ARTICLE 6.4 MECHANISM

A6.4-AMT-010

Methodological tool

Sampling and surveys

Version 01.0

United Nations
Framework Convention on
Climate Change

TABLE OF CONTENTS **Page**

1. INTRODUCTION 4
1.1. Scope 4
1.2. Entry into force and validity 4
2. DEFINITIONS 4
2.1. General terms 4
2.2. Methodological terms and definitions 4
3. NORMATIVE REFERENCES 6
4. APPLICABILITY 6
5. METHODOLOGICAL APPROACHES 6
5.1. General requirements 7
5.2. Sampling plan 8
5.2.1. Sampling design 9
5.2.2. Data collection 9
5.2.3. Procedural sampling arrangements 10
5.2.4. Reporting of data analysis and results 10
5.3. Sample size determination 11
5.4. Sampling methods and sample size determination 12
5.4.1. Simple random sampling 12
5.4.2. Stratified random sampling 14
5.4.3. Systematic sampling 17
5.4.4. Cluster sampling 18
5.4.5. One-stage cluster sampling 18
5.4.6. Multi-stage cluster sampling 19
5.5. Failure to achieve the reliability requirements 21
5.6. Survey and data collection methods 21
5.6.1. Physical survey methods 22
5.6.2. Remote survey methods 22
5.6.3. Remote data collection methods 23

APPENDIX 1. ADDITIONAL GUIDANCE ON SAMPLING METHODS AND SURVEY DESIGN 24

APPENDIX 2. BEST-PRACTICE EXAMPLES FOR A SINGLE SAMPLING PLAN FOR A HOMOGENEOUS PROGRAMME OF ACTIVITIES 28

APPENDIX 3. BEST-PRACTICE EXAMPLES FOR A SINGLE SAMPLING PLAN FOR A HETEROGENEOUS PROGRAMME OF ACTIVITIES USING A STRATIFIED SAMPLING APPROACH 39

APPENDIX 4. BEST-PRACTICE EXAMPLES FOR CONFIDENCE INTERVALS AND RELIABILITY ASSESSMENTS 58

APPENDIX 5. SAMPLE SIZE CALCULATOR 76

1. Introduction

1.1. Scope

1. This methodological tool establishes requirements for the application of sampling and surveys for Article 6.4 projects and programmes of activities (PoAs), hereinafter referred to as “Article 6.4 activities”. It specifies default reliability requirements and the information to be included in a sampling plan.

1.2. Entry into force and validity

2. This methodological tool enters into force on 9 October 2026.

3. This methodological tool remains valid for five years, until 8 October 2031, unless an earlier date applies if the methodological tool is revised or withdrawn in accordance with the “Procedure: Development, revision and clarification of methodologies and methodological tools” (A6.4-PROC-METH-001).[1]

2. Definitions

2.1. General terms

4. The following general terms apply in this methodological tool:

   (a) "Shall" is used to indicate requirements that must be followed;

   (b) "Should" is used to indicate that, among several options, one course of action is recommended as particularly suitable; and

   (c) "May" is used to indicate what is permitted.

2.2. Methodological terms and definitions

5. The following methodological terms and definitions apply in this methodological tool:

   (a) **Bias (systematic error):** a systematic deviation of a sample-based estimate from the true population parameter, resulting in consistent overestimation or underestimation. Bias typically arises from non-sampling errors, including survey design flaws, measurement errors, non-response, or incomplete coverage of the sampling frame;

   (b) **Cluster**: a naturally occurring grouping of sampling units within a population, such as households within a village or appliances within a building, that is used as the primary sampling unit in cluster sampling or multi-stage sampling;

   (c) **Confidence level:** the probability that a confidence interval, constructed from repeated sampling, contains the true population parameter (e.g., 95 per cent);

   (d) **Mean**: a numerical parameter representing the arithmetic average of a continuous variable measured across a sample, typically used to estimate the corresponding population mean;

[1] See https://unfccc.int/sites/default/files/resource/A6.4-PROC-METH-001.pdf.

(e) **Multi-stage sampling**: a sampling approach in which selection is carried out in two or more steps, typically involving the selection of clusters at an initial stage followed by sampling of units within selected clusters;

(f) **Non-sampling error**: any error not attributable to random sampling variation, including measurement errors, data processing errors, survey non-response, or coverage errors;

(g) **Parameter**: a quantitative characteristic of a population that is used in the calculation of greenhouse gas emissions or removals;

(h) **Population**: the complete set of units to which a parameter from the results of a sample applies, such as all households, installations, pieces of equipment (e.g., cookstoves, solar home systems), included within the boundary of an Article 6.4 activity (project or programme of activities);

(i) **Proportion**: a parameter representing the fraction of units in a population that possess a specified attribute, expressed as a value between 0 and 1;

(j) **Reliability:** the extent to which a sample-based estimate meets the specified target precision at the specified confidence level;

(k) **Sample**: a subset of a population, selected for the purpose of estimating one or more population parameters;

(l) **Sampling frame**: a complete, accurate and up-to-date listing or representation of all eligible sampling units in the population from which a probability sample can be drawn. For cluster or multi-stage sampling, the frame shall identify all eligible primary sampling units and, where required by the design, the lower-stage sampling units;[2]

(m) **Sampling method**: the method used to select sampling units from a sampling frame or target population, for inclusion in a sample;

(n) **Sampling plan**: a documented description of the sampling design, sample size determination, data collection procedures, quality assurance and quality control measures, and analytical methods applied to estimate parameters;

(o) **Sampling unit**: the basic element selected from the population for inclusion in the sample, such as a household, installation, piece of equipment, or facility;

(p) **Stratum (plural: strata)**: a sub-population into which a target population is divided for the purpose of stratified sampling. The strata divide the population into a set of mutually exclusive and exhaustive sub-populations;

(q) **Survey**: a data collection method that gathers information from respondents;

(r) **Target precision:** the maximum allowable sampling error for a sample-based estimate, specified as a margin of error at a specified confidence level;

(s) **Unbiased estimate**: a sample-based estimate whose expected value equals the true population parameter, such that it does not systematically overestimate or underestimate that parameter.

[2] A suitable map with the sampling units marked on it and properly delineated may also be regarded as a sampling frame and used in drawing samples.

6. Furthermore, the terms in the “Glossary: Article 6.4 mechanism terms” (A6.4-GLOS-GOV-001) shall apply.

3. Normative references

7. The following document is indispensable for the interpretation of Article 6.4 mechanism terms used in this methodological tool. When applying this tool, a valid version of the document listed below shall be used:

   (a) Glossary: “Article 6.4 mechanism terms” (A6.4-GLOS-GOV-001).[3]

4. Applicability

8. This methodological tool is applicable to Article 6.4 activities where its use is explicitly referenced in the applicable mechanism methodology.

9. This methodological tool is only applicable where the referring mechanism methodology specifies:

   (a) The parameters that are to be determined through sampling;

   (b) The sampling method to be used;

   (c) For each parameter, whether it involves mean or proportion values;

   (d) What actions are to be undertaken by activity participants if the reliability requirements are not met, such as the application of conservative default values or discount factors; and

   (e) The approach followed for addressing the uncertainty associated with the estimated parameters determined through sampling, such as using the lower or upper bound of the confidence interval, whichever results in a more conservative estimate.

10. Mechanism methodologies may provide further specifications and requirements for how this methodological tool shall be applied by activity participants in the context of the type of mitigation activities that are covered by the methodology. This may, for example, include reliability requirements other than those used in this methodological tool.

11. Where the mechanism methodology referring to this methodological tool specifies approaches or requirements that differ from those described in this methodological tool, the requirements contained in the mechanism methodology shall take precedence.

5. Methodological approaches

12. Activity participants may use sampling to obtain unbiased and reliable estimates of parameter values used in the calculation of greenhouse gas (GHG) emission reductions or net GHG removals where it is not practical or cost-effective to monitor all units within a defined population, while ensuring environmental integrity and methodological consistency.

[3] See https://unfccc.int/sites/default/files/resource/A6.4-Glossary.pdf.

5.1. General requirements

13. For each parameter to be estimated, activity participants shall identify the variable type (e.g. continuous, count or binary) type and the statistic to be estimated. The applicable mechanism methodology shall specify whether the parameter is to be estimated as a mean, proportion, total, ratio or another statistic. A binary variable may be summarized either by its mean or its equivalent proportion. The sampling design, sample-size calculation and analysis method shall be appropriate to the specified parameter, variable type, and statistic to be estimated.

14. Unless otherwise specified in the applicable mechanism methodology, activity participants shall achieve, as a minimum, a precision of ±10 per cent at a 95 per cent confidence level for each sampled parameter. A 95/10 requirement means that the half-width of the 95 per cent confidence interval should not exceed 10 per cent of the estimated parameter value, where precision is expressed in relative terms.

15. Activity participants shall select a sampling method appropriate to the characteristics of the population. Where the population is sufficiently homogeneous with respect to the parameter of interest, simple random sampling may be applied. Where the population contains distinct sub-populations that are expected to differ materially with respect to the parameter of interest, activity participants shall apply stratified random sampling or another statistically valid sampling approach that appropriately addresses such heterogeneity.

16. For Article 6.4 PoAs, parameter values shall be determined through sampling at the component project (CP) level and each CP be treated as a separate population, unless a common sampling plan covering two or more CPs within the same PoA is justified in accordance with paragraph 17 below.

17. A common sampling plan covering two or more CPs within the same Article 6.4 PoA may be applied only where activity participants demonstrate that either:

    (a) The CPs included in the group are sufficiently homogeneous with respect to the parameter of interest and the factors likely to affect that parameter. The basis for grouping CPs shall be objective, documented and verifiable and be based on technology type, implementation model, geographical or climatic conditions, user category, fuel type or other factors that may materially affect the sampled parameter; or

    (b) Any material heterogeneity among the CPs is appropriately accounted for through the sampling design either using stratified random sampling or another statistically valid approach.

18. Where a common sampling plan is applied across CPs:

    (a) Simple random sampling may be applied only where the grouped CPs are demonstrated to be sufficiently homogeneous with respect to the parameter of interest;

    (b) Where material differences exist among CPs, activity participants shall apply stratified random sampling or another statistically valid approach that ensures the relevant CPs or strata are adequately represented and that material CP-specific differences are not masked in the resulting estimate;

    (c) The grouping of CPs shall be justified using objective, documented and verifiable criteria relevant to the parameter of interest;

(d) The sampling frame shall identify the CP to which each sampling unit belongs and shall be sufficiently disaggregated to assess the representativeness of the sample across the CPs;

(e) The sample allocation across CPs or strata shall be statistically justified and shall not systematically under-represent any CP; and

(f) The sampling plan shall include procedures for identifying and addressing CP-specific non-response, missing units, non-operational units and other factors that could bias the resulting estimate.

19. Where CP-specific parameter values are applied, the reliability requirements shall be achieved separately for each relevant CP. Where a common parameter value is applied to a group of CPs, the reliability requirements shall be achieved for the grouped population.

20. Activity participants shall ensure that samples are drawn randomly from a complete and up-to-date sampling frame. Random selection may be undertaken using random number tables or an appropriate software-based random number generator. Where systematic sampling is applied, the ordering of sampling units shall be demonstrably free from trends, periodicity or cyclical patterns that could introduce bias (e.g., systematic sampling with random start). The procedure used to randomly select sampling units, including the randomization method, seed or equivalent documentation, shall be recorded so that the selection can be reproduced, e.g. for validation or verification purposes.

21. The sampling frame shall include all units within the population, including units that are inactive, non-operational, inaccessible, abandoned, removed or otherwise unavailable at the time of sampling, unless their exclusion is justified, documented and treated conservatively in accordance with the applicable mechanism methodology.

22. Activity participants shall document and justify any replacement of selected sampling units, adjustment of the sampling frame, treatment of non-response, or exclusion of records. Such measures shall aim to reduce any bias that could result in higher estimates of emission reductions or net removals.

5.2. Sampling plan

23. Activity participants shall develop a sampling plan and include it in the project design document (PDD), programme of activities design document (PoA-DD), or component project design document (CP-DD), as applicable. [4,5]

[4] Revisions to a sampling plan included in a registered PDD or PoA-DD or CP-DD, for example to shift from separate sampling efforts in each CP to a common sampling plan across multiple CPs within the PoA, shall be submitted and assessed in accordance with the applicable “Procedure: Article 6.4 activity cycle procedure for projects” (A6.4-PROC-AC-02) or the “Procedure: Article 6.4 activity cycle procedure for programmes of activities” (A6.4-PROC-AC-03). Such revisions shall be justified and shall demonstrate that the revised sampling approach remains statistically valid and capable of achieving the required reliability requirements for the relevant sampled parameters.

[5] As per the requirements in the relevant sections of the “Procedure: Article 6.4 activity standard for projects” (A6.4-PROC-AC-02) or “Procedure: Article 6.4 activity standard for programmes of activities (A6.4-PROC-AC-03)”, activity participants may submit the monitoring plan (of which the sampling plan is a part) for the proposed Article 6.4 project, PoA or CP either at the time of validation or: (a) at any time prior to the submission of a request for issuance for the first monitoring period; or (b) together with the request for issuance for the first monitoring period.

5.2.1. Sampling design

24. Activity participants shall describe the sampling design, including the following elements:

    (a) Objectives: Description of the objectives of the sampling, including the timeframe;

    (b) Parameters: Identification and description of all parameters to be estimated through sampling, including specification of whether each parameter represents a mean (from a continuous variable) or a proportion (from a discrete variable);

    (c) Reliability requirements: The relative precision and confidence level used for each parameter;

    (d) Population: Definition of the population under study for each parameter and description of the key characteristics that define inclusion. Where a population is divided into subgroups defining strata or clusters, the strata or clusters shall be described by their key characteristics;

    (e) Sampling method: Identification and justification of the sampling method to be used for each parameter in accordance with the requirements in paragraph 15 and paragraph 35 in section 5.4 below;

    (f) Sample size: Description and justification of the assumptions used to determine the minimum sample size for each parameter, in accordance with the requirements in paragraph 28 of section 5.3 below and the selected sampling method described in section 5.4; and

    (g) Sampling frame: A complete and accurate listing of all units within the population. The sampling plan shall describe how the sampling frame is constructed, maintained, updated and verified before sample selection in accordance with the requirements in paragraph 26 below.

5.2.2. Data collection

25. Activity participants shall develop and document appropriate procedures for data collection or field measurements to ensure that samples are selected in a manner that avoids bias and to minimize non-sampling errors, including:

    (a) Identification of all variables to be measured and the methods by which data will be collected (e.g., use of equipment and/or surveys);

    (b) Determination of the appropriate timing (e.g., seasonality) and frequency of measurements;

    (c) Demonstration of the representativeness of the measurements, including the application of conservative measures or corrections where appropriate;

    (d) Provisions to maximise response rates and corrective measures to address (i.e. sampling units found, after selection, not to belong to the target population), sampling units found, after selection, not to belong to the target population such as duplicate entries, ineligible units, or units outside the activity boundary. Corrective measures may include oversampling, replacing non-respondents with similar sampling units, and imputing missing responses (only where this leads to conservative results), as appropriate. Where originally selected

sampling units cannot be located, contacted, accessed or measured, the status of such units and the reason for non-response or non-measurement shall be documented;

(e) Provisions for data entry, including documentation of the steps taken to avoid and minimise the introduction of errors;

(f) Quality assurance and quality control measures, including quality checks performed on entered data and provisions to identify and address errors; and

(g) Provisions to guarantee data integrity, including, for example, methods to prevent fraud and accidental deletion.

5.2.3. Procedural sampling arrangements

26. Activity participants shall describe how the sampling design will be implemented in practice. The implementation plan shall include:

(a) An implementation schedule covering key steps such as preparation, data collection, processing, and analysis;

(b) Definition of roles and responsibilities of the personnel involved, including the required skills, technical expertise, equipment, and logistical resources necessary to carry out the sampling;

(c) Procedures to ensure the consistent and reliable execution of sampling activities, including the training and supervision of field staff, as well as data management procedures covering data collection, entry, validation, storage, and security to ensure data integrity and traceability;

(d) Procedures for addressing practical challenges encountered during implementation, such as non-response, inaccessible sampling units, or deviations from the sampling design. Appropriate contingency measures shall be defined, including provisions for replacement, re-sampling, or corrective actions, to ensure that the sampling results remain reliable and suitable for verification; and

(e) Where there is a material risk of attrition, non-operation or deterioration of the sampling frame over time, measures to identify and address such risks. These may include rotating panel sampling, fixed-panel sampling with refresh samples, sampling-frame verification checks or risk-based stratification.

5.2.4. Reporting of data analysis and results

27. The sampling plan shall include provisions for the recording and reporting of the sampling results and the data analysis methods used, including:

(a) How the data collected and the measurements are recorded (in a spreadsheet or other appropriate format that enables the reader to reproduce the results);

(b) Description of how the collected data will be used, including the analytical methods used to determine parameter estimates;[6]

[6] The treatment of missing or non-contactable units shall be reflected, as appropriate, in the parameter estimate, uncertainty assessment, reliability calculation or conservative adjustment.

