the stock under each metric option and for each country if we use the proposed indicators. Overall, one can see that the exits are minimal and substantial exits only take place under the metric options 1 (Full composite), 5 (Regression-based) or 6 (Empirical distribution).

Figure 11: Percentage of IDPs exiting the stock under each metric option for the proposed indicators



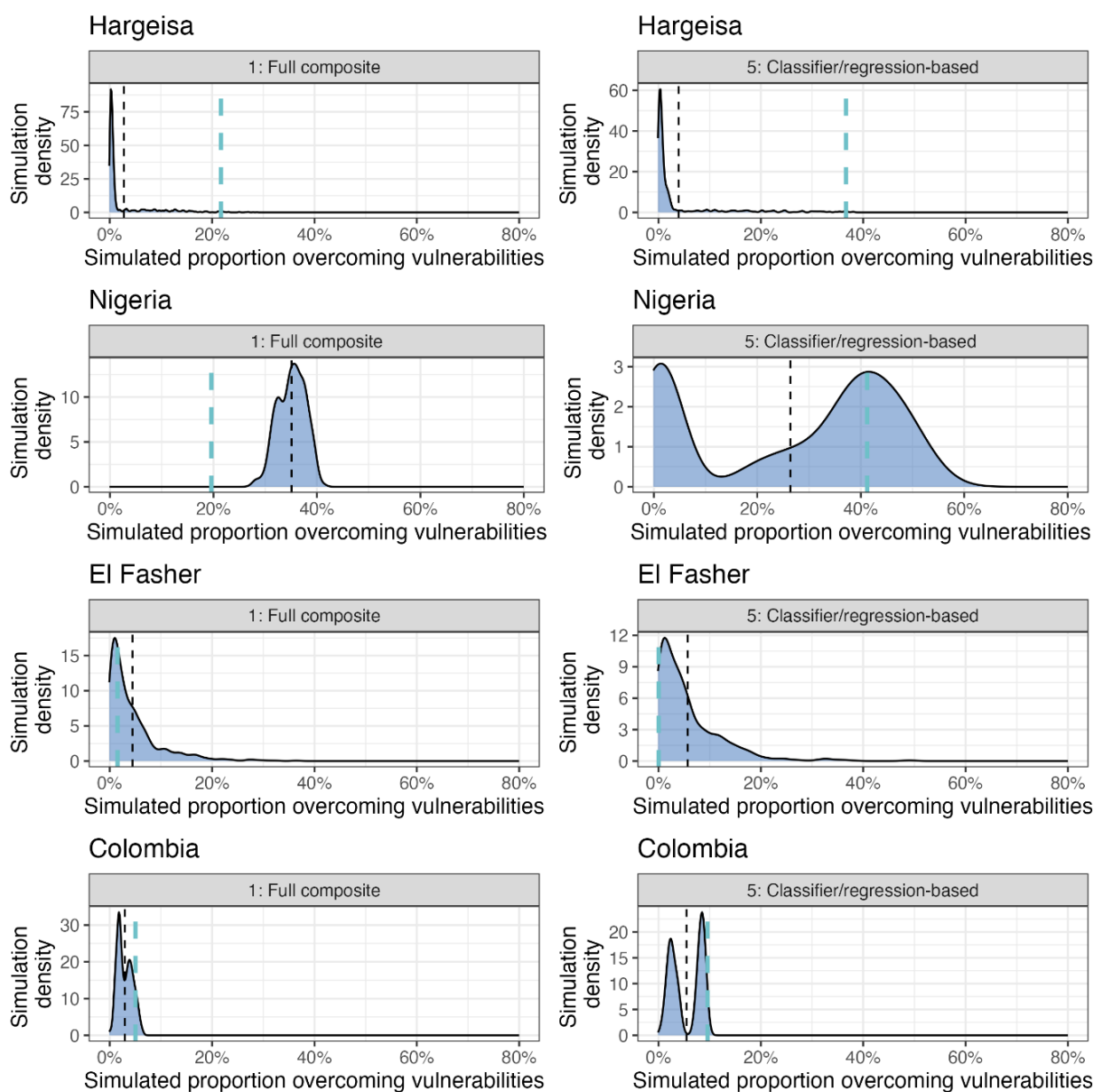
- 136. It is also important to understand if the proposed indicators perform atypical to other indicators - that means the proposed indicators lead to unusually high or unusually low exits from the stock than other indicator combinations. This would be an indication that the indicators are not well-

¹¹ For a full overview of how each indicator could be implemented, see the appendix IV.

chosen to represent the sub-criterion. Figure 12 displays the initial simulation densities for the metric options 1 and 5. In these plots, the vertical green and thicker line displays the exit from the IDP stock under the proposed indicators while the dashed thinner line displays the mean exit from the stock across all indicators.

137. For example, in Nigeria, if we compare the point estimate of around 19.6% of exits under a full composite metric (achieved by the proposed indicators) with the original simulations for this metric option, one can see that the estimated exit is slightly higher than the mean but it is not an extreme outlier. This is also the case if we look into a regression-based metric option for Nigeria. With the exception of Sudan, we generally find that the proposed indicators for the exit measure yield a slightly higher estimates of exits than the average simulation. For Sudan, the proposed indicators yield a smaller estimate. For Hargeisa, the difference between the mean simulation and the proposed indicators is the most striking. For the other datasets, the proposed indicators seem to provide a good and robust estimate of the IDP exit given that they also suffer less from data missingness than other indicators.

Figure 12: Simulation density for metric option 1 (left side) and 5 (right side). Highlighted in teal is the IDP exit from the stock under the proposed indicators.



Conclusions on challenge 2

Summary of findings

138. This study on the effect of indicator selection on IDP exits demonstrates that the exit measure is not particularly sensitive or volatile to the selection of one indicator over another, and this is true across countries and contexts. This provides some justification for EGRISS to make a global recommendation on the indicators that can be used to operationalize the sub-criteria of the exit measure and may facilitate and de-politicize the debate as EGRISS moves towards such a recommendation.
139. The notable exception to this finding is sub-criterion 4.1., which is discussed in greater detail in the chapter below on overcoming challenge 3. The analysis shows the large potential effect introduced in practical application by the recommendation currently provided in the IRIS, which is that households that have data missing in any sub-criterion (even if “by design”) should not exit the IDP stock. This issue is further discussed in the chapter below on overcoming challenge 4. Furthermore, and to a lesser degree than the above two issues, this chapter also revealed that IRIS lacks concrete suggestions for how to address the difference between data collected at the household versus data collected at the individual level. Failure to provide concrete guidance on how to aggregate individual-level data to the household level for the exit measure can easily result in incomparability across countries and/or datasets. This issue is further discussed in the chapter below on overcoming challenge 9. Finally, the simulations generally confirmed the recommendations made in IRIS to use commonly available, standardized and high-quality indicators that maximize coverage in the IDP population.
140. Based on these findings and a review of the SDG indicators and the Durable Solutions Library, a suggestion for proposed indicators has been made. Table 17 specifies the proposed indicators for the IRIS exit measure as well as a possible alternative where applicable. First empirical implementations of the proposed indicators in the four displacement examples used in this paper are promising: the proposed indicators seem to generate estimates of the IDP exit from the stock that are slightly higher than the mean estimate from the universe of potential indicators. This should be understood as a promising characteristic of the proposed indicator because these indicators suffer less from missing data or non-applicability than other indicators in the universe of options. Another strength of the proposed indicators is that they are general enough to be applicable in locations of displacement and locations of return (see further discussion in Box 7). Nevertheless, further work is needed to compare the proposed indicators with other possible operationalisations.

Box 7: Scope of the IRIS exit measure: assessments in locations of displacement and return

Estimating exits from the IDP stock in locations of displacement and return

According to IRIS, taking persons out of the IDP stock should be guided by the exit measure irrespective of whether a displaced person is found in location of displacement or location of return. The latter is often overlooked when discussing the IRIS exit measure. In other words, IDP households returning to their place of origin should not automatically be taken out of the stock – neither immediately nor after a fixed amount of time – but only when they have overcome their displacement-related vulnerabilities. The four datasets analyzed for this methodological paper only cover IDPs in locations of displacement; hence, this paper sheds no light on exit patterns from the stock in location of displacement. In principle, this should not affect the further maturation of the IRIS exit measure, as IRIS promulgates that the method for assessing the overcoming of displacement-related vulnerabilities should be identical in locations of displacement and locations of return.

Recommendation

141. It is suggested that EGRISS standardizes the indicators to operationalize the 10 sub-criteria as much as possible to increase comparability across contexts. To do so, it is suggested to rely on the SDG indicators and the Durable Solutions Library and a list of indicators has been proposed (see Table 17).

