TECHNICAL BRIEF

Purposive sampling and site selection in Phase 2
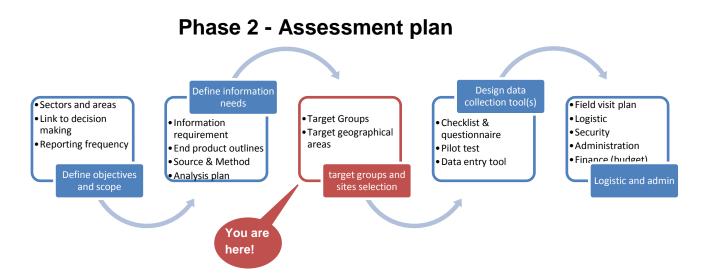
Contents

1. Introduction

This Technical Brief is for humanitarian workers planning a Phase 2 assessment. It provides guidance on how to select relevant groups and identify the most appropriate sites to assess. The site selection activity builds on information collected during the Phase 1 assessment (secondary data review and a few field visits) and identified information gaps. Site selection mainly depends on the context and the objectives of the field assessment, but will always aim to:

1) Verify or refute assumptions and impressions generated from the Phase 1 assessment
2) Get information on issues which were not fully or clearly understood from the Phase 1 assessment
3) Gain the perspective of beneficiaries on their priority needs.

To achieve these information goals and to optimize the field level assessment period during phase 2, it is necessary to carefully choose the most relevant sites to be visited. The Technical Brief focuses on the use of **purposive sampling** during phase 2 of assessments and makes use of a case study to facilitate the comprehension of meaningful site selection. As some technical terms are used, it is recommended to refer to the Glossary in Annex 4 for further clarification.

2. You are here!

The diagram below illustrates where site selection and sampling fits into the assessment process, i.e. between defining information needs and designing data collection tool(s).

Phase 2 - Assessment plan

Before deciding on target groups and sites, the following activities will have been completed:

1) Phase 1 assessment[1]
2) Objectives and scope of the rapid assessment agreed and defined[2]
3) Information needs defined[3].

---

[1] Phase 1 assessment is based on Secondary data review (See ACAPS Technical brief on Secondary Data Review) and a few community level assessment and generally produces a Preliminary Scenario Definition
[2] See ACAPS technical Brief on Phase 2 assessment - Definition of objectives and scope
[3] See ACAPS Technical Brief on Phase 2 assessment – Information needs and analysis plan

Once sites and groups for the rapid assessment are selected, the data collection tool can be designed and logistical organisation for the assessment resourced and planned.

3. Why purposive sampling?

Two major constraints apply to Phase 2 of assessment:

- During the first two weeks following a major emergency, primary data can only be realistically collected at the community level. Given the time, access and logistical constraints, collecting meaningful information from households or individuals is usually not feasible, or useful. Attempts to do so have significantly contributed to the failure of rapid assessment in the past. The recommended **sampling unit** for Initial and Rapid assessment is the **community** level.

- Because it is normally neither feasible nor desirable to survey every location affected by an emergency, a sample must be drawn[4]. The list from which we sample (the **sampling frame**) includes the known and relevant affected groups / categories / social strata. The sample therefore will be a cross-section of diverse affected groups, such as displaced persons, host communities, returnees, etc located in selected geographical areas.

In any assessment process, there will be a trade-off between the representativeness and diversity of the sample and the efficiency and timeliness with which data can be collected. Assessments in Phase 1 and 2 do not need to be as representative as they need to be rapid[5]. Time and information constraints will normally not allow for random or statistically representative sampling, consequently **purposive sampling**[6] is the most appropriate solution for these phases.

Purposive sampling enables an initial understanding of the situation, and to identify and differentiate the needs of one or more relevant groups. It produces a sample where the included groups are selected according to specific characteristics that are considered to be important as related to vulnerability (e.g. IDPs in camps, host population, etc.). With such a sample, group differences can be compared and contrasted and a range of experiences can be summarized (for example highest and lowest access to food, water or health services accessible to communities in the sample, etc.). Using this approach, the assessment team will select a sample of sites which represents a cross-section of affected areas or groups.

When using purposive sampling, it is important to seek sites that will provide an understanding of the situation of a wider group of affected people (i.e., sites "representative" of the given group). The criteria for selecting sites will depend largely on the context of the emergency. Meaningful stratification of localities or groups (in essence, defining which groups are important to consider, and defining which localities or sites belong to which group) is advised to ensure that different types and levels of impact are captured and systematic comparisons among relevant groups are possible.

---

[4] Annex 1 provides details on usual sampling method for needs assessment
[5] A comprehensive sampling of affected households or individuals won't be carried out until phase 3 whereupon full diversity strata and representation of all affected groups can be included in the sample.
[6] Purposive sampling can be useful for situations where there is a need to quickly reach a sample including targeted groups and where sampling for proportionality is not the primary concern. In purposive sampling, we sample with a *purpose* or one or more specific predefined groups in mind

4. Identification of target group(s) and site(s)

The following criteria serve as a basis for meaningful group targeting to guide the site selection process. Criteria may be combined where necessary to increase the variety within the sample.

1. **Group characteristics:**

a. Social and cultural characteristics: residence (i.e. displaced vs non displaced), ethnicity, religion, age, gender, disability, etc.

b. Homogeneity[7] vs. (meaningful) heterogeneity[8] of affected groups: IDPs living in camp settlements, host population, affected population, etc.
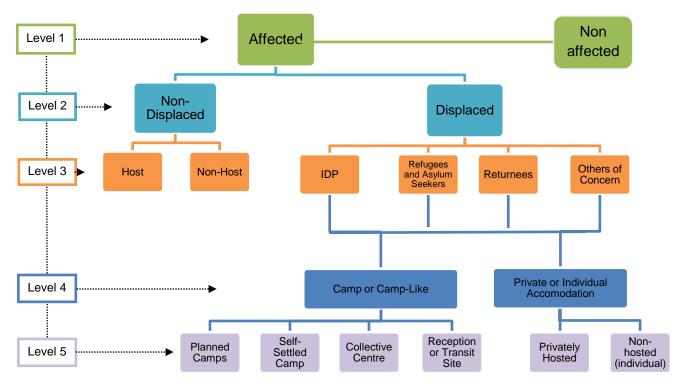
The term "affected population" during crises is generally associated with a large and diverse population. It is often poorly defined or refers to different groups of the population who have been affected by the crisis in very different ways and having different needs.