(c) The values estimated for each parameter with their corresponding 95 per cent confidence interval; and

(d) Compliance with reliability requirements (including any corrective actions taken).[7]

5.3. Sample size determination

28. Activity participants shall calculate the minimum sample size required to achieve the reliability requirements. In addition to the reliability requirements, the sample size may be determined based on:

(a) The sampling method or design;

(b) Whether the parameter of interest represents a mean value or a proportion;

(c) The expected value of the parameter; and

(d) The expected variance (or standard deviation) of the variable under consideration.[8]

29. The expected value of the parameter and the expected variance (or standard deviation) shall be based on results from similar studies, including other similar Article 6.4 activities or previous monitoring periods, pilot studies, or conservative estimates based on the activity participants' knowledge of the data. [9,10]

30. For sample-size planning, an expected standard deviation shall be obtained from a comparable study, previous monitoring period, pilot study or a conservative assumption. For assessment of an achieved mean estimate where the population standard deviation is unknown, the confidence interval shall use the Student's t-distribution, unless use of a normal approximation is statistically justified.

31. Where a normal approximation is used for estimating a proportion, the planned sample size shall provide adequate expected numbers of observations in both outcome categories, namely units with the characteristic being estimated and units without that characteristic, taking into account the expected proportion (p) and its complement (q = 1 – p). Where this condition is not met, including where either outcome category is expected to have very few observations, activity participants shall use an alternative statistically justified method for determining the sample size and assessing reliability. Where either (p) or (q) may be used for sample size determination, the value resulting in the larger required sample size shall be applied.

---

[7] The treatment of missing or non-contactable units shall be reflected, as appropriate, in the parameter estimate, uncertainty assessment, reliability calculation or conservative adjustment.

[8] The population variance is denoted by σ^2 and the population standard deviation by σ. In practice, these parameters are typically unknown and are estimated using sample data, where SD^2 denotes the sample variance and SD denotes the sample standard deviation. The standard deviation is the square root of the variance.

[9] Activity participants may conduct a sampling campaign before the monitoring period to determine the standard deviation.

[10] Depending on the sampling method, if the parameter of interest is a proportion, then the expected variance can be derived directly from the target value.

32. Where more than one parameter is to be estimated through sampling, activity participants shall determine the required sample size:

    (a) Separately for each parameter; or

    (b) Where a single sampling plan is applied for determining several parameters, based on the most conservative requirement, i.e., the largest sample size required among the parameters.

33. Where the total population size is smaller than the required sample size, including the minimum sample size, activity participants shall apply a census of the population.

34. The sample size shall be calculated according to the relevant sampling method and depending on the type of parameter as described in section 5.4.

5.4. Sampling methods and sample size determination

35. Activity participants shall select an appropriate sampling method based on the characteristics of the population and other activity-specific circumstances. The selected approach shall ensure unbiased estimates of parameters and enable the achievement of the reliability requirements. Activity participants shall justify the appropriateness of the selected method.

36. Unless otherwise specified, the sample-size equations in this section are based on large-sample approximations, asymptotic normality of the relevant estimator and symmetric confidence intervals.

5.4.1. Simple random sampling

37. Simple random sampling is suited to populations that are homogeneous, where each element of the population has an equal probability of selection into the sample and where there is no prior knowledge of the population's structure or characteristics.

38. When a variable is discrete (binary or attribute-based variables), generating a proportion parameter of interest, and the total population size is known, the sample size shall be calculated as follows:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times p \times q}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times p \times q} \qquad \text{Equation (1)}$$

Where:

n	=	Estimated minimum sample size
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval e.g. 1.96 for 95 per cent confidence interval[11]
N	=	Total population size
p	=	Expected proportion of successful cases

[11] For a two-sided confidence interval, *α* represents the total probability outside the confidence interval. For 95 per cent confidence level, *α=0.05,* with *α/2* being on each tail of the standard normal distribution curve. The corresponding critical value is therefore 97.5th percentile of the standard normal distribution, commonly expressed as *Z α/2=1.96*

q	=	Expected proportion of unsuccessful cases, $q = 1 - p$
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., $0.1 \times p$

39. For sample-size calculations, if the expected proportion of successful cases (p) exceeds 0.5, a conservative approach to relative precision shall be used: set both the expected proportion of successful cases (p) and the expected proportion of unsuccessful cases (q) to 0.5, and set $e_{abs} = 0.1 \times p = 0.05$. Where a requirement is expressed in relative terms, the absolute precision used in the sample-size equation shall be calculated from the parameter defined in the applicable mechanism methodology: $e_{abs} = r \times p$, where r is the target relative precision; by default, $r = 0.1$.

40. When a variable is discrete (binary or attribute-based variables), generating a proportion parameter of interest, and the total population size is unknown (e.g. infinitely large), the sample size shall be calculated as follows. In this case the simple random sampling method will need to be adapted to accommodate the unknown population size (e.g. systematic sampling).

$$n \geq \frac{Z_{\alpha/2}^2 \times p \times q}{e_{abs}^2} \qquad \text{Equation (2)}$$

Where:

n	=	Estimated minimum sample size
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval e.g. 1.96 for 95 per cent confidence interval
p	=	Expected proportion of successful cases
q	=	Expected proportion of unsuccessful cases, $q = 1 - p$
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., $0.1 \times p$

41. When a variable is continuous (e.g. electricity generated or measured temperature), generating a mean parameter of interest, and the total population size is known, the sample size shall be calculated as follows:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times SD^2}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD^2} \qquad \text{Equation (3)}$$

Where:

n	=	Estimated minimum sample size
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
N	=	Total population
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., $0.1 \times p$
SD	=	Expected standard deviation. SD^2 = *Expected variance*

42. When a variable is continuous (e.g. electricity generated or measured temperature), generating a mean parameter of interest, and the total population size is unknown (e.g. infinitely large), the sample size shall be calculated as follows. In this case the simple random sampling method will need to be adapted to accommodate the unknown population size (e.g. systematic sampling).

$$n \geq \frac{Z_{\alpha/2}^2 \times SD^2}{e_{abs}^2} \qquad \text{Equation (4)}$$

Where:

n	=	Estimated minimum sample size
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., *0.1 x p*
SD	=	Expected standard deviation. SD^2 = *Expected variance*

5.4.2. Stratified random sampling

43. Stratified random sampling is particularly relevant where the population is heterogeneous, and consists of different sub-populations, referred to as strata, which are more internally homogeneous.

44. In such cases, a simple random sample shall be drawn from each and every stratum, ensuring that the strata divide the units or elements in the population into a set of mutually exclusive and exhaustive sub-populations. For example, the population of participants in a commercial lighting programme is grouped according to building type, e.g., restaurants, food stores, and offices.

45. Stratified random sampling is most applicable where the population is heterogeneous overall, but can be divided into sub-populations that are relatively homogeneous with respect to the parameter of interest. Within each stratum, population elements are expected to be more similar to one another than to elements in other strata, e.g., restaurants may exhibit more similar lighting usage patterns compared to offices or food stores.

46. Stratified random sampling requires that the stratification variable be known for all elements in the sampling frame. For example, to stratify by building type, the sampling frame would need to identify the building type for each element in the population.

47. Stratification can improve the precision of parameter estimates when meaningful differences exist between strata. By accounting for such differences (e.g., lower average lighting use in office buildings compared to that in food stores), stratified sampling enables more representative estimation of population-level parameters. Where elements within each stratum are relatively homogeneous, stratified random sampling generally yields more precise estimates than a simple random sample of the same size drawn from the entire population.

48. For proportional allocation, the sample-size equations use the weighted within-stratum variance. Between-stratum variance is not added because separate stratum estimates are combined using known population weights. If another allocation rule is used, the sample-size and variance equations shall be adapted and justified.

49. For a discrete variable, generating a proportion parameter of interest, the minimum sample size (n)[12] across all strata shall be calculated as follows:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times SD^2}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD^2} \quad \text{Equation (5)}$$

and

$$SD^2 = \frac{g_1 \times p_1 \times q_1 + g_2 \times p_2 \times q_2 + \cdots + g_k \times p_k \times q_k}{N} \quad \text{Equation (6)}$$

$$\bar{p} = \frac{g_1 \times p_1 + g_2 \times p_2 + g_3 \times p_3 + \cdots + g_k \times p_k}{N} \quad \text{Equation (7)}$$

Where:

n	=	Minimum sample size across all strata
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
N	=	Total population size across all strata
SD	=	Expected weighted within-stratum standard deviation used as overall standard deviation. *SD^2 = Expected overall variance*
e_{abs}^2	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., $0.1 \times \bar{p}$
$\bar{p}$	=	Expected overall proportion
g_i	=	Size of stratum *i*
p_i	=	Expected proportion of successful cases in stratum *i*
q_i	=	Expected proportion of unsuccessful cases in stratum *i*. *($q_i = 1 - p$)*
i	=	Index identifying a stratum (where *i* = 1,…,k)

50. The number of samples from each stratum shall be allocated proportionally as follows:

$$n_i \geq \frac{g_i}{N} \times n \quad \text{Equation (8)}$$

Where:

n_i	=	Minimum estimated sample size allocated to stratum *i*
g_i	=	Size of stratum *i*
N	=	Total population size across all strata
n	=	The total sample size required across all strata

[12] The normal approximation may be applied, in accordance with the central limit theorem, provided that the sample size and expected numbers of observations within each stratum are sufficient. Where the population is non-normal or the sample size is limited, the confidence interval may be estimated using bootstrap methods rather than relying on normality assumptions.

i	=	Index identifying a stratum (where *i* = 1,…,k)

51. When the parameters are continuous, generating a mean parameter of interest and the total population size is known, the minimum sample size (*n)* shall be calculated as follows:[13]

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times SD^2}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD^2} \quad \text{Equation (9)}$$

Where:

n	=	Estimated minimum sample size across all strata
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
N	=	Total population across all strata
SD	=	Expected weighted within-stratum standard deviation used as overall standard deviation. *SD^2 = Expected overall variance*
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., 0.1×$\bar{p}$
$\bar{p}$	=	Overall proportion

52. The expected overall variance shall be calculated as follows:

$$(SD)^2 = \frac{g_1 \times (SD_1)^2 + g_2 \times (SD_2)^2 + g_3 \times (SD_3)^2 + \cdots + g_k \times (SD_k)^2}{N} \quad \text{Equation (10)}$$

and

$$Mean = \frac{g_1 \times m_1 + g_2 \times m_2 + g_3 \times m_3 + \cdots + g_k \times m_k}{N} \quad \text{Equation (11)}$$

Where:

SD	=	Expected weighted within-stratum standard deviation used as overall standard deviation. *SD^2 = Expected overall variance*
g_i	=	Size of stratum *i*
SD_i	=	Expected variance of the continuous variable under consideration in stratum *i*
N	=	Total population across all strata
$Mean$	=	Expected overall mean
m_i	=	Expected mean of the continuous variable under consideration in stratum *i*
i	=	Index identifying a stratum (where *i* = 1,…,k)

[13] The sample-size calculation and allocation to each stratum must be adapted accordingly. The core principle remains unchanged: strata that exhibit greater internal variability require proportionally larger samples, while more homogeneous strata can be represented with fewer observations. The equations below follow the same principles as those for proportion parameters above, but use the variance of the continuous parameter rather than binomial variance.

53. The expected number of samples from each stratum shall be allocated proportionally as follows:

$$n_i \geq \frac{g_i}{N} \times n \qquad \text{Equation (12)}$$

where:

n_i	=	Minimum estimated sample size allocated to stratum i
g_i	=	Size of stratum *i*
N	=	Total population size across all strata
n	=	The total sample size required across all strata
i	=	Index identifying a stratum (where *i* = 1,…,k)

5.4.3. Systematic sampling

54. *Systematic sampling* is a statistical method involving the selection of elements from an ordered sampling frame. This method may be used in combination with any other sampling design in this tool, replacing simple random selection as it only affects how a sample is selected. Using systematic sampling in place of simple random sampling does not affect any calculations, but might introduce bias through periodicity, for example.

55. Where systematic sampling is used, activity participants shall apply the form of systematic sampling that is the most appropriate in the context of the Article 6.4 activity and justify the choice. The most common form of systematic sampling is an equal-probability method, in which every k^{th} element in the sampling frame is selected for inclusion in the sample after choosing an initial random starting point within the first *k elements with*:

$$k = N / n \qquad \text{Equation (13)}$$

Where

k	=	Sampling interval
N	=	Total population size
n	=	Sample size to be selected

56. Activity participants shall ensure that the chosen sampling interval does not hide a pattern. Any pattern would threaten randomness. A random starting point *r* must be determined by generating a random integer between 1 and *k*. This value establishes the position of the first selected unit in the sampling frame and, since all subsequent selections follow at fixed intervals of *k*, it defines the entire sample. Systematic sampling shall only be applied where the population is sufficiently homogeneous, as the selected units are distributed uniformly across the population.

57. Systematic sampling is applicable in several situations. If there is a natural ordering or flow of sampling units in the population, such as output of bricks in a manufacturing process, then it is typically easier to sample every k^{th} unit to test for quality control purposes as the units are produced. In all cases, it is important that the list of sampling

units or the process itself is naturally random, in the sense that there is no pattern to its order that could bias the selection process.

58. Systematic sampling functions as an operational sampling technique that selects units at fixed intervals from what is considered to be a randomly ordered sampling frame.

5.4.4. Cluster sampling

59. *Cluster sampling* is a survey design which is useful when a population naturally divides into sub-groups (clusters). In the design, the clusters are randomly selected (sampled), rather than the individual elements to be studied. Data are then collected for all individual elements within the selected clusters, or for a randomly selected subset within each cluster.

60. Cluster sampling is used where hierarchical groupings are evident in a population, such as villages and households within villages, or appliances within buildings. For example, suppose a project installs high-efficiency devices in new apartment buildings, with several units typically in each building. In order to estimate the operating hours of the devices, one might take a sample of the buildings instead of the units, and then meter all the devices in the selected buildings.

61. In contrast to stratified sampling, where the equipment of interest is grouped into a relatively small number of homogeneous strata, cluster sampling involves many clusters (i.e., apartment buildings), and there is no expectation that the devices within each building are more homogeneous than the overall population of efficient devices.

62. Cluster sampling can be useful when no sampling frame is available at the lowest level of the hierarchy but there is one at the cluster level, as in the example in the paragraph 61, where a complete list of all devices may not be available, but a list of all new apartment buildings would be.

63. In many applications to monitor efficient equipment, the units occur naturally in clusters, with a different number of elements per cluster. For example, a building or plant location might constitute a natural cluster, with varying numbers of pieces of equipment per location.

64. A cluster sampling approach can offer cost advantages. For instance, if a significant component of the cost of data collection is travel time between buildings, but there is minimal cost to collect data on units within a building, then it is more cost-effective to collect data on all units within a sample of buildings than to take a simple random sample across all units in the study. It will, however, usually be necessary to evaluate more pieces of equipment (by sampling more clusters) to achieve the same level of precision as simple random sampling, but the reduction in costs may more than offset this increase in sampling effort.

65. Cluster sampling may only be used when the applicable mechanism methodology explicitly allows this method.

5.4.5. One-stage cluster sampling

66. In some settings, it is impractical or impossible to sample individuals directly. When a population naturally divides into clusters, e.g. households within villages, it may be more practical to randomly sample the clusters. In some instances, a sampling frame is available at cluster level but not for the individual units within the clusters making it

impossible to sample individuals directly. In such cases, one-stage cluster sampling which is also known as single-stage cluster sampling, may be used.

67. Rather than sampling individuals, the population is first divided into clusters, such as villages or buildings, and a sample of clusters is then selected. In one-stage cluster sampling, all individuals within the selected clusters are surveyed, eliminating the need for a second stage of sampling. This approach can significantly reduce logistical complexity and costs, particularly in geographically dispersed populations.

68. When a variable is discrete (binary or attribute-based variables), generating a proportion parameter of interest, assuming equal or near equal cluster sizes, the minimum sample size shall be calculated as follows:

$$n_C \geq \frac{Z_{\alpha/2}^2 \times M \times SD_B^2}{(M-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD_B^2} \qquad \text{Equation (14)}$$

and

$$SD_B^2 = \sum_{i=1}^{k} \left((p_i - \bar{p})\right)^2 \Big/ (k-1) \qquad \text{Equation (15)}$$

Where:

n_C	=	Estimated minimum number of clusters to be sampled
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
M	=	Total number of clusters in the population (e.g., villages)
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., $0.1 \times \bar{p}$
SD_B^2	=	Expected variance between cluster proportions from the pilot sample
$\bar{p}$	=	Expected overall proportion
p_i	=	Proportion within cluster *i*
k	=	Number of clusters included in the preliminary (pilot) sample, typically about 5 per cent of the total clusters
i	=	Index identifying a cluster (primary sampling units)

69. The one-stage cluster sampling design requires that all units within sampled clusters be evaluated, i.e. a census is conducted within each sampled cluster.