Table 17: proposed indicators for the IRIS exit measure

Sub-criteria IRIS	Recommended indicator	Possible alternative(s)
1.1 Victims of violence	SDG 16.1.4: Proportion of population that feel safe walking alone around the area they live	
1.2 Freedom of movement	DS 1.4.1 Target population facing restrictions to their freedom to move (implemented with a survey question on the household/individual level)	
2.1 Food security	SDG 2.1.2: Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)*	WFP's reduced Coping Strategies Index (rCSI)
2.2 Shelter and housing	SDG 11.1.1: Proportion of urban population living in slums, informal settlements**	
2.3 Medical services	DS 2.1.8 Target population who accessed essential health care services (including mental health care) the last time they needed it in the past 12 months	
2.4 Education	SDG 4.1.2: Completion rate (primary education)	DS 2.1.12: Primary/ secondary school net attendance rate in target population (i.e., proportion of school-aged children who are attending school)
3.1 Employment and livelihoods	SDG 8.5.2: Unemployment rate	
3.2 Economic security	SDG 1.2.1: Proportion of population living below the national poverty line***	DS 3.2.4: Target population who in the last 12 months was not able to pay for basic expenses
4.1 Property restitution and compensation	SDG 1.4.2: Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documents, and (b) who perceive their rights to land as secure	
5.1 Documentation	DS 5.1.1: Target population currently in possession of valid birth certificates, national ID cards, or other personal identification documents.	

* Instead of implementing a Rasch model, an ordinal scale might be used.

** Depending on EGRISS' decision regarding sub-criterion 4.1, this recommendation may require adjustments.

*** For when the national poverty line has not been recently updated (e.g., Sudan last updated in 2009), EGRISS may recommend using the proportion of the population living below the international poverty line.

Overcoming challenge 3: Sub-criterion on property restitution and compensation

142. The above analysis has shown that the exit measure is particularly volatile in the choice of indicators for the sub-criterion *4.1 Property restitution and compensation*. This was the case in both Hargeisa and Nigeria, where the choice of a more or less permissive HLP indicator can affect large percentages of the IDP stock. In Sudan, the indicator used was security of tenure rather than access to property restitution as such mechanisms for property restitution were not available to all IDPs and would not allow a comparison between hosts and IDPs. In the dataset in Colombia, HLP rights were not covered.

Simulations with and without the HLP sub-criterion

143. To further investigate how the access to property restitution and compensation mechanisms affects the simulation results, the figures 13 and 14 plot the simulation densities for an assessment of all 10 sub-criteria and for an assessment based on 9 sub-criteria only (i.e. excluding the HLP indicators). The blue shaded curves exclude the HLP indicators; the grey shaded curves are the original assessment. The dashed vertical lines in these charts 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, across all metric choices. This means that, in the four countries analyzed, 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 exit the stock.

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

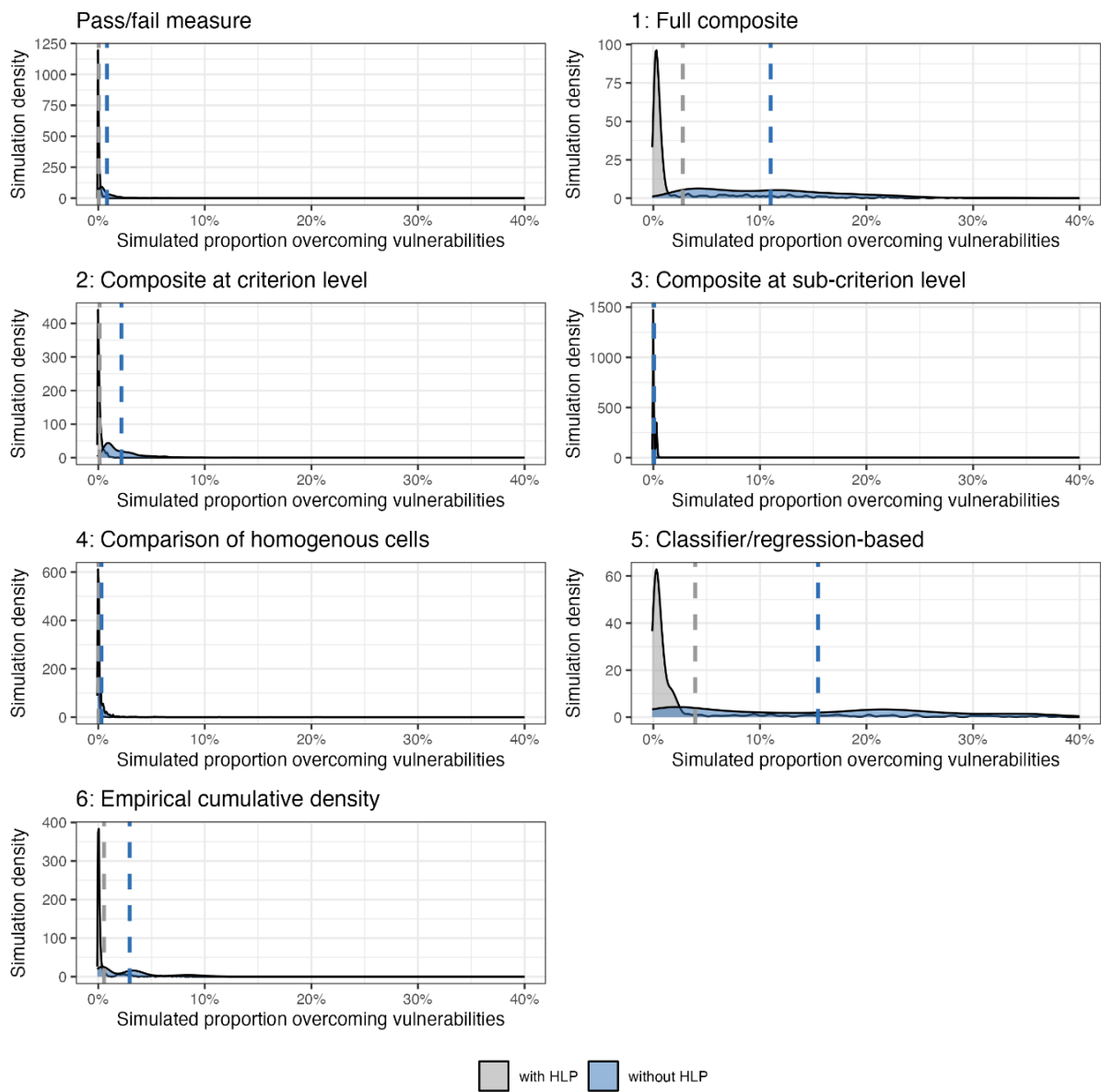
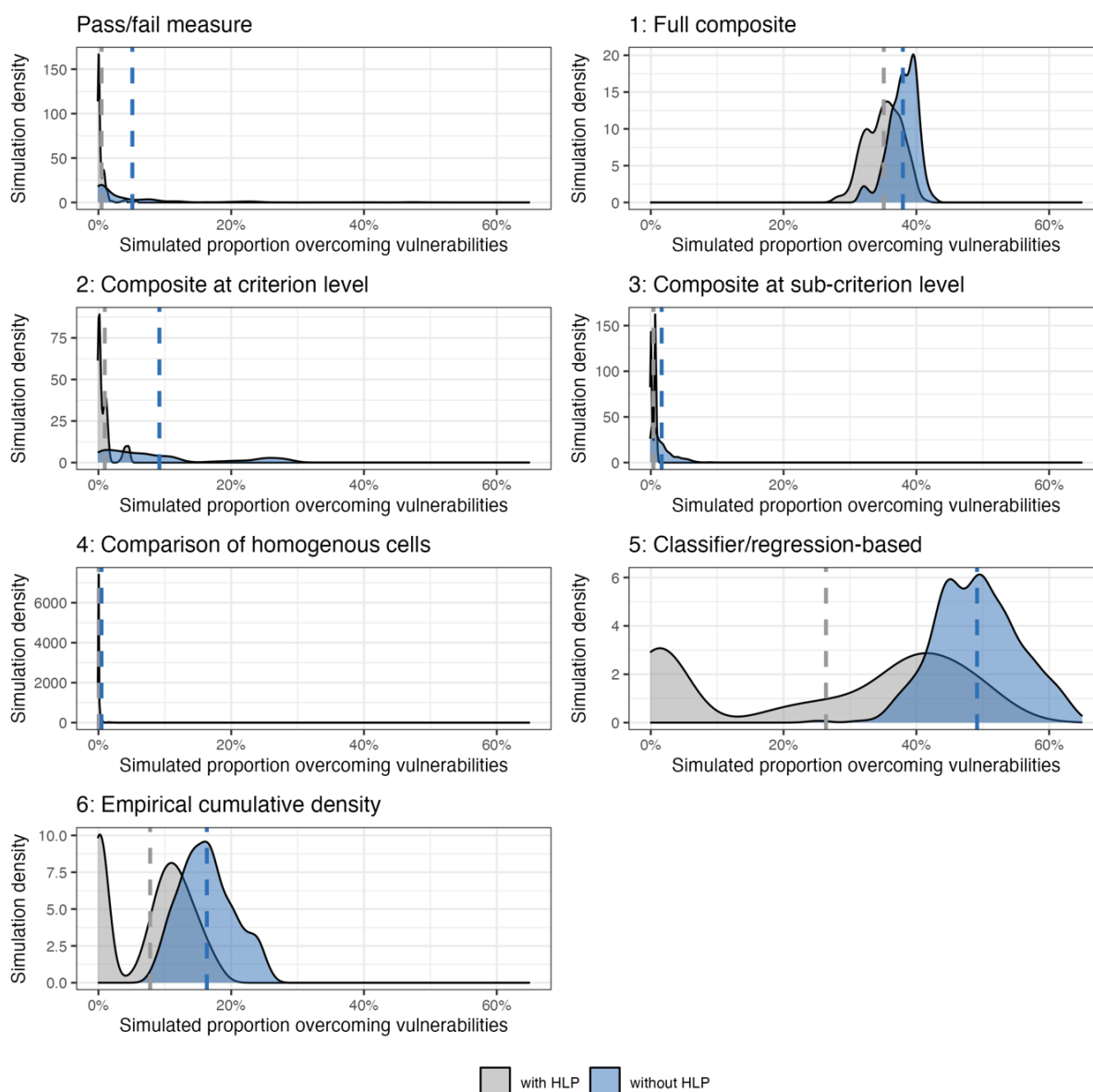


