Research Methodology Note
Assessment of HARD-TO-REACH districts (HSM)
SOM1901
Somalia

March 2023
V2

REACH *Informing more effective humanitarian action*

1. Executive Summary

Country of intervention	Somalia						
Type of Emergency	x	Natural disaster	x	Conflict	□	Other *(specify)*	
Type of Crisis	□	Sudden onset	x	Slow onset	x	Protracted	
Mandating Body/ Agency	Office for the Coordination of Humanitarian Affairs (OCHA), Inter-Cluster Coordination Group						
IMPACT Project Code	27ELN						
Overall Research Timeframe (from research design to final outputs)	01/02/2023 to 15/05/2023						

Research Timeframe *(from research design to final outputs / M&E)*	1. Pilot/ training: 27/02/2023 (Training of Trainers – ToT); 01/03/2023 and 02/03/2023 (Enumerator Training)	6. Start output drafting: 06/04/2023
	2. Start collect data: 06/03/2023	7. Outputs sent for validation: 01/05/2023
	3. Data collected: 21/03/2023	8. Outputs published: 15/05/2023
	4. Data sent for validation: 28/03/2023 (cleaned and aggregated)	9. Final presentation: May 2023
	5. Data analysis on validated data: 04/04/2023	

Humanitarian milestones *Specify **what** will the assessment inform and **when** e.g. The shelter cluster will use this data to draft its Revised Flash Appeal;*	**Milestone**		**Deadline**
	X	Somalia Humanitarian Fund (SHF) Allocations	01 /12/2023
	X	Humanitarian Needs Overview (HNO)	01 /12/2023
	□	Cluster plan/strategy	01 /12/2023
	□	NGO platform plan/strategy	01 /12/2023

Audience Type & Dissemination *Specify **who** will the assessment inform and **how** you will disseminate to inform the audience*	Audience type		Dissemination
	x Strategic		**x** General Product Mailing (e.g. mail to Cluster leads; Donors)
	x Programmatic		□ Cluster Mailing (Education, Shelter and WASH) and presentation of findings at next cluster meeting
	□ Operational		**x** Presentation of findings (to Inter-Cluster Coordination Group - ICCG)
	□ [Other, Specify]		**x** Website Dissemination (Relief Web & REACH Resource Centre)
			□ [Other, Specify]
Detailed dissemination plan required	□	Yes	x	No Results will be shared proactively with coordination bodies, at the national for wider sharing with operational partners.
General Objective	To inform humanitarian planning and context by providing information on emergency needs (Food Security & Livelihoods, Health, and WASH), displacement dynamics and access to services in HARD-TO-REACH districts ahead of post Gu discussions and the Post-Gu Integrated Phase Classification (IPC).		
Specific Objective(s)	- To understand emergency needs in HARD-TO-REACH districts. - To identify the factors impacting displacement dynamics from the HARD-TO-REACH districts. - To understand what services are accessible to households in HARD-TO-REACH districts and the barriers that impede access to services in HARD-TO-REACH districts. - To identify primary livelihoods in HARD-TO-REACH districts and how climatic hazards (including lack of rain, drought) and economic conditions influence primary livelihoods in HARD-TO-REACH districts.		
Research Questions	- Are some population groups more likely to move from Hard-to-Reach districts? What are the factors that DRIVE displacement from the HARD-TO-REACH districts? What are the factors that PREVENT displacement from the HARD-TO-REACH districts? Are some population groups unable to move, but would like to do so? If yes, whom? - What are the food security needs and coping strategies of households in HARD-TO-REACH districts? To which foods, sources of livelihood and markets do households in HARD-TO-REACH districts have access to? How are climatic hazards and economic conditions affecting access to food and livelihoods? Are some population groups more food insecure? If so, whom and why? - What are the needs and coping strategies of the households in HARD-TO-REACH districts regarding Water, Sanitation and Hygiene (WASH)? What are the constraints to access to WASH? How does the WASH situation affect the vulnerability of households in HARD-TO-REACH districts? What are the impacts of climatic hazards and economic conditions on WASH in HARD-TO-REACH districts? Are some population groups in higher WASH need? If so, whom and why? - What are the needs and coping strategies of the households in HARD-TO-REACH districts regarding Health? What are the constraints to access to Health services How does the health situation affect the vulnerability of households in HARD-TO-REACH districts? What are the impacts of climatic hazards and economic conditions on WASH in HARD-TO-REACH districts? Are some population groups in higher Health need? If so, whom and why? - What are the needs and coping strategies of the populations in Hard-to-Reach areas regarding Protection?		