5.4.6. Multi-stage cluster sampling

70. *Multi-stage sampling* is a more complex form of cluster sampling than one-stage. Measuring all units within the selected clusters may be prohibitively expensive or unnecessary. In multi-stage sampling, the cluster units are often referred to as primary sampling units, and the elements within these clusters as secondary sampling units. In contrast to one-stage cluster sampling, where all secondary sampling units are measured within a selected cluster, multi-stage (or more specifically two-stage) sampling involves collecting data from only a randomly selected subset of the secondary sampling units.

71. Multi-stage sampling may be extended to three or more stages if permitted by the applicable mechanism methodology. For example, the population may be grouped or clustered into building complexes, then buildings, and finally fixtures. This approach may be appropriate where variability exists at all three stages and needs to be reflected in the sample size determination.

72. So far, the methods described above assume simple random sampling at one level or another. However, another approach is sampling with probability proportional to size, under which larger units have a greater probability of selection. This approach is sometimes applied in cluster sampling, where clusters differ in size, and in multi-stage sampling.

73. For multi-stage cluster sampling with two stages, clusters constitute the primary sampling units. Secondary sampling units are randomly selected within each randomly sampled cluster. The required sample size depends on variability between clusters (between-cluster-level variance) and variability among secondary sampling units within clusters (within-cluster variance).

74. When a variable is discrete (binary or attribute-based variables), generating a proportion parameter of interest, the minimum sample size shall be calculated as follows[14]:

$$n_c \geq \frac{SD_B^2 \times \frac{M}{M-1} + \frac{1}{n_s} \times SD_W^2 \times \frac{(\bar{N} - n_s)}{(\bar{N} - 1)}}{\frac{e_{abs}^2}{Z_{\alpha/2}^2} + \frac{1}{M-1} \times SD_B^2}$$ Equation (16)

Where:

n_c	=	Estimated minimum number of clusters to be sampled
M	=	Total number of clusters in the population (e.g., villages)
n_s	=	A stated number of secondary sampling units to be sampled within each cluster, e.g., 10
$\bar{N}$	=	Average number of secondary sampling units per cluster (e.g., households per village)
SD_B^2	=	Expected variance between clusters (cluster-level variance)
SD_W^2	=	Expected value for average of variance within clusters
$\bar{p}$	=	Expected overall proportion
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., $0.1 \times \bar{p}$

[14] Note assumptions are (a) equal cluster sizes, or nearly equal cluster sizes for an approximation and (b) the number of secondary sampling units to be selected with a cluster is specified and identical across clusters.

5.5. Failure to achieve the reliability requirements

75. Activity participants shall assess compliance with the reliability requirements by estimating the sampling error and comparing it with the target precision. If the estimated sampling error exceeds the target precision,[15] activity participants shall:

    (a) Perform additional data collection to supplement the original sample or take a new sample to reach the reliability requirements;

    (b) Undertake any corrective actions specified in the mechanism methodology referring to this methodological tool, such as applying a conservative default value or discount factor;

    (c) Consider the relevance of post-stratification for improving the precision of estimated parameters resulting from the data analysis; or

    (d) Take other actions as specified in the mechanism methodology referring to this methodological tool, such as using different approaches to quantify emission reductions or net removals.

76. Where corrective actions are undertaken, the selected corrective actions shall be justified and documented, including the attained confidence level and precision, the reason the reliability requirements were not achieved, the corrective actions applied, and an explanation of how the selected approach ensures the conservative estimation of emission reductions or net removals.

5.6. Survey and data collection methods

77. Surveys are data collection methods used to obtain information from, or about, sampled units through interaction with respondents or direct observation. Surveys may be conducted physically, including through, on-site visits or face-to-face interviews; or remotely, including through telephone, teleconference, online platforms or SMS. When surveys are required, physical surveys are the generally preferred where feasible and appropriate, as they allow direct observation, clarification of responses and better control of respondent identification and non-response risks.

78. Other data collection and measurement tools may also be used, where appropriate, including hard-copy questionnaires, smartphones, or tablet application modules connected to data platforms, electronic sensors, meters, stove use monitors or other digital measurement tools, web-based platforms, telephone, and mailing (post). Remote self-reported surveys shall be distinguished from digital measurement tools. Self-reported surveys rely on information provided by respondents, while objective digital measurement tools generate measured data directly from the technology, equipment or system being monitored., web-based platforms, telephone, and mailing (post).

79. Activity participants shall select the survey and data collection methods that comply with the applicable mechanism methodology and are appropriate for the parameter being determined and the particular circumstances of the proposed Article 6.4 activity. The selected methods shall be capable of producing reliable, representative and verifiable data and shall be justified in the sampling plan, considering the sampling

[15] The precision of an estimated parameter at the 95 per cent confidence level should be calculated to determine whether the requirements in paragraph 14 are met.

frame, risks of non-response, measurement error and bias, data quality controls, and cost-effectiveness.

5.6.1. Physical survey methods

80. Physical methods involve on-site visits, in person surveys and face-to-face interviews. Physical observation and recording of data present a clear advantage over remote surveys, and on-site visits allow for the purpose of the study and the relevance of the data to be better explained to participants, leading to lower non-response rates. Any contingency measures to address potential barriers to on-site visits shall be considered during the survey planning stage.

81. Data can be collected through hard-copy questionnaires or electronic means, which may be particularly effective where a large number of samples are involved. Advanced data collection systems employ multiple mobile digital methods that automatically upload captured data to a central data server for storage and may also support data sorting, management, processing and reporting.

82. Data collected may be linked to a geographical location through a global positioning system (GPS) module, enabling the identification of each participant. Photographs may also be used to document specific circumstances relating to the participants.

5.6.2. Remote survey methods

83. Remote survey methods involve data collection without physical on-site interaction with sampled units. They may include: (a) e-mail, web-based platforms, or SMS; (b) telephone interviews; and (c) mail (post), as elaborated below:

    (a) E-mail or web-based platform or SMS: the use of information technology (IT) tools for data collection requires access to a computer or portable device, and a reliable internet or wireless data connection;

    (b) Telephone interview: this method requires a reliable telephone (mobile or landline) network to enable the interview. In addition, bias may arise where the statistical design of the survey relies on the random sampling of telephone numbers, as households without access to a telephone would be excluded from the survey. Telephone interviews may not be suitable for lengthy interviews, complex questions, or the collection of technical data that may require specialised training; and

    (c) Mailing (post) (questionnaires sent by regular mail): respondents complete a questionnaire received by regular mail. This method assumes a certain level of understanding of the issues under study and requires clear instructions and a simple format. However, this method does not allow for further probing or clarification of responses. Furthermore, it may not be possible to determine whether an appropriate person completed the questionnaire. Response rates may also be low, which can result in bias and reduced precision.

84. Remote survey methods may only be used where they are permitted explicitly by the applicable mechanism methodology and are appropriate for the parameter being measured. Activity participants shall justify the use of remote survey methods in the sampling plan and demonstrate that the method can produce reliable, representative and verifiable data, considering the coverage of the target population, respondent identification, non-response, measurement error, data quality controls, feasibility and cost-effectiveness. The applicable mechanism methodology may restrict the use of

remote survey methods, including for parameters that materially affect the calculation of emission reductions or net GHG removals.

5.6.3. Remote data collection methods

85. Remote data collection methods may involve data sensors and digital monitoring technologies, including for example, stove use monitoring systems (SUMs), internet of things (IoT) enabled devices, and other data logging systems.

86. The remote data collection devices shall be connected to appropriate communication technologies to transmit data securely. Where remote transmission is not available, complementary methods such as physical download and transfer shall be used.

87. Where digital measurement tools are used, their correct installation, calibration, functioning shall be ensured, together with data completeness and access for verification. Any non-functioning device installed at a sampled unit shall be promptly replaced with a working one to ensure the sample remains representative.

88. Data collected using remote data collection methods may be complemented by data obtained through surveys, where necessary, to confirm contextual factors not captured by the sensors. For example, supporting surveys may be used to collect complementary information not captured by the digital measurement tools, such as household characteristics, technology status, user identification or other contextual factors.

Appendix 1. Additional guidance on sampling methods and survey design

1. Table 1 below provides a summary of the advantages and disadvantages of the different sampling methods.

Table 1. Advantages and disadvantages of different sampling methods

Sampling Method	Advantages	Disadvantages
Simple Random Sampling: taking a random sample from the whole population	• Easiest method to understand and therefore use. • Suitable if there is little heterogeneity amongst the units being sampled	• Requires knowledge of the entire population before a sample can be selected. • If the population covers a large geographical area, then it can often lead to sampling units that are spread out over the area. Such a situation can often be costly to sample. • Only suitable where the population being studied is relatively homogeneous with respect to the parameter of interest
Systematic Sampling: taking a sample every *n* units	• Easy to apply. • Commonly used because it ensures that there is sufficient distance between sampled units	• May result in sample units being distributed across a large geographic area, which can increase data collection costs
Stratified Random Sampling: randomly sampling a different number of units from each stratum according to the weight of each stratum in the population	• Improves the precision of estimates (compared to simple random sampling) where meaningful differences exist between the strata	• More complicated • The appropriate stratification factors may not always be obvious

Sampling Method	Advantages	Disadvantages
One-stage Cluster Sampling: sampling every unit in a sample of *n* clusters from the population	• Often the most economical form of sampling because sampling units grouped according to a common characteristic (often geographical). • In some cases, it may be the only practical approach because a complete list of all households may not be available, whereas a list of villages may be available. Once the villages are selected, households can be surveyed	• Estimates are generally less precise than those obtained through simple random sampling because characteristics within clusters may be relatively homogeneous. • Larger sample sizes may be required to compensate for this effect
Multi-stage Sampling: Randomly sampling a number of units within a number of randomly selected clusters	• Enables sampling at multiple levels of the population hierarchy. • Allows different combinations of clusters and sampling units to be evaluated in order to identify the most cost-effective and reliable sampling design	• Analysis and sample size calculations are more complex than other sampling methods

.

2. Table 2 below provides an overview of survey and data collection methods

Table 2. Survey and data collection methods and preference for use

Type	Method or technology	Order of Preference	Comments or key considerations	Anticipated response rate
Survey method	Physical on-site: hard-copy questionnaires; or smartphone or tablet app modules	1st	• Higher-quality data; • Physical inspection by trained personnel; • Longer and more complex questions can be administered; and • Additional contextual data may be collected. However, for simple parameters such as survival rates (retention rates), well-designed remote monitoring methods, such as data loggers, sensors, or pay-as-you-go systems, may be equally effective or may perform better than physical on-site visits, particularly where monitoring is conducted frequently	High
	Remote: telephone interview	2nd	• Access to a telephone necessary, which may introduce sampling bias; • Reduced response rates; and • Not suitable for lengthy or complex questions. Identity of the respondent may not be verifiable	Medium
	Remote: e-mail or web-based platform or SMS	3rd	• Access to a computer or other portable device required, which may introduce a sampling bias; • Response rates may be low because e-mails may be ignored or filtered as spam; • Different households may use different IT systems, which may affect the display or usability of the questionnaire; • Questionnaire fatigue may require fewer, well-structured questions that are simple to answer; and	Low

Type	Method or technology	Order of Preference	Comments or key considerations	Anticipated response rate
			• Further probing or clarification of responses is not possible. It may be difficult to ascertain who completed the questionnaire	
	Remote: Mailing (post) (questionnaires sent by regular mail)	3rd	• Respondents should be able to read, understand and complete the questionnaire; • Questionnaire fatigue may require fewer, well-structured questions that are simple to answer; • Further probing or clarification of responses is not possible; and • It may be difficult to ascertain who completed the questionnaire. Response rates are typically low	Low
Monitoring Technology	Data sensors and other data logging systems (general)	1st	• Accurate and objective data collection; and • Appropriate application conditions should be ensured. May involve significant costs unless incorporated into the activity design (e.g., pay-as-you-go systems)	High
	Stove Use Monitoring Systems (SUMs)	1st	• Provides objective, high-frequency usage data; • Reduces reliance on self-reported data; • Particularly suitable for monitoring usage patterns (e.g., cookstove use); and • Requires installation, calibration, and data management. Should be integrated into project design	High

Appendix 2. Best-practice examples for a single sampling plan for a homogeneous programme of activities

1. Definition of criteria to establish the homogeneity of the CPs included for the single sampling plan

1. This appendix provides illustrative guidance to support the application of a single sampling plan across multiple component projects (CPs) within an Article 6.4 programme of activities (Article 6.4 PoA), in accordance with the provisions set out in section 5 of this methodological tool.

2. A single sampling plan may be applied where:

   (a) CPs are sufficiently homogeneous with respect to the parameter(s) of interest; or

   (b) Heterogeneity among CPs is appropriately addressed through the sampling design and analysis.

3. Where heterogeneity exists among CPs, activity participants may apply a stratified sampling approach, as described in section 5.4.2 of this methodological tool. Under this approach, CPs or their populations are grouped into strata based on relevant characteristics, and estimates are derived using stratified analysis within a single sampling plan. Stratification can improve the precision of parameter estimates where meaningful differences exist between strata, particularly where strata are internally homogeneous but differ from one another.

4. Activity participants may use one or more of the following non-exhaustive criteria to demonstrate homogeneity and justify the application of a single sampling plan:

   (a) **Project technology or equipment characteristics:** CPs use technologies or equipment with comparable input and output characteristics, including efficiency, performance and service levels. For example, compact fluorescent lamps (CFLs) operating on grid electricity may be considered comparable where there are no significant differences between the electricity grids within the boundary. However, for certain parameters, such as retention rates, results from stable grids may differ from those obtained in weaker grids that experience voltage or frequency fluctuations;

   (b) **End-user characteristics:** end users of the project technology or equipment have comparable socioeconomic characteristics (e.g., middle-income households);

   (c) **Geographic and environmental conditions**: CPs are located in areas with similar geographic or climatic conditions that do not significantly influence the parameter(s) of interest. For example, biogas digesters installed in cold climates may have different output rates from those installed in warmer climates, in which case stratification by geographical area may be appropriate; and

   (d) **Installation timeframes:** CPs are implemented within comparable timeframes such that temporal variations do not significantly affect the parameter(s) of interest.

2. Description of an example Article 6.4 PoA "Improved cookstoves in country X"

5. An example Article 6.4 PoA involves the dissemination of improved cookstoves to households in country X. Three CPs have already been included in the PoA: CP 1, "Improved cookstoves in province A"; CP 2, "Improved cookstoves in province B"; and CP 3, "Improved cookstoves in province C". The PoA applies the relevant Article 6.4 mechanism methodology for energy efficiency measures in thermal applications of non-renewable biomass.

6. The Article 6.4 PoA aims to reduce greenhouse gas (GHG) emissions by replacing traditional three-stone stoves with improved cookstoves. Improved cookstoves increase the efficiency of heat transfer to the cooking utensil, thereby reducing the amount of non-renewable biomass (fuel) required for cooking. As a result, GHG emission reductions are achieved through reduced consumption of non-renewable biomass resulting from the higher thermal efficiency of the improved cookstoves compared to traditional three-stone stoves.

7. Improved cookstoves were distributed under three CPs. Relevant dates and monitoring periods for each CP and for the Article 6.4 PoA as a whole are presented in table 1 below. Each installed improved cookstove is assigned a unique serial number to facilitate identification and prevent double counting.

Table 1. Relevant dates and monitoring periods for each CP

CP	Number of improved cookstoves	Date of first improved cookstove installation under the CP	Date of last improved cookstove installation under the CP	Monitoring period
CP 1	30 000	1 April 2026	31 May 2026	1 June 2026 until 31 December 2026
CP 2	10 000	1 April 2026	30 June 2026	1 July 2026 until 31 December 2026
CP 3	20 000	1 April 2026	31 July 2026	1 August 2026 until 31 December 2026
PoA Total	60 000	1 April 2026	31 July 2026	1 June 2026 until 31 December 2026

8. To calculate emission reductions achieved during the monitoring period for each CP shown in table 1, the following parameters were determined through ex post sampling surveys:

   (a) Percentage of improved cookstoves that remain in use (hereinafter referred to as the "retention rate of improved cookstoves");

   (b) Percentage of displaced traditional cookstoves that continue to be used (hereinafter referred to as the "continued-use rate of displaced traditional cookstove"); and

   (c) Operating efficiency of distributed improved cookstoves (η_{new}).

9. As described in section 5 of this methodological tool, a single sampling plan covering a group of CPs may be applied using a 95/10 confidence/precision criterion for sample size determination. Under this approach, the populations of all CPs in the group are combined into a single sampling frame, a common sample size is determined, and a single survey is undertaken. The data analysis approach depends on whether the included CPs are considered homogeneous or heterogeneous with respect to the parameter(s) of interest. Where the CPs are considered homogeneous, a simple random sampling plan may be applied. Where heterogeneity exists, a stratified sampling plan may be applied.