Rapid assessment in phase 2 is about comparing the situation between different groups of affected persons and describing how the experience and priority needs of one group differs from the experience and priority needs of another.   Therefore, group characteristics need to be clearly defined before the assessment starts to ensure appropriate targeting. The following graph provides distinct standard categories organized across five levels of hierarchies that may be of interest when breaking down the overall "affected population" into relevant "affected groups" of interest for the rapid assessment.

The categories of affected population will vary depending on the crisis. Phase 1 assessment will provide the number (estimation), definition and characteristics of the different groups and sub-groups affected, ensuring that each category described within the same level of hierarchy is mutually exclusive from the other (one person or population can't belong to the category "displaced" and "non-displaced" at the same time). Moreover, the sum of all people in each category at a given level in the hierarchy should equal the total number of affected people. Annex 6.2 provides standard definition of the different affected groups.

---

[7] A homogeneous community or group has a uniform structure or composition throughout and/or comprises people of the same or similar kind of nature and/or share the same characteristic on the way they have been or they are believed to be affected.
[8] A heterogeneous community or group consists of dissimilar or diverse ingredients or constituents. It would typically comprise a mix of different population groups (residents, internally displaced persons (IDPs), refugees and returnees) with different degree and variations on the way they have been or are believed to be affected.

Affected groups – Standard categorization 2011

Adapted from IASC Guidelines on Humanitarian profile, June 2011

For rapid assessments in Phase 2 and for large scale disasters, the decision to assess groups that belong to a lower hierarchical level than level four needs to be made with caution if the overall number of groups are too numerous or if resources are scarce. At a lower level, the necessary stratification required will increase exponentially the data volume necessary to be gathered in order to represent the variety of needs across the different groups. If time and resources are limited during Phase 2, further assessment in Phase 3 may be more appropriate to explore priority needs for groups beyond level four.

Defining affected groups and choosing groups of interest for phase 2 assessment will always be the starting point. Once identified and as detailed in the following sections, additional characteristics can be chosen that better reflect the diversity of situation being assessed. New elements of stratification are generally based on the setting or geographical location where the considered groups are living.

2. Site characteristics:

a. Density characteristics: population located in urban vs. rural

b. Livelihood or Agro Ecological Zones[9]: areas where people share broad common livelihood-sustaining activities (farming, pastoralism, fishing)

c. Geographical characteristics (altitude and/or topography): population located in coastal, riverine, plains, mountains, etc.

---

[9] 2008, FAO, ILIA: In most rural areas, a useful starting point for sampling is to divide up affected areas into livelihood 'zones', within which people share broad common livelihood-sustaining activities and goals.

d. Severity of impact: population located close from the "epicenter" of the disaster, in most affected areas, in directly or indirectly affected areas.

e. Pre-existing vulnerabilities: population located in areas with higher/lower access to services, level of poverty, prevalence of chronic malnutrition, etc.

f. Administrative units: population within a given district, department, province, etc.

3. **Gaps in existing knowledge:** Locations about which little is known, or key information is lacking, particularly where there are yet no relief agencies operating.

Identifying target groups and determining site locations is a combined process. Groups are chosen in lieu of their location because the situation of similar groups located in different geographical areas (urban vs rural, areas with information vs areas without information, etc.) are anticipated to exhibit different needs to be assessed. However, note that identifying and selecting sites are two different processes. When selecting, the principles discussed in the following section will apply.

5. Key Principles for site selection

Building on Phase 1 assessment results: Findings from Phase 1 assessment, including clear indications about most affected areas, groups and information gaps, will help to make an appropriate and informed stratification for site selection in Phase 2 assessments. Without clear recommendations from Phase 1 assessment, there is a risk that assessment teams will rush to do field work without knowing what they are looking for.

Expertise: Purposive sampling requires much more intellectual and strategic thinking than the simple demographic stratification of a representative survey. Therefore people with emergency setting experience, knowledge of the local context and skilled in rapid assessment techniques and methodologies are required to design relevant and useful sampling.

Number of sites: It is a critical issue, and there is no single answer. The sampling size to assess will be determined by the availability of staff, time, logistical support, as well as geographic spread of the disaster, and heterogeneity versus homogeneity of populations. The number of sites to visit will mainly depend on the diversity of the region and your target groups numbers:

- If the affected area is quite homogeneous, with few diversity groups, a similar geographic situation across the zone, comparable production patterns, etc, only a few sites may be needed (after a couple of sites, it may be evident that the same type of information is being repeated and it will be pointless to go further).
- If, however, the affected area is heterogeneous in terms of the factors listed above, more sites will be needed since different situations will probably be encountered which experience different circumstances.

When using purposive sampling, there is no mathematical calculation to determine the ideal number of sites, instead this will be based on informed judgment and an emphasis on securing information from a cross-section of areas and population groups affected. In all cases, the sample size must be small enough to be manageable but large enough to generate useful information.