Figure 14: Density of simulations when including and excluding HLP indicators (Nigeria)



Addressing housing, land and property issues: shifting focus to security of tenure

144. Consultations with experts on HLP confirm what the above simulations from four countries have already indicated: restitution or compensation mechanisms exist only in a very limited number of cases across the world, and even where those mechanisms exist, there are issues in terms of equitable access to them and their effectiveness. Beyond the factual situation on restitution or compensation mechanisms, there is also a broader shift in the humanitarian approach for the response to displacement towards a focus on security of tenure. The focus in HLP on access to effective restitution and compensation mechanisms for lost or abandoned housing, land and property – based on well-established principles of the right to an effective remedy and the restitution of pre-displacement homes and lands – has evolved in recent years. It is acknowledged that situations of protracted displacement, typically characterised by unresolved conflicts which rule out both restitution and voluntary return, have necessitated a shift towards a broader perspective of security of tenure, as an aspect of the right to adequate housing relevant in humanitarian and development contexts.

145. This shift in focus recognises that even where restitution programmes are in place they are primarily oriented towards freehold documented ownership which does not reflect the situation of the majority of the displaced, outside of the Balkans. This is because most housing and land assets are not formally registered in the rest of the world, and furthermore that most displaced people do not hold or have lost HLP ‘ownership’ documentation. Most significantly, HLP restitution programmes have failed to consider the specific circumstances and barriers for women and particularly marginalised groups within their design and have in many cases effectively excluded the most vulnerable. An example of this is the lack of civil documentation needed for HLP restitution claims and related inheritance disputes (e.g. ID cards) that affects women significantly more than men; as well as displaced groups at risk of statelessness.
146. Furthermore, the circumstances in which people are forced to flee often involve a combination of conflict and disasters. Security of tenure is relevant across this range of contexts. For the purpose of this exit measure, it is therefore proposed that a focus is placed on indicators that demonstrate whether IDPs have somewhere to live in safety in the longer-term – without the fear of forced eviction. This approach is in line with the basic principle in IRIS that a benchmarking against the non-displaced population should take place.
147. Given that the wording of sub-criterion 4.1 of the exit measure is very narrowly focused on restitution/compensation, and that the right to an adequate standard of living / the right to adequate housing is already considered under sub-criterion 2.2 of the exit measure, there may be a case for dropping sub-criterion 4.1 altogether (as opposed to “reinterpreting” it in a way that aims at security of tenure) – always provided that a clear focus on tenure security in the indicator selection under sub-criterion 2.2 can be guaranteed. This methodological and pragmatic proposal does not imply that access to restitution and compensation mechanisms for displaced persons is less central to the IASC framework. Access to restitution and compensation mechanisms remain a key component of durable solutions.

Conclusions on challenge 3

Summary of findings

148. Out of the ten sub-criteria currently specified in the IRIS exit measure, the only one that stands out as particularly restrictive in terms of allowing any exits from the IDP stock in practice is sub-criterion 4.1 on restitution/compensation. Consultations with HLP experts corroborate that the quantitative finding from four countries are “representative” of the wider fact that restitution or compensation mechanisms exist only in a very limited number of cases across the world, and even where those mechanisms exist, there are issues in terms of equitable access to them and their effectiveness. A direct implication of this is that the IRIS exit measure, as it currently stands, will barely allow any exits from the IDP stock in practice. However, there is a broader shift in focus from compensation/restitution towards security of tenure. Bringing the IRIS exit measure in line with this recent shift might thus be advisable.

Recommendation

149. It is suggested that the focus on HLP restitution for the purpose of the IRIS exit measure is reconsidered, and is reinterpreted as security of tenure. This would require re-labeling the sub-criterion, as its wording is currently very narrowly focused on compensation/restitution. This would also imply operationalising the sub-criterion 2.2 covering adequate standards of living with indicators that do **not** exclusively measure security of tenure but rather the broader right to an adequate standard of living/the right to adequate housing.
150. Appropriate indicators to operationalize the sub-criterion 4.1 in line with a tenure security focus would include: **SDG Indicator 1.4.2** “Proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure” (see recommendations under challenge 2). Alternative relevant indicators are tenure arrangements, experiences of evictions, or the fear of eviction.

151. Appropriate indicators to operationalize the sub-criterion 2.2 to avoid complete duplication with a revised sub-criterion 4.1. could be: **SDG Indicator 11.1.1**: Proportion of urban population living in slums, informal settlements or inadequate housing (see recommendations under challenge 2).
152. However, EGRISS might decide that this creates a degree of redundancy between 4.1 and 2.2. An alternative course of action could be to drop sub-criterion 4.1 from the exit measure. This will need to be accompanied by methodological certainty that the indicator selection under sub-criterion 2.2 indeed guarantees a clear focus on tenure security. In other words, dropping 4.1 from the IRIS exit measure can only be recommended if EGRISS at the same time makes an authoritative global recommendation on the indicators that should be used to apply sub-criterion 2.2. of the exit measure.

Overcoming challenge 4: Dealing with missing data points

153. As seen in the chapters above on overcoming challenges 1 and 2, a key problem in applying the IRIS exit measure in practice is that IRIS posits that “[t]here may be situations where data are not available for all criteria or sub-criteria. In this case the composite measure cannot be fully implemented and a complete assessment cannot be made about whether or [not] the population has overcome key-displacement related vulnerabilities” (IRIS, para. 168). In other words, IDP households remain in the stock if data is missing on one of the 10 sub-criteria. This strict approach may exclude some portions of the IDP population from ever being assessed. On average across the simulations, 64.52% of IDP households cannot be assessed and remain in the IDP stock because of definitional issues rather than specific vulnerabilities. In the country examples used in this report, the problem is particularly severe for Sudan, Hargeisa and Colombia.

Data missingness due to design and non-response

154. Missing values normally enter survey datasets through one of two channels: (i) *by design*, i.e. when a question is not administered to a certain respondent based on the questionnaire’s skip-logic; and (ii) *due to (item) non-response*, which happens when the respondent declines to answer a certain question for any reason. The former is a common design feature of most survey datasets and not a technical problem per se. The latter can be a source of bias if the likelihood of non-response is systematically correlated with individual characteristics, but the scale of item non-response is negligible in most cases. Yet, IRIS makes no explicit distinction between the two even though 4 of the 10 exit measure assessment criteria introduce missing values into the assessment by design (see Table 18).¹² For example, a questionnaire skips questions around education for childless IDP households since their educational vulnerability is undefined.

Table 18: Data missingness by design in the exit measure

Sub-criterion	Missing values by design
1.1 Victims of violence	
1.2 Freedom of movement	
2.1 Food security	
2.2 Shelter and housing	
2.3 Medical services	Only asked to households that have medical needs within a reference period.
2.4 Education	Only asked to households with school-age children/youth
3.1 Employment and livelihoods	Only asked to households with working-age members participating in labor force
3.2 Economic security	
4.1 Property restitution and compensation	Only asked to households that owned property prior to displacement
5.1 Documentation	

Options for dealing with missing data

155. Having established that missing data impedes the application of the IRIS exit measure for a substantial portion of IDP households, the question becomes how to deal with this challenge. Three techniques are commonly employed in the statistical literature for handling missing data.
- The most basic approach is to conduct **complete case analysis**. This entails maintaining the IRIS approach of only assessing households with no missing information while extrapolating the share of assessed IDPs exiting the stock to the full IDP population. While the approach is conceptually simple and stays true to the wording of the IRIS, the validity of the calculation rests on the

¹² In the process of developing the IRIS, the report “Statistical Measuring of Overcoming Internal Displacement-Related Vulnerabilities” already outlines the distinction between missingness by design and non-response but this has not been incorporated into the main recommendations.

assumption that the data are missing completely at random – i.e., that the characteristics of households with missing data are no different from those for whom complete data is available. This is clearly not the case here where questionnaire skip-logic creates missing data patterns that vary systematically based on respondent household characteristics.

- A variation of the complete-case analysis, which is only applicable for the homogeneous cell approach, is **available-case analysis**. One could calculate cell-wise indicator averages based on the households with non-missing data for the respective indicator only. This means that different indicators will be calculated based on different subsets of households within each cell thereby allowing households with missing data to exit the stock if the whole cell performs better on average than the host population benchmarks. The downside of this method is that it only works for a single metric option and makes the same missing at random assumption as complete-case analysis.
- The third option to deal with missing data is **imputation** or the idea to infer plausible values for missing data points which could have been observed were the data not missing. Imputation comes with its own set of challenges. First is whether to impute the missing raw survey data or the derived binary solutions indicator. If the former, there is nothing FDP-specific about the procedure to justify its standardization in IRIS and the issue is best left for NSOs to apply the methods they deem appropriate. Second, how would the uncertainty introduced by the imputation process be reflected in the final estimate of IDPs exiting the stock? Third, and most importantly, is the difficulty of interpreting implausible imputations. To continue with the running example, what would it mean to impute a “no pass” on the education criterion for a childless household? Another challenge with imputation is that it adds analytical complexity, which will make the exit measure more difficult to apply for low-capacity NSOs.
- Finally, EGRIS may simply opt to **reinterpret missing values as the absence of vulnerability**. The idea finds justification in the fact that a household with no school-aged children cannot experience education-related vulnerabilities. This is essentially a reversal of the IRIS approach: whereas IRIS can be seen as implicitly replacing missing data points with a “no pass” to prevent IDP households from exiting the stock, this approach would basically assign IDP households a “pass” on the missing indicators to allow them to be assessed on the other indicators for which data is available. This redefinition should focus only on missingness by design.