	• To which services and types of humanitarian assistance, if any, do households in HARD-TO-REACH districts have access to? What are the constraints to accessing services and humanitarian assistance?			
Geographic Coverage	For the March 2023 round, 23 districts will be covered through the HARD-TO-REACH approach. These districts present either "extreme access constraints" or "high access constraints" by the OCHA Humanitarian Access Working Group (HAG). The final list of districts can be found in section 3 - Methodology.			
Secondary data sources	- Somalia Livelihood Zones - Map - Somalia access severity map - OCHA, September 2022 - Humanitarian Response Plan (HRP), 2023 - Somalia - Humanitarian Needs Overview (HNO), 2023 - Somalia - IOM DTM (Displacement Tracking Matrix) - SWALIM (Somalia Water and Land Information) - Satellite imagery, NDVI - Food Early Warning Systems Network (FEWSNET), Somalia, publications - Somalia Acute Food Insecurity Snapshot l October 2022 - June 2023			
Population(s)	□	IDPs in camp	x	IDPs in informal sites
Select all that apply	□	IDPs in host communities	□	IDPs
	□	Refugees in camp	□	Refugees in informal sites
	□	Refugees in host communities	□	Refugees [Other, Specify]
	x	Host communities (Households in HtR areas)	□	[Other, Specify]
Data collection tool(s)	x	Structured (Quantitative)	□	Semi-structured (Qualitative)
	Sampling method		**Data collection method**	
Structured data collection tool *Quantitative tool*	**x Purposive** □ Probability / Simple random □ Probability / Stratified simple random □ Probability / Cluster sampling □ Probability / Stratified cluster sampling **x Snowballing**		**x Key informant interview (Target #): 2271 KI interviews**[1] **(Minimum 2 KIs per settlement, maximum 3)** □ Group discussion (Target #):_ _ _ _ _ □ Household interview (Target #):_ _ _ _ _ x Individual interview (Target #): 2271 □ Direct observations (Target #):_ _ _ _ _ □ [Other, Specify] (Target #):_ _ _ _ _	
Target level of precision if probability sampling	NA		NA	
Data management platform(s)	x	IMPACT	□	UNHCR
	□	[Other, Specify]		

---

[1] Four teams of enumerators (3 teams with 4 enumerators each, 1 team with 8 enumerators) will be collecting quantitative data over a period of 15 working days. Three teams of 2 enumerators each will be collecting qualitative data over a period of 2 days. In total, 26 enumerators will be involved in data collection, for a period of 15 (+ 1 if required) days.

Expected ouput type(s)	x	Situation overview #: 1	□	Report #: _ _	□	Profile #: _ _
	□	Presentation (Preliminary findings) #: _ _	x	Presentation (Final) #: 4 (TBD) – at the regional level	□	Factsheet #: _ _
	□	Interactive dashboard #: _	□	Webmap #: _ _	x	Map #: 5 (TBD)
	□	[Other, Specify] #: _ _				
Access	x	Public (available on REACH resource center and other humanitarian platforms)				
	□	Restricted (bilateral dissemination only upon agreed dissemination list, no publication on REACH or other platforms)				
Visibility *Specify which **logos** should be on outputs*	*REACH*					
	Donor: USAID					
	Coordination Framework: NA					
	Partners: OCHA					

2. Rationale

2.1 Background

Somalia's prolonged, complex, and multi-faceted humanitarian crisis is characterised by ongoing conflict, climate-related shocks, communicable disease outbreaks and fragile social protection mechanisms. Droughts in a fifth failed rainy season continue to exacerbate the effects of periodic natural disasters and insecurity caused by armed conflicts. The recent escalation of military offensives against Al Shabaab has significant humanitarian consequences including reprisal attacks, increased displacement and implications on humanitarian access. Overall, it is expected that up to 450,000 additional civilians will be newly displaced due to conflict in 2023, while several protection concerns are likely to persist including indiscriminate attacks against civilians, destruction of civilian infrastructure, recruitment, freedom of movement restrictions and widespread family separation[2]. With possibly continuing rising food prices, the risk of localized famine is heightened in several areas in Somalia. An estimated 8.2 million people will require multisectoral humanitarian assistance in 2023, a 400,000 increase from 2022. The number of people requiring humanitarian assistance has been increasing since 2020, the first season of the failed rains. Humanitarian needs grow both in magnitude and severity denoting the significant deterioration of the situation in Somalia. Of the 8.2 million people in need, 6.3 million people (77 per cent) are non-displaced shock-affected people. The majority of the people in need (60 per cent) live in rural areas. Displaced people make up 23 per cent of the people in need. IDPs, especially the newly displaced, have more severe needs than any other population group. It is worth nothing that 80 per cent of those displaced are women and children.[3]

While information about the severity of needs in accessible areas is often available, thanks to partners' assessment efforts in the context of rapid assessments and periodic country-wide assessments[4], information on needs in HARD-TO-REACH districts remains very limited. Furthermore, while country-wide analyses on climatic shocks are available, these are rarely cross-referenced with primary data on households' needs in the field.