Selecting a large number of sites to visit will inevitably introduce delays and make the implementation of the assessment in the prescribed timeframe impossible as well as obstructing

analysis. <u>In practice, there is seldom a situation when it is useful to include more than 30 sites for purposive sampling. In most cases, assessing 15-20 sites is sufficient.</u>[10]

Adapted to the scale: The degree to which a holistic picture of impact can be obtained will obviously be different in the case of a large scale disaster with a large and diffuse impact (e.g. tsunami, flooding) than for a small or medium scale disaster with a more geographically focused impact (e.g. landslide). In the latter case, it will be easier to be more comprehensive. If heterogeneity is high, units are very different from one another or the affected area is huge (Pakistan Floods 2010), expert inputs and more careful consideration will be needed when defining strata that would be feasible to assess.

Diversity driven: As mentioned earlier, if the impact seems uniform across the affected area, stratify using criteria such as urban/rural, in camps/outside camps, etc. If impact is not uniform, map out the areas where impacts are believed to be different and establish travel itineraries that take in different typical affected areas and population groups. At this point you want to avoid using many different characteristics to create the piles and focus on the ones you think will make the most difference in terms of the assessment results. By dividing the whole set into subsets of this type (stratification) you are ensuring that even though your sample is small, groups or communities with certain characteristics fall in the sample (e.g. that you have at least one site with group or livelihood "x" and another from group "y"). Sites to visit should not be limited to the worst-affected localities or areas that are easiest to reach. If these are the first to reach, subsequent assessments should reach other groups and areas to contrast with these first ones.

Be prepared: Selected sites may sometimes be inaccessible when the assessment team get to the field. The assessment teams should respect a pre-defined set of rules to replace communities that turn out to be inaccessible or irrelevant while in the field and under time pressure. Team leaders needs to receive a special briefing on the site selection rationale and be able to quickly react in case of access issues.

Fewer sites…: Chances to specify new vulnerabilities and needs increase with sample size, but after common pictures are seen repeatedly, there is little benefit to seeking more areas to sample[11]. Purposive sampling is therefore most successful when data review and analysis are done in conjunction with data collection. The sample size can be adjusted during the assessment based on information that becomes available.

…And more skilled assessment teams: The chance of capturing key information on diverse needs between different groups depends more on the skill, thoroughness, and consistency of information gathering from a few areas than on large scale low quality data gathering from a large sample of communities. Fewer well-chosen sites visited but with **more qualified assessment teams** are key to solid and relevant interpretation. Previous assessments have shown that using assessors without appropriate background, expertise and training allow for visiting more sites, but often produce data that are poor quality and not amenable to meaningful analysis.

Don't extrapolate: A purposive sample cannot be directly or quantitatively extrapolated to the wider population of interest from which the sample came. It can nonetheless provide critical information on areas and groups in greater or lesser need. Its purpose is to identify the most pressing issues/concerns/needs in order to set up priorities for immediate action. Further assessment in Phase 3 can always collect a more representative population-based sample while ensuring a sufficiently large sample to include newly identified vulnerable groups.

---

[10] The number of sites and the number of "questionnaires" are not related. Within one site, you may have several groups of interest and one questionnaire will need to be completed for each of those groups.
[11] Purposive sample sizes are often determined on the basis of theoretical saturation (the point in data collection when new data no longer yields additional insights to the research questions)

Example:
- *Results of the rapid assessment using purposive sampling show that 63% of the assessed sites have no access to health facilities.*
- *Reliable government sources state that overall, 100,000 people are affected by the crisis.*
- ***It is not possible** to state that 63,000 people have no access to health facilities.*

Transparency: Purposive sampling can produce a reasonably accurate picture of a given situation. However, results must be used with caution. The main limitation of purposive sampling – the impossibility to extrapolate the results to the whole population - needs to be explained to partners and key decision makers prior to the assessment to avoid misunderstanding and false expectations about the assessment findings. The methodology section of the final assessment report also needs to describe those limitations and how the data can reasonably be interpreted.

Communication of results: To avoid misinterpretation, it is recommended not to use percentages to describe the results and findings of the assessment, but rather to say: *8 out of 14 visited sites reported not having access to.....*

Box 1 – Purposive Sampling Terminology

The following terms may be useful to describe and interpret information from purposive samples:

- *Likely...*
- *...indicates that...*
- *...suggests that...*
- *...Is consistent with....*
- *...were surprised to find...*
- *It appears....*
- *What we observed is consistent with...*

6. Annex 1 - Case study

6.1 Earthquake in Larsilandia Island

A case study will be used to support the understanding of site selection process during multi sector/cluster rapid assessment, using purposive sampling. It is recommended to print the map and to keep it close to be able to follow the instructions.

The Larsilandia island has been severely hit by an earthquake of magnitude 7.2 three days ago. The main tremor occurred at 02.41 a.m. when the majority of the inhabitants were sleeping. The earthquake happened one month before the beginning of the winter season.

Pre disaster situation:

The island is 700 km long and 300 km wide at its broadest point. The total population is estimated to be 547,000 inhabitants. The east of the country is covered by a mountain chain of between 2,500 and 3,000 meters while the west is constituted of large plains around the Carolina river basin. The main road crosses the country from the West coast where the capital Andy is located (134.000 inhabitants) to the East coast across the mountain, where the second largest city of the country, Wily (77,000 inhabitants), was raised around the main commercial port. Most of the island houses and buildings are made from concrete, except in remote rural areas and mountainous areas, where they are made of wood.

The major source of income is agriculture (78%) in rural areas and in plains, especially in the Carolina river basin. Corn, soya beans and rice are the main crops. On the coastline in the East, fishing and tourism are the main source of income.

Post disaster situation:

The **preliminary scenario definition** of Phase 1 assessment, based on secondary data review and a few UNDAC field visit, provided the following information:

- The government reports more than 10,000 dead, 13,500 injured and more than 6,700 missing. There are still a high number of inaccessible areas in the western part of the country which were not able to report on potential damages.

- The communication across the island is largely disrupted. The new communication antenna situated at the top of the mountain has been damaged and is out of service.

