Draft general guidance on sampling and surveys for SSC projects

Introduction

1. The purpose of this document is to provide guidance on applying sampling methods when using small scale CDM methodologies, as well as to provide direction on what is expected in the proposed sampling plan. Several approved small scale CDM methodologies require estimates of key inputs based on samples of affected equipment or facilities. SSC projects that use sampling shall include a description of the scope and approach to sampling and a justification for the selection of the chosen approach in the project design document. Project participants are encouraged to refer to standard references on probability sampling techniques or consult with experts to resolve questions that arise in the context of their specific project requirements. A list of references on probability sampling methods and issues is provided in the annex to this report.

2. The purpose of sampling as it applies to SSC projects is to obtain 1) unbiased and 2) reliable estimates of the mean or total values of key variables to use in the calculations of greenhouse gas emission reductions. Recognizing that measurements taken from a subset (sample) of subjects will typically differ from the means for the entire population, the estimates must be unbiased in the sense that measurements taken from repeated, independent samples would be, on average, equal to the population values.

3. Second, the estimates must be reliable in terms of how closely they are likely to fall around the true population value in repeated samples. The statistical reliability or precision of an estimate based on a sample is typically expressed in terms of the probability that the sample value falls within a specified interval around the population value. For example, one might describe a sample-based estimate as having a 90% probability of falling in a range of ± 10% of the true population value (often denoted as 90/10 precision). All accepted probability sampling methods provide formulas for calculating the precision of an estimate, i.e., the probability that it falls within a given range of the true population value, based on the variability of individual measurements in the sample.

4. The precision of a sample-based estimate increases directly with its size. Prior to drawing a sample, project participants can calculate the size required to achieve a given precision level based on a forecast or expectation of the variability of the characteristic in the population. In proposing how large a sample to use to obtain a given estimate, project applicants are expected to justify the number of measurements needed to achieve a target level of precision established by the Executive Board. That justification should be based on previous studies or sound engineering judgments. If the actual sample fails to achieve the target minimum precision level set by the Board, project implementers may be required to take a supplemental sample to achieve it.

5. Subject to these two requirements (unbiased estimates and minimum precision levels), project participants have broad discretion in the sampling approach they propose to use to obtain the estimates. The choice of which type of sample to propose depends on several considerations, including the types of information to be collected through sampling, the known characteristics of the population, and the cost of information gathering. Some of the most commonly used sampling methods are summarized below, along with guidelines on circumstances where each is most applicable. More complete descriptions, together with formulas, are presented in an appendix to this document.

Sampling Precision Requirements

6. Small scale project activities are often homogenous (i.e. share a common technology with similar operating characteristics) but dispersed (i.e., the technology is implemented at a large number of sites, households or facilities). An example is a solar cooking project. Monitoring of

the efficiency of every single solar cooker in a given project activity may not be practical or economical. Another example could be a project activity that installs compact fluorescent lamps in a large number of residential dwellings. A sampling approach could be used to collect data on retention rates and operating hours. In situations such as these, sampling is likely to be an appropriate approach for baseline determination and monitoring provided the approved methodology applied to the project does not explicitly state that sampling should not be used.

7. SSC methodologies specify minimum required levels of precision and confidence for various categories of variables collected by way of sampling. The samples should be chosen so as to meet or exceed these minimum levels. Project proponents may request a revision of these requirements in the methodology or request a deviation from the approved methodology in accordance with the procedures (see <http://cdm.unfccc.int/Reference/Procedures/methSSC_proc02_v01.pdf> and <http://cdm.unfccc.int/Reference/Procedures/reg_proc03_v02.pdf> providing sufficient justifications as to why a lower level is suitable for the planned application. If the estimates from the actual samples fail to achieve the target minimum levels of precision, project participants shall perform additional data collection on a supplemental sample.

8. Where there is no specific guidance in the approved methodology, project proponents shall choose 90/10 precision as the minimum precision targets for the most important data collection efforts on the most important data variable affecting the emission reductions of the project activity (For example, consider a project activity that has installed household biogas digesters in numerous distributed locations to displace fossil fuel use for cooking. The number of annually operating biogas digesters directly impacts the emissions reductions of the project activity; therefore the number of households for the sample should be chosen so as to achieve a 90% confidence interval with 10 per cent error margin for the collected data. On the other hand a 90/30 precision may be adopted for parameters of outside impact, indirect impact and verification analysis (For example positive spill over effect of a biogas digester project activity i.e., the number of households outside the boundary of the project activity, who are not project participants but nevertheless installed biogas digesters on their own may be assessed at 90/30 precision).

9. Approved SSC methodologies define variables whose values will specifically be obtained though sampling. Moreover, project implementers may propose to obtain estimates of other variables using sampling techniques if that is the only practical or cost effective means to obtain them. For example, several methodologies require that estimates for key variables be obtained through “monitoring”, without specifying the extent of such monitoring.

10. In broad terms, the methodologies require sampling in the following types of applications:

- Obtaining a point estimate for a variable to be used in a definitional or engineering formula. For example, the average annual hours of operation of lighting is used to estimate savings, where savings equal the change in wattage (determined at installation) multiplied by the average hours of operation (based on a sample estimate). ;
- Estimating the baseline penetration or characteristics of an equipment technology. For example, several methodologies require estimates of the average efficiency of replaced equipment, such as heating or lighting systems;
- Estimating whether the penetration or operating characteristics of an efficient technology or process have changed over time. For example, the refrigerator replacement program requires an annual survey to estimate the percent of units still in operation over time.

In addition, future methodologies or amendments may call for estimating whether a field value is significantly different from a value based on laboratory tests or previous studies.

Table 1 summarizes the minimum precision requirements for different applications.

Table 1 SSC Precision Requirements

Type of Sampling Estimate	Minimum Confidence Level	Maximum Error Bound	Minimum Sample Size
Point Estimate for Engineering Calculation	90%	± 10%	50
Baseline Penetration or Equipment Characteristic	90%	± 10%	50
Change in Technology Penetration or Performance	80%	± 20%	50

Summary of Sampling Approaches and Applicability

11. The following is a summary of some of the most common types of sampling approaches and situations where each is recommended. Formulas for calculating standard errors of estimates from each sampling technique and associated sample sizes are provided in the annex to this report.

