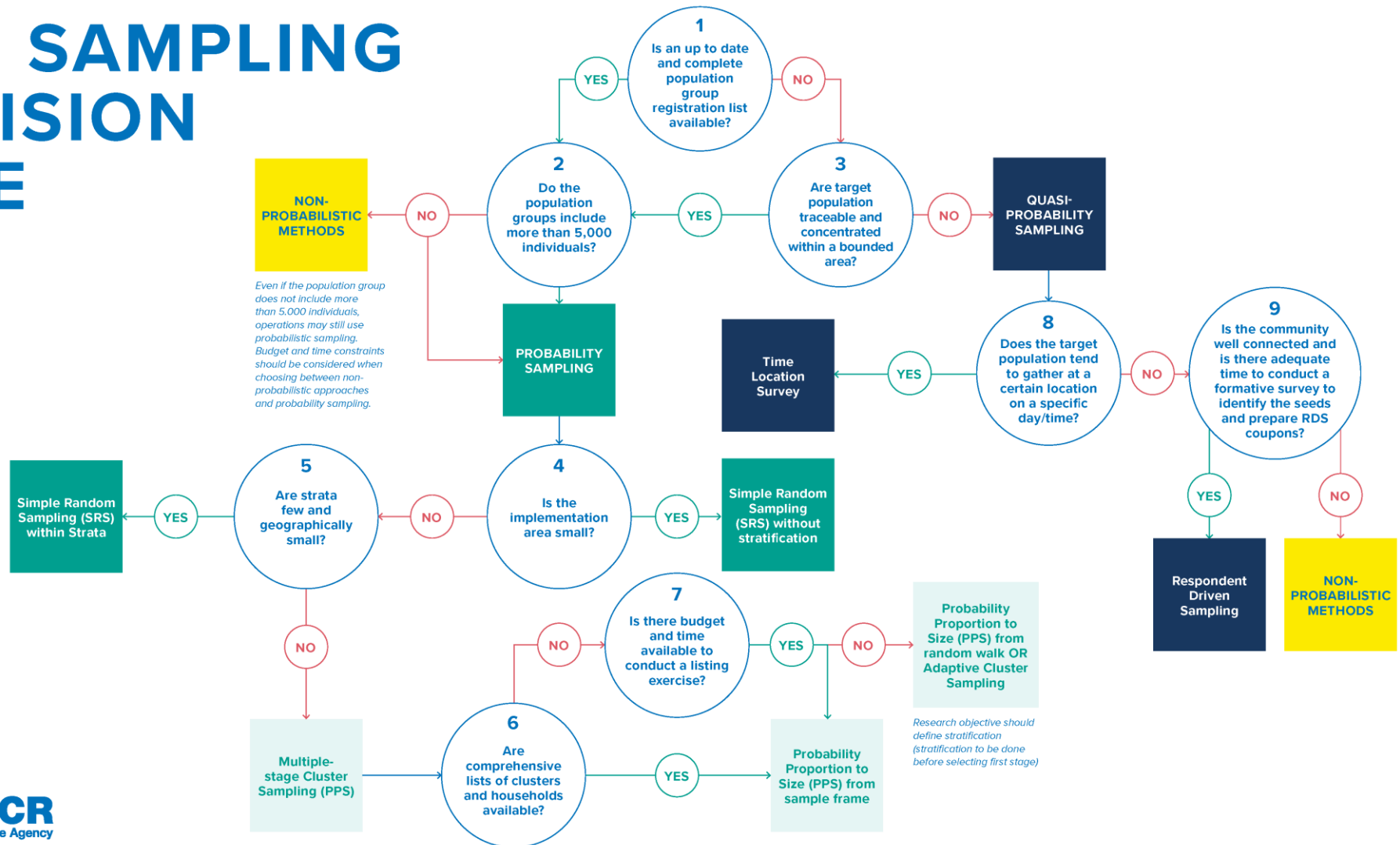


RMS SAMPLING DECISION TREE



Please note: The RMS sampling decision tree serves as a methodological framework for RMS quality assurance as well as being a tool for colleagues with expert statistical / survey expertise to decide on most appropriate methodological approaches. Please consult with your DIMAs.

Also see the [RMS basic sampling guidance](#).