- One field visit report available from the civil protection authority in Herby reports that access is feasible with a motorcycle. Members of the affected population are not currently using their houses because of the fear of aftershocks. Most of the inhabitants reported that they would go down to the plains in a few days when the cold weather arrives and return to their land and property after the winter.

- The National Disaster Management Agency reports that large population movements were observed into Andy and Wily cities. Local volunteers have registered more than 24,000 IDPs in public buildings such as schools or in informal camp settlements in parks and open spaces in the capital Andy. Various localities around the affected areas such as Cary, Dady, Judy and Sandy have also been reporting influx of IDPs. The situation in visited public buildings used by IDPs as temporary shelters or camp settlements is reported to be critical in terms of access to basic services, congestion, and protection. Some IDPs also report being hosted by their relatives in
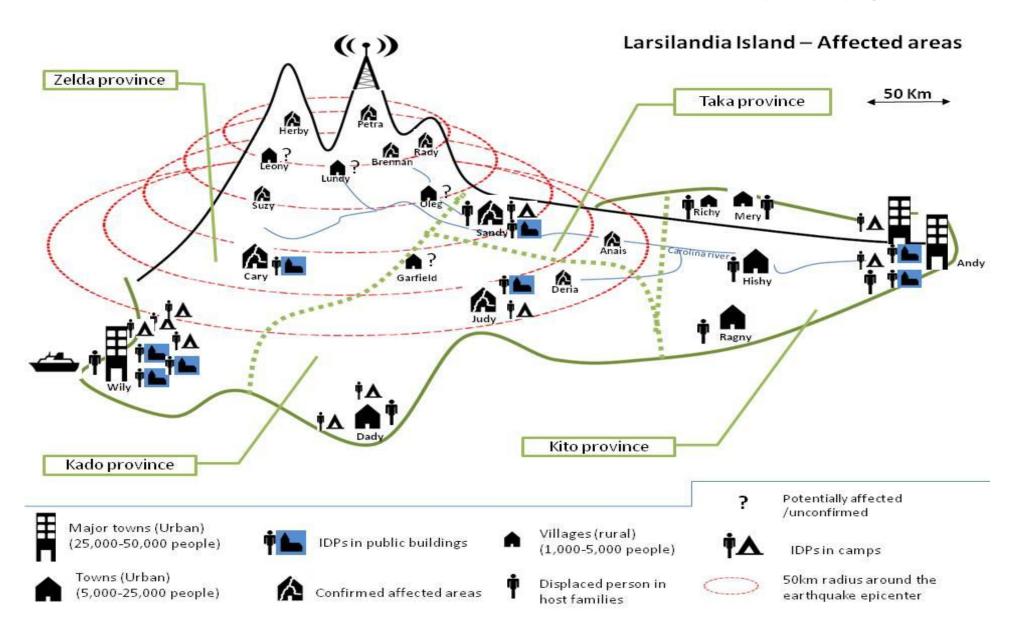
Wily and Andy. Emergency stocks in the capital are exhausted as a result of major flooding in the capital two months ago. There has been no opportunity to replenish the contingency stocks.

- The main road between the capital and the affected areas is disrupted in several points. The Ministry of Infrastructure said it would take several months before the road could be passable by car or truck. However, various trucks were seen crossing the affected areas using back roads and small forest roads.

- The following map was made available with the Phase 1 assessment report.

A multi sector rapid assessment has been agreed upon between NGOs and the government civil protection units. The objective of the rapid assessment is as follow:

To assess the needs of the affected population both in rural and urban areas and across the four provinces of Larsilandia island.

…Now, design the sampling for the rapid assessment.

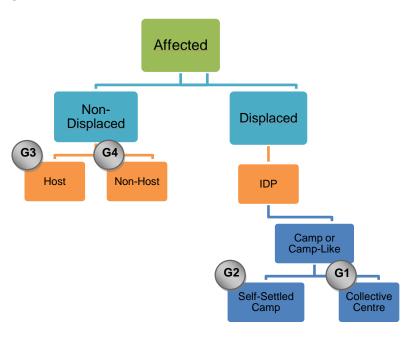Larsilandia Island – Affected areas

6.2 Step 1 - Define target groups

Larsilandia Case study: Select group of interest – Define target groups

A large area of Larsilandia has been affected by the earthquake, and there is report of affected populations in both urban and rural areas. From Phase 1 assessment, there is information about five potential Groups (G) of interest:

G1: Displaced population/IDPs/Camp or camp like/collective center (in schools)
G2: Displaced population/IDPs/Camp or camp like/Self settled camp (in public parks)
G3: Affected population/Non displaced/Host
G4: Affected population/Non displaced/Non host[12]
G5: Non affected population

According to the agreed upon objectives of the assessment, the "non-affected" population group is not considered as a group of interest, thus will not be retained as a target group.

Larsilandia's target groups for the multi sector rapid assessment are as follow:

Key notes:
- In this case, the rapid assessment focuses only on affected groups. No comparison with the situation of non-affected population will be possible.
- Each of those groups has been chosen "on purpose", which is why this sampling method is called "purposive sampling".
- "Affected groups" need to be clearly identified during Phase 1 to allow relevant sampling in Phase 2 of assessment. It is without any doubt one of the most important added value of Phase 1 assessment, i.e. to provide indication on the variety of existing affected groups, their characteristics and their location.
- However, if new groups of interest emerge during Phase 2 assessment, it is still possible to include them as appropriate and if time and resources allow.

---

[12] Affected/Non-Displaced/Non-Host refers to populations that have remained in their communities or houses that have been impacted by the disaster but have not fled their locality, e.g. sometimes households remain by their destroyed houses after an earthquake to ensure that they can keep what little possessions they have left after the damage caused by the earthquake.

6.3 Step 2 - "Map" the existing groups as per their location

Knowing which specific groups will be assessment, it is necessary to localize them. The map of affected areas as well as the assumptions provided by Phase 1 assessment support this mapping exercise.