10. In the following example, the three CPs are assumed to be homogeneous with respect to the three parameters of interest at the time of the survey as per the requirements in para 18 of the tool, for the following reasons:

    (a) Only one type of improved cookstoves is distributed;

    (b) End users of the improved cookstoves are domestic households; that is, the Article 6.4 PoA does not target commercial users; and

    (c) The CPs were implemented within a three-month period, such that temporal differences are not expected to significantly affect the parameter(s) of interest.

11. Since the three parameters of interest are assumed to be comparable across the three CPs at the time of the survey, a single survey is undertaken using a simple random sampling plan. The populations of all three CPs are combined into a single sampling frame (i.e., a total population of 60,000 improved cookstoves), and the sample size is then determined using a simple random sampling approach, as illustrated below.

3. Sample size calculations using a simple random sampling

12. According to section 5.3 of this methodological tool, where more than one parameter is to be estimated through sampling, a separate sample size calculation shall be undertaken for each parameter. The largest resulting sample size may then be selected for a single survey and common sampling effort, or separate sampling efforts and surveys may be undertaken for each parameter.

13. The selection of the appropriate equation for sample size determination depends on the type of parameter of interest. Parameters may be classified as:

    (a) A proportion (or percentage) – in this example, the retention rate of improved cookstoves and the continued-use rate of displaced traditional cookstoves; or

    (b) A mean value – in this example, the mean operating efficiency of improved cookstoves.

14. For all examples, a confidence level of 95 per cent and a target precision of ±10 per cent are applied, consistent with the reliability requirements for a single sampling plan.

3.1. Retention rate of improved cookstoves (proportion parameter)

15. The required sample size for this parameter depends on:

    (a) The expected value of the proportion parameter; and

(b) The required confidence level and target precision (95/10 confidence/precision criterion for a single sampling plan).

16. Based on similar studies undertaken in the region, the expected value of this proportion during the monitoring period (table 2) is assumed to be 0.85 (85 per cent). Since the CPs are assumed to be homogeneous with respect to the retention rate of improved cookstoves, a simple random sampling plan is applied to estimate the proportion parameter of interest.

Table 2. Estimation of the proportion parameter of interest

CP	Number of improved cookstoves	Expected proportion
CP 1	30 000	
CP 2	10 000	
CP 3	20 000	
Total	**60 000**	**0.85**

17. The equation to give us the required sample size is:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times p \times q}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times p \times q} \qquad \text{Equation (1)}$$

Where:

n	=	Estimated minimum sample size
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
N	=	Total population size
p	=	Expected proportion of successful cases
q	=	Expected proportion of unsuccessful cases (*q = 1 – p*)
e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., 0.1 x p

18. Substituting the values into the equation gives:

$$n \geq \frac{1.96^2 \times 60{,}000 \times 0.85 \times 0.15}{(60{,}000-1) \times (0.1 \times 0.85)^2 + 1.96^2 \times 0.85 \times 0.15} = 67.7 \qquad \text{Equation (2)}$$

19. Therefore, the minimum required sample size is 68.

20. Note that the sample size of 68 means that data are required from 68 cookstoves for analysis. If the expected the response rate is only 70 per cent, the initial sample size should be increased accordingly. In this example, the adjusted sample size would be 68/0.70 = 98 cookstoves.

3.2. Continued-use rate of displaced traditional cookstoves (proportion parameter)

21. The required sample size for this parameter depends on:

    (a) The expected value of the proportion parameter; and

    (b) The required confidence level and target precision (95/10 confidence/precision criterion for a single sampling plan).

22. Based on similar studies undertaken in the region, the expected value of this proportion is assumed to be 0.10 (10 per cent) during the monitoring period (table 3). Since the CPs are assumed to be homogeneous with respect to the continued-use rate of displaced traditional cookstoves, a simple random sampling plan is applied to estimate the proportion parameter of interest.

Table 3. Estimation of the proportion parameter of interest

CP	Number of improved cookstoves	Expected proportion
CP 1	30 000	
CP 2	10 000	
CP 3	20 000	
Total	**60 000**	**0.10**

23. According to this methodological tool, a proportion may be expressed either as the proportion of units exhibiting a characteristic or as the complementary proportion of units not exhibiting the characteristic. For example, the parameter may represent either: (i) the proportion of cookstoves that remain operational; or (ii) the proportion of cookstoves that are no longer operational. For sample size determination, activity participants may use the larger of the two proportions, namely p or $(1 - p)$.

24. In this example, a continued-use rate of 10 per cent is equivalent to a discontinued-use rate of 90 per cent. The sample size calculation is therefore based on an anticipated discontinued-use rate of 90 per cent, as follows:

$$n \geq \frac{1.96^2 \times 60{,}000 \times 0.9 \times 0.1}{(60{,}000 - 1) \times (0.1 \times 0.9)^2 + 1.96^2 \times 0.9 \times 0.1} = 42.7 \qquad \text{Equation (3)}$$

25. Therefore, the minimum required sample size is 43. Note that a sample size of 43 means that data are required from 43 cookstoves for analysis. If the expected response rate is only 70 per cent, the initial sample size shall be increased accordingly. In this example, the adjusted sample size would be 43 / 0.70 = 62 cookstoves.

3.3. Operating efficiency of distributed improved cookstoves (mean value parameter)

26. The required sample size for this parameter depends on:

    (a) The expected mean;

    (b) The expected standard deviation; and

(c) The required confidence level and target precision (95/10 confidence/precision criterion for single sampling plan).

27. Based on similar studies undertaken in the region, the expected mean is assumed to be 0.20 during the monitoring period (table 4) and the expected standard deviation is assumed to be 0.05. Since the CPs are assumed to be homogeneous with respect to the operating efficiency of improved cookstoves, a simple random sampling plan is applied to estimate the mean parameter of interest.

Table 4. Estimation of the mean parameter of interest

CP	Number of improved cookstoves	Expected mean	Expected standard deviation
CP 1	30 000		
CP 2	10 000		
CP 3	20 000		
Total	**60 000**	**0.20**	**0.05**

28. The equation to give us the required sample size is:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times SD^2}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD^2} \quad \text{Equation (4)}$$

Where:

n	=	Estimated minimum sample size
N	=	Total number of improved cookstoves in the population (60,000 in this example)
$Mean$	=	Expected mean
SD	=	Expected standard deviation
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
e_{abs}	=	Relative precision. By default 10 per cent

29. Substituting the values into equation (4) gives:

$$n \geq \frac{1.96^2 \times 60{,}000 \times 0.05^2}{(60{,}000-1) \times (0.1 \times 0.20)^2 + 1.96^2 \times 0.0025} = 24.0 \quad \text{Equation (5)}$$

30. Therefore, the minimum required sample size is 24 cookstoves.

31. If the expected response rate is only 70 per cent, the initial sample size should be increased accordingly. In this example, the adjusted sample size would be 24 / 0.70 = 34.29, which is rounded up to 35 cookstoves.

32. Oversampling (sampling in excess of the estimated minimum sample size) is strongly encouraged at the sampling-design stage, not only to compensate for potential attrition, outliers or non-response, but also to reduce the risk that the required reliability is not achieved during data analysis, thereby necessitating additional sampling efforts. As

described in section 5.3 of this methodological tool, where more than one parameter is determined through sampling, the required sample size shall be based on the largest sample size obtained from the relevant calculations. In this example, the largest sample size is 98. Therefore, a single sample of 98 improved cookstoves is selected for the estimation of all three parameters of interest (table 5).

33. Table 5 summarizes the response-rate-adjusted sample sizes determined for each parameter. In accordance with section 5.3 of this methodological tool, the largest sample size is selected where a single sampling plan is used to estimate multiple parameters. Hence, in table 5 below, 98 becomes the common sample size since it is the largest.

Table 5. Sample size determination for multiple parameters

Parameter	Required sample size adjusted for the expected response rate
Retention rate of improved cookstoves	98
Continued-use rate of displaced traditional cookstoves	62
Operating efficiency	35

4. Sampling results and reliability calculations

34. The examples presented below illustrate how to estimate a mean parameter and a proportion parameter, and how to assess compliance with the reliability requirements. The sampling method applied is simple random sampling.

35. An adjusted sample of 98 units was selected as per paragraph 20, to cover the combined population of CP 1, CP 2 and CP 3. The required reliability criterion is 95/10, corresponding to a 95 per cent confidence level and a precision of ±10 per cent.

4.1. Retention rate of improved cookstoves (proportion parameter)

36. The parameter of interest is the proportion of improved cookstoves that remain operational in the population, estimated using data collected from the sample. The population consists of the 60,000 improved cookstoves distributed across the three CPs. A simple random sample of 98 improved cookstoves was selected, and the operational status of each cookstove was recorded. A random number generator was used to select the sample.

37. In this example, 92 of the 98 sampled cookstoves were found to be operational during the monitoring period. The sample proportion is therefore p = 92 / 98 = 0.9388. Rounded to three decimal places, the estimated proportion is 0.939. In other words, 93.9 per cent of the sampled cookstoves were operational.

38. To assess compliance with the reliability requirements, the results are expressed as a confidence interval. In this example, the 95 per cent confidence interval for the population proportion is calculated using equation 6 below.

$$p \pm Z_{\alpha/2} \times \sqrt{1 - \frac{n}{N}} \times \sqrt{\frac{p \times q}{n}} \qquad \text{Equation (6)}$$

Where:

n = Sample size

N	=	Total population
p	=	Sample proportion
q	=	Expected proportion of unsuccessful cases (*1 - p*)
$Z_{\alpha/2}$	=	Percentile of the standard normal distribution corresponding to a (1 − *α* / 2) x 100 per cent confidence level (e.g., 1.96 for 95 per cent confidence level)

39. Substituting the relevant values into equation (6) gives:

$$0.939 \pm 1.96 \times \sqrt{1 - \frac{98}{60{,}000}} \times \sqrt{\frac{0.939(1 - 0.939)}{98}} = 0.939 \pm 0.047$$ Equation (7)

40. This results in a 95 percent confidence interval of (0.892 to 0.986) for the population proportion, i.e., 89.2 percent to 98.6 percent.

41. The 95 per cent confidence interval indicates that the true population proportion of improved cookstoves that remain operational during the monitoring period is estimated to lie between 89.2 per cent and 98.6 per cent.

42. Half the width of the confidence interval (also referred to as the margin of error) may be used to assess whether the required reliability criterion has been achieved. Reliability is achieved where the margin of error does not exceed 10 per cent of the estimated proportion in relative terms.

43. Reliability may therefore be assessed using the following calculation:

$$\frac{\frac{1}{2}\ width\ of\ the\ confidence\ interval}{estimated\ proportion}$$ Equation (8)

44. In this example, the estimated proportion of improved cookstoves that remain operational is 0.939 and the 95 per cent confidence interval is 0.892 to 0.986. The achieved reliability is therefore:

$$\frac{\frac{1}{2}\ (0.986 - 0.892)}{0.939} = 0.05$$ Equation (9)

45. The achieved precision is 5 per cent of the estimated proportion. Therefore, the estimated proportion of 93.9 per cent operational cookstoves meets the required reliability criterion of 95/10.

4.2. Continued-use rate of displaced traditional cookstoves (proportion parameter)

46. The parameter of interest is the proportion (or percentage) of displaced traditional cookstoves that are no longer used in the population, estimated using the sample proportion. The population consists of 60,000 cookstoves. A simple random sample of 98 cookstoves was selected, and for each sampled unit it was recorded whether the displaced

traditional cookstove continued to be used. The same sample was used for the estimation of the previous parameter.

47. In this example, 88 of the 98 sampled units indicated that the displaced traditional cookstove was no longer used. The sample proportion is therefore p = 88/98 = 0.898. In other words, 89.8 per cent of the displaced traditional cookstoves were no longer used.

48. To assess compliance with the reliability requirements, the results are expressed as a confidence interval. In this example, the 95 per cent confidence interval for the population proportion is calculated using equation 6 above.

49. Substituting the values into equation (6) gives:

$$0.898 \pm 1.96 \times \sqrt{1 - \frac{98}{60{,}000}} \times \sqrt{\frac{0.898(1 - 0.898)}{98}} = 0.898 \pm 0.060$$ Equation (10)

50. This results in a confidence interval of (0.838 to 0.958) for the population proportion, i.e., from 83.8 per cent to 95.8 per cent).

51. The 95 per cent confidence interval indicates that the true population proportion of displaced traditional cookstoves that are no longer used during the monitoring period is estimated to lie between 83.8 per cent to 95.8 per cent.

52. As in the previous example, reliability may be assessed as follows:

$$\frac{\frac{1}{2}\,(0.958 - 0.838)}{0.898} = 0.067$$ Equation (11)

53. The achieved precision is 6.7 per cent of the estimated proportion. Therefore, the estimated proportion of 89.8 per cent of displaced traditional cookstoves no longer being used meets the required reliability requirements of 95/10.

4.3. Operating efficiency of distributed improved cookstoves (mean value parameter)

54. The parameter of interest is the mean operating efficiency of the distributed improved cookstoves in the population, estimated using the sample mean. The population consists of the 60 000 improved cookstoves distributed across the three CPs. A simple random sample of 98 improved cookstoves was selected, and the operating efficiency of each sampled cookstove was measured. The same sample of 98 cookstoves was used to estimate the two previous parameters.

55. In this example, the sample mean operating efficiency of the 98 cookstoves is 0.1971 and the sample standard deviation is 0.06.

56. To assess compliance with the reliability requirements, the results are expressed as a confidence interval. In this example, the 95 per cent confidence interval is 0.185 to 0.209.

$$Sample\ Mean \pm t_{\alpha/2,(n-1)} \times \sqrt{1-\frac{n}{N}} \times \frac{SD}{\sqrt{n}}$$ Equation (12)

Where:

n	=	Sample size
N	=	Total population
$Mean$	=	Sample mean
SD	=	Sample standard deviation
$t_{\alpha/2,(n-1)}$	=	Critical value of Student's t-distribution with *(n – 1)* degrees of freedom for which the cumulative probability equals *1 – α / 2* (e.g., $t_{0.025,\ 97}$ = 1.985 for a 95 per cent confidence level).

57. Substituting our values into equation (13) gives:

$$0.1971 \pm 1.985 \times \sqrt{1-\frac{98}{60{,}000}} \times \frac{0.06}{\sqrt{98}} = 0.1971 \pm 0.012$$ Equation (13)

$$= (0.185, 0.209)$$

58. This results in a confidence interval of (0.185, 0.209) for the population mean. The 95 per cent confidence interval indicates that the true population mean operating efficiency of the improved cookstoves during the monitoring period is estimated to lie between 0.185 and 0/209.

59. A symmetric confidence interval for the true population parameter may also be expressed as the sample mean ± the margin of error. The estimate of 0.1971 is considered reliable where the margin of error is within the required precision (i.e., 10 per cent of the sample mean).

60. Reliability may therefore be assessed using the following calculation:

$$\frac{\frac{1}{2}\ width\ of\ the\ confidence\ interval}{mean}$$ Equation (14)

61. In this example, the mean operating efficiency is 0.1971 and the 95 per cent confidence interval is 0.185 to 0.209. The achieved reliability is therefore:

$$\frac{\frac{1}{2}\ (0.209 - 0.185)}{0.1971} = 0.061$$ Equation (15)

62. The achieved precision of 6.1 per cent of the sample mean. Therefore, the estimated mean operating efficiency of 0.1971 meets the required reliability criterion of 95/10.

5. Conclusions

63. The sample estimates for all three parameters meet the reliability requirements of 95/10.

Appendix 3. Best-practice examples for a single sampling plan for a heterogeneous programme of activities using a stratified sampling approach

1. Introduction

1. Where improved cookstoves are deployed across multiple provinces under different CPs, the parameter(s) of interest may not be homogeneous across the CPs. Levels of technology adoption may differ significantly between communities due to factors such as demographic characteristics, cultural practices, or differences in programme implementation. For example, in the illustrative example in appendix 2, three times as many cookstoves were distributed in CP 1 as in CP 2, which could reasonably result in differing adoption patterns.

2. Where homogeneity across CPs cannot be demonstrated, a stratified sampling approach may be applied to account for differences among CPs and improve the precision of parameter estimates. As in the simple random sampling example, a single survey may be used to collect data for the three parameters of interest: retention rate of improved cookstoves, continued-use rate of displaced traditional cookstoves, and operating efficiency of improved cookstoves across all CPs. The key distinction is that, under a stratified sampling design, the allocation of samples to each CP is determined in advance as part of the sampling design. The required sample size for each CP and parameter of interest is illustrated in section 2 below.

2. Sample size calculations for multiple parameters

3. According to section 5.3 of this methodological tool, where more than one parameter is to be estimated through sampling, a separate sample size calculation shall be undertaken for each parameter. Where a single sampling plan is applied, the required sample size shall be based on the most conservative requirement, namely the largest sample size required among the parameters. Alternatively, separate sampling efforts may be undertaken for each parameter, where this is preferred.

4. Different equations are used for sample size determination depending on the type of parameter of interest. Parameters may be classified as:

   (a) A proportion (or percentage), such as the retention rate of improved cookstoves and the continued-use rate of displaced traditional cookstoves; or

   (b) A mean value, such as the average operating efficiency of improved cookstoves.