Simple Random Sample

12. A **simple random sample** is a subset of individuals (a sample) chosen from a larger set (a population). Each individual is chosen randomly and entirely by chance, such that each individual has the same probability of being chosen at any stage during the sampling process, and each subset of *k* individuals has the same probability of being chosen for the sample as any other subset of *k* individuals (Yates, Daniel S.; David S. Moore, Daren S. Starnes (2008). *The Practice of Statistics, 3rd Ed.*. Freeman. ISBN 978-0-7167-7309-2.). Simple random sampling is the most straightforward method for designing a sample.

13. Under simple random sampling, each case in the sample frame (an exhaustive list of all the cases in the population) has an equal probability of being selected into the sample. The mean value of the measurement from a random sample is an unbiased estimate of the true population mean, which means that repeated independent samples will provide estimates that are, on average, equal to the population mean.

14. Simple random sampling has the advantages that it is the most straightforward way of obtaining a representative estimate based on a random subset of the population. One simply assigns a random number to each case in the sample frame and selects the cases with the highest numbers corresponding to the target sample size. (For practical reasons discussed below, it is always advisable to oversample from the frame.) Using random sampling methods are recommended when more efficient sampling techniques are infeasible, impractical, or where the population is relatively homogeneous with respect to the object of the study. For example, other sampling methods typically require more information from the sample frame, such as a stratification variable.

15. Simple random sampling is free of classification error, and it requires minimum advance knowledge of the population. Its simplicity also makes it relatively easy to interpret data collected.

For these reasons, simple random sampling best suits situations where there is limited information available about the population and data collection can be efficiently conducted on randomly distributed items, or where the cost of sampling is small enough to make efficiency less important than simplicity. If these conditions are not true, stratified sampling or cluster sampling may be a better choice.

Systematic Sampling

16. Systematic sampling is a **statistical method** involving the selection of elements from an ordered **sampling frame**. The most common form of systematic sampling is an equal-probability method, in which every k^{th} element in the frame is selected, where *k*, the sampling interval (sometimes known as the 'skip'), is calculated as:

k = population size (N) / sample size (n)

17. Using this procedure each element in the population has a known and equal probability of selection. This makes systematic sampling functionally similar to **simple random sampling**. It may be much more efficient, however, if variance of the characteristic of interest within the systematic sample is greater than its variance in the population.

18. The researcher must ensure that the chosen sampling interval does not hide a pattern. Any pattern would threaten randomness. A random starting point must also be selected. Systematic sampling is to be applied only if the given population is logically homogeneous, because systematic sample units are uniformly distributed over the population.

19. Systematic sampling is applicable in a number of situations. If there is a natural flow of subjects in the population, such as output of bricks in a manufacturing process, then it is typically easier to sample every kth unit to test for quality as they are produced. If personnel are expected to take field measurements from a sub-sample of subjects based on information gathered in the course of each survey, then systematic sampling may be easier to implement. That would be the case, for example, if a surveyor takes an inventory of lighting fixtures in a building and then installs meters on a subset of them. In all cases, it is important that the list of subjects or the process is naturally random, in the sense that there is no pattern to its order.

Stratified Random Sample

20. Another method is called stratified random sampling. When sub-populations vary considerably, it is advantageous to group cases into relatively homogeneous subpopulations and sample each subpopulation, called a stratum, independently. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. For example, the population of participants in a commercial lighting program might be grouped according to building type. The stratification requires that information on the stratification variable, e.g., building type, be contained in the sample frame. Then random or systematic sampling is applied within each stratum.

21. Stratification can increase the efficiency, i.e., produce a gain in precision for a given sample size, if the cases within each stratum are more homogeneous than across strata. For example, if lighting usage within building types (office buildings, retail stores, etc.) varies less than across building categories, then estimates of hours of operation using a stratified sample will produce an estimate with lower variance for a given sample size.

22. If population density varies greatly within a region, stratified sampling can also ensure that estimates will be made with equal accuracy in different parts of the region, and that comparisons of sub-regions can be made with equal statistical power. For example, a survey taken throughout a

particular province might use a larger sampling fraction in the less populated north, since the disparity in population between north and south is so great that a sampling fraction based on the provincial sample as a whole might result in the collection of only a handful of data from the north. Randomized stratification can therefore be used to improve population representativeness in a study.

23. Stratified random sampling is most applicable to situations where there are natural groupings of subjects whose characteristics are more similar within group that across groups. It requires that the grouping variable be known for all subjects in the sample frame. For example, the sampling frame would require information on the building type for each case in the population to allow stratification by that characteristic.

Cluster Sampling

24. Clustered sampling refers to a technique where the population is divided into sub-groups (clusters), and the sub-groups are sampled, rather than the individual elements to be studied. Cluster sampling is used when “natural” groupings are evident in a population. In this technique, the total population is divided into sub-groups (clusters), and a sample of the groups is selected. For example, suppose a project installs high efficiency motors in buildings, with several motors typically in each building. If one is interested in estimating the operating hours of the motors, one might take a sample of the buildings instead of the motors, and then meter all of the motors in the selected buildings. In contrast to stratified sampling, where the equipment of interest is grouped into a relatively small number of homogeneous segments, there are many clusters of motors (i.e., buildings), and there is no expectation that the motors in each building are more homogeneous than the overall population of efficient motors.

25. One version of cluster sampling is **area sampling** or **geographical cluster sampling**. Clusters consist of geographical areas. Because a geographically dispersed population can be expensive to survey, greater economy than simple random sampling can be achieved by treating several respondents within a local area as a cluster. It is usually necessary to increase the total sample size to achieve equivalent precision in the estimates, but cost savings may make that feasible.

26. There are at least two reasons for using a clustered sampling approach to collect data. The first is cost. If a significant component of the cost of data collection is travel time between sites, then it may make sense to monitor all of the equipment at individual locations to reduce that cost component. Under that approach, it will typically be necessary to meter more pieces of equipment than under random sampling to achieve a given level of precision. But the reduction in cost may more than offset any negative effects on sample precision, allowing one to take a larger sample for a given budget, with an increase in precision.

27. The second reason is the ease of constructing the sample frame. In some cases where the program participant is collecting baseline information, it may be impossible to enumerate the population of pieces of equipment from which to draw the sample. But it is possible to enumerate the clusters, e.g., buildings. In that situation, the program participant could sample the buildings and conduct an inventory of the equipment in the chosen units.