Design a table where the different stratification levels selected to serve the purpose of the assessment objectives are represented. In the Larsilandia case, consider both urban and rural location where IDPs (both in collective centers and self-settled camps), host population and affected residents can be found.

A simple way to identify those sites is to draw a matrix with in columns the name of the different affected administrative entities (or any other relevant stratification chosen) and in rows e.g. the types and characteristics of the target groups.

Box 2: Larsilandia Case study – Mapping the affected groups						
Mapping matrix for Larsilandia Earthquake affected areas:						
			Province			
Setting	**Group**		**Zelda**	**Kado**	**Taka**	**Kito**
Rural	G1: IDPs in collective centres					
	G2: IDPs in self-settled camps					
	G3: Host population					Richy Mery
	G4: Affected pop non displaced/non host		Susy Herby Lundy? Leony?	Deria Garfield?	Brennan Rady Petra Oleg? Anais	
Urban	G1: IDPs in collective centres		Wily Cary	Judy	Sandy	Andy
	G2: IDPs in self-settled camps		Wily	Dady Judy	Sandy	Andy
	G3: Host population		Wily	Dady	Sandy	Andy Ragny Hishy
	G4: Affected pop non displaced/non host		Cary	Judy	Sandy	
Total sites where groups of interest can be found			**9**	**7**	**9**	**7**

Key notes:
- **For rural areas:** In one hand, we have no indication from Phase 1 reports that there are IDPs in collective centers or self-settled camps in rural areas, nor host population (except within Kito Province). In the other hand, there are a lot of reports about affected population (not displaced) still living in their villages, close to their houses and properties.

- No information is available on Lundy, Garfield and Leony villages in rural mountainous areas. Still, due to their proximity to the epicenter and to the fact that close settlement have been reported to be affected, it is very likely that they are also affected and can definitively be included within eligible sites, at least to verify if they have been or not affected.

- **For urban areas:** Some urban settlements have been directly affected by the earthquake (Cary, Sandy, Judy) and in the same time are reporting presence of IDPs in camp settlements or collective centers and are hosting affected population.

Quick analysis of the previous table indicates that some locations are mentioned several times, meaning that several groups are reported at a same location. Mapping "group diversity" at site level is also recommended when designing purposive sampling, as diversity of situation is what we intend to measure. The following tables detail the different groups represented at one site level, based on the findings of the phase 1 assessment:

Group diversity in urban affected location:

Province	Sites	G1	G2	G3	G4
Taka	Sandy	X	X	X	X
Zelda	Cary	X			X
Kado	Judy	X	X		X
Zelda	Wily	X	X	X	
Kito	Andy	X	X	X	
Kito	Hishy			X	
Kito	Ragny			X	
Kado	Dady		X	X	
Total		**5**	**5**	**6**	**3**

Group diversity in rural affected locations:

Province	Sites	G1	G2	G3	G4
Zelda	Petra				X
Zelda	Leony?				X?
Zelda	Lundy?				X?
Zelda	Brennan				X
Zelda	Rady				X
Zelda	Suzy				X
Zelda	Herby				X
Zelda	Oleg?				X?
Kado	Garfield?				X?
Taka	Anais				X
Kado	Deria				X
Kito	Richy			X	
Kito	Mery			X	
Total		**0**	**0**	**2**	**11**

While in rural areas, group representation at site level is not very diverse, urban areas are showing important variations related to the number and type of groups present at site level. Site selection process will need to ensure this diversity is reflected.

In top of the pre-defined stratification (Cf. Objectives of the assessment), rural areas in the Larsilandia case may be divided into two specific settings: mountainous and plains, where the situation is likely to be very different for the affected groups, specifically considering the upcoming winter in locations above 1500m.

Introducing a new stratification (mountainous vs. plain) on top of the ones related to administrative units (the four provinces) and the setting (rural vs urban) at this stage would only increase the data volume. When analyzing the results, we will be able to see if a significant difference exists between groups located in mountainous and plain areas, confirm this hypothesis and plan for further assessment in Phase 3 if relevant.

Generally speaking, distinction according to administrative boundaries may be of interest for planning purposes but will most of the times not reflect the environmental, socio economic, topographic or livelihood characteristics of the area and sometimes not be adapted to the requirements of the sample.

6.4 Step 3 - Select most appropriate sites for assessment

It is not possible to assess all the affected sites that fit with the pre-defined criteria. As a result, focus should be addressed to assessing target groups in key areas where impacts are believed to be "typical" of that group and may be different from other defined groups, and establish travel itineraries that take in those different typical affected areas and population groups.

For example, affected populations in rural area (Group 4) are potentially present in 11 sites across three provinces. It is not necessary to visit all those locations to have an understanding of the impact of the crisis on this specific group in rural areas. Therefore choices need to be made regarding the sites to be assessed and why one site more than another. Proceeding by elimination, you will reduce the number of sites to be visited and come out with a final list for your field assessment plan.

As detailed in section 4 of the Technical Brief, the eligibility of one site for assessment will depend on a mix of the following criteria:

1) **Lack (or abundance) of information.** *Do I already have information available on this specific site from other sources (Government, UNDAC, Civil protection, Red Cross)?*

   If information about one group or one site is sufficient, there is no reason to assess twice, just use the already available information as secondary data.

   - In the Larsilandia case, there is already information available from the Civil Protection Units about the situation in Herby.
   - Conversely, there is no information available about Lundy, Leony and Garfield villages while there is clear suspicion that they are affected. Visiting one or two of those locations should be considered as a priority.

2) **Homogeneity and similarities between sites.** *If the situation of one specific group in one area is likely to be similar to another site where a same group is present, there is no reason to assess systematically every single site.*

   In Larsilandia, the situation of affected residents (G4) in Herby, Leony, Lundy, Susy, Petra, Brennan and Rady villages is likely to be similar (all are small villages in mountainous areas located within a 100 km radius from the epicentre, probably sharing same type of livelihood and sources of incomes). There is no need to visit all those sites to understand how residents of small villages in mountainous rural areas are affected. A small sample of two or three sites will be enough to specify needs and impact of affected residents (G4) in those locations.