5. For all examples, the sampling design is based on the reliability requirements specified in this methodological tool, namely a confidence level of 95 per cent and a precision of ±10 per cent. Retention rate of improved cookstoves (proportion parameter).

6. The required sample size for this parameter depends on:

   (a) The expected value of the proportion; and

   (b) The required confidence level and target precision, (95/10 confidence/precision criterion for a single sampling plan).

7. To illustrate the application of stratified random sampling, the estimated retention proportions for each CP, based on historical records, have been selected such that the overall retention proportion is approximately 0.85, while allowing for heterogeneity among CPs (table 1).

Table 1. Illustration of stratified random sampling

CP	Number of improved cookstoves	Expected retention proportion
CP 1	30 000	0.84
CP 2	10 000	0.73
CP 3	20 000	0.91
Total	**60 000**	**0.845**[1]

8. The formula for the total sample size is:

$$n \geq \frac{1.96^2 \times N \times SD^2}{(N-1) \times (0.1 \times 0.845)^2 + 1.96^2 \times SD^2} \quad \text{Equation (1)}$$

Where:

SD^2	=	Expected variance
n	=	Sample size
N	=	Total population size

9. The values in table 1 above, are used to calculate the overall variance[2] and weighted overall proportion of improved cookstoves that remain in use, using the equations (2) and (3) below, respectively:

$$SD^2 = \frac{(g_1 \times p_1 \times q_1) + (g_2 \times p_2 \times q_2) + (g_3 \times p_3 \times q_3)}{N} \quad \text{Equation (2)}$$

$$\bar{p} = \frac{(g_1 \times p_1) + (g_2 \times p_2) + (g_3 \times p_3)}{N} \quad \text{Equation (3)}$$

Where:

g_i	=	Size of stratum *i* (i.e., CP_i) where *i* covers the *k* strata of the population
p_i	=	Proportion of success in stratum *i* (i.e., CP_i)
$\bar{p}$	=	Overall sample proportion
q_i	=	Expected proportion of unsuccessful cases in stratum *i* (i.e., CP_i)
N	=	Total population

[1] The overall retention proportion is derived from the weighted average of the CP-level values, based on the number of improved cookstoves in each CP.

[2] The variance of a proportion is calculated as: *p x (1 – p).*

SD	=	Expected standard deviation (SD^2 = expected variance)
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
i	=	Index identifying a stratum (where *i* = 1,…,*k*), i.e., *k* = 3

10. Substituting our values into equations (4) and (5) gives:

$$SD^2 = \frac{(30{,}000 \times 0.84 \times 0.16) + (10{,}000 \times 0.73 \times 0.27) + (20{,}000 \times 0.91 \times 0.09)}{60{,}000} = 0.12735 \quad \text{Equation (4)}$$

$$\bar{p} = \frac{(30{,}000 \times 0.84) + (10{,}000 \times 0.73) + (20{,}000 \times 0.91)}{60{,}000} = 0.845 \quad \text{Equation (5)}$$

11. Substituting the results from equations 2 and 3 into equation 6 gives:

$$n \geq \frac{1.96^2 \times 60{,}000 \times 0.12735}{(60{,}000 - 1) \times 0.0845^2 \times 1.96^2 \times 0.12735} = 68.44 \quad \text{Equation (6)}$$

12. Therefore, the required sample size across all three CPs is 69.

13. The total sample size may then be allocated across the CPs using the following equation:

$$n_i = \frac{g_i}{N} \times n \quad \text{Equation (7)}$$

Where:

n_i	=	Sample size allocated to stratum *i* (i.e., CP_i)
g_i	=	Size of the stratum *i* (i.e., CP_i)
N	=	Total population size
n	=	Total sample size

14. Substituting the CP population sizes, the total population size, and the total sample size into equation (8) gives the following results presented in table 2.

Table 2. Required sample size allocation across CPs

CP	Number of improved cookstoves	Equation	Sample size allocated
CP 1	30 000	$n_{CPA_1} = \frac{30\ 000}{60\ 000} \times 69 = 34.50$	35
CP 2	10 000	$n_{CPA_2} = \frac{10\ 000}{60\ 000} \times 69 = 11.50$	12

CP	Number of improved cookstoves	Equation	Sample size allocated
CP 3	20 000	$n_{CPA_3} = \frac{20\,000}{60\,000} \times 69 = 23.00$	23
Total	**60 000**		**70**

15. **Note:** The allocated sample sizes are rounded to whole numbers. As a result, the total allocated sample size is 70 rather than the initial estimate of 69 across all CPs.[3]

16. Note that a sample size of 70 requires valid data from 70 cookstoves. If the expected response rate is 70 per cent, the sample size should be adjusted accordingly within each CP by dividing the required sample size by the expected response rate. This approach also allows for different response rates to be applied to different CPs, where justified. The sample sizes adjusted for an expected response rate of 70 per cent in each CP are presented in table 3.

Table 3. Required sample sizes adjusted for an expected response rate of 70 per cent in each CP

CP	Sample size required (not accounting for non-response)	Equation	Sample size required, adjusted for non-response[4]
CP 1	35	$\frac{35}{0.7} = 50.00$	50
CP2	12	$\frac{12}{0.7} = 17.14$	18
CP 3	23	$\frac{23}{0.7} = 32.86$	33
Total	**70**		**101**

17. After adjusting for non-response within each CP, the total sample size required to achieve the 95/10 confidence/ precision criterion for the retention-rate estimate is 101 cookstoves.

[3] The calculated sample size of 69 represents minimum theoretical sample required to achieve the 95/10 reliability target. However, when this sample is allocated proportionally across strata, the resulting stratum-specific sample sizes are not always whole numbers (e.g. 34.5 and 11.5). Since only whole units can be sampled, the allocations must be rounded. Rounding down would reduce the sample size in one or more strata below the proportional allocation required by the design. Therefore, the affected strata are rounded up, resulting in a total sample size of 70. This is the smallest practical sample size that preserves the proportional allocation and ensures that the 95/10 reliability requirement is achieved.

[4] Non-response adjustment should ideally be applied to the proportional allocations before final rounding. This would result in sample sizes of 50, 17 and 33 (total 100). An approach that applies the adjustment after rounding the proportional allocations, as in table 3, may result in a total sample size greater than the initially determined total across all CPs e.g., >100 in table 3. This slightly increases the required sample size, but it is a more conservative approach would still satisfy the minimum 95/10 reliability requirement.

2.1. Continued-use rate of displaced traditional cookstoves (proportion parameter)

18. The required sample size for this parameter depends on:

    (a) The expected value of the proportion; and

    (b) The required confidence level and target precision (95/10 confidence/precision criterion for a single sampling plan).

19. Based on studies conducted in the region, the overall continued-use proportion is assumed to be 0.10 (10 per cent) during the monitoring period (table 4). To illustrate the application of stratified random sampling, the expected continued-use proportions for each CP have been selected such that the overall continued-use proportion is approximately 0.10, while allowing for heterogeneity among CPs.

Table 4. Illustration of stratified random sampling for the continued-use rate of displaced traditional cookstoves

CP	Number of improved cookstoves	Expected continued-use proportions
CP 1	30 000	0.11
CP 2	10 000	0.18
CP 3	20 000	0.06
Total	**60 000**	**0.105**[5]

20. In this example, an overall continued-use proportion of 0.10 is equivalent to a an overall discontinued-use proportion of 0.90. As described in this methodological tool, project proponents may use the larger of the two complementary proportions in the sample-size calculation. Therefore, the sample-size calculations below are based on the expected discontinued-use proportions for each CP (table 5).

Table 5. Expected discontinued-use proportions by CP

CP	Number of improved cookstoves	Expected discontinued-use proportions
CP 1	30 000	0.89
CP 2	10 000	0.82
CP 3	20 000	0.94
Total	**60 000**	**0.895**

21. The equation used to determine the total sample size across all three CPs is:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times SD^2}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD^2} \qquad \text{Equation (8)}$$

Where:

n	=	Estimated sample size across all strata
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval

[5] The overall continued use proportion is obtained using population-weighted averages based on the number of improved cookstoves in each CP.

e_{abs}	=	Target precision expressed as the maximum allowable absolute sampling error, i.e., 0.1 x p
SD	=	Expected standard deviation (SD^2 = expected variance)

22. The values in table 5 above are used to calculate the weighted overall variance[6] and weighted overall proportion for the discontinued-use cookstove using equations (9) and (10) as follows:

$$SD^2 = \frac{(g_1 \times p_1 \times q_1) + (g_2 \times p_2 \times q_2) + (g_3 \times p_3 \times q_3)}{N} \quad \text{Equation (9)}$$

$$\bar{p} = \frac{(g_1 \times p_1) + (g_2 \times p_2) + (g_3 \times p_3)}{N} \quad \text{Equation (10)}$$

Where:

g_i	=	Size of stratum *i* (i.e. CP)
p_i	=	Expected proportion of successful cases in stratum *i* (i.e. CP)
$\bar{p}$		Expected overall proportion
SD	=	Sample standard deviation (SD^2 = overall variance)
N	=	Total population across all strata
i	=	Index identifying a stratum (where *i* = 1,…,*k*), i.e., *k* = 3

23. Substituting the values into equations (9) and (10) gives:

$$SD^2 = \frac{(30{,}000 \times 0.89 \times 0.11) + (10{,}000 \times 0.82 \times 0.18) + (20{,}000 \times 0.94 \times 0.06)}{60{,}000} = 0.09235 \quad \text{Equation (11)}$$

$$\bar{p} = \frac{(30{,}000 \times 0.89) + (10{,}000 \times 0.82) + (20{,}000 \times 0.94)}{60{,}000} = 0.8950 \quad \text{Equation (12)}$$

24. Substituting V and N into equation (9) gives:

$$n \geq \frac{1.96^2 \times 60{,}000 \times 0.09235}{(60{,}000 - 1) \times (0.1 \times 0.895)^2 + 1.96^2 \times 0.09235} = 44.26 \quad \text{Equation (13)}$$

25. Therefore, the required sample size is 45.

[6] The variance of a proportion is calculated as: *p(1-p).*

26. The total sample size may then be allocated across the CPs using the following equation:

$$n_i = \frac{g_i}{N} \times n \qquad \text{Equation (14)}$$

Where:

n_i	=	Sample size allocated to stratum *i* (i.e. CP)
g_i	=	Size of stratum *i* (i.e. CP)
N	=	Total population size
n	=	Total sample size
i	=	Index identifying a stratum (where *i* = 1,…,*k*), i.e., *k* = 3

27. Substituting the CP population sizes, the total population size and the total sample size into equation (14) gives the following results presented in table 6.

Table 6. Sample size allocation by CP

CP	Number of improved cookstoves	Equation	Sample size allocated
CP 1	30 000	$n_{CP_1} = \frac{30\,000}{60\,000} \times 45 = 22.50$	23
CP 2	10 000	$n_{CP_2} = \frac{10\,000}{60\,000} \times 45 = 7.50$	8
CP 3	20 000	$n_{CP_3} = \frac{20\,000}{60\,000} \times 45 = 15.00$	15
Total	**60 000**		**46**[7]

28. A sample size of 46 cookstoves may need to be adjusted upwards to account for non-response. If an expected response rate of 70 per cent is applied, the sample size within each CP should be increased by dividing the allocated sample size for each CP by the expected response rate (see table 7).

Table 7. Required sample sizes adjusted for an expected response rate of 70 per cent in each CP

CP	Sample size required before adjustment for non-response	Equation	Sample size required, adjusted for non-response
CP 1	23	$\frac{23}{0.7} = 32.86$	33
CP 2	8	$\frac{8}{0.7} = 11.43$	12
CP 3	15	$\frac{15}{0.7} = 21.43$	22
Total	**46**		**67**

[7] The allocated sample sizes are rounded up to whole numbers. As a result, the total allocated sample size increases from 45 to 46 due to the rounding up of the sample sizes within each CP.

29. After adjusting for non-response within each CP, the total sample size required to achieve the 95/10 confidence/precision criterion for the continued-use rate estimate is 67 cookstoves.

2.2. Operating efficiency of distributed improved cookstoves (mean value parameter)

30. The required sample size for this parameter depends on:

    (a) The expected mean, since the desired reliability is expressed relative to the mean;

    (b) The expected standard deviation; and

    (c) The required confidence level and target precision (95/10 confidence/precision criterion for a single sampling plan).

31. Based on similar studies in the region, the expected mean operating efficiency is assumed to be 0.20 and the expected standard deviation is assumed to be 0.05. To illustrate the application of stratified random sampling, the expected means and standard deviations for each CP have been selected such that the overall mean and weighted overall standard deviation are approximately 0.20 and 0.05, respectively, while allowing for heterogeneity among CPs (see table 8).

Table 8. Expected mean and standard deviation by CP

CP	Number of improved cookstoves	Expected mean	Expected standard deviation
CP 1	30 000	0.20	0.06
CP 2	10 000	0.05	0.09
CP 3	20 000	0.30	0.03
Total	**60 000**	**0.2083**	**0.0587**

32. The equation used to determine the total sample size across all three CPs is:

$$n \geq \frac{Z_{\alpha/2}^2 \times N \times SD^2}{(N-1) \times e_{abs}^2 + Z_{\alpha/2}^2 \times SD^2} \qquad \text{Equation (15)}$$

33. The values in table 8 above are used to calculate the weighted overall mean and weighted overall variance:

$$\overline{m} = \frac{(g_1 \times m_1) + (g_2 \times m_2) + (g_3 \times m_3)}{N} \qquad \text{Equation (16)}$$

$$SD^2 = \frac{(g_1 \times SD_1^2) + (g_2 \times SD_2^2) + (g_3 \times SD_3^2)}{N} \qquad \text{Equation (17)}$$

Where:

g_i = Size of the stratum *i* (i.e. CP)

m_i = Mean of the stratum *i* (i.e. CP)

$\overline{m}$	=	Weighted overall mean
SD_i	=	Standard deviation of stratum *i* (i.e. CP)
N	=	Total population size
e_{abs}		Target precision expressed as the maximum allowable absolute sampling error (i.e. 0.1 x $\overline{m}$)
i	=	Index identifying a stratum (where *i* = 1,…,*k*), i.e., *k* = 3

34. Substituting the values into equations (16) and (17) gives:

$$\overline{m} = \frac{(30{,}000 \times 0.20) + (10{,}000 \times 0.05) + (20{,}000 \times 0.30)}{60{,}000} = 0.2083$$ Equation (18)

$$SD^2 = \frac{(30{,}000 \times 0.06^2) + (10{,}000 \times 0.09^2) + (20{,}000 \times 0.03^2)}{60{,}000}$$ Equation (19)

$$= 0.00345$$

35. Substituting the values into equation (17) gives:

$$n \geq \frac{1.96^2 \times 60{,}000 \times 0.00345}{(60{,}000 - 1) \times (0.1 \times 0.283)^2 + 1.96^2 \times 0.00345} = 30.5211$$ Equation (20)

36. Therefore, the required total sample size is 31 cookstoves.

37. The total sample size may then be allocated across the CPs using the following equation:

Where: Equation (21)

$$n_i = \frac{g_i}{N} \times n$$

and

n_i	=	Sample size allocated to the stratum *i* (CP)
g_i	=	Size of the stratum *i* (CP)
N	=	Total population
n	=	Total sample size
	=	Index identifying a stratum (where *i* = 1,…,k)

38. Substituting the CP population sizes, the total population size, and the total sample size into equation (21) gives the following results presented in table 9.

Table 9. Sample size allocation by CP

CP	Number of improved cookstoves	Equation	Sample size allocated
CP 1	30 000	$n_{CP_1} = \frac{30\,000}{60\,000} \times 31 = 15.50$	16
CP 2	10 000	$n_{CP_2} = \frac{10\,000}{60\,000} \times 31 = 5.17$	6
CP 3	20 000	$n_{CP_3} = \frac{20\,000}{60\,000} \times 31 = 10.33$	11
Total	**60 000**	**31**	**33**

39. Note: The allocated sample sizes are rounded to whole numbers. As a result, the total allocated sample size increases from 31 to 33 due to the rounding up of the sample sizes within each CP.

40. If an expected response rate of 70 per cent is applied, the sample sizes should be increased by dividing the allocated sample size within each CP by the expected response rate to ensure that the required reliability criterion can be achieved. The sample sizes adjusted for an expected response rate of 70 per cent are presented in table 10.

Table 10. Required sample sizes adjusted for an expected response rate of 70 per cent

CP	Sample size required before adjustment for non-response	Equation	Sample size required, adjusted for non-response
CP 1	16	$\frac{16}{0.7} = 22.86$	23
CP 2	6	$\frac{6}{0.7} = 8.57$	9
CP 3	11	$\frac{11}{0.7} = 15.7$	16
Total	**33**		**48**

41. After adjusting for non-response within each CP, the total sample size required to achieve 95/10 confidence/precision criterion for the estimate of mean operating efficiency is 48.