Simulations on the problem of missing data

156. To arrive at an evidence-informed decision on the way forward, the simulations were re-run by reinterpreting missing values as the absence of vulnerability.¹³ Table 19 shows that when data missingness is redefined as absence of vulnerability the number of IDPs exiting the stock increases. The reason is that more IDP households can be assessed. The greater the share of missing data (e.g. due to skip patterns), the larger the increase in IDP exits gained from the redefinition across displacement contexts. For Colombia, for example, this treatment of missing data means that over 28% of IDP households may exit the stock. This is a more realistic estimate given the displacement context. This finding is also consistent across metric options (see Table 20).

Table 19: Average percentage of IDPs exiting the stock across datasets and indicator combinations

Dataset	% Missing (avg.)	Missing as not vulnerable	Original IRIS approach
Hargeisa	83.91%	19.26%	1.06%
Colombia	82.68%	28.30%	2.17%
Sudan	81.04%	11.39%	1.72%
Nigeria	10.45%	14.83%	10.16%
Average	64.52%	18.44%	3.78%

¹³ A complete case analysis does not allow to assess all IDP households, available-case analysis is only an option for homogeneous cells, and data imputation can take many - often complex - forms. Hence, the focus is here on discussing the impact of treating missing data as absence of vulnerability.

Table 20: Average percentage of IDPs exiting the stock across metrics and indicator combinations

Metric	Missing as not vulnerable	Original IRIS approach
Pass/fail measure	10.49%	0.66%
1: Full composite	37.03%	11.31%
2: Composite at criterion level	11.63%	0.97%
3: Composite at sub-criterion level	3.41%	0.18%
4: Comparison of homogenous cells	0.68%	0.24%
5: Classifier/regression-based	38.09%	10.38%
6: Empirical cumulative distribution	27.78%	2.71%

157. The metric showing the highest increase (in relative terms) is the composite index at sub-criterion level. Since this option requires the largest number of indicators, it is also most susceptible to missing data in the given indicator universe. Reinterpreting missing data as the absence of vulnerability would allow all households in the stock to be assessed resulting in a three-fold increase in the exit rate for this metric option.
158. In order to advance the work on the formulation of the exit measure, it is overall advisable to distinguish more clearly between data that is missing by design and data that is missing due to non-response. Data missing due to item non-response - which usually affects only a small fraction of observations - could be explicitly left to the national statistical offices to apply their missing data treatment procedures. In the absence of such procedures, or where certain households continue to have missing data after the application of said procedures, then those households should be excluded from assessment under the exit measure as IRIS already specifies. However, for data missing by design, this should be reconsidered and clarified that missing data entries due to skip patterns and non-applicability should be interpreted as the absence of vulnerability. Figure 15 visualises the substantial difference a redefinition of missing data by design as absence of vulnerability makes for the performance of the possible metric options in simulations. The densities displayed in grey are based on simulations with missingness by design defined as non-vulnerability while the blue distributions are the original estimates. The density distributions and the mean exits (displayed as vertical lines) show that the redefinition yields more realistic estimates of the exit from the IDP stock across all 4 assessed contexts.

Conclusion on challenge 4

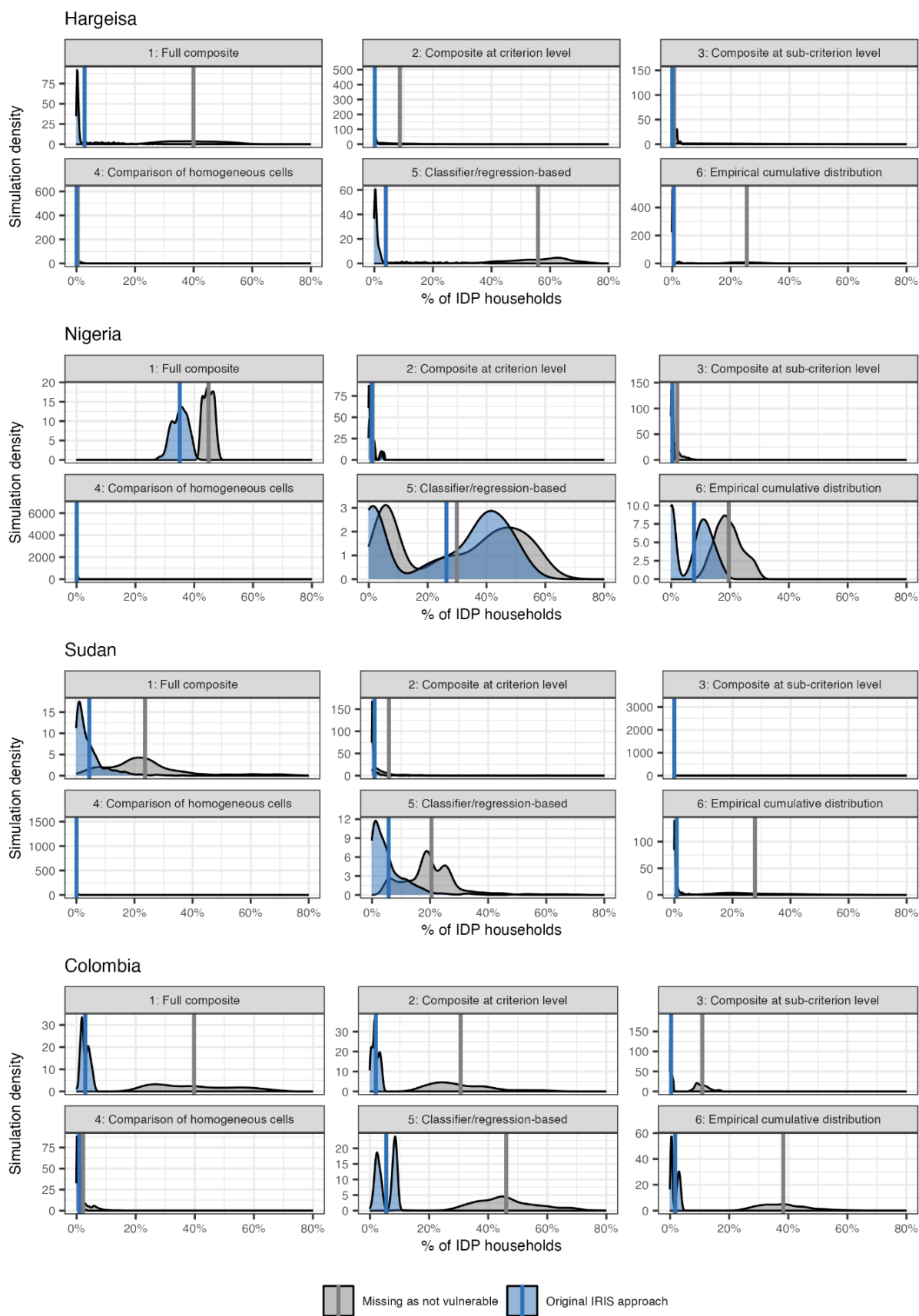
Summary of findings

159. IRIS posits that no assessment can be made and thus that households should remain in the IDP stock if data is missing for any of the exit measure's sub-criteria. This is problematic given that a substantial proportion of households – almost a third of households across our simulations - are affected by missing data. In the vast majority of cases, this data is missing “by design”, i.e. for perfectly valid reasons such as a household without children not being asked about current school attendance of children.

Recommendation

160. For data missing “by design”, the IRIS recommendation should be clarified, pointing out that missing data due to valid skip patterns and non-applicability should be interpreted as the absence of vulnerability in that particular indicator. Missingness because of other reasons (e.g. household refuses to answer or questions were not asked/ indicator not collected) should not automatically be treated as lack of vulnerability.