28. In most applications of cluster sampling to monitor efficient equipment, the sub-groupings of units occur naturally, with a different number of elements per cluster. For example, a building or plant location might constitute a natural cluster, with varying numbers of motors per location.

Multi-Stage Sampling

29. **Multistage sampling** is a complex form of cluster sampling. Measuring all the sample elements in all the selected clusters may be prohibitively expensive or not necessary. Under those

circumstances, multistage cluster sampling becomes useful. In multi-stage sampling, the units (referred to as primary units) in the population are divided into smaller sub-units (referred to as secondary units), similar to cluster sampling. In contrast to cluster sampling where all of the secondary units (elements) are measured, data are collected for only a sample of the sub-units. For example, a study of efficient lighting might first draw a sample of buildings, and then take a sample of lighting fixtures in each selected building. If the characteristics of the fixtures in a given building are very similar and the costs of measuring them is relatively high, then taking a sample of fixtures may be sufficient to achieve a target level of precision at lower cost. On the other hand, if the measurements are inexpensive once a technician is on-site, then it may make sense to monitor all of the fixtures.

Multi-stage sampling can be extended further to three or more stages. For example, one might group the population into building complexes, then buildings, and finally fixtures.

30. There are many variations in methods in applying multi-stage sampling. If the number of secondary units in each primary unit is not known in the sample frame, then one approach is to draw a sample of primary units at random, count the number of secondary units in each selected primary unit, and then take detailed measurements for a sample of secondary units. If the number of secondary units is known in the sample frame and varies only moderately across units, then one can stratify the primary unit population by size and draw successive random samples of primary and secondary units. The standard formulas for random sampling apply to the secondary unit means, and the formulas for stratified sampling apply to the grand mean. Another option is to sample the primary units with probability proportional to size, and to draw a random sample of the secondary units in the selected primary units. The relative performance of these alternatives depends on the population characteristics, the costs of data collection, and the availability of information on the primary and secondary units in the sample frame.

Sampling Practices

31. In all of the approaches, care must be taken to ensure that the samples are drawn in a manner that avoids any systematic bias and that the data collection minimizes non-sampling errors. In order to achieve those goals, practitioners are expected to observe sound practices in designing samples and administering surveys and field measurements.[1] Those practices include:

- **Defining precisely the sampling objectives, target population and the sample measurements.** The target population from which the sample will be drawn, the information that will be collected, and the methods of measurements should be clearly specified.
- **Developing the sampling frame**. The implementer should compile as complete list of the subjects in the target population as possible, along with any information needed to implement the chosen sampling technique and to contact selected subjects. In cases where the planned measurements will be taken from project participants, that list would typically have been maintained as part of the program or project tracking. In cases where the measurements are aimed at determining baseline penetrations or technology characteristics, the implementer may need to construct the list from other sources, such as electric utility account records, vehicle registrations, or business directories. The implementer should identify where the sample frame may differ from the target population and establish procedures on how that problem will be handled. For example, if the target population is residential households or dwellings with electric service and utility

[1] For a very comprehensive treatment of issues surrounding sample/survey design, see Household Sample Surveys in Developing and Transition Countries, United Nations, 2005, ISBN 92-1-161481-3.

billing records are used to construct the frame, then the sampling may need to address such issues as non-residential and master metered accounts. In situations where it is impossible to construct a sample frame that accurately represents the target population, the implementer may need to use area cluster sampling or another approach that is feasible, while applying the corresponding estimation and statistical confidence formulas.

- **Randomizing cases and drawing sample**. The implementer should ensure that the sample is drawn at random from the sample frame. If a systematic sampling is chosen, then the ordering of subjects on the sample should be random and free of any trend or cyclical pattern.
- **Selecting the most effective information gathering method**. The implementer should decide on what would be the most reliable and cost effective method for collecting the data, depending on the variables of interest. Alternative methods include visual inspections, physical measurements, respondent self-reports, and operational logs. For example, equipment penetrations and retention rates may be determined by inspections or self-reports. Estimates of electric consumption could be based on different metering technologies depending on the characteristics of the equipment. Vehicle travel miles or equipment operating schedules could be drawn from odometers or operation logs.
- **Conducting surveys/measurements.** The project implementer is expected to establish and implement procedures to ensure that the field data collection is performed properly and that any potential errors are minimized and documented. Such procedures include developing field measurement protocols, training personnel, establishing contact procedures, documenting coverage problems, missing cases, and non-response, minimizing non-sampling measurement errors, and quality control for data coding errors.

Requirements for Sampling Plan in PDD

32. Project design documents submitted to the Executive Board should include plans for collecting information from samples. Those plans should cover the topics summarized in the previous section:

- **Field Measurement Objectives and Data to be collected**. The plan should clearly describe the variables to be collected, the scope and method of the survey or field measurements, their frequency, and how the data will be used;
- **Target Population**. The plan should describe the target population for the survey or field measurements and summarize its known characteristics.
- **Sampling Frame**. The plan should summarize the sampling frame and the information it contains that will be required to implement the proposed sample method. If the frame is not already at least partially constructed when the proposal is submitted, the plan should describe how it will be developed. The plan should also describe how the characteristics of the sampling frame may differ from those of the target population and whether such differences may seriously affect how representative the estimates may be of the desired variables. Methods for dealing with sampling frame problems, including possibly supplementing it for known sources of non-coverage, should be described. (See Kish, pp. 53-59 for a treatment of frame problems.)
- **Sample Method**. The sampling method should be presented. That method should be consistent with the information contained in the frame.

- **Desired Precision/Expected Variance and Sample Size**. The plan should present and justify the target number of completed surveys or field measurements (the sample size). That justification should include a prediction of the variance of the variables of interest and basis for that prediction.
- **Procedures for Administering Data Collection and Minimizing Non-Sampling Errors**. The plan should describe the procedures for conducting the data collection, including training of field personnel, provisions for maximizing response rates, documenting out-of-population cases, refusals and other sources of non-response, and related issues.

Sampling Plan Evaluation Criteria

33. The proposed sampling plans will be evaluated based on whether they adequately address all of the issues and topics outlined above. Assessment includes whether the proposed approach to sample is practical given the available information about the population and the feasibility of developing the sample frame. The sampling approach will be evaluated for its adequacy in dealing with the range of sampling and non-sampling errors that may arise. The basis for the forecasts of the variance will assessed, along with the sufficiency of the proposed sample size given the minimum precision/confidence levels.