The goal of the assessment is to draw attention to the severity of the needs in HARD-TO-REACH districts, and to demonstrate that if the situation is not monitored, it may deteriorate even worse especially taking into account the limited humanitarian interventions available for those areas. Despite the fact that different humanitarian actors have done ad-hoc sectoral assessments in various parts of the South-Central regions, research that would give a regular multi-sectoral overview of the humanitarian situation in those areas has so

---

[2] OCHA, Somalia: Humanitarian Response Plan – February 2023
[3] OCHA, Somalia: The Humanitarian Needs Overview, February 2023
[4] REACH Somalia MSNA Results Table, Somali Joint Market Monitoring Initiative (JMMI) Situation Overview

far been lacking. The March 2023 round of data collection will also continue to focus on the influence of climate hazards (including lack of rain, drought). Analysis from last round indicates that water sources are drying up, water prices are rising, and there is a definitive impact on agricultural and livestock activities which is placing immense pressure on already vulnerable communities' ability to access food and water. People in these areas are exhausting their coping mechanisms with displacement as the ultimate result, with protection and security incidents rising in prevalence in drought impacted areas.[5]

2.2 Intended impact

This round of the HARD-TO-REACH assessment will happen in March 2023, after the post *Deyr* harvest season. The assessment will contribute to the IPC to better understand evolving emergency needs (Food Security & Livelihoods, Health and WASH). This assessment will also help inform displacement dynamics and the influence of climatic hazards/economic conditions within HARD-TO-REACH districts which can then inform humanitarian actors (clusters, donors and partners including Protection, Shelter and CCM clusters in Somalia) when planning for the post Gu rain season and the post Gu IPC.

3. Methodology

3.1 Methodology overview

The HARD-TO-REACH methodology will follow a different procedure than rounds prior. The March 2023 HARD-TO-REACH cycle will aim to employ two different approaches by strengthening the area of knowledge(AoK) KII methodology with the use of mobile phone interviews to capture data from KIs who are living in HARD-TO-REACH districts at the time of data collection. Mobile phone interviews are to be acquired via snowball through the AoK KIs interviewed in both this round and the OCT 22 – NOV 22 round, and through external contacts. HARD-TO-REACH will continue to employ the Area of Knowledge (AOK) methodology, which offers settlement-level data on needs to shape the response in a situation where direct household surveys are currently limited or unavailable due to security constraints. The AoK approach will consist of identifying key informants (KIs) who can testify about the humanitarian needs of a specific area. These KIs are identified based on their knowledge of the specific area that is being assessed, either because they have been displaced from this area recently or because they travel a lot in this area, or because they still have family or friends residing in the area that they can contact. This technique is meant to support strategic planning and contribute to a more focused and evidence-based humanitarian response in areas where humanitarian workers cannot go on a regular basis due to logistics and security constraints.

Data will be collected from the 06th, March 2023 to the 21st, March 2023, by the enumerator teams in the informal or formal sites[6] around the HARD-TO-REACH districts identified as part of the Humanitarian Access Working Group (AWG) under the Humanitarian Coordination Team (HCT), while some enumerators will be assigned to take on the mobile phone interviews. Both groups are being supervised by one field officer in each of the four base locations – Baidoa, Kismayo, Garowe and Mogadishu. IDP camps that received new arrivals from HARD-TO-REACH settlements[7] in the previous three months, or new IDP settlements (set up in the last three months by people who arrived from HARD-TO-REACH districts) will be identified through REACH field team, local authorities, partners (CCCM cluster, IOM DTM), and REACH field networks.

The methodology is articulated into a quantitative component, whereby enumerators will conduct interviews with KIs (target of 2,271 interviews in total), pre-identified by field officers (FOs), through a structured KOBO questionnaire.

---

[5] REACH, Hard-to-reach Assessment, formatted analysis, October-November 2022

[6] Site where more than 15 displaced households have settled collectively. Sites not built to accommodate people, but serving that purpose, set up on state-owned or private land/buildings – CCCM Cluster Somalia.

[7] Settlements are eligible to be covered by this assessment if they are listed under OCHA's directory of inaccessible settlements. However, additional settlements may become eligible based on REACH Field Officers' feedback from the field; this is notably the case when the shifting access situation results in new locations becoming inaccessible shortly before, or during, data collection.

3.2 **Population of interest**

Geographical area assessed: The assessment targets the inaccessible areas of Somalia that are located within the inaccessible districts identified as part of the MSNA assessment (see table 1 below).

Table 1 – List of districts presenting extreme/high access constraints or considered inaccessible, and the data collection method used in 2022:

S/N	State	Region	District	OCHA Access list	Covered in DSA 2022	Covered in HARD-TO-REACH OCT 2022	Covered in MSNA 2022	Targeted for HARD-TO-REACH March 2023
1.	Hirshabelle	Middle Shabelle	Adan Yabaal	Extreme access constraints	No	Yes	No	Yes
2.	Somaliland/Puntland	Sanaag	Badhan	High access constraints	No	No	No	Yes
3.	Hirshabelle	Middle Shabele	Balcad	High access constraints	Yes	No	Yes	Yes
4.	Jubaland	Middle Juba	Bu'aale	Extreme access constraints	No	Yes	No	Yes
5.	Hirshabelle	Hiraan	Bulo Burto	High access constraints	No	Yes	No	Yes
6.	SWS	Bay	Buur Xakaba	High access constraints	No	No	Yes	Yes
7.	Galmudug	Galgaduud	Ceel Buur	High access constraints	No	Yes	No	Yes
8.	Galmudug	Galgaduud	Ceel Dheer	Extreme access constraints	No	Yes	No	Yes
9.	Jubaland	Gedo	Ceel waaq	High access constraints	Yes	No	Yes	Yes
10.	SWS	Bay	Diinsoor	High access constraints	Yes	No	Yes	Yes
11.	Hirshabele	Hiraan	Jalalaqsi	High access constraints	No	No	Yes	Yes
12.	Jubaland	Lower Juba	Jamaame	High access constraints	No	Yes	No	Yes
13.	Jubaland	Middle Juba	Jilib	Extreme access constraints	No	Yes	No	Yes
14.	SWS	Lower Shabelle	Kurtunwaarey	High access constraints	No	Yes	No	Yes
15.	Puntland	Bari	Qandala	High access constraints	No	Yes	No	Yes
16.	SWS	Bay	Qansax Dheere	High access constraints	Yes	No	Yes	Yes
17.	SWS	Bakool	Rab Dhuure	High access constraints	No	Yes	No	Yes
18.	SWS	Lower Shabelle	Sablaale	Extreme access constraints	No	Yes	No	Yes
19.	SWS	Bakool	Tayeeglow	Extreme access constraints	No	Yes	No	Yes
20.	Galmudug	Mudug	Xarardheere	High access constraints	No	Yes	No	Yes
21.	Jubaland	Middle Juba	Saakow	Extreme access constraints	No	Yes	No	Yes
22.	SWS	Bakool	Xudur	High access constraints	Yes	No	Yes	Yes
23.	SWS	Lower Shabelle	Wajid	High access constraints	Yes	No	Yes	Yes

Population assessed: The target population for this assessment is the host community households currently living in HARD-TO-REACH districts and HARD-TO-REACH settlements around these districts.

Unit of measurement and analysis: The unit of measurement and analysis will be the settlement. However, given that some indicators cannot be collected at the settlement level, some individual-level questions will be also asked.[8]

3.3. Secondary data review

In addition to the secondary data outlined in the original ToR of the assessment[9], most of which will be used for sampling purposes, additional sources will be mobilized:

- Food Security and Nutrition Analysis Unit (FSNAU) publications;
- Somalia Water and Land Information Management (SWALIM) Combined Drought Index;
- Integrated food security Phase Classification (IPC) reports
- WASH Cluster Drought Update – Somalia, January 2023
- OCHA, Drought Situation reports;

- Humanitarian Response Plan (HRP), 2023 - Somalia
- Humanitarian Needs Overview (HNO), 2023 - Somalia
- United Nations University - Institute for Environment and Human Security, Understanding and reducing agricultural drought risk: Examples from South Africa and Ukraine, 2018;
- United Nations Office for Disaster Risk Reduction, Global Assessment Report on Disaster Risk Reduction, Special Report on Drought 2021
- FEWSNET, Somalia publications
- World Food Program, Annual Country Report 2021, Somalia
- Somalia Health Cluster, 4W Dashboard, 2022
- UN OCHA, Somalia drought response and famine prevention, 2023
- World Bank, HDX Data
- WASH Cluster, Somalia WASH Cluster Drought Update, 2023
- Somalia: Drought and Famine Displacement Monitoring Dashboard (November 2022)

Broadly speaking, as much as possible secondary sources will be used to contextualise findings, such as the Integrated food security Phase Classification (IPC) reports providing information on food security and nutrition needs as well as projections, per region. The World Food Programm (WFP) Annual country report adds more context to the food security situation in further context and complement the IPC. The Somalia Water and Land Information Management (SWALIM) can provide climatic data such as the combined drought index capturing information on crops, pasture, fire danger, water shortages, livestock migration. FEWSNET's seasonal monitoring adds important value to the monitoring of seasons, especially during drought conditions, contributing with information on rainfall during Somali seasons. In particular, this and the SWALIM data will be triangulated with reported impact of drought and floods in HARD-TO-REACH districts. The WASH and CCCM clusters' products will be used to provide key definitions. For the Health sector, the Somalia Health cluster Dashboard provide an interactive alternative to better understand the distribution of health care services and the reach of the health-related programming in Somalia. Finally, stock satellite imagery will be used to triangulate findings (e.g., Normalized Difference Vegetation Index "NDVI",rainfalls).

A selection of secondary data resources, including existing REACH tools, informed indicator, and questionnaire design.

3.4 Primary Data Collection

Sampling

Given that physical access to the target locations is limited and there is no possibility of drawing a representative sample, purposive sampling will be adopted for both the Area of Knowledge and the mobile phone data collection methods.

---

[8] These are included indicators related to demographics of the KI as well as eligibility indicators and KIs departure or arrival time.
[9] REACH, Research Terms of Reference, Assessment of Hard-to-reach districts, October 2021.