3. Sampling results and reliability calculations: combining all three parameters of interest in a single survey

42. When more than one parameter is estimated through a single survey, the sample size used should be based on the largest sample size required among the parameters. This ensures that the required reliability requirements are achieved for all parameters. Based on the sample size calculations presented in the three preceding stratified sampling examples, a single survey designed to collect data for all three parameters would require a sample size of 101 cookstoves.

Table 11. Required sample sizes for multiple parameters

Parameter	Required sample size adjusted for an expected response rate of 70 per cent
Retention rate of improved cookstoves	101
Discontinued-use rate of displaced traditional cookstoves	67
Operating efficiency of distributed improved cookstoves	46

43. The allocation of the total sample size of 101 cookstoves across the three CPs is the same as that presented in the retention-rate example.

44. A sample of 101 cookstoves allocated across the three CPs will produce combined population estimates for the three parameters at a 95 per cent confidence level and a precision of at least ±10 per cent. Therefore, in this example, sampling 101 cookstoves results in precision greater than ±10 per cent for the estimates of the discontinued-use rate and operating efficiency.

3.1. Retention rate of improved cookstoves (proportion parameter)

45. For this example, a stratified sample of 101 improved cookstoves allocated across the three CPs was selected. The parameter of interest is the proportion of improved cookstoves that remain in operation within the population of 60,000 improved cookstoves distributed across the three CPs. Table 12 below presents the number of improved cookstoves planned to be surveyed within each CP and the number of cookstoves actually surveyed, after accounting for attrition and non-response.

Table 12. Planned and actual survey sample sizes by CP

CP	Number of cookstoves planned to be surveyed	Number of cookstoves actually surveyed
CP 1	50	42
CP 2	18	12
CP 3	33	25
Total	**101**	**79**

46. Table 13 presents illustrative survey results for the retention-rate parameter.

Table 13. Illustrative survey results for the retention-rate parameter

CP	Number of cookstoves surveyed	Number of cookstoves still in operation	Retention proportion
CP 1	42	35	$p_{CP_1} = \frac{35}{42} = 0.833$
CP 2	12	9	$p_{CP_2} = \frac{9}{12} = 0.75$
CP 3	25	23	$p_{CP_3} = \frac{23}{25} = 0.92$

47. The overall proportion of improved cookstoves that remain in operation across all three CPs is estimated using a weighted average of the CP-level proportions, calculated as follows:

$$p_{Strat} = \sum_{i=a}^{k} \frac{g_i}{N} \times p_i \qquad \text{Equation (22)}$$

Where:

p_{Strat}	=	The stratified estimated overall proportion
g_i	=	Size of stratum *i* (CP)
N	=	Total population
p_i	=	Proportion observed in stratum *i* (CP)
i	=	Index identifying a stratum (where *i* = 1,…,*k*)

48. Substituting the observed proportions and population sizes for each CP into the equation gives:

$$p_{Strat} = \left(\frac{30{,}000}{60{,}000} \times 0.83\right) + \left(\frac{10{,}000}{60{,}000} \times 0.75\right) + \left(\frac{20{,}000}{60{,}000} \times 0.92\right) \qquad \text{Equation (23)}$$
$$= 0.8467$$

49. Therefore, the estimated overall proportion of improved cookstoves still in operation across the three CPs is 0.8467, or 84.67 per cent if expressed as a percentage.

50. To assess whether the required reliability criterion has been achieved, a confidence interval is calculated for the estimated overall proportion as follows:

$$p_{Strat} \pm Z_{\alpha/2} \times s.e.(p_{Strat}) \qquad \text{Equation (24)}$$

Where:

p_{Strat}	=	The stratified estimated overall proportion
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
$s.e.(p_{Strat})$	=	Standard error of the stratified estimated overall proportion

51. The standard error of the stratified estimated overall proportion is calculated as follows:

$$s.e.(p_{Strat}) = \sqrt{\sum_{i=a}^{k} \left(\frac{g_i}{N}\right)^2 \times \left(1 - \frac{n_i}{g_i}\right) \times \frac{p_i \times q_i}{n_i}} \qquad \text{Equation (25)}$$

Where:

$s.e.(p_{Strat})$	=	Standard error of the stratified estimated overall proportion

g_i	=	Size of stratum *i* (CP)
N	=	Total population
n_i	=	Number of sampled units in stratum *i* (CP)
p_i	=	Proportion observed in stratum *i* (CP)
q_i	=	Expected proportion of unsuccessful cases in stratum *i* (1 – p)
i	=	Index identifying a stratum (where *i* = 1,…,*k*)

52. Substituting the values from this example into the equation gives the following standard error:

$$s.e.(p_{Strat}) = \sqrt{\left(\left(\frac{30{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{42}{30{,}000}\right) \times \frac{0.83 \times 0.17}{42}\right) + \left(\left(\frac{10{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{12}{10{,}000}\right) \left(\left(\frac{20{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{25}{20{,}000}\right) \times \frac{0.92 \times 0.08}{25}\right)\right)} \quad \text{Equation (26)}$$

Substituting the values from this example into the equation gives:

$$s.e.(p_{Strat}) = 0.0398 \quad \text{Equation (27)}$$

53. Substituting the results from the previous calculations into the confidence-interval equation gives:

$$0.8467 \pm 1.96 \times 0.0398 = (0.7687, 0.9247) \quad \text{Equation (28)}$$

54. Therefore, the estimated overall proportion of improved cookstoves that remain in operation across the three CPs is 0.8467 or 84.67 per cent, with a corresponding 95 per cent confidence interval of 0.7687 to 0.9247.

55. Expressed in relative terms, the relative margin of error (RME) or achieved relative precision may be calculated using the following equation:

$$RME = \frac{0.5 \times Width\ of\ Confidence\ Interval}{Estimated\ Stratified\ Overall\ Proportion} \quad \text{Equation (29)}$$

56. Substituting the values of previous equations gives:

$$RME = \frac{0.5 \times (0.9247 - 0.7687)}{0.8467} = \frac{0.078}{0.8467} = 0.0921 \quad \text{Equation (30)}$$

57. The achieved relative precision is 9.21 per cent of the estimated proportion. Therefore, the estimate of 84.67 per cent for the proportion of improved cookstoves that remain in operation meets the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

3.2. Continued-use rate of displaced traditional cookstoves (proportion parameter)

58. Although the sample size calculations for this parameter indicated that a sample size of 67 would be sufficient to achieve the required 95/10 confidence/precision criterion, a sample size of 101 cookstoves was planned in order to satisfy the requirements of the retention-rate parameter, which required the largest sample size among the three parameters. Table 14 below presents the number of cookstoves planned to be surveyed and the number of cookstoves actually surveyed after accounting for non-response.

Table 14. Planned and actual survey sample sizes by CP

CP	Number of cookstoves planned to be surveyed	Number of cookstoves actually surveyed
CP 1	50	42
CP 2	18	12
CP 3	33	25
Total	**101**	**79**

59. As noted in the sample size calculations, the sample-size determination for this parameter was based on the discontinued-use proportion rather than the continued-use proportion. Accordingly, the estimates presented below relate to the discontinued-use proportions.

60. Table 15 below presents illustrative survey results for the discontinued-use parameter.

Table 15. Illustrative survey results for the discontinued-use parameter

CP	Number of cookstoves actually surveyed	Number of traditional cookstoves still in use	Number of traditional cookstoves no longer in use	Discontinued-use proportion
CP 1	42	6	36	$p_{CP_1} = \frac{36}{42} = 0.86$
CP 2	12	2	10	$p_{CP_2} = \frac{10}{12} = 0.83$
CP 3	25	1	24	$p_{CP_3} = \frac{24}{25} = 0.96$

61. Although table 15 presents the estimated discontinued-use proportions for each CP, the parameter of interest is the overall discontinued-use proportion across all three CPs. This is estimated using the following equation:

$$p_{Strat} = \sum_{i=a}^{k} \frac{g_i}{N} \times p_i \qquad \text{Equation (31)}$$

Where:

p_{Strat}	=	Stratified estimated overall proportion
g_i	=	Size of stratum *i* (CP)
N	=	Total population
p_i	=	Proportion observed in stratum *i* (CP)

i	=	Index identifying a stratum (where *i* = 1,…,*k*)

62. Substituting the observed proportions and population sizes for each CP into the equation gives:

$$p_{Strat} = \left(\frac{30{,}000}{60{,}000} \times 0.86\right) + \left(\frac{10{,}000}{60{,}000} \times 0.83\right) + \left(\frac{20{,}000}{60{,}000} \times 0.96\right) = 0.8883$$ Equation (32)

63. Therefore, the estimated discontinued-use proportion across the three CPs is 0.8883, or 88.83 per cent.

64. The precision associated with this estimate may be assessed using a confidence interval. The equation for the 95 per cent confidence interval is:

$$p_{Strat} \pm Z_{\alpha/2} \times s.e.(p_{Strat})$$ Equation (33)

Where:

p_{Strat}	=	Stratified estimated overall proportion
$Z_{\alpha/2}$	=	Critical value of the standard normal distribution corresponding to the selected confidence interval, e.g., 1.96 for 95 per cent confidence interval
$s.e.(p_{Strat})$	=	Standard error of the stratified estimated overall proportion

65. The standard error of the stratified estimated overall proportion is calculated as follows:

$$s.e.(p_{Strat}) = \sqrt{\sum_{i=a}^{k} \left(\frac{g_i}{N}\right)^2 \times \left(1 - \frac{n_i}{g_i}\right) \times \frac{p_i \times (1 - p_i)}{n_i}}$$ Equation (34)

Where:

$s.e.(p_{Strat})$	=	Standard error of the stratified estimated overall proportion
g_i	=	Size of stratum *i* (CP)
N	=	Total population
n_i	=	Number of sampled units in stratum *i* (CP)
p_i	=	Proportion observed in stratum *i* (CP)
i	=	Index identifying a stratum (where *i* = 1,…,*k*)

66. Substituting the values from this example into the equation gives:

$$s.e.(p_{Strat}) = \sqrt{\left(\left(\frac{30{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{42}{30{,}000}\right) \times \frac{0.86(1-0.86)}{42}\right) + \left(\left(\frac{10{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{12}{10{,}000}\right) \times \frac{0.83(1-0.83)}{12}\right) + \left(\left(\frac{20{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{25}{20{,}000}\right) \times \frac{0.96(1-0.96)}{25}\right)}$$ Equation (35)

Therefore:

$$s.e.(p_{Strat}) = 0.0349 \quad \text{Equation (36)}$$

67. Substituting the results from the previous calculations into the confidence-interval equation gives:

$$0.8883 \pm 1.96 \times 0.0349 = (0.8199, 0.9567) \quad \text{Equation (37)}$$

68. Therefore, the estimated discontinued-use proportion across the three CPs is 0.8883, with a corresponding 95 per cent confidence interval of 0.8199 to 0.9567.

69. Expressed in relative terms, the relative margin of error (RME) or achieved related precision is calculated using the following equation:

$$\text{RME} = \frac{0.5 \times Width\ of\ Confidence\ Interval}{Estimated\ Stratified\ Ovelerall\ Proportion} \quad \text{Equation (38)}$$

70. Substituting the values from the previous calculations gives:

$$\text{RME} = \frac{0.5 \times (0.9567 - 0.8199)}{0.8883} = \frac{0.0684}{0.8883} = 0.077 \quad \text{Equation (39)}$$

71. The achieved precision is 7.7 per cent of the estimated proportion. Therefore, despite attrition and non-response reducing the achieved sample size, the estimate of 88.83 per cent for the discontinued-use proportion meets the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

3.3. Operating efficiency of distributed improved cookstoves (mean value parameter)

72. The third parameter of interest is the mean operating efficiency of the distributed improved cookstoves. As a single survey was undertaken to collect data for all three parameters of interest, the planned and achieved sample sizes presented in Table 16 are the same as those used in the two preceding examples.

Table 16. Planned and actual survey sample sizes by CP

CP	Number of cookstoves planned to be surveyed	Number of cookstoves actually surveyed
CP 1	50	42
CP 2	18	12
CP 3	33	25
Total	**101**	**79**

73. Table 11 presents the mean operating efficiency and sample standard deviation observed for each CP.

Table 17. Sample mean operating efficiency and sample standard deviation by CP

CP	Number of cookstoves actually surveyed	Sample mean operating efficiency	Sample standard deviation
CP 1	42	0.22	0.07
CP 2	12	0.06	0.06
CP 3	25	0.28	0.04

74. To estimate the overall mean operating efficiency across all three CPs, a weighted average of the CP-level means is calculated using equation:

$$m_{Strat} = \sum_{i=a}^{k} \frac{g_i}{N} \times m_i \qquad \text{Equation (40)}$$

Where:

m_{Strat}	=	Stratified estimated overall mean
g_i	=	Size of stratum *i* (CP)
N	=	Total population
m_i	=	Mean observed in stratum *i* (CP)
i	=	Index identifying a stratum (where *i* = 1,…,k)

75. Substituting the observed means and population sizes for each CP into the equation gives:

$$m_{Strat} = \left(\frac{30{,}000}{60{,}000} \times 0.22\right) + \left(\frac{10{,}000}{60{,}000} \times 0.06\right) + \left(\frac{20{,}000}{60{,}000} \times 0.28\right) \qquad \text{Equation (41)}$$
$$= 0.2133$$

76. Therefore, the estimated overall mean operating efficiency across the three CPs is 0.2133.

77. The precision associated with this estimate may be assessed using a confidence interval. The equation for the 95 per cent confidence interval is:

$$m_{Strat} \pm 1.96 \times s.e.(m_{Strat}) \qquad \text{Equation (42)}$$

Where:

m_{Strat}	=	Stratified estimated overall mean
1.96	=	Critical value of the standard normal distribution for which the cumulative probability equals $1 - \alpha/2$(e.g. $Z_{0.025} = 1.96\ for$ a 95 per cent confidence level)
$s.e.(m_{Strat})$	=	Standard error of the stratified estimated overall mean

78. The standard error of the stratified estimated overall mean is calculated as follows:

$$s.e.(m_{Strat}) = \sqrt{\sum_{i=a}^{k} \left(\frac{g_i}{N}\right)^2 \times \left(1 - \frac{n_i}{g_i}\right) \times \frac{SD_i^2}{n_i}}$$ Equation (43)

Where:

$s.e.(m_{Strat})$	=	Standard error of the stratified estimated overall mean
g_i	=	Size of stratum *i* (CP)
N	=	Total population
n_i	=	Number of sampled units in stratum *i* (CP)
SD_i^2	=	Variance observed in stratum *i* (CP)
i	=	Index identifying a stratum (where *i* = 1,…,*k*)

79. Substituting the values from this example into the equation gives:

$$s.e.(m_{Strat}) = \sqrt{\left(\left(\frac{30{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{42}{30{,}000}\right) \times \frac{(0.07)2}{42}\right) + \left(\left(\frac{10{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{12}{10{,}000}\right) \times \frac{(0.06)2}{12}\right) + \left(\left(\frac{20{,}000}{60{,}000}\right)^2 \times \left(1 - \frac{25}{20{,}000}\right) \times \frac{(0.04)2}{25}\right)}$$ Equation (44)

Therefore:

$$s.e.(m_{Strat}) = 0.00667$$ Equation (45)

80. Substituting the results from the previous calculations into equation (43), gives a 95 per cent confidence interval for the estimated overall mean operating efficiency:

$$0.2133 \pm 1.96 \times 0.00667 = (0.2003, 0.2264)$$ Equation (46)

81. Therefore, the estimated overall mean operating efficiency across the three CPs is 0.2133, with a 95 per cent confidence interval of 0.2003 to 0.2264.

82. Expressed in relative terms, the relative margin of error (RME) or the achieved relative precision may be calculated using the following equation:

$$RME = \frac{0.5 \times \text{Width of Confidence Interval}}{\text{Estimated Stratified Overall Mean}}$$ Equation (47)

83. Substituting the values from the previous calculations into the equation gives:

$$RME = \frac{0.5 \times (0.2264 - 0.2003)}{0.2133} = \frac{0.0130}{0.2133} = 0.061$$ Equation (48)

84. The achieved precision is 6.1 per cent of the estimated mean. Therefore, despite attrition and non-response reducing the achieved sample size, the estimate of 0.2133 for the mean operating efficiency of improved cookstoves meets the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

4. Conclusion

85. In summary, the estimates of all three parameters satisfy the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

Appendix 4. Best-practice examples for confidence intervals and reliability assessments

1. Introduction

1. The examples presented in this appendix illustrate how to estimate a mean value and a proportion, and how to assess whether the resulting estimates satisfy the required reliability criterion. For simplicity, both examples assume simple random sampling. Unless otherwise specified in the applicable mechanism methodology, the required reliability criterion is 95/10, i.e., 95 per cent confidence and ±10 per cent precision, in accordance with this Methodological tool.

2. The confidence intervals and reliability calculations presented in this appendix rely on standard statistical approaches, including normal-theory approximations, which are generally appropriate where sample sizes are sufficiently large or where the underlying data do not exhibit substantial skewness. Alternative approaches may be applied where justified.