The sampling plan submitted by project proponents will be reviewed and assessed based on how effectively they address the following issues and questions:

- Does the sampling plan present a reasonable approach for obtaining unbiased, reliable estimates of the variables?
- Is the data collection method likely to provide reliable data given the nature of the variables and project, or is it subject to measurement errors?
- Is the population clearly defined and how well does the proposed approach to developing the sampling frame represent that population? Does the frame contain the information necessary to implement the sampling approach?
- Is the sampling approach suitable, given the nature of the variables, the data collection method, and the information in the sample frame?
- Is the proposed sample size adequate to achieve the minimum confidence/precision requirements? Is the ex ante estimate of the population variance needed for the calculation of the sample size adequately justified?
- Are the procedures for the data measurements well defined and do they adequately provide for minimizing non-sampling errors?

Example

Baseline Penetration of Compact Fluorescent Lamps (CFLs), Average Annual Operating hours, and Survival Rates for Project Lamps

34. **Project Description**. The project provides CFLs to residential households with low electric use through direct installations. The project targets low use households because the utility provides service to those customers at a discount to its marginal cost of electricity under its inverted block tariff. Teams of installers inspect dwellings and identify fixtures suitable for CFLs. The bulbs in each of those fixtures are replaced with a CFL of comparable lumens.

35. **Target Population, Measurement Objectives and Methods**. The target population for the project and the field data collection is residential dwellings with electric service whose average

monthly consumption falls below a given level. Three types of measurements will be taken. The first is a baseline inventory of lighting fixtures in each dwelling, including the percentage of screw-in fixtures already using CFLs. The information from the baseline survey will be used for future program planning, but not for the current project, because the requirement under the direct installation program that 100% of the retrofitted fixtures use incandescent bulbs. Once customers become more familiar with CFLs, the project will transition to offering incentives through normal retail channels. The baseline survey will provide critical information for designing that program.

36. The second measurement is the hours of operation of the CFLs. Those will be measured using light loggers that record the time intervals when the fixtures are turned on. The loggers will be placed in fixtures for a minimum of ninety days and moved periodically to capture any seasonal variations in lighting use. The primary objective of the measurements is to gain a reliable estimate of the average annual hours of operation of retrofitted fixtures for the purpose of calculating electricity savings.

37. The third set of measurements is aimed at determining the retention rates or effective useful lives of the CFLs. That will be accomplished be inspecting a sample of retrofitted fixtures annually to determine if the CFL is still operating or has been replaced by an incandescent bulb or comparable CFL.

38. **Sample Frame**. The sample frame for the baseline survey will be developed from the utility's customer account records. The frame consists of currently active accounts with a residential service code. The frame includes information on the customer's service and billing address, as well as electricity consumption for the past twelve billing periods.

39. The sample frames for the hours of operation and the CFL retention rates will be developed as part of the project tracking system. Each retrofitted fixture will be enumerated in the tracking system, along with information on its characteristics (e.g location, lumens).

40. **Sample Method**. The baseline survey will be performed on a cluster sample of dwellings. The dwellings will be drawn at random from the sample frame with each case having an equal probability of selection. For each chosen sample that participates in the survey, a complete inventory of fixtures will be taken. Information on fixtures location, type, wattage, and other relevant characteristics will be recorded. The baseline penetration of CFLs will be calculated as the total number of installed CFLs divided by the total number of screw-in fixtures that are CFL compatible.

41. The fixtures for the hours of operation measurements will be selected using two stage sampling. First, a sample of participating dwellings will be drawn at random from the project tracking system with the probability of selection proportional to the number of retrofitted fixtures in the dwelling. Then a single retrofitted fixture will be chosen at random from each dwelling in the sample for metering. This sampling procedure will be repeated every quarter and the meters will be moved to the new sample.

42. **Desired Precision/Expected Variance and Sample Size**. The target levels of precision for the baseline penetration and the average annual CFL operating hours are both ±10% with a 90% percent confidence level (critical t value of 1.64). For the purpose of determining the sample size for the CFL penetration rate, the project planners expect that approximately 20% of all residential fixtures already have CFLs. That expectation is based on a non-representative pilot survey and anecdotal information from project planners. If the fixtures for the baseline survey were selected totally at random from the residential population, the sample size needed to estimate the 20% penetration within ± 2% with a 90% confidence level would be 1076 (equals $.2\times.8\times(1.64/.02)^2$). But because the sample is clustered, the expected variance is higher. The planning purposes, it is assumed that the actual variance is 2.5 times that for a random sample. That yields a total sample

size of fixtures equal to 2690. The planners conservatively estimate that there is an average of at least six screw-in fixtures per dwelling, resulting in a total number of surveyed dwellings equal to approximately 450.

43. For the purpose of determining the sample size for the metering of hours of operation, the variance of the estimate used in planning is approximated by the formula for a simple random sample. That is a good approximation, given that the dwellings are selected with probability proportional to the number of retrofitted fixtures. For planning, it is assumed that the average annual lighting use is 1250 hours, or slightly less than 3.5 hours per day. Based on this, the target precision bound is =/- 125 hours per year. Previous studies of residential lighting usage have found that the standard deviation of usage is on the order of 500 hours per year, which implies that 95% of usage lies in the range of 250 to 2250 hours per year. The sample size needed to estimate the hours of usage within the target range at a 90% confidence level is less than 50 (equals $(500 \times 1.64/125)^2 = 43$). To be conservative, as well as to capture seasonal variations in lighting usage, four groups of fifty fixtures will be metered for successive ninety day periods. Each of the four sub-samples will be drawn independently, allowing seasonal comparisons of usage (although at a lower confidence level than average daily usage). The total sample of 200 fixtures is very conservative and is intended, in part, to compensate for any increase in the sampling error due to the two stage sampling approach.

44. The CFL attrition rate will be estimated by means of a longitudinal annual survey of participating dwellings. A random sample of participants will be drawn, and their dwelling fixtures will be inspected annually to verify continued CFL use. For planning purposes, we expect that the attrition rate will be approximately 10% in the first year, due to early failure and dissatisfaction with CFL performance. Afterwards, the removal rate is expected to fall to around 4%. Using the higher sample size needed to estimate the 4% failure rate =/- 20% at the 80% confidence level (critical t value of 1.282), a minimum of almost 986 fixture must be inspected annually. Using an average number of four retrofitted fixture per dwelling, the minimum number of housing units to be inspected is 250. To be conservative, the project will inspect 300 dwellings for annually.

