

- Thematic Section IV: Multi-Sector Questionnaire -



RS30 // PURPOSITIVE SAMPLING, CLUSTERIZATION AND SEGMENTATION TECHNIQUES

The present document provides a series of sampling techniques for selecting and defining communities as targeted areas for the CPA implementation. These techniques ensure the operational feasibility of the CPA by tailoring the step of community selection and redefining the definition of community to the context under study. In fact, what a community encompasses varies from one context to another.

The application of these techniques is not compulsory and actually only recommended on specific occasions to overcome operational limitations or to ensure representativeness. The sampling techniques here proposed help to deal with the following situations:

- ▶ Purposive sampling: When limited resources are available to target all intended communities
- ▶ Clusterization: When size of communities is too small to make the implementation of the CPA efficient
- ▶ Segmentation: When size of communities is too big to ensure an adequate representation within the community.

Purposive Sampling

Purposive sampling methods are non-probability techniques and are thus, not considered as the ideal methods for satisfying representativeness: extrapolating result obtained in one community to other not assessed communities (in contrary to probability sampling frequently used in household surveys). Nevertheless, when probability methods are not feasible or when time and human resources are limited, purposive sampling represents an alternative for obtaining in-depth and useful information on the matter under study (Etikan, Musa, & Alkassim, 2016).

In the frame of the CPA, sampling in the sense of selecting a reduced number of communities to represent a wider range of communities is not contemplated. The use of purposive sampling methods in the frame of the CPA should be limited to pilot tests prior a full implementation of the CPA with the aim of obtaining a deeper understanding of the context with the Multi-Sector Questionnaire (MQ).

The reasoning to limit the usage of sampling to pilot test lies on the CPA being a community-based approach with an action-research methodology which focus on specificities and unique characteristics of the community. Therefore, if the full CPA (MQ + Narrated Community Perspective) is intended and resources are not available to cover all targeted communities, the first step should envisage reducing the number of communities.

Following Flyvbjerg (2006) and Patton (1990), there are up to seven methods of purposive sampling. Here we discuss three of the methods which can be used depending on the objectives and the context where the MQ is applied:

▶ **Homogeneous Sampling:** This method selects only communities that respond to a specific criterion of interest. For example, if a program aims at responding to the needs of internally displaced persons specifically, only communities where an important number of internal displaced persons reside will be selected.

This approach helps to centre resources on what is most relevant for achieving specific goals. Therefore, while final results only address a specific group/characteristic, it effectively represents the most relevant group for the program.

▶ **Maximum Variation Sampling:** This method consists of selecting a small number of communities covering the different levels of protection risk found in a specific context. For example, three communities are selected a priori, based on hypothesis drafted on secondary sources and field experience; one is expected to have a very high level of protection risk; one would be expected to have a neutral level of protection risk, and the third one as having very low level of protection risk.

Following this approach, first would allow you to confirm or reject your hypothesis drafted based on secondary information and field experience and second would provide you an estimated range of the protection risk variations and potential specific problematics found for each type of community.

▶ **Extreme/Deviant Case Sampling:** The method proposes to select only communities that are located at one extreme of the Protection Risk spectrum. In other words, to select only communities with the, a priori, highest or lowest protection risks levels. In this sense, this approach allows to hypothesize that anything remaining will be above/worst or below/better than the communities previously assessed.

Tips for Purposive sampling

- ▶ When applying the full CPA (MQ + NCP), purposive sampling is NOT recommended as a sampling method for representing a wider number of communities by selecting only a few.
- ▶ The use of purposive sampling methods in the frame of the CPA should be limited to pilot tests prior a full implementation of the CPA with the aim of obtaining a deeper understanding of the context and an added value to the pilot test.

Clusterization

Clusterization refers to the concept of cluster sampling widely used in household surveys and also applicable to the CPA¹, but with the limitation that the CPA does not apply probability sampling. Therefore, clusterization implies the grouping together of a number of communities to form a “cluster”, with the objective of increasing the efficiency of the MQ implementation.

Redefining the Community

An important consequence of clusterization (and the segmentation – see below) is that it redefines the initial definition of a community. When clusterization is applied, the cluster becomes the unit of analysis and thus the Community in the frame of the CPA, it actually refers to the cluster itself, rather than to the units within the cluster (i.e. the original communities). As a consequence, the results of the CPA represent the situation of the cluster, not anymore of the original communities.