KIs will be selected based on their knowledge of their settlement of origin in the HARD-TO-REACH districts, as well as their current residency OR the period they have been displaced from the HARD-TO-REACH district.

If the KI is not living in the HARD-TO-REACH district, then the following eligibility criteria will apply (and will be integrated in data collection tools):

1) Being newly displaced from the HARD-TO-REACH district within the past 30 days before the start of data collection **or** having visited the HARD-TO-REACH district in the last 30 days prior to the start of data collection **or** having been in contact with friends/family living in the settlement of origin in the last 30 days prior to the start of data collection.

2) **And** come from a HARD-TO-REACH district where at least one household still remains. Since the assessment of HARD-TO-REACH districts aims to fill in gaps in understanding the humanitarian context, targeting settlements that are no longer inhabited would not contribute to this.

Additionally, for REACH to assess the robustness of the information given by the interviewed KI, the questionnaire will ask the duration of the stay of the KI before leaving the HARD-TO-REACH settlement. This won't be considered as criteria of eligibility but rather as an indication of the level of knowledge of the interviewed KI.

Finally, the questionnaire includes a section to allow a snowballing approach. The interviewees will be asked if they can refer REACH team to another KI that matched the eligibility criteria, from their settlement of origin or any other settlement in a HARD-TO-REACH district. The contact details of these additional KIs will be collected for sampling purposes. FOs will make sure to keep an up-to-date contact list of potential KIs and local guides in order to build a strong network at the field level. This information will not be shared externally and will be stored only on REACH assets, protected by a password.

Table 2: Targets per HARD-TO-REACH district are as follows.

District	Total Number of Settlements	Target Number of Settlements (25% of the total settlements)	Target Number of Interviews (Maximum of 3 interviews per settlement)
Adan Yabaal	34	9	26
Badhan	51	13	38
Balcad	248	62	186
Bu'aale	133	33	100
Bulo Burdo	182	46	137
Buur Xakaba	634	159	476
Ceel Buur	45	11	34
Ceel Dheer	71	18	53
Ceel Waaq	61	15	46
Diinsoor	211	53	158
Jalalaqsi	88	22	66
Jamaame	123	31	92
Jilib	164	41	123
Kurtunwaarey	80	20	60
Qandala	71	18	53
Qansax Dheere	93	23	70
Rab Dhuure	88	22	66
Saakow	127	32	95
Sablaale	96	24	72
Tayeeglow	129	32	97
Wajid	63	16	47

Xarardheere	33	8	25
Xudur	203	51	152
Totals		**757**	**2271**

Structured KI interviews

Based on information from UNHCR PRMN (Protection and Return Monitoring Network) and CCCM cluster settlement verifications, as well as based on information obtained in the field, Field Officers (FOs) will identify eligible KIs among new arrivals in the sites around Baidoa, Garowe, Kismayo and Mogadishu and also identify KIs who are living in the HARD-TO-REACH settlements to be interviewed through mobile phone interviews.

Table 3: Targets per base are as follows:

Location	Target number of KIIs
Baidoa	1'066
Garowe	91
Kismayo	456
Mogadishu	658
Total	**2,271**

The quantitative tool will include questions relating to the following clusters: Food Security and Livelihoods, Health, and WASH. The tool will also relate to the following topics: displacement dynamics, access to/barriers to services, markets, priority needs and access to humanitarian assistance. The tool has been adapted to specifically capture the influence of climate hazards and economic conditions.

Most indicators will be collected at the settlement level, except for a selection of individual-level indicators regarding the KI's profile (including eligibility questions).

For most indicators, KIs will be reporting at the time of data collection unless otherwise specified.
Data collection will be organized as follows:

- **Field Officers (FOs) and enumerator training:** 4 full days for the training for the upcoming round in March, including 1 day of pilot.
- **Data collection:** 3 weeks of daily face-to-face data collection in informal IDP sites/ Mobile phone call interviews around Baidoa, Garowe, Kismayo and Mogadishu.
- **Data cleaning:** Daily data checking and cleaning will be conducted by the field and assessment team during data collection.

3.5 Data Processing & Analysis

IMPACT data cleaning minimum standards checklist will be followed.[10] Detailed data cleaning procedures will be outlined in the data cleaning Standard Operating procedures in **Annex 1: Data Cleaning SOPs.**

Every day, the surveys are uploaded on the REACH/IMPACT Kobo-server and downloaded by the Database Officer (DO) at the end of data collection. The DO anonymises and subsequently checks the dataset before it goes through to Field and Assessment Officers who will be conducting data checking and cleaning with log changes and deletions. The Assessment Officer will oversee and do the data cleaning templates for the Field Officers, who are in turn responsible for data checking and the supervision of field teams. The following protocols will be in place to ensure the quality of data collected:

- Daily data cleaning by Field Officers, who identify outliers, anomalies, and logical inconsistencies, and give regular feedback to enumerators through daily briefings and ad-hoc training. Data points that cannot be resolved through follow-ups with the enumerators or respondents will be deleted. If survey

---

[10] IMPACT Memo Data-Cleaning-Min-Standards-Checklist 28012020-1.pdf (reachresourcecenter.info)

records have more than three outliers that cannot be checked, the entire record is deleted from the dataset. Also, if the duration of the survey taken is very long or short and the enumerators couldn't provide concise and clear justification, the entire survey will be deleted.
- Weekly data cleaning will be conducted by the Assessment Officer, who reviews data cleaning conducted by Field Officers and provides additional feedback to the data collection teams in regular communication, briefings, and training.
- The GIS and Database Officers do data aggregation and spatial verification, who provide feedback to ascertain settlement coverage.