3. If calculations are performed manually an appropriate number of decimal places should be retained throughout the intermediate steps, with rounding applied only to the final result. To illustrate this principle, the calculations presented in the examples retain several decimal places throughout and apply rounding only to the final reported values.

2. Example 1: LED Project – mean value parameter

4. The parameter of interest in this example is the mean daily usage of a light-emitting diode (LED), measured in hours per day, across the population of LEDs distributed in a particular region.

5. The population consists of 420,000 households, each of which received one LED. A simple random sample of 140 households was selected, and the daily usage (in hours) of each LED was recorded. The sample data are presented in table 1.

Table 1. Average LED daily usage (hours)

LED lamp	Daily usage	LED lamp	Daily usage	LED lamp	Daily usage	LED lamp	Daily usage	LED lamp	Daily usage	LED lamp	Daily usage	LED lamp	Daily usage
1	3.78	21	3.63	41	2.81	61	4.17	81	3.62	101	2.24	121	0.58
2	3.12	22	3.17	42	4.57	62	4.68	82	2.46	102	4.79	122	6.09
3	4.42	23	3.26	43	3.56	63	2.99	83	6.14	103	4.59	123	0.39
4	4.09	24	6.97	44	4.41	64	3.34	84	0.67	104	3.27	124	3.69
5	1.15	25	0.48	45	3.26	65	5.37	85	4.73	105	1.86	125	2.04
6	2.87	26	2.50	46	0.30	66	2.17	86	1.03	106	0.00	126	4.51
7	4.79	27	2.92	47	5.48	67	2.36	87	2.34	107	6.70	127	4.39
8	4.20	28	6.82	48	1.75	68	3.12	88	4.66	108	3.36	128	3.58
9	1.13	29	0.92	49	3.38	69	4.69	89	2.40	109	5.39	129	4.23
10	3.68	30	2.35	50	1.24	70	5.40	90	5.28	110	2.04	130	5.28
11	2.91	31	0.19	51	3.62	71	4.22	91	5.90	111	3.58	131	3.71
12	2.47	32	4.19	52	7.41	72	1.27	92	0.60	112	6.27	132	2.41
13	3.46	33	3.15	53	1.74	73	2.93	93	5.85	113	0.41	133	1.58
14	2.19	34	3.19	54	3.60	74	2.17	94	1.22	114	4.55	134	3.96
15	2.25	35	7.15	55	2.18	75	4.24	95	7.76	115	2.61	135	5.86
16	2.37	36	1.70	56	4.12	76	6.07	96	4.50	116	6.37	136	5.46
17	2.38	37	2.98	57	4.88	77	5.26	97	5.68	117	4.30	137	2.90
18	3.23	38	5.00	58	2.92	78	2.46	98	2.81	118	3.08	138	3.17
19	1.78	39	0.99	59	0.82	79	1.33	99	4.03	119	3.17	139	4.17
20	3.57	40	6.54	60	3.16	80	2.55	100	0.24	120	6.24	140	6.93

6. The parameter of interest – the mean daily usage of a LED (hours) across the population of distributed LEDs – is estimated using the sample mean. The sample mean is often denoted by $\bar{y}$ and is calculated as:

$\frac{1}{n}(y_1 + y_2 + \cdots + y_n)$, or the shorthand form $\frac{1}{n}\sum_{i=1}^{n} y_i$.

Where:

n = Sample size (140 in this example)

7. The sample mean daily usage for the 140 LEDs is 3.4686 hours. For reporting purposes, this may be rounded to an appropriate number of decimal places. In this example, the mean daily usage is estimated to be 3.47 hours. The sample standard deviation is 1.7557 hours.

3. Confidence, precision and reliability

8. Rather than presenting only a single estimate, the results should be summarized using a confidence interval. In this example, the 95 per cent confidence interval is 3.18 to 3.76 hours. This means that there is 95 per cent confidence that the true population mean daily usage of a LED lies between 3.18 and 3.76 hours. While the sample mean is the estimate used in calculations, it is good practice to present the estimate together with its confidence interval when reporting results.

9. The 95 per cent confidence interval for the population mean is calculated as:

Sample mean ± t-value$_{g/2}$ x standard error of the mean.[1]

10. The estimate of 3.47 hours is considered to satisfy the required reliability criterion if the achieved precision, defined as the margin of error (t-value x standard error of the mean), does not exceed the required precision threshold. For small-scale mechanism activities, this threshold is ±10 per cent of the estimated mean.

11. Detailed calculations are presented below. In this example, the achieved precision is 8.4 per cent of the estimated mean. Therefore, the estimate of 3.47 hours satisfies the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

4. Checking reliability

4.1. Standard error of the mean

12. The standard error of the mean for a simple random sample is calculated as:

$$\sqrt{(1-f) \times \frac{SD^2}{n}} \qquad \text{Equation (1)}$$

[1] These calculations assume that the sampling distribution of the mean is approximately normal, which is generally appropriate where the sample size is sufficiently large or where the underlying data do not exhibit substantial skewness.

Where:

f	=	Sampling fraction, calculated as the proportion of the population included in the sample (140 / 420,000 = 0.00033)
SD^2	=	Sample variance (3.0826 in this example)
n	=	Sample size (140 in this example)

13. Substituting the values into the equation gives:

$$\sqrt{(1-f)\times\frac{s^2}{n}} = \sqrt{\left(1-\frac{140}{420000}\right)\times\frac{3.0826}{140}} = \sqrt{0.99967\times\frac{3.0826}{140}} = \sqrt{0.0220} = 0.1484 \quad \text{Equation (2)}$$

14. Therefore, the standard error of the mean is 0.1484.

4.2. t-value

15. This value depends on (i) the selected confidence level; and (ii) the sample size. The appropriate value may be obtained from Student's t-distribution tables or calculated using standard statistical software. In Microsoft Excel, the value may be calculated using the TINV.2T[2] function.

16. For a sample size of 140 and a 95 per cent confidence level, the corresponding t-value is 1.9599.

4.3. Precision

17. The achieved precision associated with an estimate is calculated as:

t-value x standard error of the mean.

18. For this example, assuming 95 per cent confidence level, the achieved precision for the estimated mean LED usage is:

$$\pm(1.96\times 0.1484), i.e., \pm 0.2908 \quad \text{Equation (3)}$$

19. Expressed relative to the estimated mean, the achieved precision is:

$$\frac{0.2908}{3.4686} = 0.0838 \text{ or } 8.38 \text{ per cent} \quad \text{Equation (4)}$$

20. Therefore, the estimate satisfies the required reliability criterion, as the achieved precision is less than ±10 per cent of the estimated mean. Alternative method for assessing reliability

21. The limits of the confidence interval are calculated as:

(a) Sample mean ± (t-value x standard error of the mean);

[2] The t-value corresponding to a 95 per cent confidence level may be calculated in Microsoft Excel using the function T.INV.2T(0.05, n- 1), where n is the sample size. For example, T.INV.2T(0.05,139) returns the t-value for a sample size of 140 at the 95 per cent confidence level.

(b) This may be expressed more generally as: sample mean ± precision.

22. Where the lower confidence limit is equal to the sample mean minus the precision and the upper confidence limit is equal to the sample mean plus the precision.

23. The relative margin of error (RME) or the achieved relative precision may therefore be calculated using the following equation:

$$RME = \frac{0.5 \times width\ of\ confidence\ interval}{mean} \quad \text{Equation (5)}$$

24. In this example, the mean LED daily usage is 3.4686 hours and the 95 per cent confidence interval is 3.1778 to 3.7595 hours. The achieved relative precision is therefore:

$$\frac{0.5 \times (3.7595 - 3.1778)}{3.4686} = \frac{0.5 \times 0.5817}{3.4686} = 0.084 \quad \text{Equation (6)}$$

25. The achieved precision is 8.4 per cent of the estimated mean and therefore satisfies the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

26. This approach may be particularly useful where statistical software has been used to generate both the confidence interval and the sample mean.

5. Example 2: Cookstove project – Proportional parameter

27. The parameter of interest in this example is the proportion (or percentage) of cookstoves in a particular region that remained operational at the end of the third year following distribution. The population consists of 640,000 households, each of which received one cookstove. A simple random sample of 274 households was selected, and for each household it was recorded whether the cookstove was still operational.

28. The parameter of interest – the proportion (or percentage) of cookstoves that remained operational in the population – is estimated using the sample proportion. as the sample proportion is commonly denoted by *p* and is calculated as follows:

$$\hat{p} = \frac{r}{n} \quad \text{Equation (7)}$$

Where:

$\hat{p}$	=	Sample proportion
r	=	The number of cookstoves that remain operational (successful cases); and
n	=	Sample size

29. In this example, 159 of the 274 sample cookstoves remained operational. The sample proportion is therefore:

$$\hat{p} = \frac{159}{274} = 0.5803 \quad \text{Equation (8)}$$

30. For reporting purposes, this may be rounded up to an appropriate number of decimal places. In this example, the estimated proportion is 0.58, or 58 per cent. In other words, 58 per cent of the cookstoves were estimated to remain operational after the third year.

5.1. Confidence, precision and reliability

31. Rather than presenting only a single estimate, the results should be summarized using a confidence interval. In this example, the 95 per cent confidence interval for the proportion is 0.5218 to 0.6387. This means that there is a 95 per cent confidence that the true proportion of cookstoves in the population that remain operational lies between 0.5218 and 0.6387 (52.18 per cent and 63.98 per cent).

32. The 95 per cent confidence interval for the population proportion is calculated as: sample proportion ± 1.96 x standard error of the proportion.[3]

33. The estimate of 58 per cent is considered to satisfy the required reliability criterion if the achieved precision, defined as the margin of error (1.96 x standard error of proportion), does not exceed the required precision threshold. For small-scale mechanism activities, this threshold is ±10 per cent of the estimated proportion. In this example, the required precision threshold is ±0.058 in absolute terms (±5.8 percentage points).

34. Detailed calculations are presented below. In this example, the achieved precision is 10.06 per cent of the estimated proportion. Therefore, the estimate of 58 per cent operational cookstoves does not satisfy the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

5.2. Checking reliability

5.2.1. Standard error of the proportion

35. The standard error of the proportion for a simple random sample is calculated as follows:

$$\sqrt{(1-f)\times\frac{p\times q}{n}} \qquad \text{Equation (9)}$$

Where:

f	=	Sampling fraction, calculated as the proportion of the population included in the sample ($\frac{274}{640000}=0.00043$)
p	=	Sample proportion (0.5803 in this example)
q	=	*1 – p*, representing the proportion of cookstoves that were not operational after three years (0.4197 in this example)
n	=	Sample size (274 in this example)

[3] A confidence interval for a proportion is: sample proportion ± z-value x standard error of the proportion. The z-value depends on the level of confidence. For 95 per cent confidence it is 1.9599.

36. Substituting the values into the equation gives:

$$\sqrt{(1-f)\times\frac{pq}{n}}=\sqrt{(1-0.00043)\times\frac{0.5803\times0.4197}{274}}=0.0298$$ Equation (10)

37. The same result may be obtained by expressing the sample proportion directly in terms of the observed numbers of operational and non-operational cookstoves:

$$\sqrt{(1-f)\times\frac{p\times q}{n}}=\sqrt{\left(\frac{640000-274}{640000}\right)\times\frac{\left(\frac{159}{274}\right)\times\left(\frac{115}{274}\right)}{274}}=0.0298$$ Equation (11)

38. Therefore, the standard error of the proportion is 0.0298. Expressed in percentage terms, the standard error is 2.98 percentage points.

5.2.2. Precision

39. The achieved relative precision associated with a proportion is calculated as: z-value × standard error of the proportion.

40. For this example, assuming a 95 per cent confidence level, the achieved precision for the estimated proportion of operational cookstoves is ±(1.9599 x 0.0298), i.e., ±0.0584.

41. Expressed relative to the estimated proportion, the achieved relative precision is 0.0584 / 0.5803 = 0.1006 or 10.06 per cent.

42. The data does not meet the required specification. Therefore, the estimate does not satisfy the required reliability criterion of 95 per cent confidence and ±10 per cent precision as the achieved precision exceeds the required threshold by a small margin.

5.2.3. Alternative method for assessing the achieved relative precision

43. The limits of the confidence interval are the sample proportion ± (z-value x standard error of the proportion). This may be expressed more generally as: sample proportion ± precision, where the lower confidence limit is equal to the sample proportion minus the precision, and the upper confidence limit is equal to the sample proportion plus the precision.

44. The achieved precision may therefore be calculated using the following equation:

$$\frac{{}^{1}/_{2}\times\text{ width of confidence interval}}{\text{proportion}}$$ Equation (12)

45. In this example, the proportion of cookstoves that remain operational is 0.5803 and the 95 per cent confidence interval is 0.5219 to 0.6387.

46. The achieved relative precision is therefore:

$$\frac{{}^{1}/_{2} \times (0.6387 - 0.5219)}{0.5803} = \frac{{}^{1}/_{2} \times 0.1168}{0.5803} = 0.1006$$ Equation (13)

47. The achieved relative precision is 10.06 per cent of the estimated proportion and therefore does not satisfy the required reliability criterion of 95 per cent confidence and ±10 per cent precision.

48. This approach may be particularly useful where statistical software has been used to generate both the confidence interval and the sample proportion.

5.2.4. Comments

49. The calculations presented above assume that the sampling distribution of the proportion is approximately normal. This assumption is generally appropriate where the proportion of interest is neither very small nor very large and where the sample size is sufficiently large.

50. Where the sampling fraction (*f) is* small, the finite population correction factor (*1 – f*) will be close to 1. In such cases, the standard error of the proportion is often approximated using the more conservative equation:

$$\sqrt{\frac{p \times q}{n}.4}$$ Equation (14)

51. If statistical software is used to calculate the confidence interval, the software may apply an exact method based on the binomial distribution rather than a normal approximation. In such cases, reliability may be assessed using the confidence interval approach based on the width of the confidence interval.

6. How to deal with failure to achieve reliability

6.1. Introductory notes on addressing failure to achieve reliability

52. This section outlines a series of steps to follow where the required reliability criteria are not achieved. The illustrative scenario assumes a project for which the required reliability criterion is 95/10 (i.e., 95 per cent confidence and ±10 per cent relative precision), and where the parameter of interest is a mean value parameter.

6.2. Scenario: LED project – Mean value parameter

53. The parameter of interest in this example is the mean daily usage of a LED (in hours) across the population of LEDs distributed in a particular region. The population consists of 420,000 households, each of which received one LED. A simple random sample was to be selected.

54. The sample size calculation assumed an expected population mean LED daily usage of 3.5 hours and an expected population standard deviation (SD) of 2.5 hours. Based on these assumptions, the required sample size was calculated to be 196 households.

6.3. Example 1

55. A simple random sample of 140 households was selected, and the daily usage (in hours) of each LED was recorded. Summary statistics and the results of the reliability assessment are presented in table 2 below.

Table 2. Summary statistics and results of the reliability assessment

Summary statistics	Sample data
Population size	420 000
Sample size (n)	140
Mean	3.7230
Standard deviation	3.7838
Standard error of the mean	0.3197
Absolute precision	± 0.6266
Relative precision	± 0.1683 i.e., 16.83 per cent

56. In the example, the required reliability criterion is not achieved. Therefore, the estimated mean daily usage of 3.72 hours is not sufficiently reliable for the required level of confidence and precision. This also indicates that the confidence interval associated with the estimate is wider than permitted under the required reliability criterion.

57. The 95 per cent confidence interval is 3.10 to 4.35 hours. This indicates that there is a 95 per cent confidence that the true population mean LED daily usage lies between 3.10 to 4.35 hours.

58. Possible steps that may be considered where the required reliability criterion is not achieved are presented below. However, before undertaking any recalculations or additional sampling, the data should first be carefully reviewed to ensure that they are of the highest possible quality and that the planned analyses remain appropriate.

59. The approaches outlined below are presented in a suggested order. First, the underlying data should be scrutinized, if this has not already been undertaken. Next, one or more statistical analysis approaches may be considered using the existing dataset. If the required reliability criterion is still not achieved, additional sampling may be undertaken. No particular order is implied among the statistical analysis approaches:

    (a) Scrutinize the raw data;

    (b) Possible analysis approaches:

        (i) Scrutinize the summary statistics;

        (ii) Post-stratification; and

    (c) Undertake additional sampling.

60. If none of these approaches results in the required reliability criterion being achieved, then the estimate should be considered not to satisfy the required reliability criterion.[4]

6.4. Analysis approach 1: scrutinize the raw data

61. It is important to scrutinize the raw data carefully prior to estimating the parameter and checking its reliability. This can be done using graphical summaries such as histograms, boxplots, and normal probability plots. These plots may help identify outliers in the data, as well as any skewness in the distribution.

62. An outlier may be the result of a mistake (e.g., incorrectly recorded or entered into the database), in which case it may be corrected. Alternatively, it may be a genuine observation, in which case it should be retained and included in the analysis. If the data are highly skewed, it may be appropriate to transform them prior to analysis. The reliability should then be assessed using the transformed data. Common transformations include logarithmic and square-root transformations.