Clusterization is particularly recommended in contexts where a large amount of small size targeted communities are found (e.g. in rural contexts). As an indicative rule of thumb, when communities are smaller than 50 persons, then clusterization should be considered. However, the final decision on clusterization should be taken according to the operational capacities of the organization and the characteristics of the context.

For the clusterization to be effective, all cluster units (i.e. the original communities) should be represented during MQ interviews by selecting at least one representative of each cluster unit. This way, it ensures that the information recorded through the MQ reflects the perspective of each cluster unit.

Given the importance of having the right information reflected at the cluster level, its definition need to be based on the following premises (Rodríguez, Ferreras, & Núñez, 1991):

- ▶ The units of the cluster should be as homogeneous as possible (i.e. similar between them) in line with the criteria defined in the **Definition of Community**.

¹ Although the term “cluster sampling” usually refers to a penultimate stage where units are grouped, and from there a subsample of unit is selected randomly to be assessed, its traditional term refers to the situation where all units of the cluster are assessed . In the frame of the CPA, the later definition of cluster sampling is used.

- ▶▶ Cluster should be of reduced size; meaning not exceeding a certain number of units by cluster. The rule of thumb in the context of the CPA is to set the limit at 7 units per cluster.

This way we ensure that the cluster itself acts as a new community, with: 1. a certain degree of homogeneity between its different units; 2. where all units are represented during interviews; and 3. where the total number of clusters over the total number of units is kept at a reasonable rate.

Tips for Clusterization

- ▶▶ Clusterization should only be applied in context where there are limited resources to interview all originally targeted communities.
- ▶▶ Clusterization better fits in rural contexts where a large number of targeted communities of reduced size are found.
- ▶▶ Clusters should be homogeneous, guided by the **Definition of Community**, and of reduced size.
- ▶▶ All cluster units embedded within one cluster have to be represented during the MQ interview.
- ▶▶ Note that either way a group of communities, all of them having a reduced population, can be clustered together or a community with reduced population can be clustered together with a large one.

Segmentation

As defined by ICF International (2012) segmentation allows the splitting of initially targeted big sized communities into smaller segments which become the new Communities. For this reason, segmentation appears as a handy technique in urban contexts where large cities can be found or in highly dispersed communities where it is needed to cover long distances from one side to another of the community.

As a rule, when the size of an initially targeted community is greater than 3.000 persons, segmentation is to be applied. Unlike clusterization, where its application principally relies on operational and contextual criteria, segmentation can also be applied when the population is smaller than 3.000, but is required when the population is greater than 3.000 persons. In fact, in the latter case, the MQ cannot guarantee the reliability of results.

The principles for segmentation are:

- ▶▶ The population within segments has to be as homogeneous as possible in line with the criteria defined in the **Definition of Community**.
- ▶▶ The number of segments to be selected is determined by first, the premise of homogeneity and second, that no segments should be greater than 3.000 persons.

In addition, when segmenting initial communities, it is important to ensure that representatives can be identified in each segment to answer the questions of the MQ. All defined segments should be treated as independent community. As a consequence, the results of the CPA represent the situation of each segment. However, unlike with clusterization, results can be further aggregated to represent the original community.

Tips for Segmentation

- ▶▶ Segmentation should always be applied when the initial targeted community has a population greater than 3.000 persons.
- ▶▶ Segmentation better fits in urban contexts or highly dispersed areas.
- ▶▶ To form a segment, its population should be homogeneous (as guided by the **Definition of Community**) and not exceed 3.000 persons.
- ▶▶ Each segment is treated as an independent community, where the MQ is implemented.
- ▶▶ Results of the segments can be aggregated to represent the whole original community.

References

- ▶ Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- ▶ Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- ▶ ICF International. (2012). *Demographic and Health Survey Sampling and Household Listing Manual*. <https://dhsprogram.com/publications/publication-dhsm4-dhs-questionnaires-and-manuals.cfm>
- ▶ Patton, M. Q. (1990). Purposeful sampling. In *Qualitative evaluation and research methods* (2nd ed., pp. 169–186). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/nur.4770140111>
- ▶ Rodríguez, J., Ferreras, M. L., & Núñez, A. (1991). Inferencia estadística, niveles de precisión y diseño muestral. *Reis: Revista española de investigaciones sociológicas*, (54), 139–166. <https://dialnet.unirioja.es/servlet/articulo?codigo=249348>



The work of authorship is licensed under Creative Commons.