Data collected with structured tools will be aggregated at the settlement level (except for the selection of key indicators that will be disaggregated by KI length of displacement).

Given that more than one quantitative survey will be collected for a given settlement, data from key informants reporting on the same settlement is aggregated to the settlement level using an R script which employs the following logic to calculate settlement-level responses:

- Single response questions: Most survey questions only allow a KI to select a single response. For this type of question, mode aggregation is used, whereby "I don't know" responses are dropped and then the most reported response is taken for each settlement. Should several KIs from the same settlement provide different responses to the same question, the result is reported as "No consensus".
- Multiple response questions: Mode aggregation is used, whereby "I don't know" responses are dropped and then all other responses reported by the KIs are presented.

4. Key ethical considerations and related risks

For detailed guidance on how to complete this section, see also Step 5 of the IMPACT Research Design Guidelines

The proposed research design meets / does not meet the following criteria:

The proposed research design...	*Yes/ No*	*Details if no (including mitigation)*
... Has been coordinated with relevant stakeholders to **avoid unnecessary duplication** of data collection efforts?	Yes	
... **Respects respondents, their rights and dignity** (*specifically by: seeking informed consent, designing length of survey/ discussion while being considerate of participants' time, ensuring accurate reporting of information provided*)?	Yes	
... Does not **expose data collectors to any risks as a direct result** of participation in data collection?	Yes	
... Does not **expose respondents / their communities to any risks as a direct result** of participation in data collection?	Yes	
... Does not involve **collecting information on specific topics which may be stressful and/ or re-traumatising** for research participants (both respondents and data collectors)?	Yes	

… Does not involve **data collection with minors** i.e. anyone less than 18 years old?	Yes	
… Does not involve **data collection with other vulnerable groups** e.g. persons with disabilities, victims/ survivors of protection incidents, etc.?	Yes	
… Follows IMPACT SOPs for management of **personally identifiable information**?	Yes	

5. Roles and responsibilities

Table 2: Description of roles and responsibilities

Task Description	Responsible	Accountable	Consulted	Informed
Research design	*REACH Assessment Officer (AO)*	*REACH Assessment Officer (SAO), REACH Research Manager (RM)*	*REACH Deput Country Coordinator (DCC), IMPACT HQ Research Desig and Data Unit (RDDU)*	*E OCHA Somalia Information Management and Assessment Working Group (IMAWG), Inter-Cluster Coordination Group (ICCG), Drought Operations Coordination Center (DOCC)*
Supervising data collection	REACH Field Officers (FOs)	REACH Senior Field Officer REACH AO REACH SAO	REACH RM REACH DCC	OCHA Somalia, ICCG, IMAWG, DOCC
Data processing (checking, cleaning)	REACH AO, FOs, GIS Officer (GISO) DO	REACH SAO	REACH RM RDDU	OCHA Somalia, ICCG, IMAWG, DOCC
Data analysis	REACH AO, GISO and DO	REACH SAO	REACH DCC REACH RM RDDU	OCHA Somalia, ICCG, IMAWG, DOCC

Output production	REACH AO	REACH SAO	OCHA Somalia REACH DCC REACH RM Research and Reporting Unit (RRU)	ICCG, DOCC
Dissemination	REACH AO	REACH SAO	REACH DCC REACH RM	OCHA Somalia, ICCG, IMAWG, DOCC
Monitoring & Evaluation	REACH AO, REACH SAO	REACH RM	REACH DCC Humanitarian partners, OCHA Somalia HQ Research Department	OCHA Somalia, ICCG, IMAWG, DOCC
Lessons learned	All team members involved in the assessment (field team, data team, assessment team), partners if possible	REACH AO, REACH SAO	REACH DCC REACH RM HQ Research Department Humanitarian partners	OCHA Somalia, ICCG, IMAWG, DOCC

Responsible: *the person(s) who executes the task*

Accountable: *the person who validates the completion of the task and is accountable of the final output or milestone*

Consulted: *the person(s) who must be consulted when the task is implemented*

Informed: *the person(s) who need to be informed when the task is completed*

Annex 1: Data Cleaning SOPs

Data Cleaning Procedures for Key Informant (KI) Surveys

Below are the data cleaning standard operating procedures (SoPs) for KI surveys. SoPs are subject to change depending on feedback from Assessment Officers, the Database Manager or Field Coordinator during data collection.