45. Data Collection Procedures. The baseline survey, metering, and annual inspections will be carried out by a professional survey firm. The firm will use experienced field inspectors who are fully trained in proper techniques. The project will prepare the field inspectors with information about the project, CFL characteristics, contact procedures, treatment of out-of-population cases, refusals, and other issues arising in on-site inspections of this type. The recipients of the CFLs will be required to agree to being surveyed as a condition of project participation. All of the data collected under each survey component will be coded into an electronic database and checked for accuracy. Complete reports summarizing the results of each survey component will be prepared.

Annex 1

SAMPLING FORMULAS

Definitions

- N denotes the number of projects or devices in the population and assume that the projects are labeled $i = 1, \ldots, N$;
- *y* denotes any measurable variable of interest, such as hours of operation, and y_i denotes the value of y for project i;
- Y denotes the true total of y for all N projects in the population, i.e., $Y = \sum_{i=1}^{N} y_i$;
- $\overline{Y}$ denotes the population mean of y, $\overline{Y} = \frac{Y}{N} = \frac{1}{N}\sum_{i=1}^{N} = y_i$ [2];
- S^2 denotes the (true) population variance of y, $S^2 = \frac{1}{(N-1)}\sum(y_i - \overline{Y})^2$ [3];
- n denotes the sample size;
- s^2 denotes the estimate of the population variance based on the sample[4].

Simple Random Sampling

1. Under simple random sampling, each case in the sample frame has an equal probability of being selected into the sample. The estimate of the mean value from the sample is given by the formula:

$$\bar{y} = \sum^{n} \frac{y_i}{n}$$

2. The sample mean is an unbiased estimate of the true population mean, which means that repeated independent samples will provide estimates that are, on average, equal to the population mean.

3. The sample estimate for the population variance is given by the formula:

---

[2] Some texts use the notation μ to denote the population mean.
[3] Some texts use the notation σ^2 to denote the population variance rather than (uppercase) S^2.
[4] It is important to note the differences among the uses of the term "variance" in this document. The first is the **population variance**, which is the true variance of the variable of interest in the population and is unknown, unless a complete census is taken. The second is the **estimated variance of the variable of interest** from the sample. The third is variance of the mean, which is the **variance of the mean value $\bar{y}$**. Its square root is the standard error of the mean. The last is the **expected variance**. This is the researcher's expectation (or prior guess) of what the sample variance will turn out to be prior to taking the measurements from the sample.

$$S^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$ [5]

The standard error of *y*, denoted by the symbol *s*, is simply its square root.

4. The standard error of the mean $\bar{y}$, a measure of the dispersion of the estimate of the mean value from the sample around the true population mean, is:

$$S_{\bar{y}} = \frac{s}{\sqrt{n}} \sqrt{1 - n/N}$$ [6]

5. Under the fairly mild assumption that $\bar{y}$ is normally distributed, which "is adequate in most practical situations"[7], one can state the probability that $\bar{y}$ falls within a specified range of the population mean. That formula is given by:

$$Prob\left\{ \bar{Y} - \frac{ts}{\sqrt{n}}\sqrt{1-n/N} \le \bar{y} \le \bar{Y} + \frac{ts}{\sqrt{n}}\sqrt{1-n/N} \right\} = Prob(t)$$

Where t is the normal deviate corresponding to $Prob(t)$, i.e., the percentage of the normal probability density that falls within t standard deviations of the mean. Suppose, for example, that one obtains an estimate of mean operating hours of 800 per year from a sample, with a standard error ($S_{\bar{y}}$) of ± 40. Then one can state with 90% confidence that the true population mean lies in the range of ± 1.64*40 around 800, where 1.64 is the value of t corresponding to 90%. One can state with approximately 95% confidence that average operating hours falls in the range of 722-878 hours per year (1.96 standard deviations around 800 in each direction).

6. The formula also provides the basis for specifying the size of a sample, given a target level of precision for the estimate of the mean value and a prior estimate of the variance in the variable of interest in the population. Suppose, for example, that one wished to estimate the average operating hours with a precision of ± 100 hours per year, with a confidence of 90%. Further suppose that previous studies showed that the standard deviation of operating hours in the population is on the order of 500 hours. (For the purpose of this example, assume that the population (N) is large, so that one can ignore the term $\sqrt{1-n/N}$, known as the finite population correction term or fpc.) Given that the sample estimates of the mean are approximately normally distributed, 90% of its density falls within 1.64 standard deviations of its mean. To meet the precision target of 100, one must obtain an estimate of the mean with a standard error of 100/1.64, or approximately 61. Solving for the value of n results in a sample size of 67.24. Rounding up to the nearest whole number, a sample of 68 will produce an estimate of the operation hours that lies within 100 of the true population average over 90% of the time.

7. The formula for the required minimum sample size ($n_{\min}$) to estimate the mean value of a variable ($\bar{y}$) within an interval (l_t) at a designated probability P, where $t_{crit,}P$ is the t value corresponding to P, is given by:

---

[5] Cochran William G., (1977). Sampling Techniques, 3rd Ed.. John Wiley and Sons. ISBN 0-471-16240-X, p. 26.

[6] Cochran (1977), p.27

[7] Cochran (1977), p. 39

$$n_{\min} = \frac{t^2_{crit,p} \times S^2_y}{\left(l^2_t + (S^2_y \times \frac{t^2_{crit,P}}{N}) \right)}$$

8. The second term in the denominator, $(S^2_y \times t_{crit,p})/N$, approaches zero for a large population, so that the formula can be approximated by:

$$n_{\min} = \frac{t^2_{crit,p} \times S^2_y}{l^2_t}$$

9. When the variable of interest is binary (yes/no), the variance s^2_y is simply the frequency (percent "yes") times one minus the frequency (percent "no"). That would be the case, for example, in estimating the attrition rate of installed equipment, where "yes" indicates the equipment was removed and "no" indicates the equipment is still operating. If one uses an ex ante estimate of 10% attrition, the expected variance is .09 (equals .1×.9). Suppose one wants to calculate the minimum sample size necessary to estimate the attrition rate, ± 20%, at the 80% confidence level. Then $t_{crit,80\%}$= 1.282. The precision is .02 (20% of .1). The formula then is sample size = $.1\times.9\times(1.282/.02)^2$. That is equal to 370.