Figure 15: Density of simulations when redefining missingness as non-vulnerability



Overcoming challenge 5: Aggregation of individual-level indicators to household level

161. Related to the problem of missing data is the fact that some of the sub-criteria are more naturally measured on the individual-level than the household level. For example, employment is easiest to assess on the individual level. If no clear aggregation rule is defined, a comparison of IDP exits across contexts risks being flawed because one data producer may make the decision to aggregate to the household level differently than another producer of official statistics. Furthermore, clear guidance on the aggregation from individual level questions to the household level clarifies if households are applicable for a certain sub-criterion (e.g. aggregation of information from school-age children to the household level clarifies if the household is vulnerable on the dimension of education or not).
162. To make a meaningful assessment of the household-level exit of IDPs from the overall stock, EGRISS should hence make clear how data collected from individual respondents can be aggregated to the household level. As Table 21 summarises, this aggregation challenge is most acute for the sub-criteria *2.4 Education*, *3.1 Employment and livelihoods*, and *5.1 Documentation*, but may also arise in other sub-criteria depending on the indicator chosen.¹⁴

Table 21: Sub-criteria heavily affected by the problem of aggregating indicators to the household level

Sub-criterion	Asked at the individual level
1.1 Victims of violence	
1.2 Freedom of movement	Commonly assessed at community-level
2.1 Food security	
2.2 Shelter and housing	
2.3 Medical services	
2.4 Education	Asked at individual-level
3.1 Employment and livelihoods	Asked at individual-level
3.2 Economic security	
4.1 Property restitution and compensation	
5.1 Documentation	Asked at individual-level

Available data aggregation rules

163. There are different ways in which data can be aggregated from the individual to the household level:
- **All individuals must pass:** The first relatively straightforward way to aggregate individuals' survey responses to the household level is to determine that all individuals must pass a sub-criterion for the household to pass the sub-criterion. In the example of employment, all adults available for work within a household must be employed. For many indicators, this aggregation rule might not make sense. For employment, many scenarios are possible why not all household members are employed without this being an indication of vulnerability.
 - **At least one individual must pass:** The second aggregation rule could be to define that at least one household member must pass. In the context of employment, this would mean that at least one adult in the household available for work must be employed for the household to be considered "employed". This aggregation rule has been used for the simulations presented in this paper. However, for some of the sub-criteria, this aggregation may be less sensible. For example, one may want all adult household members to have identification documents to count the overall household as being able to identify themselves.

¹⁴ Note that *1.2 Freedom of movement* is commonly assessed at the community-level in previous IDP profilings. For the exit measure, as discussed under challenge 2, breaking this down to the household level would be crucial. Also see discussion box for more information.

- **Mean value across individuals:** Less relevant in the context of the exit measure in its current form is the option to average across mean values of household members to generate a household level estimate (e.g. average income in household). This option is more relevant for metric indicators than for binary or categorical indicators that are more realistically used for the exit measure.
- **Extreme values:** Another option less relevant in the context of the exit measure is to pick the extreme value amongst individual estimates for the household value. This option is more relevant for numeric data, where minimum and maximum values are more meaningful (e.g. highest income or lowest income amongst household members).

Simulations on the data aggregation from individuals to the household

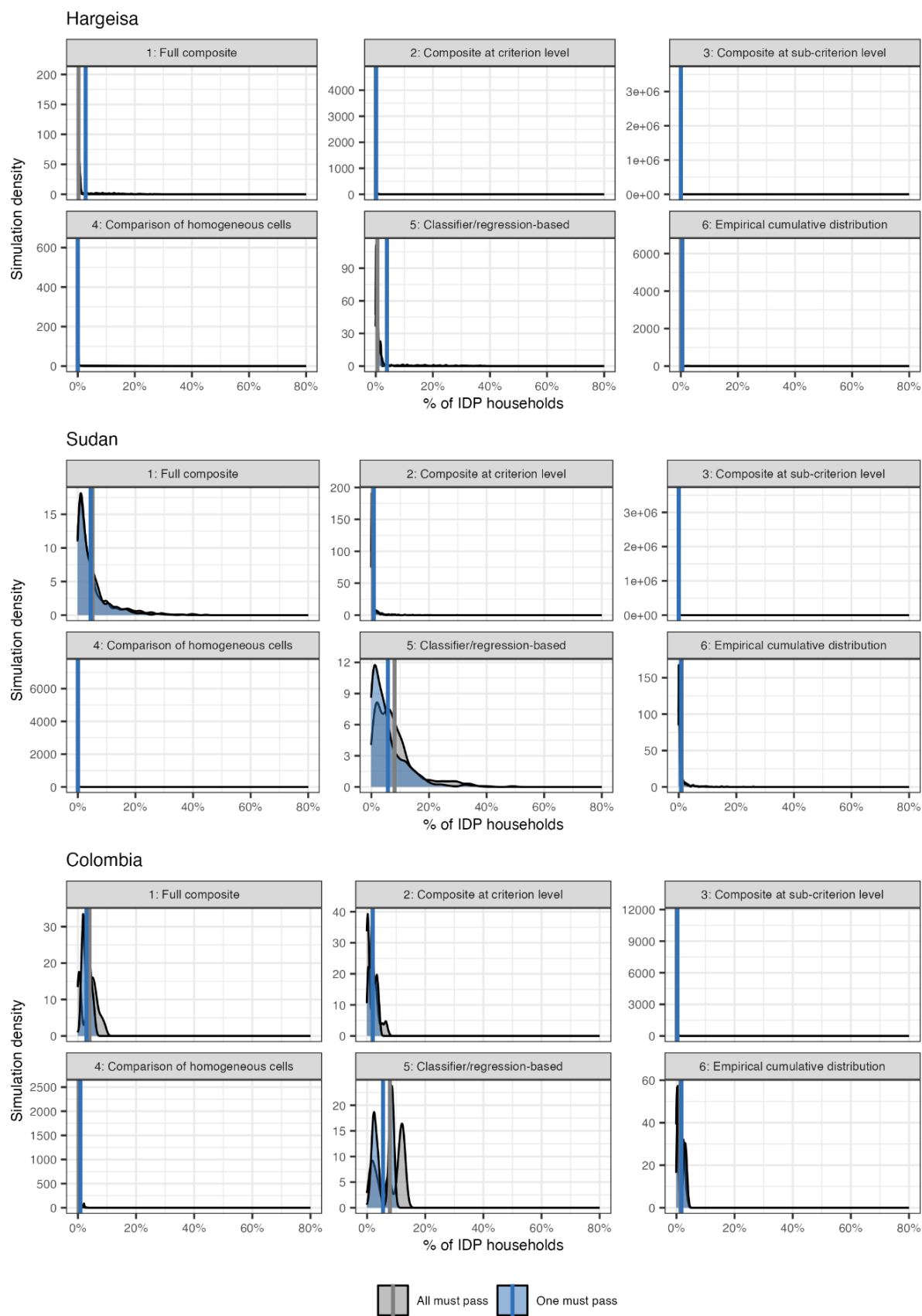
164. While many more options for data aggregation from the individual to the household level are available than the four outlined here, the most relevant choice in the context of the exit measure is likely between the choice that all applicable individuals in a household must pass a sub-criterion and at least one individual must pass a sub-criterion because these aggregation rules are meaningful for binary indicators.
165. For the simulations, the choice was made to use the “at least one” rule across all indicators when an aggregation was necessary. However, this choice has a potentially big impact on the percentage of households meeting a certain sub-criterion. To illustrate this, Table lists the 8 possible indicators for the sub-criteria on health, education and documentation that were aggregated from the individual level to the household level in the case of Hargeisa. For these indicators, the table shows the percentage of households that would pass this sub-criterion when an at least one rule is used or an all must pass rule. For example, while around 55% of households have at least one adult with access to ID documentation, only around 6.4% of households possess ID documentation for all members of the household. Given these stark differences in the performance of households depending on the aggregation rule, it is crucial that EGRISS provides clear guidance for data producers.
166. To make a more systematic comparison, Figure 16 displays the density of simulations in blue when only one household member must pass the subcriterion. In gray, the same graph displays the simulations and the resulting exits from the IDP stock if all household members must pass the subcriterion to overcome a vulnerability related to displacement. In the data setup used in these simulations, the differences between the two data aggregation rules is not very strong. This is likely due to the fact that data missingness affects the results.

Table 22: Percentage of households in Hargeisa achieving a pass on an individual sub-criterion when an at least one or an all must pass rule is used for data aggregation. Only aggregated indicators are shown.

Indicator	Percentage (At least one rule)	Percentage (All pass rule)
2.3. Immunized children	24.15	2.62
2.4 Children read/write	26.51	6.09
2.4 Children ever went to school	26.93	6.49
2.4 Children currently go to school	25.58	6.06
2.4 Children attend secondary school	2.88	0.33
5.1 ID documentation	55.29	6.41
5.1 Access to ID replacement	55.29	6.41
5.1 Birth certificate	8.04	0.53

167. On the one hand, an “all must pass” rule is stricter, leading to a lower estimate of IDP exits from the stock in the case of a full composite or regression-based metric in Hargeisa. On the other hand, the rule that one person must pass may still suffer from missing data issues, leading to less exits from the IDP stock under this rule in the case of regression-based estimates in Sudan or Colombia. Overall, the rule that all individuals must pass appears to be stricter and less easily implementable in practice.

Figure 16: Density of simulations when aggregating to the household level with an all must pass or one must pass rule



Summary of findings

168. Some of the sub-criteria for the exit measure are more naturally measured on the individual level than the household-level. For example, employment is easiest to assess on the individual-level. For indicators commonly collected on the individual level, it is hence crucial to identify an aggregation rule from the individual to the household level. The two plausible available aggregation rules for the exit measure are that all individual members must pass or that at least one member must pass for the household to be counted as achieving a certain sub-criterion. The simulations show that this choice is important on the indicator level. Aggregating data to the household level with the “at least one” rule is slightly more lenient. The approach would also partially address the problem of missing data for one person in the household as - as long as another person overcomes vulnerabilities - the household can still exit the stock. For some indicators, like employment, it might suffice that one adult household member overcomes their vulnerability and enters employment. For other indicators, a rule that all household members must enjoy their rights before this vulnerability is overcome, appears more appropriate.