OVERVIEW OF DAILY RESPONSIBILITIES

Field Coordinator (FC)
- Responsible for communicating security concerns from Senior/Field Officers (S/FOs) and implementing partners to larger Senior Management Team (SMT), who can then decide on whether or not to change the sampling framework and communicate that to Assessment Officers (AOs) and Database Manager (DM)
- Knowledgeable of AO to SFO/FOs communications regarding data cleaning issues via Teams group and e-mails

Senior Field Officer (SFO) & Field Officers (FOs)
- Responsibility for Somalia regions divided between SFO/FOs, each assigned to contact enumerator team leaders, individual enumerators if any issues with the data
- In constant communication between the database manager (DM) and field SFO / FOs / individual enumerators regarding issues with data collection and data quality issues
- Ultimately responsible for progress tracking in Excel sheets
- Responsible for following up on feedback given to enumerators via Excel sheets

Database Manager (DM)
- Responsible for downloading, deleting and anonymizing raw data and passing to S/DBOs for cleaning
- Runs daily R data checking script with clean data, and raw data from most recent day to identify errors for SFO/FOs to follow up on
- Reviews daily cleaned data outputs
- Updates Daily Progress Tracker (HTML)
- Makes final call on survey deletions. Updates deletion columns in online tracking sheets
- Responsible for overseeing changes to sampling framework and adjusting sampling targets accordingly

Database Officer (DBO)
- Responsible for supervising the visual check, split data for visual checks for S/DBOs and consolidate all the visual check results in single file
- Runs data cleaning and checking scripts on daily basis
- Sends cleaning result to AOs for providing feedback for and follow up on it
- Consolidate all cleaning logs which feedback provided for them
- Visually check for patterns in the data in questions to see if there is any evidence of enumerators developing habits of always entering the same values
- DBO's responsible for ensuring daily backups of their cleaned data occur and that proper file naming protocol is followed for cleaned data and cleaning log
- DBOs communicate all data issues to Lead DBO, who communicates feedback and issues to responsible DM to contact responsible SFO, who contacts field teams and individual enumerators to clarify any issues with the data

Assessment Officer (AO)
- Responsible for reviewing cleaning result on daily basis and providing feedback to either DM, SFO/FOs, GISO
- Responsible for compiling all subsequent feedback from DM, SFO/FOs, and GISO and sending to DM.
- In communication regarding security or logistical concerns that change sampling framework – communication lead by FC, DM, and SMT

Individual Enumerators
- Ensure phones are fully charged prior to next day of data collection.
- Ensure phones are set to the correct time and date prior to data collection. Achieved with steps below: Settings→General management→Date and time→Automatic date and time AND Use 24-hour format ON

- Ensure phones used for interview calls are recharged with airtime.
- Collect coordinates and finalize survey after asking final questions
- Upload forms to the kobo server daily

Additional information for DM and DBO:

- We've been automating most of the data processing stages and encourage our DBOs to implement their new ideas for developing this process. In addition, the Data Team should familiarize themselves with all data cleaning scripts, in particular, translation, replacing cleaning log and pattern check logger, time check and etc.
- Data Team is responsible for familiarizing themselves with the constraints of the kobo data collection tool.
- The Data Team is to have the latest version of the kobo tool and this SOP open to understand each question code.
- The column structure must be followed in order
- Any consistent pattern issues related to specific enumerators should be noted by DBOs
- Translations from Somali are harmonized to match existing categories - if possible. If answers do not fall into already existing categories but occur more than once, answers are harmonized as new categories

DATA PROTECTION
To maintain the safety and security of both respondents and enumerators, the following procedures will be followed:

- GIS points, settlement location, enumerator information and interview times are deleted from final dataset.
- Settlement location will be removed from processed raw dataset that is shared among AO and FC
- Province codes, district codes, and enumerators will be used for raw dataset that is shared among AO and FC
- Database Manager is the only individual with full access to raw dataset.

DELETION OF DATA
- Considering the deletion criteria, the deletion command will be coded in R script to apply on daily data processing
- Database Ofifcer is responsible for deletion of surveys

Criteria for deletion:
- All surveys completed in under 20 minutes or more than 80 minutes
- All surveys with a pattern match of 90% or higher to another survey
- All surveys with a respondent 17 years of age or younger
- All surveys without consent
- All surveys with 6 or more flags, without reasonable explanation

DATA CHECKING SUMMARY

A – DUPLICATE CHECKS: PARENT WORKSHEET
Duplicate uuid's should not be present

#	STEP
1.	**Duplicate surveys are flagged for deletion**. We have already coded this in our data processing script. If there was any duplicate, it will be flagged for deletion.

B – TIME CHECKS: PARENT WORKSHEET

	Survey should take between 20 and 70 minutes

#	STEP
1.	This survey should take between **20 and 70 minutes** to complete.
	Under 20 minutes: Surveys will be deleted as invalid as they are of suspected poor quality/to be fake.
	80 minutes or more: Surveys will be deleted as invalid as they are of suspected poor quality/to be fake.
	Surveys taking **longer than 70 minutes will be flagged for Area FC/FOs review.** If there is no valid explanation provided, then the interview will be deleted. Operations will be notified of this deletion.
2.	We're using audit files to check the duration of each survey and using R script for doing this operation.