10. Estimates for variables with greater variability require larger samples to achieve comparable confidence/precision levels. One measure of the variability relative to the population mean value is called the coefficient of variation (cv). It is defined as the ratio of the standard deviation to the mean (the standard deviation is the square root of the variance). Using standard notation, this is σ/μ, where σ denotes the standard deviation and μ denotes the population mean. The cv is a useful measure of variability if the range of the variable is always positive. That would be the case, for example, for hours of operation of equipment. In that situation, the cv less than one might be reasonable. Average lighting use might be in the range of 1250 hours per year, with a standard deviation of 500 hours. Then 95% of all use lies roughly between 250 and 2250, two standard deviations of the mean. The cv in this example is 0.4. The cv for a binary variable whose that is between 20% and 80% lies between .5 and 2.0. But when the binary variable is close to zero, the cv grows much larger and is very sensitive to small changes.

11. If the precision range is defined in terms of a percentage of the mean value (e.g., 10% of the mean), the formula for the minimum sample size can be re-stated in terms of the cv and the percent (pct). It is:

Min sample size = $(\sigma/\mu)^2\times(t_{crit}/pct)^2$.

For example, if one wished to estimate the average hours of operation for lighting in a large population, with a cv of 0.4, within ±10% at the 90% confidence level ($t_{crit,\ 90\%}$=1.645), then one would need a minimum sample size of 44 (43.3, rounded up).

12. The following tables provide minimum sample size values by confidence precision levels for representative cases where the cv = 1 (Table A) and where the cv = 0.5 (Table B). In both examples, it is assumed that the population is sufficiently large, so that there is no need to use the finite population correction factor. Also, the minimum sample size of 50 is substituted in cases where the formula yields a number below that threshold. That avoids any adjustment in the critical values due to small sample sizes.

Table A: Minimum Samples Sizes for Coefficient of Variation = 1

		Confidence Level			
		80%	90%	95%	99%
Precision as Percent of Mean	1%	16435	27060	38416	66358
	5%	657	1082	1537	2654
	10%	164	271	384	664
	20%	50	68	96	166

Table B: Minimum Samples Sizes for Coefficient of Variation = 0.5

		Confidence Level			
		80%	90%	95%	99%
Precision as Percent of Mean	1%	4109	6765	9604	16589
	5%	164	271	384	664
	10%	50	68	96	166
	20%	50	50	50	50

Systematic Sampling

13. Systematic sampling is a **statistical method** involving the selection of elements from an ordered **sampling frame**. The most common form of systematic sampling is an equal-probability method, in which every k^{th} element in the frame is selected, where *k*, the sampling interval (sometimes known as the 'skip'), is calculated as:

k = population size (*N*) / sample size (*n*)

14. Using this procedure each element in the population has a known and equal probability of selection. This makes systematic sampling functionally similar to simple random sampling. It is however, much more efficient (if variance within systematic sample is more than variance of population).

15. The researcher must ensure that the chosen sampling interval does not hide a pattern. Any pattern would threaten randomness. A random starting point must also be selected. Systematic sampling is to be applied only if the given population is logically homogeneous, because systematic sample units are uniformly distributed over the population. Example: Suppose a brick manufacturer wants to validate the quality of its bricks, then using systematic sampling it can choose every 500^{th} brick being produced and perform the tests on this sample. This is random sampling with a system. From the sampling frame, a starting point is chosen at random, and choices thereafter are at regular intervals. For example, suppose you want to sample 20 bricks from a production run of 10000 bricks, so every 500^{th} brick leaving the kiln is chosen after a random starting point between 1 and 500. If the random starting point is 11, then the bricks selected are 11, 511, etc.

16. Systematic sampling may also be used with non-equal selection probabilities. In this case, rather than simply counting through elements of the population and selecting every k^{th} unit, we

allocate each element a space along a number line according to its selection probability. We then generate a random start from a uniform distribution between 0 and 1, and move along the number line in steps of 1.

Example: We have a population of 5 units (A to #). We want to give unit A a 20% probability of selection, unit B a 40% probability, and so on up to unit E (100%). Assuming we maintain alphabetical order, we allocate each unit to the following interval:

A: 0 to 0.2 B: 0.2 to 0.6 (=0.2+0.4) C: 0.6 to 1.2 (=0.6+0.6) D: 1.2 to 2.0 (=1.2+0.8) E: 2.0 to 3.0 (=2.0+1.0)

17. If our random start was 0.156, we would first select the unit whose interval contains this number (i.e., A). Next, we would select the interval containing 1.156 (element C), then 2.156 (element E). If instead our random start was 0.350, we would select from points 0.350 (B), 1.350 (D), and 2.350 (E).

18. Systematic sampling is a relatively simple method to apply, which makes it easily understood and useful when field personnel are asked to sample a subset of units based on their inspections. It is also useful when the population is organized in some temporal or other natural order without a trend or cyclical pattern. That would be the case for the production of bricks or other manufacturing processes.

Stratified Random Sampling

19. Another method is called stratified random sampling. When sub-populations vary considerably, it is advantageous to sample each subpopulation (stratum) independently. Under stratified random sampling, the population is divided into relatively homogenous subgroups, called strata. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. For example, the population of participants in a commercial lighting program might be grouped according to building type. The stratification requires that information on the stratification variable, e.g., building type, be contained in the sample frame. Then random or systematic sampling is applied within each stratum.

20. Allocation can be proportionate or optimum. Proportionate allocation uses a sampling fraction in each of the strata that is proportional to that of the total population. If the population consists of 60% in the male stratum and 40% in the female stratum, then the relative size of the two samples (three males, two females) should reflect this proportion. Optimum allocation (or Disproportionate allocation) - Each stratum is proportionate to the standard deviation of the distribution of the variable. Larger samples are taken in the strata with the greatest variability to generate the least possible sampling variance.

21. If population density varies greatly within a region, stratified sampling will ensure that estimates can be made with equal accuracy in different parts of the region, and that comparisons of sub-regions can be made with equal statistical power. For example, a survey taken throughout a particular province might use a larger sampling fraction in the less populated north, since the disparity in population between north and south is so great that a sampling fraction based on the provincial sample as a whole might result in the collection of only a handful of data from the north. Randomized stratification can therefore be used to improve population representativeness in a study.