Recommendation

169. It is recommended that EGRISS provides clear guidance on how the aggregation from individual level indicators to household level exits from the IDP stock should be made. This should be standardized across contexts. It would be necessary to provide a manual to compilers of official statistics on whether a positive outcome is achieved if at least one household member achieves a certain criterion or all of them. The recommendation on the appropriate aggregation rule should depend on what makes the most common sense for the chosen indicator to cover IDP vulnerabilities well. However, to allow a functioning and realistic exit measure, the application of an “at least one” rule in most indicators seems more important to avoid being implausibly restrictive on exits from the statistical stock.

Overcoming challenges 6-9: the road towards a comparable exit measure

170. As outlined in the section on the state of play on the “IRIS exit measure”, four more challenges should be addressed during the development phase of a robust and generally applicable exit measure. The challenges of statistical uncertainty, the definition of a final comparator population, changing benchmark values over time, and measurement problems due to assistance are downstream problems that have to be addressed once agreement on the challenges 1-4 is achieved. The following sections briefly outline what possible scope for standardization and recommendations are available for each downstream challenge. It is also briefly highlighted what kind of methodological analyses may be needed to overcome the challenges in an empirically informed way.

Overcoming challenge 6: Addressing statistical uncertainty in benchmark values

171. The sixth challenge is that each comparator value in the host community or the national average will come with statistical uncertainty, i.e. there is a confidence interval around the comparator point estimate. While some data collections on IDPs and their progress towards solutions will yield relatively precise estimates and can be compared to precise estimates in comparator populations, other data collections or displacement contexts may only have information available with higher uncertainty as this heavily depends on sample sizes, variation in responses, and data collection processes. For example, when a national average from a representative large-scale survey is used as a benchmark, the point estimate of the benchmark values may be more precise than estimates generated from a smaller host community sample.
172. In the IRIS, there is no clear guidance on how to account for this statistical uncertainty when IDP households are compared to the benchmark. In terms of scope for EGRIS to provide recommendations, it may be advisable that the Expert Group highlights considerations for statistical uncertainty prominently in a future revision of the recommendations while giving practical and more detailed guidance on accounting for uncertainty in the Compilers’ Manual. Reporting on statistical uncertainty is not only important to better understand what the “true” size of the exit from the IDP stock is. Statistical uncertainty is also an indication of the quality of official statistics and can help to identify displacement contexts where better and more targeted solutions assessments are required.
173. A range of computationally demanding options are available to deal with statistical uncertainty (e.g. resampling techniques). However, in the context of the exit measure, the scope for multi-step procedures to deal with statistical uncertainty is limited by the capacity of producers of official statistics. Reasonable suggestions can include to conduct statistical hypothesis testing to assess if the mean value in the IDP population and the benchmark population are significantly different from one another. Depending on the local capacity, it may also make sense to recommend that producers of official statistics use the lower bound and upper bound of the confidence interval around the benchmark estimate to provide not only one “best estimate” of IDP exits but a “lower and higher estimate”. More complex resampling techniques to deal with statistical uncertainty seem less realistic to recommend.
174. Each of the proposed metric options may require slightly different strategies to deal with uncertainty in the estimate of IDP exits from the stock. These strategies are summarised in Table 23. To sum up the table, while composite metrics for all criteria, at criterion level, or at sub-criterion level may allow for comparisons of upper and lower estimates of the IDP stock (based on variance in the samples), the metric options 4 and 5 (homogeneous cells and regression-based approaches) can directly incorporate statistical uncertainty by drawing confidence intervals around the estimates. For the option to use an empirical cumulative distribution, confidence

intervals are commonly generated by drawing subsamples from the data, making this more challenging. Overall, accounting for uncertainty is the most natural and easy to implement in the case of regression-based approaches and a comparison of homogeneous cells.

175. Once challenges 1 to 4 have been addressed, further methodological work with the use of well-sampled data collections could help to provide evidence on the scale of the difference between lower, best and highest estimate in the context of the exit measure.

Table 23: Strategies to deal with statistical uncertainty in the available metric options

Metric option	Strategy to deal with statistical uncertainty
Composite measure across all criteria, across criteria and across sub-criteria	To account for statistical uncertainty in composite indices, one would first calculate the mean estimate of the composite index for the full benchmark population. For example, for a composite measure at the criterion level, one would calculate the mean value amongst the host community for each composite index under each criterion. In a second step, one would then calculate the standard deviation of this composite index for this population. While the population mean constitutes our "best estimate", we can use the mean and the population standard deviation, to then calculate the 95% confidence interval around the mean composite index. This would give us a "lower estimate" and a "higher estimate". Following this, we would compare individual IDP households no longer simply against the best estimate, but we would also benchmark households against the lower and higher estimate. This provides the opportunity to account for the fact that the benchmark value comes with statistical uncertainty. This can be done for one index across all criteria but can also be applied to multiple composite indices at the criterion or sub-criterion level.
Comparison of homogenous cells	In the case of a homogeneous cells approach, one could calculate not only the confidence intervals around the mean of the benchmark population but also the confidence intervals around the mean of the homogeneous cells. This allows us to compare the lowest estimate in a homogeneous cell of IDP households with the lowest estimate in the benchmark population. At the same time, we can compare the best and the highest estimate in the IDP households in one cell to the best and highest estimate in the benchmark population. Again, this would result in three estimates of the IDP exit to account for statistical uncertainty in our measurement.
Classifier/regression-based approach	In a regression-based framework, each prediction whether an IDP household is closer to the other IDP households or closer to the host community comes with statistical uncertainty. The regression produces a confidence interval around the prediction that can be used to estimate a low and high estimate of the exit from the IDP stock.
Empirical cumulative distribution	In the framework of an empirical cumulative distribution, a common way to produce uncertainty estimates is to use some form of resampling or bootstrapping. In other words, instead of just calculating the exit from the IDP stock for the whole sample, one could do this for randomly selected subsamples. After repeating the calculations on various samples of the data, one can generate the mean exit from the stock as well as the high and low estimate based on the distribution of IDP exits from the stock across the various samples.

Overcoming challenge 7: Defining the comparator population

176. Regarding the challenge to define the final comparator population, the IRIS points out that *"for the purpose of official statistics, and because it is the approach often taken in this context, it is recommended that the general/national population is used as the comparison group when deciding on the targets or thresholds for scoring each sub-criteria"* and that *the average of the general/national population is used as the target*". This recommendation is very relevant as information about the general population is more often available and simplifies analysis.¹⁵ The simulations in Colombia, which were based on a comparison of IDPs to the national average rather than the host community (as in the other simulated datasets), show that this is a feasible endeavour.

¹⁵ Using the national average also lifts the burden of finding a definition for host communities that is applicable across contexts.

177. Nevertheless, further methodological work could assess if in the same displacement context, using a national average vs a host community average yields statistically significant differences in the estimate of the IDP exit. Such methodological work would help EGRISS to shape detailed guidance in the Compliers' Manual on how to produce reliable estimates when no national average is available or estimations based on the host community are more feasible. In general, the way forward is to focus on a representative sample of the national population as best possible benchmark as already stated in the IRIS.

Overcoming challenge 8: Vulnerabilities overcome merely by assistance received

178. In many displacement contexts, humanitarian assistance for IDPs, for example in the context of camps, supports displaced populations. However, this assistance provides a challenge for measuring if IDPs overcome their key displacement-related vulnerabilities for the purpose of the exit measure. IDPs that experience food security because of humanitarian assistance and not because of sustainable income sources should not exit the stock because they can still be considered as vulnerable. Hence, a standardized approach is needed to account for assistance in the exit measure. The way to factor assistance into the exit measure will likely depend on the final chosen indicators and to what extent they are affected by humanitarian assistance provided through national systems or international humanitarian partners. Broad consultations and literature reviews could be a next step to develop methodological and analytical guidance on the scale of this problem and possible solutions to the challenge.

Overcoming challenge 9: Addressing changing benchmark values over time

179. As challenges 6 and 7 have already indicated, the choice of the right benchmark value, which comes with uncertainty and has to be well-targeted, is crucial. This is also reflected in challenge 8: For the exit measure, the benchmark value should be understood as a moving target. Over time, the benchmark values may increase or decrease. This will affect the number of IDP households exiting the stock. IDPs may exit the stock without an improvement of their situation because the benchmark decreases. IDPs previously exiting the stock may re-enter the stock without new displacement event when the benchmark increases.
180. At this stage, it is not clear to what extent the empirical shifts in benchmark values has a significant effect on exits from the IDP stock. To further assess this, a methodologically driven analysis could investigate panel data on IDP solutions. A panel analysis of the same IDP households could help to assess if shifting benchmark values affect the IDP population. This analysis makes more sense to conduct after the challenges 1-5 have been addressed to be able to assess the volatility of the exit measure across time.