RMS sampling decision tree instructions:

Ask the questions in each box and follow the decision tree to determine which sampling method is most appropriate for your RMS for each population group included (e.g., refugees and asylum seekers, as well as IDPs), based on the country context (i.e., availability of sampling frame, budget restrictions, location of population group, etc.).

1. Is an up to date and complete registration list of the population group available?

- 'Up to date' refers to a database which reflects the current population in the country. If the size or demographic composition of the population group has changed more than 10% since the last registration, it cannot be considered up to date.
- 'Complete' refers to a database which contains registration information for 80% or more of the total population group.
- **If yes, ask:** (2) Does the population group include more than 5,000 individuals?
- **If no, ask:** (3) Is the study population traceable and lives in a bounded area(s)?

2. Does the population group include more than 5,000 individuals?

- **If yes:** Use probability sampling methods. Probability sampling methods use random selection to choose survey participants. This ensures results are statistically representative of the study population. Probabilistic sampling requires a defined study population and known probability for selecting each household. Ask (4) "Is the implementation area small?" to determine what type of probability sampling.
- **If no:** Either non-probabilistic or probabilistic methods may be used to gather indicator information for reporting. Non-probabilistic methods do not use random selection to choose survey participants and therefore suffer from sampling biases. Indicator estimates derived from such methods will not be statistically representative of the population.
 - (a) Even if the population group is smaller than 5,000 individuals, operations may still use probabilistic sampling. Budget and time constraints should be considered when choosing between non-probabilistic approaches and

probability sampling. Probabilistic approaches will generally require more resources.

- (b) Non-probabilistic methods do not guarantee that the results will be representative of the concerned population and should be seen as a last resort to generate indicator data. These may include conducting surveys using convenience sampling, quota sampling, or snowball sampling.¹ *Convenience sampling* requires little or no planning and involves selecting respondents who are readily available. An example of convenience sampling would include interviewing households who are present in the street or a public building. *Quota sampling*, as the name suggests, involves determining a certain number of respondents from each study group (i.e., population groups such as refugees and asylum seekers, IDPs) and conveniently interviewing households from each group. *Snowball sampling* involves selecting respondents randomly or purposefully in the beginning and asking for referrals.

3. Is the population group traceable and concentrated within a bounded area(s)?

- 'Traceable' refers to population groups whose homes can be located or who can be reached via telephone.
- 'Concentrated within a bounded area(s)' refers to population groups who live within a geographic area with clearly identified boundaries, such as towns, villages, neighbourhoods, settlements, camps or other geographic areas for which specific boundaries can be clearly identified by a field team.
- **If yes, ask:** (2) "Does the population group include more than 5,000 individuals?"
- **If no:** Use quasi-probability sampling methods. Quasi-probability methods do not use pure random sampling to select participants, but still provide population representative estimates using mathematical/statistical techniques. Ask (8) "Does the population group tend to gather at a certain location on a specific day/time?" to determine what type of quasi-probability sampling.

¹ For a simple overview of non-probabilistic sampling, see <https://comparativemigrationstudies.springeropen.com/articles/10.1186/s40878-016-0044-9/tables/1>, <https://www.questionpro.com/blog/non-probability-sampling/>, or <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm>

PROBABILITY SAMPLING

4. Is the implementation area small?

- Determine if the implementation area contains only a few (2-3 for example) villages, cities, or camps and if households within those areas are easily accessible by vehicle or on foot.
- **If yes:** Use *Simple Random Sampling (SRS) without stratification*. SRS gives each population group an equal chance of being selected for the sample.
- **If no, ask:** (5) Are strata few and geographically small?

5. Are strata few and geographically small?

- *Strata* are distinct subgroups of the target population. Each population group can be further split into different strata. For example, RAS and IDPs may be divided into two distinct subgroups, or stratum: 1. RAS living in camps; 2. RAS living outside of camps, or 1. Rural IDPs; 2. Urban IDPs, or 1. Returnees who returned before a certain; 2. after a certain date.
- **If yes:** Use *Simple Random Sampling (SRS) within Strata*. SRS gives each group within a stratum an equal chance of being selected for the sample.
- **If no:** Use a Multiple-Stage Cluster Sampling (PPS) strategy. A probability proportion to size (PPS) strategy first selects clusters (e.g., villages, enumeration areas, etc.), then selects households within the cluster. Clusters are selected based on their population size. Ask (6) “Are comprehensive lists of clusters and households available?” to determine what type of Multiple-Stage Cluster Sampling (PPS).