22. Stratification can increase the efficiency, i.e., produce a gain in precision for a given sample size, if the cases within each stratum are more homogeneous than across strata. For example, if lighting usage within building types (office buildings, retail stores, etc.) varies less than

across building categories, then estimates of hours of operation using a stratified sample will produce an estimate with lower variance for a given sample size.

Notation for Stratified Random Sampling

N_h denotes the total number of units in the stratum *h*

n_h denotes the number of units in the sample in stratum *h*

y_{hi} denotes the value for the *i*th unit in stratum *h*

$W_h = \frac{N_h}{N}$ denotes the stratum weight

$f_h = \frac{n_h}{N_h}$ denotes the sampling fraction in the stratum

$\overline{Y}_h = \frac{\sum^{N_h} yhi}{N_h}$ denotes the true population mean in stratum *h*

$\overline{y}_h = \frac{\sum^{N_h} yhi}{N_h}$ denotes the sample mean in stratum *h*

$S_h^2 = \frac{\sum^{N_h} (y_{hi} - \overline{Y}_{hi})^2}{N_h - 1}$ denotes the true population variance

23. The estimate of the mean from a stratified random sample is:

$\overline{y}_{st} = \sum W_h \overline{y}_h$

24. The estimate is unbiased, regardless of how the sample is distributed across strata, as long as the cases are drawn independently, at random from each stratum. If the stratified sample is drawn proportionately from each stratum, i.e., if $n_h / n = N_h / N$ the stratified sample mean is arithmetically the same as the random sample mean.

25. The variance of the mean estimate is given by:

$V(\overline{y}_{st}) = \sum W_h^{\ 2} \frac{S_h^2}{n_h}(1 - f_h)$. If a random sample is taken in each stratum, then an unbiased estimate of S_h^2 is

$S_h^2 = \frac{1}{(n_h - 1)} \sum^{n_h} (y_{hi} - \overline{y}_h)^2$. This produces an unbiased estimate of the overall variance of the sample mean:

$$S^2(\bar{y}_{st}) = \frac{1}{N^2} \sum N_h (N_h - n_h) \frac{S_h^2}{n_h}.$$

26. That formula can be used to calculate a sample size necessary to achieve a given level of precision in the same manner as described for a random sample. The calculation requires prior estimates of the variances in each stratum. Given those prior estimates, one can calculate an optimal allocation of the total sample across strata (n_h), and then compute the resulting variance (and standard error) for the different total sample sizes. That determines the sample necessary to achieve a specified level of precision and confidence.

27. The optimal allocation of a sample across strata depends on the relative variances in each stratum and the cost of data collection per sample point. The formula for the optimal allocation is:

$$\frac{n_h}{n} = \frac{W_h S_h / \sqrt{c_h}}{\sum \left(W_h S_h / \sqrt{c_h} \right)}$$, where c_h is the cost per sample point in stratum h.

28. Even where it is impossible to obtain accurate estimates of relative stratum variances and data collection costs, the formula provides guidance for improving the efficiency of sample estimates through stratification. Those are to allocate larger portions of the sample to strata where:

- the subpopulation is larger;
- there is more variability within the stratum; and
- Sampling is less expensive.

Cluster Sampling

29. Clustered sampling refers to a technique where the population is divided into sub-groups (clusters), and the sub-groups are sampled, rather than the individual elements to be studied. Cluster sampling is used when "natural" groupings are evidence in a statistical population. In this technique, the total population is divided into sub-groups (clusters) and a sample of the groups is selected. For example, suppose a project installs high efficiency motors in buildings, with several motors typically in each building. If one is interested in estimating the operating hours of the motors, one might take a sample of the buildings instead of the motors, and then meter all of the motors in the selected buildings. In contrast to stratified sampling, where the equipment of interest is grouped into a relatively small number of homogeneous segments, there are many clusters of motors (i.e., buildings), and there is no expectation that the motors in each building are more homogeneous than the overall population of efficient motors.

30. One version of cluster sampling is **area sampling** or **geographical cluster sampling**. Clusters consist of geographical areas. Because a geographically dispersed population can be expensive to survey, greater economy than simple random sampling can be achieved by treating several respondents within a local area as a cluster. It is usually necessary to increase the total sample size to achieve equivalent precision in the estimators, but cost savings may make that feasible.

31. There are at least two reasons for using a clustered sampling approach to collect data. The first is cost. If a significant component of the cost of data collection is travel time between sites, then it may make sense to monitor all of the equipment at individual locations to reduce that cost component. Under that approach, it will typically be necessary to meter more pieces of equipment than under random sampling to achieve a given level of precision. But the reduction in cost may more than offset any negative effects on sample precision, allowing one to take a larger sample for a given budget, with an increase in precision.

32. The second reason is the ease of constructing the sample frame. In some cases where the program participant is collecting baseline information, it may be impossible to enumerate the population of pieces of equipment from which to draw the sample. But it is possible to enumerate the clusters, e.g., buildings. In that situation, the program participant could sample the buildings and conduct an inventory of the equipment in the chosen units.

33. In most applications of cluster sampling to monitor efficient equipment, the sub-groupings of units occur naturally, with a different number of elements per cluster. For example, a building or plant location might constitute a natural cluster, with varying numbers of motors per location.

34. There are many different techniques for sampling clusters, each with different formulas for estimating the means and their variances. One can sample each cluster with equal probability or with a probability proportional to the size of the cluster. No single method is always superior to the others in all applications. Their relative precisions depend on the characteristics of the population and the clusters. If the variable of interest is unrelated to the size of the clusters, then sampling with a probability proportional to size generally works best. For example, if the average hours of operation do not depend on the number of pieces of equipment in each cluster, then sampling clusters with a probability proportionate to the size of the cluster tends to work better than other sampling techniques.

For a sample of clusters drawn at random with probabilities proportional to the size of the cluster, the unbiased estimate of the mean is:

$$\bar{\bar{y}}_{pps} = \frac{1}{n}\sum\left(\frac{y_i}{M_i}\right) = \frac{1}{n}\sum(\bar{y}_i)$$

Where:

n denotes the number of sampled clusters

M_i denotes the size of cluster *i*

$\bar{y}_i$ denotes the mean value of *y* in cluster *i*

So the unbiased estimate of the overall mean is simply the unweighted mean of the values in each cluster, under probability sampling proportional to the size (number of elements) in M.