6.5. Analysis approach 2(a): scrutinize the summary statistics

63. The two key elements in the sample size calculation (other than the reliability criterion and confidence level) are the expected population mean value and the expected population standard deviation. If the sample data do not satisfy the required reliability criterion, this could be due to the sample estimates differing substantially from the expected values. For instance:

    (a) If the sample mean is lower than the expected population mean, but the standard deviation is the same as expected for the population, then the achieved relative precision will be greater than 10 per cent; or

    (b) If the sample mean is the same as the expected population mean, but the standard deviation is larger than expected, then the achieved relative precision reliability will be greater than 10 per cent.

64. This is illustrated in table 3 below, where the sample size calculation is the same as that described in the scenario above (mean LED daily usage of 3.5 hours and standard deviation of 2.5 hours, giving a required sample size of 196 households). However, only 140 households were sampled. Table 3 below shows how the achieved relative precision is affected by different combinations of sample mean and sample standard deviation for a sample of 140 households.

Table 3. Sample size of 140 households

Scenario	Population size (N)	Sample size (n)	Sample mean	Sample standard deviation	Absolute precision	Relative precision (in per cent)
(i)	420 000	140	3.0 hours	2.5	0.41	13.80
(ii)	420 000	140	3.5 hours	3.0	0.50	14.20

[4] A generalized discounting approach to address deficiencies in reliability is not provided in this methodological tool. However, where project proponents can demonstrate that discounting emission reduction estimates or applying a conservative estimate of the parameter (e.g., using the lower or upper confidence bound, as appropriate) is the only practical option, the procedures for request for deviation shall be followed.

65. In scenario (i), the absolute precision is as anticipated. However, because the sample mean is lower than expected, the relative precision is greater than the required 10 per cent. In scenario (ii), because the sample standard deviation is larger than expected, both the absolute precision and the relative precision are greater than anticipated in the sample size calculation.

66. It may be reasonable to accept data arising from scenarios such as (i). The following criterion is therefore proposed:

    (a) Provided that the sample standard deviation is not more than 10 per cent greater than the standard deviation used in the sample size calculation, and the sample size is at least 100 units, the sample data may be accepted.

67. In example 1 the sample size exceeds 100 units, but the sample standard deviation is considerably larger than the value used in the sample size calculation (3.78 compared with 2.5). The sample standard deviation is therefore more than 50 per cent greater than the value assumed in the sample size calculation. Consequently, the sample data cannot be accepted under the criterion proposed above.

68. The rationale for the conclusions in the preceding paragraphs is explained below.

69. The distribution of the sample variance, SD^2, is given by:

$$\frac{(n-1) \times SD^2}{\sigma^2} \times x^2_{1-\alpha,n-1} \qquad \text{Equation (15)}$$

70. From probability tables for the Chi-squared distribution, it can be shown that, for samples of size ≥100, approximately 95 per cent of sample variances would be expected to be less than 1.24 times the population variance (i.e., less than 1.24 σ^2). It may therefore be argued that, provided the sample variance is not more than 24 per cent greater than the expected population variance, the sample variance can be regarded as an acceptable estimate of the population variance.

71. Similarly, provided that the sample standard deviation is not more than approximately 12 per cent greater than the expected population standard deviation, it may be regarded as an acceptable estimate of the population standard deviation.

72. Table 4 below extends this argument to other sample sizes. As the sample size becomes smaller, the multiplier becomes larger, which is to be expected. The smaller the sample size, the greater the variability in the estimates of the population variance and, consequently, the wider the distribution of the sample variance.

Table 4. Sample size and ratio of sample standard deviation to population standard deviation

Sample size	95 per cent of ratios less than this value
30	1.21
40	1.18
50	1.16
60	1.15
70	1.14
80	1.13

Sample size	95 per cent of ratios less than this value
90	1.12
100	1.12
125	1.10
150	1.09
175	1.09
200	1.08
250	1.07
300	1.07

73. Balancing the need for a reasonable amount of information with the recognition that the sample standard deviation is unlikely to be exactly the same as the population standard deviation, then the sample data may be accepted provided that:

(a) The sample size is sufficiently large;

(b) The sample standard deviation is only marginally larger than the population standard deviation; and

(c) The minimum sample size is 100 units and that the maximum overage in the sample standard deviation is 10 per cent.[5]

6.6. Analysis approach 2(b): Post-stratification

74. If there appears to be a characteristic of the population that is responsible for the apparent increase in the variability of the data, for example, households in urban areas use their LEDs for longer periods than the households in rural areas, then this characteristic may be used as a stratification variable, and the mean LED daily usage recalculated using post-stratification techniques. In this example, geographical area (urban or rural) is the stratification variable. Other examples might include the different time of the year in which data are collected, where usage varies seasonally.

75. Taking the stratification variable into account should improve the precision of the estimate of the mean usage for the population, and hence improve its reliability. It may also provide a more accurate estimate of the population mean usage. An example of post-stratification is presented below.

76. This example uses the same dataset as example 1, for which the following data were collected (see table 5 below).

[5] When a sample fails to meet the required reliability because the observed standard deviation is higher than the value used in the sample size calculation, the data may still be accepted if the deviation reflects normal sampling fluctuation rather than true excess variability. As the estimate of a standard deviation becomes stable only with sufficiently large samples, a minimum sample size of 100 is required before such acceptance can be justified. Statistical properties of the sampling distribution show that, for samples of this size, most sample standard deviations fall within roughly 10–12 per cent of the population value; therefore, the tool allows an overage of up to 10 per cent above the assumed standard deviation. Accordingly, sample data that miss the reliability criterion solely due to a moderately higher standard deviation may still be regarded as acceptable if the sample includes at least 100 observations and the sample standard deviation does not exceed the assumed value by more than 10 per cent, ensuring both statistical defensibility and practical flexibility in data evaluation.

Table 5. Summary of sample data and reliability calculations

Summary statistics	
Population size	420 000
Sample size (n)	140
Mean	3.7230
Standard deviation	3.7838

77. The total population of 420,000 households in this region is distributed between rural and urban areas in a ratio of 60 per cent to 40 per cent respectively, and it is believed that LED daily usage is lower in rural areas than in urban areas. Of the 140 households included in the sample, 104 were from rural areas and 36 were from urban areas. The proportions in the sample from rural and urban areas (74 per cent and 26 per cent, respectively) do not match those of the population. As a result, the estimate of average daily usage based on all 140 LEDs may be lower than the true population average.

78. Summary statistics for the two sub-groups are presented in table 6.

Table 6. Summary statistics for the two sub-groups

Summary statistics	Rural	Urban
Population size	252 000	168 000
Sample size (n)	104	36
Mean	1.96	8.816
Standard deviation	0.55	3.875

79. Post-stratification can be used to estimate the mean LED daily usage while reflecting the rural-to-urban proportions in the population. This is done by calculating a weighted average of the stratum means, as follows:

80. Post-stratification mean:

$$\bar{y}_{post-st} = \frac{1}{N}\sum_{i=1}^{k} g_i \times m_i \qquad \text{Equation (1)}$$

Where:

N	=	Total population size across all strata
g_i	=	Population size in each stratum *i*
$\bar{y}_i$	=	Mean for stratum *i*
k	=	Number of strata
i	=	Index identifying a stratum (where *i* = 1,…,k)

81. Substituting the values from this example gives:

$$i.e. \frac{(252000 \times 1.96 + 168000 \times 8.816)}{420000} = 4.7024 \qquad \text{Equation (16)}$$

82. The standard error of this estimate, which can then be used to determine the precision, is:

$$se = \sqrt{\sum_{i=1}^{L} \left[\frac{g_i}{N}\right]^2 \times \left[\frac{g_i - n_i}{g_i}\right] \times \frac{{SD_i}^2}{n_i}}$$ Equation (17)

Where:

se	=	Standard error
N	=	Total population size
g_i	=	Population size in each stratum *i*
n_i	=	Sample size from stratum *i*
SD_i	=	Standard deviation for stratum *i*
i	=	Index identifying a stratum

83. Substituting the values from this example gives:

$$i.e. \sqrt{\left(\frac{252000}{420000}\right)^2 \times \left(\frac{252000 - 104}{252000}\right) \times \frac{0.55^2}{104} + \left(\frac{168000}{420000}\right)^2 \times \left(\frac{168000 - 36}{168000}\right) \times \frac{3.875^2}{36}} = 0.2603$$ Equation (18)

84. Consequently, the absolute precision (assuming a 95 per cent confidence interval) is 1.96 x 0.2603 = 0.5102.

85. The relative precision is therefore: 0.5102 / 4.7024 = 10.85 per cent.

86. The achieved relative precision is 10.85 per cent and therefore the required reliability criterion has not been achieved.

6.7. Analysis approach 3: take an additional sample

87. Another option for improving the precision of the study data is to take an additional sample. The formula used to calculate the size of the additional sample would be the same as that used for the original study. However, the values used for the mean and standard deviation may need to be revised in light of the sample results. It is also advisable to explore different combinations of plausible mean and standard deviations values in order to identify a total sample size that is sufficient to address the reliability concern. The additional sample size would then be the difference between the revised total sample size and the number of observations already collected.

88. For example, in example 1, the mean and standard deviation used in the original sample size calculation were 3.5 and 2.5 hours, respectively. The sample of 140 LEDs produced a mean of 3.72 hours and a standard deviation of 3.78. These values may represent more realistic estimates of the population mean LED daily usage and population standard deviation and can therefore be used in revised sample size calculations. Table 7 below presents several combinations of mean and standard deviation values and the corresponding sample size requirements.

Table 7. Comparison of sample size estimates under different mean and standard deviation assumptions

Sample size for 95/10 reliability (95 per cent confidence and 10 per cent precision)		
Standard deviation	**Mean**	**Sample size**
2.5	3.5	196
2.5	3.72	174
3.78	3.5	448
3.78	3.72	397

89. If the activity participant believes that the population mean is approximately 3.5 hours, but that the standard deviation was underestimated and is likely to be closer to 3.78, then the required total sample size would be 448. In this case, an additional sample of 308 observations (448 – 140) would be required. However, if the population mean is also believed to be closer to 3.72 hours, then the required total sample size would be 397 and the additional sample required would be 257 observations (397 – 140).

90. Note that the illustration above uses only two values for the mean and two values for the standard deviation. In practice, a range of plausible mean and standard deviation values should be considered. This is consistent with the approach recommended when undertaking the original sample size calculation.

6.8. Alternative post-adjustment methods to address failure to achieve the required reliability criterion

91. Where the required reliability criterion cannot be achieved despite scrutinizing the data, revising the summary statistics, applying post-stratification, or considering the option of taking an additional sample, the activity participant may apply a conservative post-estimation adjustment to the parameter of interest. This approach recognises that, in some circumstances, further sampling may not be feasible, and statistical refinements may not sufficiently reduce variability. In such cases, a conservative adjustment may be applied to ensure that the resulting estimate does not overstate emission reductions.

92. This method consists of selecting either the upper or lower confidence bound corresponding to the required confidence level. The choice of bound shall reflect the direction that results in a conservative estimate of emission reductions. Where a lower parameter value results in a more conservative estimate of emission reductions (e.g., mean usage hours, retention rates, or continued-use proportions), the lower confidence bound shall be used. Conversely, where a higher parameter value results in a more conservative estimate of emission reductions (e.g., emissions factors or leakage parameters), the upper confidence bound shall be used. The application of this method shall be clearly justified and documented, including the confidence interval applied and the rationale for the selected confidence bound.

93. Where the required reliability criterion has not been achieved, the activity participant may also choose to apply a confidence interval corresponding to a higher level of conservativeness (e.g., a 95 per cent confidence interval rather than a 90 per cent confidence interval), provided that the selected confidence bound results in a conservative estimate of emission reductions and the approach is appropriately justified and documented.

94. The selected confidence bound shall be used as the adjusted parameter value for the purposes of calculating emission reductions.

6.8.1. Scenario: LED project – mean value parameter

95. This example uses the same example as example 1, for which the following data are presented in table 8.

Table 8. Updated sample statistics and associated reliability assessment

Summary statistics	Sample data
Population size	420 000
Sample size (n)	140
Mean	3.7230
Standard deviation	3.7838

96. The required reliability criterion is not achieved by the sample data and therefore the estimated mean of 3.72 hours is not sufficiently reliable. This also means that the confidence interval associated with the estimate is wider than required.

97. In this example, the 95 per cent confidence interval is calculated as:

$$mean \pm t_{\alpha/2,(n-1)} \times \frac{SD}{\sqrt{n}} = 3.7230 \pm 1.96 \times \frac{3.7838}{\sqrt{140}} \quad \text{Equation (19)}$$

Where:

n	=	Sample size
$mean$	=	Sample mean
SD	=	Sample standard deviation
$t_{\alpha/2,(n-1)}$	=	Critical value of Student's t-distribution with *(n – 1)* degrees of freedom for which the cumulative probability equals *1 – α / 2* (e.g., $t_{0.025,97}$ = 1.985 for a 95 per cent confidence level)

98. The resulting 95 per cent confidence interval is 3.10 to 4.35 hours. There is therefore 95 per cent confidence that the true population mean LED daily usage lies between 3.10 and 4.35 hours. While the sample mean is the estimate used in calculations, it is advisable to present the estimate together with its confidence interval.

99. As an alternative to taking an additional sample, the activity participant may use a confidence bound from the 95 per cent confidence interval. The choice of confidence bound shall be made in a manner that ensures conservativeness in the estimation of emission reductions.

100. The lower and upper confidence bounds of the 95 per cent confidence interval are shown in table 9.

Table 9. Selection of a conservative confidence bound in lieu of additional sampling

	Lower confidence bound	Mean	Upper confidence bound
95 per cent confidence interval	3.10	3.72	4.35

101. In this example, a lower value of the parameter results in a more conservative estimate of emission reductions. Accordingly, the lower confidence bound of 3.10 is selected for use in the calculations.

6.8.2. Scenario: cookstove project – proportion parameter

102. This example illustrates the application of a conservative confidence-bound adjustment where the parameter of interest is a proportion.

103. The parameter of interest is the proportion (or percentage) of improved cookstoves that remain in use (hereinafter referred to as "the retention rate of improved cookstoves") within a population of cookstoves distributed in a particular region. The population consists of 80,000 households, each of which received one cookstove. A simple random sample was to be selected.

104. The sample size calculation assumed an expected retention rate of 0.85 (85 per cent). This resulted in a required sample size of 68 cookstoves. As the expected response rate was 70 per cent, the sample size was increased to 98 cookstoves (68 / 0.70 = 98). A simple random sample of 98 households was then selected, and whether the improved cookstoves remained in use was recorded. Summary statistics and the results of the reliability assessment are presented in table 10.

Table 10. Summaries of the data and results of the reliability assessment

Summary statistics	Sample data
Population size	80 000
Sample size (n)	98
Number of distributed improved cookstoves still operational	74
Retention rate of improved cookstoves	0.7551
Relative precision (in per cent)	11.28

105. The required reliability criterion is not achieved by the sample data and therefore the estimated retention is not sufficiently reliable.

106. In this example, the 95 per cent confidence interval is:

$$p \pm 1.96 \times \sqrt{\frac{p \times q}{n}} = 0.7551 \pm 1.96 \times \sqrt{\frac{0.7551 \times (1 - 0.7551)}{98}} \quad \text{Equation (20)}$$

Where:

n	=	Sample size
p	=	Sample proportion
q	=	Expected proportion of unsuccessful cases (1 - p)
1.96	=	Critical value of the standard normal distribution for which the cumulative probability equals $1 - \alpha / 2$ (e.g., $z_{0.025}$ = 1.96 for a 95 per cent confidence level)

107. The resulting 95 per cent confidence interval is 67.00 per cent to 84.02 per cent. There is therefore 95 per cent confidence that the true retention rate lies between 67.00 per cent and 84.02 per cent. While the sample proportion is the estimate that would normally be

used in calculations, it is advisable to present the estimate together with its confidence interval.

108. As an alternative to taking an additional sample, the activity participant may use a confidence bound from the 95 per cent confidence interval. The choice of confidence bound shall be made in a manner that ensures conservativeness in the estimation of emission reductions.

109. The lower and upper confidence bounds of the 95 per cent confidence interval are shown in table 11.

Table 11. Selection of a conservative confidence bound in lieu of additional sampling

	Lower confidence bound	**Mean**	**Upper confidence bound**
95 per cent confidence interval (in per cent)	67.00	75.51	84.02

110. In this example, a lower retention rate results in a more conservative estimate of emission reductions. Accordingly, the lower confidence bound of 67.00 per cent is selected for use in the calculations.

Appendix 5. Sample size calculator

1. Activity participants may use the sample size calculator available at: https://unfccc.int/documents/660516

- - - - -

Document information

Version	*Date*	*Description*
01.0	9 October 2026	SBM 023, Annex 20. Initial adoption.
Decision Class: Regulatory Document Type: Tool Business Function: Methodology Keywords: A6.4 mechanism, best practices, methodologies, sampling		