Going “full circle”: a final round of analysis implementing all of the above recommendations

181. The *International Recommendations on IDP statistics* (IRIS) provide an important framework to better capture IDP statistics and outline when a country’s stock of IDPs can be reduced. The fundamental idea that IDP households can be taken out of the stock for statistical purposes when they overcome key displacement-related vulnerabilities and perform as well or even better than the national average (or the host community) is important. However, more work is needed to develop a fully applicable statistical measure for IDP exits. This report summarises key methodological gaps and challenges that should be addressed before a workable exit measure can be applied by producers of official statistics.
182. The main challenges are certainly how to implement a feasible measure that allows comparisons between IDPs and a comparator population while also ensuring that IDP households overcome key dimensions of vulnerability. Next to the selection of an appropriate metric - such as various composite indices, a comparison of population cells, regression-based approaches or approaches based on empirical cumulative distributions - feasible statistical indicators must be chosen. Other challenges include how to deal with rights-based indicators, missing data, and data aggregation to the household level. Finally, more methodological work is needed to understand how statistical uncertainty, the definition of the comparator population, assistance to IDPs and changing benchmarks over time affect the estimates of an exit measure.
183. This report demonstrates the empirical performance of different metric options and indicators to propose possible ways forward. It is recommended that EGRISS uses the results of this work to make clear recommendations on how the statistical measure could be implemented to take IDPs out of stock data. Further methodological work on a composite measure at the criterion level or a homogeneous cells approach is important. Alternative metric options would be to use a cumulative empirical distribution. The report also proposes indicators to standardize the measurement of the exit measure across contexts, heavily relying on the SDG indicators and the Durable Solutions Indicator Library.
184. One key challenge is to develop a exit measure that is not by design implausibly prohibitive on exits from the statistical stock of IDPs. The simulations in this report demonstrate that it is important to reframe housing, land and property access as tenure security, to treat missing data by design as non-vulnerability, and to aggregate data from individuals to the household level in a sensible way. These steps are important to avoid implausibly prohibitive design features of the exit measure and make the measure widely applicable.
185. To provide some evidence on whether these recommendations improve estimates of the IDP exit from the stock, Figure 17 shows the results when the recommendations are implemented as far as possible. In the figure, the data was manipulated ...
 - ... to implement the recommended indicators,
 - ... to reframe access to HLP compensation as tenure security,
 - ... to define missing data by design as non-vulnerability,
 - ... to aggregate individual level data to the household by a “one must pass” rule.

Figure 17: Percentage of IDPs exiting the stock if all recommendations are implemented



186. Compared to previous simulations, the recommended steps lead to estimates of the exit from the IDP stock which are better aligned with the underlying principle that IDPs performing similar to the benchmark should be able to exit the stock and do not have to significantly *outperform* the comparator population. For example, under the empirical distribution approach, 27% of the Colombian IDP population are taken out of the stock for statistical purposes while about 13% achieve this in Hargeisa and 6% in El Fasher. These final simulations, implementing the full circle of recommendations, demonstrate that the composite measure at the sub-criterion level and the homogeneous cell approach are implausibly restrictive by design.
187. Further methodological work could assess the remaining challenges in the exit measure or implement the suggested indicators in various country-examples to produce better estimates of the exit from IDP stock for other countries and over time. EGRISS may want to consider standardising the selection of indicators in particular as a feasible next step while more methodological evidence is gathered. Lastly, it is crucial to discuss the feasibility of the proposed metric and indicator options with the producers of official statistics, such as National Statistical Offices, to get a sense of where further evidence and guidelines are necessary.

Annexes

Annex I: Detailed simulations methodology

A.1. Overall, the simulation approach 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. The data was then aggregated to the household level. 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 stock 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 exit measure. The following sections outline the details of the methodology.

Step 1: Identifying the universe of potential indicators

A.2. 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.¹⁶ 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.

A.3. 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.¹⁷ See Table A.2 for a breakdown of the possible indicators per sub-criterion.

A.4. 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

A.5. 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 exercise is the *Internal Displacement Profiling in Hargeisa*¹⁸ that was conducted

¹⁶ 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.

¹⁷ 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.

¹⁸ UNHCR, 2015, “Internal Displacement Profiling in Hargeisa”, <https://reliefweb.int/report/somalia/internal-displacement-profiling-hargeisa-december-2015>

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*¹⁹ 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*²⁰ 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.²¹ Finally, the case selection was complemented with the *Living Standards Measurement Study in Colombia 2019* (LSMS)²². 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 datasets - summarised in Table A.1 - 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.²³

Table A.1. 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

- A.6. The datasets were then checked to identify whether they contained the relevant 52 indicators from the Durable Solutions Library. Table A.2 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.²⁴
- A.7. 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 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 1552 possible indicator combinations that can be assessed in simulations.hhh

¹⁹ World Bank, 2018, “Profile of Internally Displaced Persons in North-East Nigeria 2018”, <https://microdata.worldbank.org/index.php/catalog/3410>

²⁰ 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>

²¹ Covering hosts and IDPs is crucial for this methodological assessment to facilitate the task of finding a comparator population.

²² DANE, 2019, <http://microdatos.dane.gov.co/index.php/catalog/678/study-description>

²³ If survey weights were available (not the case in Hargeisa), they are considered in the simulations to ensure representativeness.

²⁴ 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.

- A.8. In Nigeria, 39 indicators could be identified. However, since the indicators in Nigeria are more concentrated on two sub-criteria, only 12474 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.
- A.9. In Sudan, the profiling in El Fasher yielded 31 indicators that can be combined to 34560 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.²⁵
- A.10. In Colombia, the LSMS contains a total of 25 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.

Table A.2: Potential and empirically existent indicators for durable solutions

Sub-criterion 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	9	5	2
1.2 Freedom of movement	1	1	1	0	0
2.1 Food security	1	2	1	3	1
2.2 Shelter and housing	8	6	11	6	4
2.3 Medical services	4	3	2	2	2
2.4 Education	3	4	3	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	4	3	4	0
5.1 Documentation	4	3	1	2	1
Number of possible combinations		15552	12474	34560	1440

Step 3: Preparing the indicators and indices for simulation

- A.11. 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, this should not only be done for one set of 10 indicators but for at least 1000 combinations (if available) in the data. To iterate through at least 1000 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 outcome. For a simple

²⁵ 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.

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.²⁶

A.12. 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. 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, for example, 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 into subgroups:

Hargeisa:

- **Gender of the household head** (female, male)
- **Clan of the household** (Haw, Isa, Rah, Dar, Dir, Gav, 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 2000, between 2000 and 2010, after 2010)

Colombia:

- **Gender** (female, male)
- **Age group** (Children, Youth, Adults, Elderly)
- **Marital status** (Not married, with partner, married, divorced/separated, single, widowed, missing)
- **Region of displacement** (34 admin areas)
- **Duration of displacement** (In 5-year steps)
- **Ethnic minority** (yes, no)
- **Peasant** (yes, no)

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)

Sudan:

- **Housing type** (tent, dwelling, tukul, flat, house, incomplete)
- **Year of arrival**
- **Number of times displaced**
- **Community location** (Same district, same state, different state, outside of Sudan)
- **Arrived in company** (Alone, With family, In larger group)

A.13. 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.

A.14. After the indicators were coded, all indicators were initially aggregated from the individual to the household level with a rule that one person performs positively, the whole household of displaced persons achieve a durable solution. The note discusses other options to aggregate the data.

Step 4: Iteratively simulating the indicator and metric choice

A.15. After identifying the indicators for this assessment in each survey and preparing the indicators for analysis, iterative simulations were run. For each dataset, 1000 possible combinations of 10

²⁶ 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.

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.

Annex II: Limitations and methodological concerns

A.16. It is important to acknowledge some key limitation in this simulation study that may be relevant for EGRISS to make further decisions on the IRIS exit measure:

1. **Sample:** Only four datasets were used to test the exit 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 exit measure in mind.
2. **Data quality:** While the datasets were useful because they included multiple indicators that could be used to operationalize the IRIS exit 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 exit 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 exit measure, it would be beneficial to streamline the elements needed 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 exit 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 outlined in the IRIS 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 exit 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.

Annex III: Full mapping of all available indicators for the exit measure

Sub-criteria (IRIS)	Potential indicators from the inter-agency Durable Solution Library	IDP Profiling in Hargeisa (UNHCR 2015)	IDP Profiling in Nigeria (World Bank 2018)	IDP Profiling in El Fasher (JIPS 2019)	LSMS in Colombia (DANE 2019)
1.1 Victims of violence	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.				
	Target population who think it likely they will experience serious consequences due to a hazard.	Target population is worried about being exposed to theft, crime or vandalism in their place of residence.			Target population satisfied with the current level of security.
	Target population who feel safe walking alone around the area they live (during day or night).	Target population is feeling unsafe or insecure in their place of residence.	Target population feeling very or moderately safe.		
			Target population feeling very or moderately safe walking at night.	Target population feeling very or moderately safe walking at night.	
			Target population feeling very or moderately safe walking during the day.	Target population feeling very or moderately safe walking during the day.	
	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).		Target population experiencing non-physical or physical harm in the last 12 months.	Target population experiencing a robbery.	
	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).	Target population who experienced victimizing events in their place of residence in the past 12 months.	Target population who have experienced harm and have reported it in the last 12 months.	Target population who have experienced harm and have reported it to the police.	
	Target population who were affected by hazard in the previous 12 months (or since time of displacement, if displaced for less than 12 months).				Target population who experienced a natural disaster in the past 12 months.
	Target population who experienced violence in the previous 12 months, who reported their victimization to competent authorities or other officially recognized conflict resolution mechanism.			Target population who report thefts or disputes to formal conflict resolution mechanisms.	