3) **Group diversity and representation at same location.** *Does one site comprise several groups of interest? Do several close sites present an interesting diversity of groups?*

   For instance, in Sandy all four groups of interest for urban areas are present. In that case, more time can be spent on this site to assess different groups. This will both lessen travel time and reduce resources required for the assessment. In Sandy for instance, the four groups can be easily assessed in one or two days: IDPs in collective centres, IDPs in self settled camps, Host population and Affected population. Proximity between sites also works.

4) **Minimum usable data for comparison.** *Do I have enough data for one specific group for each pre identified setting (rural/urban) to provide a good understanding of their situation? Am I likely to find the same results for a same group in different locations? Does visiting the same group in another site will provide me with new information?*

If there is no reason to conduct assessments for each location and for each group, it is necessary to assess a minimum of times the same group across different locations (when possible) to allow relevant comparison based on sufficient data. For instance, in Zelda province, IDPs in self-settled camps can only be found in Wily town. This is an opportunity to visit at least two or three of those camps in the same day to conduct the same questionnaire and ensure that enough information has been gathered about their situation and describe their needs. If the situation is similar between the two camps, there is no need to visit a third one. If their situation show great divergence, a visit to another camp in the same town (if it exists) is necessary to ensure that all the needs of this particular group have been captured. This would need to be an on-site decision from the assessment team leader.

5) **Severity of impact.** *Are there any sites that are likely to be more affected than others? Which sites should I start with? Which site do I need information about when compared to other sites?*

Box 3: Larsilandia Case Study – Site selection

In the following tables for rural and urban settings, the list of eligible sites defined during step 2 has been reorganized into "clusters" of sites according to the following criteria:

- **Availability of information**: Sites with already available information will not be selected, while sites without information need to be considered when relevant.
- **Similarities and homogeneity between sites**: villages or towns, proximity the one to another.
- **Group diversity and representation at same location**: number and types of affected group represented at the site level.
- **Estimated severity of impact**: From "most affected" to "least affected", from "directly affected" to "indirectly affected". This provides an indication of where to go first and how to prioritise between sites.

By combining these different criteria and the rule of minimum usable data for comparison, a final list of sites for the rapid assessment can be selected. In the following tables:
- Green cells represent final sites selected for the rapid assessment.
- Blank cells represent sites that have not been selected for field assessment.
- Red cells represent sites where information already exists and therefore are not necessary to visit twice.

Priority level	Province	Rural	G1	G2	G3	G4	Rural areas - Site selection justification
1	Zelda	Petra				X	**Key note:** Petra/Leony?/Lundy?/Brennan/Rady/Suzy/Herby are all villages located in rural areas of Zelda and Kado province and where affected population (G4) can be found. We assume that Leony and Lundy are also affected, even if no information is available. Is it very likely that the situation in those different villages will be very similar (All located with a 100km radius from the epicenter), so there is no reason to visit them all. As only one affected group is represented in each sites, only one questionnaire will be performed in each chosen location. Leony or Lundy will also have to be visited as their status is unknown. We already have information about Herby (Civil Protection Units), so this site does not require a new field assessment. **Site selection:** Petra or Brennan or Rady (depending on accessibility), Suzy and Lundy. One questionnaire adapted for affected population in each site (three questionnaires total).
		Leony?				X?	
		Lundy?				X?	
		Brennan				X	
		Rady				X	
		Suzy				X	
		Herby				X	
2	Zelda	Oleg?				X?	**Key note:** Oleg/Garfield are located at greater distance from the epicenter of the earthquake than the others villages and no information about their status is currently available. However, we have good reason to believe they are affected too (all villages around are affected) and we need to find out the situation there. The distance between the two is more than 50 Km, so it is unsure if the situation is similar between the two locations and both will need to be assessed. **Site selection:** Oleg and Garfield. Most probably only affected population there, so one questionnaire adapted for affected population per site (two questionnaires total).
	Kado	Garfield?				X?	
3	Taka	Anais				X	**Key note:** Anais/Deria (Taka and Kado province) are located at greater distance from the epicenter of the earthquake than the precedent villages, but are reported directly affected. The sites are located close to each other and are both located within a 250 Km radius from the epicenter of the earthquake, so impact will be most likely similar in both sites. Only one site is recommended for assessment (Deria is located close to Judy so secondary data can be collected there and if there is a need and if the description of needs differs from the findings in Anais, then assessment team can be tasked to assess Deria). **Site selection:** Anais, one questionnaire adapted for affected population.
	Kado	Deria				X	
4	Kito	Richy			X		**Key note:** Richy and Mery (Kito Province) have not been directly affected by the earthquake but are reported to be hosting IDPs. Sites are close, so situation in one site will be very similar to the situation in the other site. Only one site would be necessary for assessment normally **BUT** you will need to have some elements of comparison for describing the needs of the Host Population in rural areas. Assessing only one site will not provide sufficient information to reach good enough conclusions about their status and needs. Therefore visiting the two locations is recommended (distance between the two sites will allow the same team to visit both location the same day). **Site selection:** Richy and Mery. One questionnaire adapted for Host population in each site (two questionnaires total).
		Mery			X		