Multiple-stage Cluster Sampling

6. Are comprehensive lists of clusters and households available?

- If lists of clusters and the households within each cluster are available, this list can be used to select households.
- **If yes:** Use *Probability Proportion to Size (PPS) from sample frame*. After selecting clusters based on their population size, use the available sample frame to randomly select households.
- **If no, ask:** (7) Is there budget and time available to conduct a listing exercise?

7. Is there budget and time available to conduct a listing exercise?

- A household listing exercise can be used to construct a sampling frame but is time consuming and expensive. A listing exercise requires listing agents that count every household within the specified boundary and create a list of all eligible households.
- **If yes:** Use *Probability Proportion to Size (PPS) from sample frame*. After selecting clusters based on their population size, conduct a listing exercise within the selected clusters and use the sample frame created from the listing exercise to randomly select households.
- **If no:** Use *Probability Proportion to Size (PPS) from random walk AND/OR adaptive cluster sampling* depending on the location of target clusters and concentration of population groups within clusters. Adaptive cluster sampling is used to select clusters while random walk is used to select households within a cluster. Note that adaptive cluster sampling may require additional technical support.
 - (a) Use *Probability Proportion to Size (PPS) from random walk if the population is concentrated within clusters and clusters tend to be found in the same region or geographic area*. After selecting clusters based on their population size, enumerators survey every *n*th household until the required number of households is selected. For example, enumerators may visit every third household within the cluster.
 - (b) Use *adaptive cluster sampling if the population's geographic locations are generally known and the population is concentrated, but the clusters are spread out within the country*. This method first divides the population into a grid of plots and randomly selects a few initial plots. The selected plots are reviewed to determine whether the population is present or not – a condition required to select the plot. If there are a minimum number of households of the population group present in the plot, the research team also reviews any neighbouring plots to ensure the minimum number of households of the population group are present. Each plot that has the minimum number of households of the population group is selected until there are no additional neighbouring plots with such households present. Clusters of plots are then generated where the survey will take place. As previously noted, this method may be most useful when the population is very concentrated in certain plots, but these plots are widely dispersed.

QUASI-PROBABILITY SAMPLING

8. Does the population tend to gather at a certain location on a specific day/time?

- The population may habitually gather for certain ceremonies, festivities, or events that are well known. In case these are known in advance and all population members have the possibility to be present at the location, these days and times can be opportunities to survey the population and. For example, the population might gather for a humanitarian distribution or a registration exercise.
- **If yes:** Use a *Time Location Survey*. Identify times and locations where the population is likely to gather and select the times and locations where the greatest number will be present. Conduct interviews on location.
- **If no, ask:** (9) Is the population a well-connected community and is there adequate time and expertise to carry out a Respondent Driven Sampling (RDS) approach, which entails conducting a formative survey to identify seeds and prepare RDS coupons?

9. Is the population a well-connected community and is there adequate time and expertise to carry out a Respondent Driven Sampling (RDS) approach, which entails conducting a formative survey to identify seeds and prepare RDS coupons?

- **If yes:** Use *Respondent Driven Sampling (RDS)*. Identify initial 'seed' population who are well connected in their community and can refer additional households of the same population group to the survey. Respondents will continue to recruit additional households until the desired sample size is reached. Additional data protection considerations need to be taken when adopting RDS as seeds will be sharing personal data of other potential survey respondents.
- **If no:** Use non-probabilistic methods. Non-probabilistic methods do not guarantee that the results will be representative of the concerned population and should be seen as a last resort to generate indicator data. These may include conducting surveys using convenience sampling, quota sampling, or snowball sampling.
 - (a) *Convenience sampling* requires little or no planning and involves selecting respondents who are readily available. An example of convenience sampling would include interviewing households who are present in the street or a public building.
 - (b) *Quota sampling*, as the name suggests, involves determining a certain number of respondents from each study group (i.e., population groups such as refugees and asylum seekers, IDPs) and conveniently interviewing households from each group.
 - (c) *Snowball sampling* involves selecting respondents randomly or purposefully in the beginning and asking for referrals.