C - TRANSLATIONS FROM Somali

Organization and Village names may be entered in Somali

#	QUESTION	ACTION
1.	**settlement_other, organisation_other** *Parent Worksheet*	**'Other' answers, entered initially in Somali.** Translations from Somali done within same cell in the *Data Checking* Tab. If possible, translations are first harmonized into groups that match existing categories. If answer falls into category that already exists, the 'other' category answer is corrected in the same cell.

D – SPECIFIC QUESTION CHECKS

Parent Worksheet

#	QUESTION	ACTION
1.	'Other' columns	Check that data entered into 'Other' column is translated, logical, and consistent with the context. Action: If data entered into 'Other' column matches any of the potential survey responses, re-classify that entry and log the change. If the entry cannot be reclassified, just translate.
2.	ig0vt16 wx6iv19	If "ig0vt16" = 'damage_losses_due_to_drought' and 'drought_prolonged_lack_of_rain' is not selected for "wx6iv19" If "ig0vt16" ='damage_losses_due_to_flooding' and 'flooding' is not selected for "wx6iv19" **1.** Flag the entry.

		2. Follow up with the enumerator on whether there is an explanation to this contradiction. Change accordingly if an explanation is provided.
		3. If an enumerator follow-up is not possible, follow up with the respondent and change.
		4. If neither of the follow-ups is possible, delete entries from the respective columns
		5. Communicate to the enumerators to make sure the question is understood and asked properly.
3.	zd0yx28 wx6iv19	If "If_crop_loss_was_experienced" = "lack_of_rain" and 'drought_prolonged_lack_of_rain' is not selected for "wx6iv19" If "If_crop_loss_was_experienced" = flooding and 'flooding' not selected under "wx6iv19" 1. Flag the entry. 2. Follow up with the enumerator on whether there is an explanation to this contradiction. Change accordingly if an explanation is provided. 3. If an enumerator follow-up is not possible, follow up with the respondent and change. 4. If neither of the follow-ups is possible, delete entries from the respective columns Communicate to the enumerators to make sure the question is understood and asked properly.
4.	py8si99 wx6iv19	If "py8si99" = drought and 'drought_prolonged_lack_of_rain' not selected under "wx6iv19" If reason_livestock_decrease = flooding and 'flooding' not selected under "wx6iv19" 1. Flag the entry. 2. Follow up with the enumerator on whether there is an explanation to this contradiction. Change accordingly if an explanation is provided. 3. If an enumerator follow-up is not possible, follow up with the respondent and change. 4. If neither of the follow-ups is possible, delete entries from the respective columns Communicate to the enumerators to make sure the question is understood and asked properly
5.	dv4hj16 gn8wo34	If "dv4hj16" = general_food_distribution, and 'food' is not selected under "gn8wo34" If reason_livestock_decrease = flooding and 'flooding' not selected under "wx6iv19" 1. Flag the entry. 2. Follow up with the enumerator on whether there is an explanation to this contradiction. Change accordingly if an explanation is provided. 3. If an enumerator follow-up is not possible, follow up with the respondent and change. 4. If neither of the follow-ups is possible, delete entries from the respective columns Communicate to the enumerators to make sure the question is understood and asked properly

E – ENUMERATOR PATTERN CHECKS

#	QUESTION	ACTION
1.	Select_one and select_multiple questions	Visual check that individual enumerators are not developing patterns for answers to these questions or answering the same answers for every survey

H – DAILY R SCRIPT FOR RAW DATA CHECK

Producing reports broken by region, the purpose of the daily R script check is to provide guidance to FOs and DBOs for enumerator follow up and provide an additional check against enumerator error.

R Script incorporates cleaned data as well as raw data from the most recent day of data collection

#	QUESTION	ACTION
1.	**progress check**	At overall and district level to determine day-to-day progress and ensure a timely data collection.
2.	**time checks**	Flags surveys with illogical time stamps including an end time that is before the start time, interviews submitted in the future, interviews that are either under 20 minutes or more than 70 minutes. Interviews under 15 minutes or more than 79 minutes will be automatically deleted.
3.	**survey frequency**	Flags enumerator productivity in most recent day of data collection. Enumerators with more than 5 surveys per day maybe notified to slow down and interviews may be deleted if number exceeds 5.

SUMMARY OF SAMPLING FRAMEWORK CHANGES
- **Step 1:**
  Individual enumerator or team leader raises security or logistical concern (including active conflict, natural disaster, change in mobile connectivity in area, hostility from local actors etc.) to FO. FC then alerts DM/AO and SMT.
- **Step 2:**
  FOs escalates security or logistical concern to Operations/SLO.
- **Step 3:**
  SMT determines sampling framework change and communicates sampling change to AO and DM. Alternatively, AO communicates sampling change to FC and FOs.
- **Step 4:**
  DM adjusts sampling framework to accommodate changes.