Priority level	Province	Urban	G1	G2	G3	G4	Urban areas - Site selection justification
1	Taka	Sandy	X	X	X	X	**Key note:** Cary/Sandy/Judy are urban locations directly affected by the earthquake with affected population (G4) and IDPs in collective centres (G1) in all locations. Sandy is also hosting IDPs (G3) and there is report of IDPs in self-settled camps (G2). The three locations have different characteristics (all located more than 100 km the one from the others, some in plains and some in mountainous areas) and are composed of different types of affected groups (G1, G2, G3 and G4). Impact is presumably greater in locations where all affected groups are represented and we need to check if this is the case. The three locations will needs to be assessed, in priority order Sandy, Cary and Judy (Judy present more affected groups than Cary but is located at a greater distance from the epicentre of the earthquake). **Site selection:** Sandy, Cary and Judy. In Sandy the 4 groups can be assessed with adapted questionnaires (4 questionnaires). In Cary, groups G1 and G4 will be assessed (2 questionnaires). In Judy, groups G1, G2 and G4 will be assessed (3 questionnaires). In total, nine questionnaires will be used in those three sites. Group G3 will be only assessed once but there is no other location with G3 presence reported in this given sites "cluster".
	Zelda	Cary	X			X	
	Kado	Judy	X	X		X	
2	Zelda	Wily	X	X	X		**Key note:** Andy/Wily are locations not directly affected by the earthquake but both reporting considerable numbers of IDPs in self-settled camps (G2), IDPs in collective centres (G1) and Host population (G3). **Site selection:** Wily and Andy will be assessed. In each site at least two questionnaires will be performed for each group to have a clear understanding of their situation and allow comparison between results of the same group. If more time is available and if results differ a lot, one more questionnaire by group can be administered. In total, a minimum of 12 questionnaires are recommended in those two sites.
	Kito	Andy	X	X	X		
3	Kito	Hishy			X		**Key note:** Ragny/Hishy/Dady are urban locations not directly affected by the earthquake and hosting IDPs (Kado and Kito Province). IDPs in self-settled camps are also reported in Dady. Hishy and Ragny are close and will probably show a similar situation for the host population, so there is no need to visit them both. Dady is isolated from any other town and no comparison is possible with any other site, so it will need to be assessed. **Site selection:** Dady and Hishy. In Hishy one questionnaire will be used for host community. In Dady 2 questionnaires will be used for G2 and G3. As there are two IDPs camps reported in Dady, the team can also complete a second questionnaire for G2. In total, four questionnaires can be performed in those two sites.
	Kito	Ragny			X		
	Kado	Dady		X	X		

> Box 4: Larsilandia Case Study – Site selection
>
> By the end of step 2, 32 typical sites where groups of interest can be found were identified. The selection process in step 3 resulted in a reduction of the number of sites to 15. Eight of the sites are in rural areas while seven are urban.
>
> A total of 33 questionnaires need to be administered to complete the assessment.

6.5 Step 4 - Check your resources

At this stage, the only remaining point is ensuring that the necessary means (logistics, skilled staff and finance) are in place to follow the site selection plan. Logistical preparation will also involve reviewing potential limitations caused by time, transport, accessibility and security constraints. If there are not enough resources or the team faces significant constraints, the process will need to be restarted from step 1 in order to:

- Reduce the number of sites and/or,
- Reduce the number of questionnaires within each site (thus affecting the rule of minimum data usable for comparison, thus reducing the accuracy of the findings) and/or,
- Reduce the number of assessed groups (thus keeping more for Phase 3 of assessment or relying more on in crisis secondary data).

Such modifications will be done in detriment of the overall sample requirement. The implications of the site number reduction will have to be communicated clearly to decision makers.

7. Annex 2 - Type of sampling for needs assessment (Phase 1 to phase 4)

Type of sampling	Description
Representative/probability sampling Recommended for Phase 3 and 4	• Based on the principle that any location or informant has an equal (or known non-zero) chance of being selected relative to any other location or informant • Can be used in humanitarian contexts when lists of targeted households are available and all selected locations are accessible **Advantages:** • Generally viewed as the most representative and rigorous type of sampling • Allows results from the sample to be extrapolated to the wider affected area and population **Disadvantages:** • Can be expensive and time consuming to implement, especially in large target areas • Requires special training for correct use • Can miss key informants, i.e., individuals who have particular knowledge about an area or issue • More appropriate for quantitative rather than qualitative type of assessment
Purposive sampling (non-probability) Recommended for Phase 1 and 2	• Uses the judgement of community representatives, project staff or assessors to select typical locations and/or informants according to certain pre-defined characteristics (purposes) **Advantages:** • Moderately rigorous if correct and clear criteria for selection are followed • Useful when targeting specific groups of affected population or specific affected areas. Less time consuming and less expensive than representative sampling **Disadvantage:** • Generalisations are biased and not recommended. Samples are not representative of population due to subjectivity of respondents
Convenience sampling (non-probability) Not recommended	• Easily accessible locations or informants are sampled **Advantage:** Can be quick and saves resources **Disadvantage:** This is the least rigorous sampling option and definitively not representative, and not clear what are these conveniently located areas are "typical" of, therefore not recommended

Adapted from Joint education needs assessment, 2010

8. Annex 3 - Standard definition of "affected groups"[13]

Casualties - The sum of dead, missing, and injured:
- **Dead** - Persons confirmed as dead and persons missing and presumed dead[14].
- **Missing** - Persons whose status during or after an emergency is not known
- **Injured** - Persons suffering from physical injuries, trauma or an illness requiring medical treatment[15]

Affected - The sum of displaced and non-displaced persons

1) **Non-Displaced** - The sum of host and non-host persons
- **Host** - Persons who are part of a host community or family receiving affected people. Due to the stress placed on the host families and communities, they are considered part of the humanitarian caseload.
- **Non-Host** - People requiring immediate assistance during a period of emergency, who have not moved from their homes or places of habitual residence.

2) **Displaced** - The sum of internally displaced persons, refugees and asylum Seekers, and others of concern
- **Internally Displaced Persons** – "persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border"[16].
- **Refugees and Asylum Seekers** - A refugee is someone who "owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality, and is unable to, or owing to such fear, is unwilling to avail himself of the protection of that country"[17]. "An Asylum Seeker is someone who says he or she is a refugee, but whose claim has not yet been definitively evaluated[18].
- **Others of Concern** - persons who have been displaced by the emergency and form part of the humanitarian caseload, but do not fall into either of the above categories .