The estimate of the variance of $\bar{y}_{pps}$ is given by:

$$v(\bar{\bar{y}}_{pps}) = \sum(\bar{y}_i - \bar{\bar{y}})^2 / n(n-1)$$

35. In situations where the size (number of elements) of each unit is not known in advance, the formulas for a simple random sampling apply to cluster sampling. The overall mean value of the estimate for elements is the average of the means across units. There is no gain in efficiency due to cluster sampling. Rather the reason for using that method is the practical need, given the absence of information about the relative sizes of each cluster.

Multi-Stage Sampling

36. **Multistage sampling** is a complex form of cluster sampling. Using all the sample elements in all the selected clusters may be prohibitively expensive or not necessary. Under these circumstances, multistage cluster sampling becomes useful. In multi-stage sampling, the units (referred to as primary units) in the population are divided into smaller sub-units (referred to as secondary units), similar to cluster sampling. In contrast to cluster sampling where all of the

secondary units (elements) are measured, data are collected for only a sample of the sub-units. For example, a study of efficient lighting might first draw a sample of buildings, and then take a sample of lighting fixtures in each selected building. If the characteristics of the fixtures in a given building are very similar and the costs of measuring them is relatively high, then taking a sample of fixtures may be sufficient to achieve a target level of precision at lower cost. On the other hand, if the measurements are inexpensive once a technician is on-site, then it may make sense to monitor all of the fixtures.

Multi-stage sampling can be extended further to three or more stages. For example, one might group the population into building complexes, then buildings, and finally fixtures.

37. There are many variations in methods in applying multi-stage sampling. If the number of secondary units in each primary unit is not known in the sample frame, then one approach is to draw a sample of primary units at random, count the number of secondary units in each selected primary unit, and then take detailed measurements for a sample of secondary units. If the number of secondary units is known in the sample frame and varies only moderately across units, then one can stratify the primary unit population by size and draw successive random samples of primary and secondary units. The standard formulas for random sampling apply to the secondary unit means, and the formulas for stratified sampling apply to the grand mean. Another option is to sample the primary units with probability proportional to size, and to draw a random sample of the secondary units in the selected primary units. The relative performance of these alternatives depends on the population characteristics, the costs of data collection, and the availability of information on the primary and secondary units in the sample frame.

In all cases for multi-stage sampling methods, as well as other sampling strategies, the steps for calculating the sample size necessary to achieve a specified level of precision and confidence are comparable. First, one decides on how the overall sample and the subsamples should be allocated among primary and secondary sampling units. The formulas for optimal allocations depend on the chosen sampling strategy, the sizes of the primary and secondary units, the expected variances in those units, and the relative costs of data collection in each unit. Second, the standard error of the sample mean estimate, as a function of the primary sample size n, is calculated. Then the size of n necessary to achieve a target standard error based on the required precision and confidence level is calculated.

Methods for Obtaining Expected Variances

38. All classical sampling methods require an expected or prior estimate of the variance of the variable of interest in the population in order to calculate the sample size necessary to achieve a target level of precision. There are several ways that one can develop values for the expected variances of population characteristics. The first method is to take the sample in two steps. The first is a random sample that is used to obtain a preliminary estimate of S^2 and the second is an additional sample to achieve a target final variance of the estimate. The formula for the final sample size (initial random sample, plus additional sample) is:

$$n = \frac{S_1^2}{V}\left(1 + \frac{2}{n_1}\right)$$

39. Where S_1^2 is the estimate from the first step, and n_1 is the sample size in the first step, and V is the target variance of the final estimate.

Under that method, the final sample size is larger than would be if S were known exactly. In that case, the required sample size would be S^2/V. The two step approach increases the sample size by $2/n_1$.

40. The second method is to use a pilot survey to estimate the variance. The pilot is often a convenience survey, either measuring units that are easy to access or that are suspected of posing problems in the general survey. In general, the cases in the pilot survey sample are not used for the final estimates.

41. A third method is to use results from previous surveys. As the number of efficiency and related programs expand, there will be an increasing number of datasets available from which to draw such estimates. In some countries, regulators are already compiling reference values for such purposes.

42. A last method is to use engineering judgments about the nature of the population characteristics under investigation to "bound" the likely range of the variance. These judgments can be supplemented with sensitivity analyses that determine how much the required samples sizes vary with respect to the assumptions. For example, in validating the installations of efficient equipment, the maximum value of the variance falls at 50%. Under a "worse case" scenario, that could be used to compute the required sample size.

43. It is important to emphasize that if the expected variances prove overly optimistic, the estimates from the final samples will fall short of the required minimum precision and confidence levels. In such cases, program participants may be required to perform supplemental monitoring to meet those requirements.

References

The following are references to textbooks on survey sampling that are generally regarded as authoritative treatments of that subject:

Cochran, W.G. 1977, *Sampling Techniques,* 3d edn, Wiley, New York.

Dillman, D.A. 2007, Mail and internet surveys : the tailored design method, 2 ., 2007

update wi new internet, visual, a mix-mode guide edn, J. Wiley, Hoboken, N.J.

Groves, R.M. 2004, Survey methodology, J. Wiley, Hoboken, NJ.

Groves, R.M. 1989, Survey errors and survey costs, Wiley, New York.

Hansen, M.H. 1953, Sample survey methods and theory, Wiley, New York.

Harkness, J., de Vijver, F.J.R.v. & Mohler, P.P. 2003, Cross-cultural survey methods, J.

Wiley-Interscience, Hoboken, N.J.

Kish, L. 1995, Survey sampling, Wiley classics library edn, Wiley, New York.

Levy, P. & Lemeshow, S. 1999, *Sampling of populations: methods and applications,*

New York ; Wiley, c1999.

Household Sample Surveys in Developing and Transition Countries, United Nations, 2005, ISBN 92-1-161481-3

Sampling Software Packages

Most statistical software packages have procedures for drawing samples. Widely used commercial packages include SAS, Stata, and SPSS. The UN report cited above reviews these packages and others. Other resources are summarized at the following websites, along with links to free software.

http://www.hcp.med.harvard.edu/statistics/survey-soft/#Online

http://www.freestatistics.info/stat.php

- - - - -

History of the document

Version	Date	Nature of revision(s)
01	EB xx, Annex # 28 May 2009	To be considered at EB xx.