			Target population who find it very easy or somewhat easy to access dispute resolution mechanisms.		
			Target population that find dispute resolution very or moderately effective.		
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.
				Target population having to borrow money for food.	
		Number of meals eaten per day.			
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 access to land and renting or owning it legally.	Target population legally owning the property.	Target population legally occupying their dwelling.
	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 overcrowded housing/shelter (> X persons per sleeping room).	Target population living in overcrowded housing/shelter (> X persons per sleeping room).	Target population living in overcrowded housing (>3 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 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 urban population living in slums, informal settlements, or inadequate housing.		Target population squatting.		
	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 access to improved drinking water sources.	Target population with access to clean drinking water.
			Target population with no obstacles for water access.		
	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 improved sanitation facilities.	Target population with improved sanitation facilities.	Target population with improved sanitation facilities.
		Target population with bath/shower in household			
2.3 Medical services	Target population covered by essential health services.		Distance to health facilities in hours.	Distance to health facilities in hours.	Target population in possession of health insurance.
	Target population who accessed essential health care services (including mental health care) the last time they needed it in the past 12 months.	Target population with access to essential health care when needed.	Target population that access essential health care when needed.	Target population satisfied with health care.	Target population satisfied with current level of health.
	Births within target population attended by skilled health personnel within the past 12 months (% of total births taken place within the past 12 months).	Target population with births or pregnancies attended by skilled health personnel.			
	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 immunized children.			

2.4 Education	Primary school net attendance ratio in targeted population (% of children of primary school age in target population).	Target population of school age that can read and write.	Time to next education facility.	Time to next education facility.	Target population that is literate.
		Target population of school age that has ever attended school (primary and secondary).	Target population being satisfied with primary education.	Target population of school age that has ever attended school.	Target population attending school.
		Target population of school age that is currently attending school.		Target population of school age that is currently attending school.	Target population attending an officially-recognized educational establishment.
	Secondary school net attendance ration in target population (% of children of secondary school age in target population).	Target population of school age that is currently attending or have completed secondary school or university.			
	Target population who own a mobile phone.	Target population which owns a mobile phone.			
3.1 Employment and livelihoods	Target population employed in formal and informal sector (employment rate).	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.
	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 aged 15-24 years not in education, employment or training.				Labor force participation.
	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.	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 capable of managing unexpected expenses without borrowing money or receiving help from others.		Target population with written employment contract.
				Target population in possession of a written tenancy agreement.
	Target population who in the last 12 months was not able to pay for basic expenses.	Target population who was not able to pay house rental / services in the last 6 months.		Target population defaulting on utility bills.
	Target population who in the last 12 months obtained a loan to cover basic expenses.			Target population satisfied with current level of income.
	Target population's average expenditure against average total expenditures, per capita.		Target population consuming more than average.	
	Ratio of average food expenditures against average total expenditures, per capita.		Ratio of food consumption against total consumption.	Target population requiring child labor.
	Target population below the poverty line.		Target population below 1.9 USD PPP 2011 Poverty Line.	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 3.2 USD PPP 2011 Poverty Line.
	Target population who own productive assets by type of assets.	Average number of assets owned by target population.		
	Target population with access to markets.		Distance to market in hours.	Distance to market in hours.
	Target population where at least one person in household has a bank account.		Target population with access to a bank account.	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 legally owning the property.	
	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.	Target population with lost HLP who have accessed restitution or compensation mechanisms.	Target population with access to compensation mechanisms.		
	Target population with housing, land and property left behind who have had their claims to assets (incl. land and property) resolved.	Target population with lost HLP who have had their claims resolved and enforced.			
	Target population with housing, land and property left behind who have had their claims to assets (incl. land and property) enforced.				
	Agricultural target population with ownership or secure rights over agricultural land.				
	Agricultural target population with use rights to agricultural land.				
5.1. Documentation	Target population currently in possession of valid birth certificates, national ID cards or other personal identification documents relevant to the context.	Target population in possession of birth certificates.		Target population in possession of birth certificates.	
	Target population with other personal documentation necessary to accessing their rights.	Target population with personal documentation or access to mechanisms to replace them if they are lost.	Target population that have not lost their documents and know how to replace them if missing.	Target population with personal documentation.	Target population with personal documentation.
	Target population registered by authorities as Internally Displaced Persons – if relevant to context.				
	Children under 5 years of age in target population whose births have been registered with a civil authority.	Children in target population with a birth certificate or registration.			

Annex IV: Presence and implementation of proposed indicators for the IRIS exit measure

Indicator	Implementation in Hargeisa	Implementation in Nigeria	Implementation in El Fasher	Implementation in Colombia
1.1 Victims of violence: SDG 16.1.4	Fully implemented	Fully implemented	Fully implemented	Approximated with satisfaction with current security situation
1.2 Freedom of movement: DS 1.4.1	Approximated with ability to visit public places	Fully implemented	Not present	Not present
2.1 Food security: SDG 2.1.2	Not available, approximated with ability to pay for food	Ordinal scale of food security rather than Rasch model; shorter reference frame of 6 months	Ordinal scale of food security rather than Rasch model; shorter reference frame of 7 days	Approximated with living above food security line
2.2 Shelter and housing: SDG 11.1.1	Simplified questions on improved water sources and facilities, no indicator for structural permanency of shelter	No indicator for the sub-component "location of housing", security of tenure not always asked	No indicator for the sub-component "location of housing"	No indicator for the sub-component "location of housing" and "permanency of housing"
2.3 Medical services: DS 2.1.8	Shorter reference frame of 3 months	Shorter reference frame of 6 months	No indicator for medical access, satisfaction with medical services used instead	No indicator for medical access, satisfaction with medical services used instead
2.4 Education: SDG 4.1.2	No measure of school completion: approximated with children ever attended school	No measure of school completion/attendance: approximated with current school attendance	No measure of school completion: approximated with children ever attended school	No measure of school completion: approximated with children ever attended school
3.1 Employment and livelihoods: SDG 8.5.2	No measure of unemployment: Approximated with breadwinner in the family	No measure of unemployment: approximated with income sources from regular work	No measure of unemployment: Approximated with reporting of having a paid job	Fully implemented
3.2 Economic security: DS 3.2.2	No measure of poverty: Approximated with capability to cover unexpected expenses	Below 1.9 USD PPP 2011 poverty line used instead of national poverty line	Below 1.9 USD PPP 2011 poverty line used instead of national poverty line	Fully implemented
4.1 Security of tenure: SDG 1.4.2	Fully implemented	Fully implemented	Fully implemented	Fully implemented
5.1 Documentation: DS 5.1.1	Fully implemented	Approximated with question whether they lost documentation in displacement	Fully implemented	Fully implemented

Annex V: Difficulty of passing proposed indicators under challenge 2

- A.17. Another question in regards to the performance of the proposed SDG and DS indicators might be if all indicators contribute equally to the exit measure. In other words, are there any proposed indicators that are significantly harder or easier to pass than other indicators and might need to be revised? Table A.3 summarises how many IDPs “pass” each proposed indicator in each country context without the comparison to any benchmark. For instance, around 96.5% of the IDP households report that they feel safe walking in their area in Nigeria (recommended SDG indicator 16.1.4). In the table, indicators that were extremely difficult to pass – with less than 50% of IDP households achieving a “one” for “not vulnerable” in a binary indicator – are marked in grey. In particular the proposed indicator for *2.2 Shelter and housing*, which is SDG indicator 11.1.1, is difficult to pass across many contexts. The indicator is characterized by a multitude of sub-indicators. One of them is the durability of housing structures. With many of the IDPs in the example datasets living in temporary tents provided by UNHCR, they often do not pass this criterion, explaining low exits. This is important to consider given that tenure security, another sub-indicator of SDG indicator 16.1.4 that is hard to pass, also is the chosen indicator for sub-criterion 4.1.
- A.18. Note, however, that this comparison of indicators must be understood as a relative concept: While shelter and housing might not be sufficiently durable for IDPs, this might also be the case for the host community to which the IDP households are compared. In the table, one should also note the high percentage of missingness for some of the proposed indicator, for example on education. Overall, many of the proposed indicators from the SDG indicators or the Durable Solutions Library could also not be fully implemented or had to be adapted.

Table A.3 Difficulty of achieving a pass per proposed indicator for the exit measure

Indicator	Hargeisa	Nigeria	El Fasher	Colombia
1.1 Victims of violence: SDG 16.1.4	62.06% pass	96.52% pass	95.21% pass	62.06% pass
1.2 Freedom of movement: DS 1.4.1	Not present	94.73% pass	Not present	Not present
2.1 Food security: SDG 2.1.2	85.96% pass	38.69% pass	30.36% pass	85.96% pass
2.2 Shelter & housing: SDG 11.1.1	84.77% pass	2.35% pass	0.07% pass	84.77% pass
2.3 Medical services: DS 2.1.8	70.34% pass	98.58% pass	63.84% pass	70.34% pass
2.4 Education: SDG 4.1.2	91.99% pass	93.75% pass	22.12% pass	91.99% pass
3.1 Employment & livelihoods: SDG 8.5.2	97.49% pass	82.31% pass	0.6% pass	97.49% pass
3.2 Economic security: DS 3.2	52.87% pass	10.38% pass	99.33% pass	52.87% pass
4.1 Security of tenure: SDG 1.4	95.64% pass	3.54% pass	62.59% pass	95.64% pass
5.1 Documentation: DS 5.1.1	99.94% pass	78.41% pass	66.8% pass	99.94% pass