Each of the above 3 sub-classes of displaced has the following potential sub-categories. Note that these lower-level classes may not be as universal as the higher level classes described above. The classes and their definitions may need to be adjusted to match operational realities.

- **Camp or Camp-like** - the sum of planned camp or settlement, self-settled camp, collective centre, and reception or transit Site
    - **Planned Camps** - Planned camps are places where displaced populations find accommodation on purpose-built sites, and a full services infrastructure is provided, including water supply, food distribution, non-food item distribution, education, and health care, usually exclusively for the population of the site.
    - **Self-settled Camps** - A displaced community or displaced groups may settle in camps, independent of assistance from local government or the aid community. Self-settled camps are often sited on state-owned, private or communal land, usually after limited negotiations with the local population or private owners over use and access.
    - **Collective Centres** - This type of settlement is where displaced persons find accommodation in pre-existing public buildings and community facilities, for example, in schools, barracks, community centres, town halls, gymnasiums, hotels, warehouses, disused factories, and unfinished buildings. They are often used when displacement occurs inside a city itself, or when there are significant flows of displaced people into a city or town. Often, mass shelter is intended as temporary or transit accommodation.
    - **Reception and Transit Camps** - Oftentimes, it is necessary to provide temporary accommodation for displaced persons. These camps might be necessary at the beginning of an emergency as a temporary accommodation pending transfer to a suitable, safe, longer term holding camp, or at the end of an operation as a staging point of return. Reception and transit camps are therefore usually either intermediate or short-term installations.
- **Private or Individual Accommodation –** The sum of privately hosted and non-hosted
    - **Hosted -** living in someone's else home with them, sharing resources and a hearth with another household group
    - **Non-hosted -** living in his own accommodation

---

[13] Adapted from IASC 2011 Guidelines on the Humanitarian Profile.
[14] EMDAT criteria, http://www.emdat.be/criteria-and-definition
[15] Ibid
[16] UN Guiding Principles on Internal Displacement UN doc E/CN.4/1998/53/Add.2
[17] Article 1, 1951 Convention relating to the status of refugees
[18] UNHCR, http://www.unhcr.org/pages/49c3646c137.html

9. Annex 4 - Glossary

Affected Groups are groups of people requiring (immediate) assistance during a period of emergency, i.e. requiring basic survival needs such as food, water, shelter, sanitation and immediate medical assistance[19].

Sample: A sample is a subset of a frame where elements are selected based on a randomised process with a known probability of selection.

Sampling unit: A sampling unit is that element or set of elements considered for selection in stage of sampling[20]. A sampling unit is one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made. The definition of unit may be made on some natural basis, e.g., household, persons, units of product, tickets, etc., or upon some arbitrary basis, e.g., areas defined by grid co-ordinates on a map, a community, a site, etc[21].

Sampling frame: The list of all areas and population from which a sample is drawn for the assessment is described as the sampling frame. The sampling frame must be defined at the start of the assessment planning process.

Stratification: Stratification consists of dividing the population into subsets (called strata) within each of which an independent sample is selected. The division of a population into parts is known as strata, especially for the purpose of drawing a sample, an assigned proportion of the sample then being selected from each stratum.

The process of stratification may be undertaken on a geographical basis, e.g. by dividing up the sampled area into sub-areas on a map; or by reference to some other quality of the population, e.g. by dividing the persons in a town into strata according to gender or into three strata according to whether they belong to upper, middle or lower income groups.

The term stratum is sometimes used to denote any division of the population for which a separate estimate is desired, i.e. in the sense of a domain of study. It is also used sometimes to denote any division of the population for which neither separate estimates nor actual separate sample selection are made[22].

Purposive Sampling (Non-Probability Sampling): A sample in which the groups for interview are selected according to the researcher's choice. It does not involve random selection, so extrapolation of results to wider populations is not possible; its value lies in selecting information-rich cases for in-depth analysis related to the issues being studied[23]. In purposive sampling, we sample with a *purpose* in mind. One of the first things done when collecting data is to verify that the respondent does meet the criteria for being in the sample. With a purposive sample, you are likely to get the opinions of your target population, but you are also likely to overweight subgroups in your population that are more readily accessible.

Theoretical saturation: The point at which new data collected and analysed no longer bring additional insights to the research questions. For example, if interviews 11 through 15 contain the same information found in the first 10 interviews, theoretical saturation has been reached[24].

---

[19] CERD Glossary, http://www.emdat.be/glossary/9
[20] University of Illinois Chicago, http:\www.uic.edu\classes\socw\socw560\Sampling1.htm
[21] OECD Statistical Glossary, http://stats.oecd.org/glossary/detail.asp?ID=2381
[22] The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", 2003
[23] WFP EFSA 2009
[24] USAID Qualitative research methods 2005

10. References

This document is mainly based on review of past Coordinated Needs Assessment and discussion with assessment experts, particularly:

- Richard Garfield, Professor at Columbia University
- Oleg Bilukha, CDC
- Muireann Brennan, CDC
- Shelley Gornall, UNHCR
- Xavier de Radigues, HNTS
- Hisham Khogali, consultant
- Aldo Benini, consultant
- Samuel Petragallo, HAC
- Susan Erb, ACAPS
- Sandie Walton-Ellery, ACAPS

Apart from the reference in footnotes in the text, the written references used for this technical brief are:

2011, HPN, Common Needs Assessments and humanitarian action, Richard Garfield available at http://www.odihpn.org/documents/networkpaper069.pdf

2000 Rapid Rural Appraisal – Manual for CRS field workers practitioners available at http://www.crsprogramquality.org/storage/pubs/me/RRAPRA.pdf